논문 2008-45SP-5-14

3GPP2 SMV의 실시간 유/무성음 분류 성능 향상을 위한 Gaussian Mixture Model 기반 연구

(Enhancement Voiced/Unvoiced Sounds Classification for 3GPP2 SMV Employing GMM)

송 지 현*, 장 준 혁**

(Ji-Hyun Song and Joon-Hyuk Chang)

요 약

본 논문에서는 패턴 인식에서 우수한 성능을 보이는 가우시안 혼합모델 (Gaussian mixture model, GMM)을 이용하여 비정상적인 잡음환경에서 3GPP2 selectable mode vocoder (SMV)의 유/무성음 분류 알고리즘 성능 향상을 위한 방법을 제안한다. 기존의 SMV에 대해서 분석하고, 이를 기반으로 유/무성음 분류 알고리즘에서 우수한 성능을 보여주는 특징 벡터를 선택하여 GMM의 입력벡터로 효과적으로 이용한다. 다양한 잡음환경에서 시스템의 성능을 평가한 결과 GMM을 이용한 제안된 방법이 기존의 SMV의 방법보다 우수한 유/무성음 분류 성능을 보였다.

Abstract

In this paper, we propose an approach to improve the performance of voiced/unvoiced (V/UV) decision under background noise environments for the selectable mode vocoder (SMV) of 3GPP2. We first present an effective analysis of the features and the classification method adopted in the SMV. And then feature vectors which are applied to the GMM are selected from relevant parameters of the SMV for the efficient voiced/unvoiced classification. For the purpose of evaluating the performance of the proposed algorithm, different experiments were carried out under various noise environments and yields better results compared with the conventional scheme of the SMV.

Keywords: Voiced/Unvoiced classification, Selectable Mode Vocoder (SMV), Gaussian Mixture Model (GMM)

I. 서 론

최근 새로운 디지털 무선 통신 시스템과 양방향 음성 통신 서비스가 비약적으로 발전하면서, 한정된 주파수 자원과 같은 디지털 무선 통신 시스템 환경의 효율적인 사용에 대한 연구가 진행 되고 있다. 특히, 네트워크와 전송 채널 상태에 따라서 전송 속도가 유연하게 변화함과 동시에 높은 품질을 보장하는 가변 전송률 음성 부호화하는 기술이 각광 받고 있다[1~2]. 실제로 우수한 가변적인 비트율을 갖는 음성 코덱의 실현을 위해서 유/무성음 구간을 분류하는 알고리즘의 성능이 중요한 요소로 작용하고 있고, 이와 관련하여 다양한 잡음 환경에서도 우수한 성능을 보이는 알고리즘의 연구가 활발히 진행되고 있다. 특히, 신호의 주기적 특성을 이용한 방법과 통계적 특성을 이용한 방법이 유/무성음 분류 알고리즘에서 매우 우수한 성능을 보여 주는 것으로 알려져 있고, 이러한 특징 벡터로 영교차율, 에너지, 피치, 상관계수, 선형 예측 계수등이 사용된다[3~7].

본 논문에서는 ETSI의 3GPP2 표준 코덱인 Selected Mode Vocoder (SMV)에 대해서 분석하고, SMV의 유/

* 학생회원, ** 정회원, 인하대학교 전자공학부
(Department of Electronics Engineering, Inha University)

※ 본 연구는 지식경제부 및 정보통신연구진흥원의 IT 핵심기술개발사업 [2008-F-045-01]과 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (IITA-2008-C1090-0804- 0007).

접수일자: 2008년1월19일, 수정완료일: 2008년8월5일

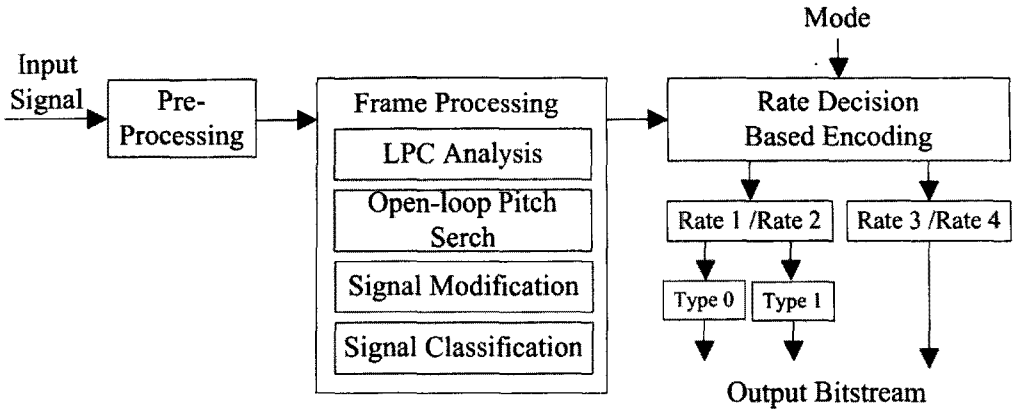

그림 1. SMV의 인코딩 과정에 대한 블록도
Fig. 1. Block diagram of the SMV encoding part.

무성음 분류 성능을 향상 시킬 수 있는 GMM 기반의 알고리즘을 제안한다. 구체적으로, 실제 유/무성음 분류에서 많이 사용되는 특징 벡터 중 SMV에 사용되는 특징 벡터를 선택하여 통계적 패턴 인식알고리즘에서 우수한 성능을 보이는 GMM의 특징 벡터로 사용하여 적은 계산량의 추가로 우수한 성능을 보여주는 알고리즘을 제안하였다. 제시된 알고리즘의 성능은 다양한 잡음 환경에서 기존의 SMV 유/무성음 분류 알고리즘 보다 향상된 결과를 보여주었다.

II. SMV (selectable mode vocoder)

SMV는 ETSI의 3GPP2 표준 코덱으로서 Extended Code Excited Linear Prediction (ex-CELP) 기반의 압축 방식을 사용하는데, 사람의 청각 특성에 최적화된 모델을 사용하여 음성을 저전송률로 압축하는데 효율적이다[8~9]. 또한, 한정된 주파수 대역을 효율적으로 사용하기 위해 가변 전송률을 갖고 이동국과 기지국 사이의 통신망 채널에 따라서 동적으로 바뀌는 4가지 모드를 제공하여 다양한 평균 전송률을 갖는 특성 때문에 시스템의 효율성과 음질간의 관계를 선택적으로 조절 할 수 있다. 표 1은 SMV에서 유성음 44%, 무성음 13.1%, 무음 42.9%로 구성된 깨끗한 음성 테스트 파일의 각 모드에 대한 결정된 전송률의 백분율을 보여준다.

표 1. 깨끗한 음성에 대한 전송률의 백분율 (%)
Table 1. Rate percentages for clean speech (%).

	mode 0	mode 1	mode 2	mode 3
Rate 1 (8.55 kbps)	55.9	28.5	11.0	5.3
Rate 2 (4.0 kbps)	4.5	18.7	36.2	42.0
Rate 3 (2.0 kbps)	0	10.8	9.7	9.7
Rate 4 (0.8 kbps)	39.6	41.9	42.9	42.9

1. SMV 인코더 개요

SMV는 8 kHz로 샘플링된 입력신호를 20 ms 길이의 프레임 단위로 처리한다. 그림 1은 SMV의 인코딩 과정에 대한 블록도를 나타낸다. 입력신호는 먼저 전처리기를 통해서 고대역 통과 필터를 통과한 후 잡음 억제기를 통과한다. 프레임 처리기는 전 처리된 신호로부터 피치, 단기 예측 오차, 선형 예측 계수 등을 계산한다. 신호 분류기는 프레임 처리기를 통해서 구해진 특징 벡터들과 각각의 문턱 값과의 비교를 통해서 프레임을 잡음, 묵음, 무성음, 비정상적 유성음, 정상적 유성음, 변화 중 한 개로 분류 되고, 통신 상태에 따라서 결정된 mode와 현재 프레임의 분류된 종류를 기반으로 전송률 결정 알고리즘에 의해 전송률이 결정된다. Rate 1 또는 Rate 1/2로 분류된 경우 다시 비정상적 유성음을 나타내는 type 0 과 정상적 유성음을 나타내는 type 1로 나누어져 비트를 할당한다. Type 0은 type 1에 비해서 적응 코드북에 더 많은 비트를 할당하고 고정 코드북에는 더 적은 비트를 할당하여 부호화 한다. Rate 1/4 또는 Rate 1/8은 Line Spectral Frequencies (LSF)와 에너지를 이용하여 부호화한다.

2. SMV 유/무성음 분류 시스템

SMV는 유/무성음 분류 알고리즘 결과를 기반으로

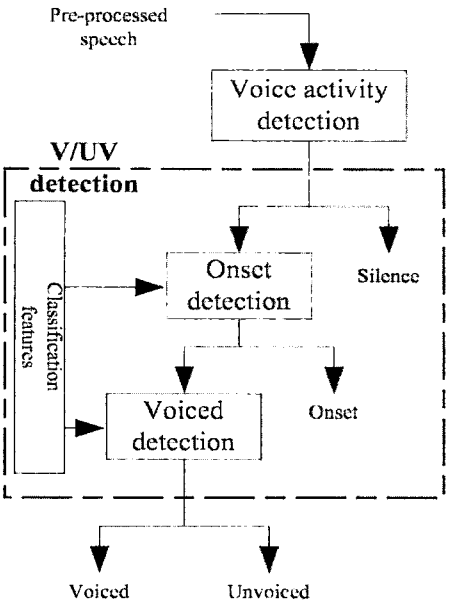

그림 2. SMV 유/무성음 분류 블록다이어그램
Fig. 2. Block diagram of voiced/unvoiced classification of the SMV.

입력된 신호의 프레임 분류 및 전송률을 결정하기 때문에 유/무성음 분류 알고리즘의 성능은 음성 코덱에서 매우 중요한 요소이다. SMV는 두 단계를 통해서 유/무성음 분류를 한다. 첫번째로 전처리된 신호와 가중된 잔여 신호를 이용하여 유/무성음의 특징을 잘 나타내는 7개의 특징 벡터를 추출하고 고정된 문턱값과 비교를 통해서 무음, 유성음, 무성음으로 분류 한다. 구체적으로 가중된 잔여 신호의 sharpness와 에너지, first order coefficient, 영 교차율 (zero-crossing rate), 전처리 신호의 평균 에너지, 샘플의 절대값이 0.1 보다 작은 샘플의 수가 유/무성음 분류 알고리즘의 특징벡터로 사용된다. 이 분류 결과를 기반으로 개회로 피치 검출을 통해서 구해진 특징 벡터와 이전 프레임의 분류 결과를 이용하여 더욱 세분화된 유/무성음 분류를 하는데, 그림 2에 유/무성음 분류 블록 다이어그램을 나타내었다[10]. 먼저, 유/무성음 분류기는 음성 검출기 (Voice Activity Detection, VAD)의 분류 결과를 통해서 입력 신호를 무음과 무성음으로 분류 한다. 무성음으로 분류된 프레임은 변화 검출기를 통해서 변화와 무성음으로 분류 되고, 무성음은 다시 유성음 검출기를 통해서 유성음과 무성음으로 재분류 된다. 유/무성음 분류 알고리즘에서 사용된 특징 벡터는 입력된 신호의 크기와 관련이 있기 때문에 깨끗한 입력 신호에서 우수한 유/무성음 분류 성능을 보여주지만 노이즈 환경에서는 성능을 보장하지 못하는 것을 알 수 있다.

Ⅲ. 제안된 GMM 기반의 유/무성음 분류 시스템

SMV의 유/무성음 분류 성능을 향상시키기 위해서 실시간 GMM 기반의 분류 알고리즘을 제안한다. 이번 연구에서 GMM을 사용하게 된 동기는 SMV 특징 벡터의 통계적 분포를 다른 평균과 공분산 행렬을 갖는 복수개의 가우시안 함수에 의해서 효과적으로 표현할 수 있기 때문이다. 제안된 방법은 SMV의 특징 벡터 중 유/무성음 분류 알고리즘에서 우수한 성능을 보여주는 특징 벡터를 별도의 계산 과정 없이 추출하여 GMM의 특징 벡터로 사용하여 분류 성능을 향상시킨다. 그림 3은 제안된 방법의 유/무성음 분류 알고리즘의 블록다이어그램을 나타내는데 구체적으로 SMV의 VAD에서 신호가 있다고 판단될 경우 제안된 GMM은 생성된 특징 벡터와 GMM의 모델을 이용하여 유성음과 무성음에 대한 우도를 생성하고, 우도비 테스트 (Likelihood Ratio Test, LRT)를 통해서 유/무성음을 분류 한다.

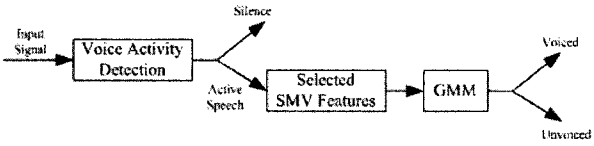

그림 3. 제안된 방법의 유/무성음 분류 블록도
Fig. 3. Block diagram of the proposed SMV method.

1. 제안된 알고리즘에 사용된 특징 벡터

특징 벡터의 통계적 편차가 클수록 더욱 우수한 성능을 보여주는 GMM의 특징벡터를 선택하기 위해서 SMV에서 사용되는 특징 벡터의 통계적 분포를 분석하였다. 그 중 그림 5에서 도시한 바와 같이 통계적 분포 특성이 우수한 에너지, 피치, 피치 상관계수, 반사 계수를 특징 벡터로 사용하였으며, 자세한 기술은 아래와 같다.

1.1. 에너지 (E)

일반적으로 에너지는 유성음은 크고, 무성음 작게 나타나는 특성 때문에 유/무성음 분류 알고리즘에서 우수한 특징 벡터로 이용된다. SMV에서는 선형 예측 부호화 (Linear Prediction Coding, LPC) 분석 과정에서 추출된 신호의 파워 ($R_1(0)$)와 LPC 윈도우의 길이 L_{lpc} (= 240)을 이용하여 얻어진다[10].

$$E = \max\left(10, 10 \cdot \log_{10}\left(\frac{R_1(0)}{L_{lpc}}\right)\right) \quad (1)$$

1.2 피치, 피치 상관계수

SMV의 개회로 피치 검출 과정은 그림 4와 같은 3개의 윈도우를 이용하여 3개의 피치와 피치 상관계수가 추출되고 고정된 문턱 값과 이전 프레임의 피치를 이용하여 각 프레임간 상관성을 고려하여 수정된다.

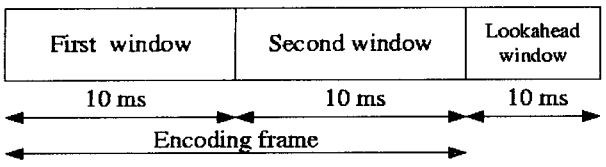

그림 4 LPC 윈도우
Fig. 4. LPC window.

1.3 반사계수 (reflection coefficients)

음성 프레임의 마지막 1/4에 중심을 둔 윈도우를 이용하여 계산된 LPC 계수를 Levinson-Durbin알고리즘에 사용하여 얻어진다[11].

2. Gaussian Mixture Model (GMM)

먼저 유/무성음 분류 시스템에서 사용되는 GMM은 가우시안 혼합성분 밀도의 가중치 합의 함수로서 다음과 같이 정의 된다[12~13].

$$P(\vec{x}|\lambda)=\sum_{i=1}^{M}\alpha_i P_i(\vec{x}) \tag{2}$$

여기서

$$P_i(\vec{x})=\frac{1}{(2\pi)^{\frac{D}{2}}\left|\sum_i\right|^{\frac{1}{2}}}exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu_i})^T(\sum_i)^{-1}(\vec{x}-\vec{\mu_i})\right\} \tag{3}$$

여기서 α_i는 혼합 성분의 가중치를 나타내고 μ_i는 평균 벡터, Σ_i는 공분산 행렬을 나타낸다.

$$\lambda=\left\{\alpha_i,\mu_i,\sum_i\right\},\quad i=1,.....,M \tag{4}$$

GMM은 크게 유/무성음에 대한 특징 벡터의 분포를 가장 잘 나타낼수 있는 모델을 찾는 훈련부와 이 모델을 이용하여 인식하는 인식부로 나눌 수 있다. 먼저 훈련부는 위와 같은 파라미터를 가지고 Expectation Maximization (EM) 알고리즘 기반의 학습을 통해서 유/무성음에 대한 혼합 가우시안 모델 λ을 추정하고 인

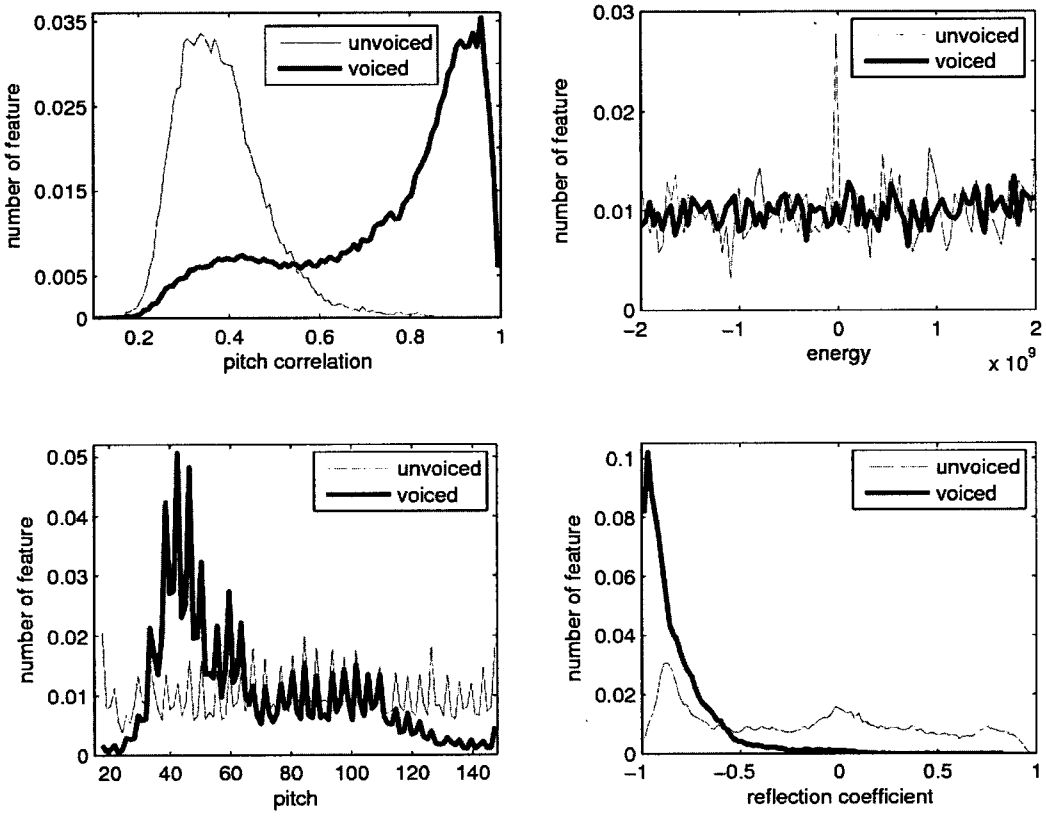

그림 5. SMV 특징 벡터에 대한 통계적 분포도
Fig. 5. Normalized distributions of the selected SMV feature vectors.

식부는 훈련부에 만들어진 λ을 이용해서 입력된 음성 신호에 대한 사후 확률을 구하여 가장 큰 확률을 갖는 모델을 찾는다.

$$\hat{S} = \arg \max_{k \quad \{v,u\}} \sum_{t=1}^{T} \log p(\vec{x}_t|\lambda_k). \quad (5)$$

본 실험에서는 GMM의 특징 벡터로서 SMV의 유/무성음 분류 알고리즘에 사용되는 피치 상관계수, 피치와 이전에 추출된 특징 벡터인 에너지와 반사계수를 사용하여 다양한 혼합성분 개수로 λ_v(유성음 모델)과 λ_u(무성음 모델)를 추정하였다. 테스트 과정에서 GMM의 모델 파라미터 λ_v과 λ_u에 실제 데이터가 입력되어 각 모델에 대한 우도를 생성하고, 아래와 같은 우도비(Likelihood Ratio)를 이용하여 유/무성음을 분류 한다.

$$\Lambda_x^{(t)} = \frac{P(\vec{x}^{(t)}|\lambda_v)}{P(\vec{x}^{(t)}|\lambda_u)} \begin{matrix} \text{voiced} \\ > \\ < \\ \text{unvoiced} \end{matrix} \eta \quad (6)$$

여기서 η 는 유/무성음 분류의 문턱값이고, t는 프레임 번호를 나타낸다.

Ⅳ 실 험

본 논문에서는 제안된 GMM 기반의 유/무성음 분류 성능을 평가하기 위해서 4명의 여자와 4명의 남자에 의해서 녹음된 NTT 음성데이터베이스가 사용되었다. 실험의 GMM 훈련에서 유성음 44.0%, 무성음 13.1%, 무음 42.9%로 구성된 총 230초의 깨끗한 음성이 사용되었고,테스트에는 총 220초의 음성이 사용되었다. 실제로 신뢰성 있는 결과 도출을 위해 훈련에 사용된 데이터는 테스트에 사용되지 않았고, 두 시스템의 실제 성능을 판단하기 위해서 20 ms 마다 유성음 (2), 무성음 (1), 무음 (0)으로 수동으로 표시한 메뉴얼을 만들었다. 잡음 환경은 car, street, office, white를 사용하였으며 SNR을 5, 10, 15, 20 dB로 부과하였다.

먼저, SMV와 제안된 방법의 유/무성음 분류 성능을 비교하기 위해서, 유/무성음 검출 확률 (P_d) 실험을 하였다.

표 2는 SMV와 제안된 알고리즘에서 실제 유성음을 유성음이라고 판단한 유성음 검출 확률 (P_v)과 무성음을 무성음이라고 판단한 무성음 검출 확률 (P_u)을 보여주고 있다. 실험 결과 SMV 유/무성음 분류 알고리즘

표 2. SMV와 제안된 알고리즘의 유/무성음 분류 성능 비교

Table 2. Comparison of voiced/unvoiced detection probability P_d between the method of the SMV and the proposed SMV technique.

environments		SMV		proposed SMV	
noise	SNR (dB)	V	UV	V	UV
clean	∞	0.85	0.80	0.95	0.93
car	5	0.81	0.90	0.95	0.81
	10	0.84	0.85	0.95	0.89
	15	0.85	0.79	0.95	0.91
	20	0.86	0.76	0.96	0.90
street	5	0.67	0.46	0.93	0.82
	10	0.77	0.55	0.94	0.87
	15	0.83	0.61	0.94	0.89
	20	0.85	0.65	0.95	0.89
office	5	0.49	0.53	0.88	0.75
	10	0.68	0.59	0.89	0.89
	15	0.79	0.66	0.92	0.91
	20	0.84	0.66	0.94	0.91
white	5	0.57	0.16	0.87	0.93
	10	0.73	0.31	0.89	0.93
	15	0.78	0.42	0.90	0.93
	20	0.85	0.55	0.95	0.91

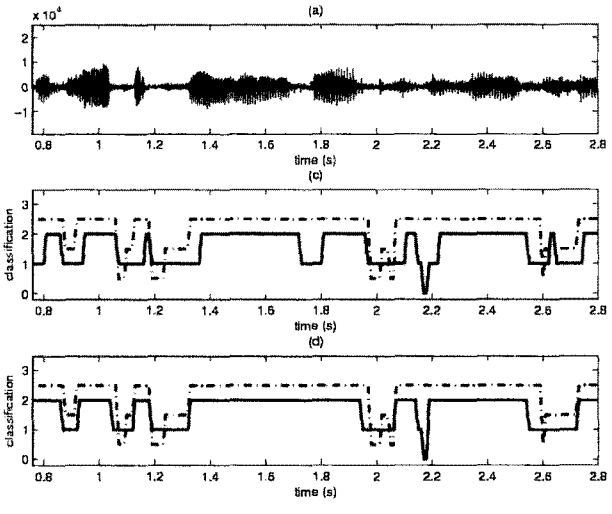

그림 6. (1) 테스트 파일의 파형 (2) SMV의 유/무성음 분류 결과 (3) 제안된 알고리즘의 유/무성음 분류 결과

Fig. 6. (1) Waveform of the test file (2) Results of the SMV (3) Results of the proposed SMV method.

의 경우 car 노이즈 같은 특수한 노이즈 환경에서 다소 우수한 성능을 보여주는 반면 제안된 알고리즘은 대체적으로 우수한 성능을 보여주는 것을 볼 수 있다. 특히, office 와 white 잡음 환경에서 매우 향상된 성능을 보였다.

그림 6은 15dB SNR을 갖는 office 잡음 환경에서 SMV와 제안된 알고리즘의 유/무성음 분류 결과를 시간축 상에서 테스트 파일의 메뉴얼과 비교하여 도시하였다. 그림 6(2)와 6(3)의 점선은 테스트 파일의 메뉴얼을 나타내고 유성음 (2.5), 무성음 (1.5), 무음 (0.5)를 나타낸다. 실험 결과로 부터 특징 벡터의 통계적 분포 특성을 이용한 제안된 실시간 GMM 기반의 유/무성음분류 기법이 더 우수함을 검증할 수 있었다.

V. 결 론

본 논문에서는 ETSI의 3GPP2 표준 코덱인 SMV의 실시간 유/무성음 성능을 향상시키기 위해 GMM 기반의 유/무성음 분류 방법을 제안하였고, 계산량을 줄이기 위해서 기존의 SMV에서 사용되는 특징 벡터만을 효과적으로 이용하여 GMM의 특징 벡터로 사용하였다. 다양한 잡음 환경에서 기존 SMV의 유/무성음 분류 성능과 비교한 결과 GMM을 이용한 제안된 방법이 유/무성음 분류에서 향상된 성능을 보여주었다.

참 고 문 헌

[1] 3GPP2 Spec., "Source-controlled variable-rate multimedia wideband speech codec (VMR-WB), service option 62 and 63 for spread spectrum systems," 3GPP2-C.S0052-A, vol. 1.0, Apr. 2005.

[2] Y. Gao, E. Shlomot, A. Benyassine, J. hyssen, Huan-yu Su, and C. Murgia, "The SMV Algorithm Selected by TIA and 3GPP2 for CDMA Applications," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 709-712, May 2001.

[3] J. -H. Chang, N. S. Kim, and S. K. Mitra, "A statistical model-based V/UV decision under background noise environments," *IEICE Trans. Inf. & Syst., vol. E87-D*, no. 12, pp.2885-2887, Dec. 2004.

[4] S. Ahmadi and A. S. Spanias, "Cepstrum-based pitch detection using a new statistical V/UV classification algorithm," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 3, May 1999.

[5] B. Atal and L. R. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with application to speech recognition," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-24, pp. 201-212, Jun. 1976.

[6] L. Siegel, "A procedure for using pattern classification techniques to obtain a voiced/unvoiced/ classifier," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP_27, pp. 83-88, Jun. 1976.

[7] L. R. Rabiner and M. R. Sambur, "Application of an LPC Distance Measure to the Voiced-Unvoiced -Silence Detection Problem," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-25, no. 4, pp. 339-343, Aug. 1977.

[8] S. C. Greer, and A. Dejaco, "Standardization of the selectable mode vocoder," Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 953-956, May 2001.

[9] C. V. Goudar, P. Rabha, M. Deshpande and A. Rao, "SMVLite: Reduced Complexity Selectable Mode Vocoder," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 701-704, May 2006.

[10] 3GPP2 Spec., "Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems," 3GPP2- C.S0030-0, v3.0, Jan. 2004.

[11] P. Vary and R. Martin, Digital Speech Transmission : enhancement, coding and error concealment, pp.182-187, 2006.

[12] G. Xuan, W. Zhang and P. Chai, "EM algorithm of gaussian mixture model and hidden Markov model," *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 145-148, Oct. 2001.

[13] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19-41, Jan. 2000.

저 자 소 개

송 지 현(학생회원)
2007년 인하대학교 전자전기 공학부 학사 .
2007년 ~ 현재 인하대학교 전자공학과 석사과정
<주관심분야 : 디지털신호처리>

장 준 혁(정회원)
1998년 경북대하교 전자공학과 학사.
2000년 서울대학교 전기공학부 석사.
2004년 서울대학교 전기컴퓨터공학부 박사.
2000년 ~ 2005년 (주)넷더스 연구소장
2004년 ~ 2005년 캘리포니아 주립대학, 산타바바라(UCSB) 박사후연구원
2005년 한국과학기술연구원(KIST) 연구원
2005년 ~ 현재 인하대학교 전자공학부 조교수
<주관심분야 : 음성 신호처리, 오디오 신호처리, 통신 신호처리, 휴먼/컴퓨터 인터페이스>