

지각 오디오 부호화기에서의 스펙트럼 에너지 기반 톤 성분 검출 알고리즘

정희원 이근섭*, 연규철**, 박영철***, 윤대희*

Tonality Detection based on Spectrum Energy in Perceptual Audio Coder

Keun-Sup Lee*, Kyu-Chel Yeon**, Young-Cheol Park***, Dae Hee Youn* *Regular Members*

요 약

지각 오디오 부호화기는 신호의 지각적 중복성과 함께 지각에 무관한 성분들을 줄이기 위해 인간의 청각 특성인 마스킹 효과를 이용하여 부호화 과정에서 발생하는 양자화 잡음을 귀에 들리지 않는 수준 이하로 낮춘다. 이때 마스킹 하는 주체인 입력 신호의 특성이 톤, 잡음 중 어떤 성분에 가까운지에 따라 주위의 작은 신호들을 마스킹하는 양이 달라지기 때문에 입력 신호의 특성을 알아내는 것은 지각 부호화기의 성능을 좌우하는 중요한 요소이다. 본 논문에서는 낮은 복잡도로 구현 가능한 새로운 톤 성분 검출 알고리즘을 제안하였다. 제안한 알고리즘들은 MPEG 심리음향 모델-II에 비해 초월함수가 적게 사용되며 사용되는 연산도 단순하기 때문에 낮은 복잡도의 부호화기 구현에 적합하다. 제안한 알고리즘은 다양한 신호들에 대해 성능평가가 수행되었으며, DSP를 사용한 구현 결과 약 3 MIPS 정도로 구현 가능하였다.

Key Words : perceptual audio; tonality; masking effect; psychoacoustic model; spectrum energy.

ABSTRACT

The goal of perceptual audio coder is to reduce redundancy and irrelevancy of audio signal based on the concept of masking. Several studies on masking effect reveal that the masking threshold varies as a function of the noise-like or tone-like nature of audio signals. Therefore, tonality of audio signal influences significantly the quality and efficiency of perceptual audio coder. In this paper, we propose a new effective algorithm for tonality measure using spectrum energy. Since the proposed algorithm consists of a few transcendental functions and simple operations, it has lower complexity than MPEG psychoacoustic model-II. The proposed algorithm was tested with some audio signals, and DSP implementation showed that the proposed algorithm could be implemented with 3 MIPS. These results illustrate the efficiency of proposed algorithm in both performance and complexity.

1. 서 론

지각 오디오 부호화기는 신호의 지각적 중복성과 함께 지각에 무관한 성분들을 줄이기 위해 인간의 청각특성을 반영한다. 지각 부호화기에 이용되는 대

표적인 청각 특성이 마스킹 효과이다. 마스킹 효과란 하나의 성분이 다른 성분에 의해 들리지 않는 현상을 말하는데, 주로 두 성분의 크기, 주파수, 시간적, 공간적 관계 등에 따라 이러한 현상이 나타난다^[1]. 마스킹 효과는 크게 주파수 영역의 순시 마스

* 연세대학교 전기전자공학과 미디어 통신 신호처리 연구실(taraji@cyclon.yonsei.ac.kr, dhyoun@yonsei.ac.kr),

** LG전자 UMTS 단말연구소(scaee@lge.com), *** 연세대학교 정보기술학부(young00@dragon.yonsei.ac.kr)

논문번호 : 040123-0317, 접수일자 : 2004년 3월 17일

킹과 시간 영역의 동시 마스킹으로 나뉘는데, 일반적으로 지각에 무관한 성분들을 줄이는 데에는 순시 마스킹이 좀더 중요한 역할을 한다^[2]. 따라서 많은 지각 오디오 부호화기들이 주파수 영역에서 부호화를 수행한다.

지각 부호화기의 목표는 마스킹 효과를 이용하여 부호화 과정에서 발생하는 양자화 잡음을 귀에 들리지 않는 수준 이하로 낮추는 것이다. 이때 마스킹하는 주체는 부호화 대상인 입력 신호이며, 입력 신호에 의해 마스킹 되는 것은 양자화 잡음이다. 이렇게 귀에 들리는 수준과 들리지 않는 수준의 경계에 해당하는 것을 마스킹 임계치라고 부르며, 이것은 마스킹 하는 주체인 입력신호의 특성이 잡음 성분에 가까우지, 톤 성분에 가까우지에 따라 값이 변한다. 일반적으로 잡음 성분이 톤 성분에 비해 좀더 큰 주변 신호들을 마스킹 할 수 있기 때문에, 잡음 성분이 많은 신호의 마스킹 임계치가 톤 성분이 많은 신호에 비해 좀더 높아진다^[3]. 따라서 입력 신호의 특성을 정확하게 알아내는 것은 마스킹 임계치를 계산하는 데에, 그리고 나아가서는 부호화기의 성능에 중요한 영향을 미친다.

본 논문은 스펙트럼 에너지를 사용하여 낮은 복잡도로 오디오 신호에서 톤 성분을 찾아내는 방법을 제안하였다. 기존에 사용되었던 MPEG 심리음향 모델-II의 알고리즘은 실시간 구현이 어려울 정도의 복잡한 연산을 요구하며, 보다 낮은 복잡도의 심리음향 모델-I 알고리즘은 톤 성분의 위치와 크기를 정확하고 자세하게 얻기 힘들었다. 본 논문에서 제안하는 알고리즘은 좁은 주파수 대역에 에너지가 집중되는 톤 성분의 특징을 이용하여 주변 에너지와의 에너지 차이를 톤 성분 검출에 사용하였다. 제안된 알고리즘은 보다 간단한 연산으로 심리음향 모델-II 만큼의 자세한 심리음향 정보를 얻을 수 있다. 논문의 구성은 다음과 같다. 2장에서는 기존의 오디오 부호화기에서 사용되고있는 톤 검출 알고리즘들에 대해 알아보고, 3장에서는 본 논문에서 제안한 톤 성분 검출 알고리즘에 대해 설명하였다. 4장에서는 제안한 알고리즘의 성능을 몇 개의 실험 결과를 통해 보였으며, 마지막으로 5장에서 결론을 맺는다.

II. 기존의 톤 성분 검출 알고리즘

어떤 신호 성분이 다른 성분을 마스킹하는 정도는 그 신호 성분의 종류에 따라 다른데, 일반적으로

협대역 잡음 신호가 톤 성분을 마스킹 하는 "잡음 마스킹-톤 임계치"가 "톤 마스킹-잡음 임계치"에 비해 더 높으며, 그 정도는 주파수 대역에 따라 다르다^[3]. 따라서 잡음 성분의 주파수 대역에서는 톤 성분의 대역에 비해 더 많은 양의 신호가 마스킹 되고, 결과적으로 그 대역은 상대적으로 더 많은 양자화 잡음이 허용된다. 그러므로 톤과 잡음 성분의 추정 정확도가 낮으면 마스킹 임계치가 부정확해 지므로, 이 마스킹 임계치에 따라 부호화를 수행하는 지각 오디오 부호화기의 성능이 저하된다.

입력 신호가 톤과 잡음 중 어떤 쪽에 가까우지를 결정하기 위해, 스펙트럼의 편평도를 나타내는 SFM (Spectrum Flatness Measure)^[4]를 사용할 수 있다. SFM은 파워 스펙트럼의 기하 평균값(Gm)을 산술 평균값(Am)으로 나눈 값으로 일반적으로 dB 값으로 계산된다. 기하평균값은 항상 산술 평균값보다 같거나 작으므로 SFM 값은 항상 0 보다 같거나 작다. 두 평균값이 같은 경우는 모든 데이터 값이 같은, 즉 가장 편평할 때이므로 SFM 값이 클수록 전 대역에 에너지가 고르게 퍼져있는 잡음에 가까운 상태를 나타내게 된다. 그러나 SFM 값은 각 주파수가 아닌 전 대역에 걸친 스펙트럼에 대해 하나의 값으로 얻어지므로 특정 성분들을 간과하는 단점이 있다. 따라서 좀더 정확한 마스킹 임계치를 얻기 위해서는 각 주파수 대역 단위로 톤 성분을 추정할 수 있어야 한다.

최근 지각 오디오 부호화기들 중 가장 많이 사용되고 있는 MPEG 오디오는 복잡도에 따라 두가지 형태의 심리음향 모델을 제시하고 있다. 두 모델 중 상대적으로 높은 정확도와 복잡도를 갖는 MPEG 심리음향 모델-II는 각 스펙트럼의 크기와 위상에 대한 예측을 통해 톤 성분을 찾는 방법을 사용한다^{[5][6]}. 각 스펙트럼에 대한 예측 값들과 원래 값들 사이의 유클리디안 거리가 구해지면 이 값은 다음과 같은 과정을 거쳐 비예측성(c_{ω})이란 값으로 변환된다. 아래 식 (1)에서 r_{ω} 와 f_{ω} 는 각각 스펙트럼의 크기와 위상을 나타내며, \hat{r}_{ω} 와 \hat{f}_{ω} 은 과거의 신호들로부터 예측을 통해 얻은 값이다.

$$c_{\omega} = \frac{\left((r_{\omega} \cos f_{\omega} - \hat{r}_{\omega} \cos \hat{f}_{\omega})^2 + (r_{\omega} \sin f_{\omega} - \hat{r}_{\omega} \sin \hat{f}_{\omega})^2 \right)^{0.5}}{r_{\omega} + \text{abs}(\hat{f}_{\omega})} \quad (1)$$

비예측성은 간단한 로그 변환을 통해 톤 지수로 변환되며, 이 톤 지수가 톤 마스킹과 잡음 마스킹

사이의 보간의 역할을 하게 된다. 이 방법은 각 스펙트럼 단위로 톤 성분을 좀더 정확하게 검출할 수 있다는 장점이 있지만, 초월 함수와 나눗셈 등의 복잡한 연산으로 인해 구현 시 복잡도가 높다. 또한 예측을 위해 하나의 오디오 프레임을 몇 개의 서브 프레임으로 나누어 처리하기 때문에 주파수 해상도가 낮아지며, 음성 신호에서 자주 발생하는 아주 짧은 시간 동안 존재하는 톤 성분의 경우 예측에 실패할 수 있는 단점이 있다⁶⁾.

상대적으로 복잡도가 낮은 MPEG 심리음향 모델-I은 스펙트럼의 지역적 최대값을 찾는 방법을 사용한다. 톤 성분은 스펙트럼 상에서 대부분 좁은 대역에 집중되기 때문에, 심리음향 모델-I은 주위의 스펙트럼들에 비해 7 dB 이상 큰 지역적 최대값을 찾아서 그 스펙트럼을 톤 성분으로 결정하고 나머지 스펙트럼 성분들은 모두 잡음 성분으로 분류한다⁷⁾. 심리음향 모델-I은 비교적 간단한 연산으로 구현이 가능하지만, 위의 두 방법과는 달리 각 스펙트럼을 톤 성분 또는 잡음 성분으로 이분법적 분류를 하므로 정확도가 낮다. 또한, 특정 주파수 스펙트럼의 하나의 오디오 프레임 내의 일시적인 에너지 증가를 톤으로 오인할 가능성도 있으므로 마스킹 임계치의 정확도가 떨어지는 단점이 있다.

III. 제안한 톤 성분 검출 알고리즘

제안한 톤 성분 검출방법은 MPEG 심리음향 모델-I처럼 입력 신호의 데시벨 에너지를 연산하고 주변 주파수의 데시벨 에너지와의 차이를 계산한다. 그러나 톤과 잡음 성분에 대해 이분법적 분류를 하는 MPEG 심리음향 모델-I과는 달리 연속적인 값의 톤 지수를 얻을 수 있는 새로운 함수를 정의하였다. 톤 성분은 좁은 주파수 대역에 에너지가 집중되기 때문에 주변의 에너지에 비해 더 높은 에너지를 갖는다. 따라서 톤 성분이 강할수록 주변 에너지와의 차이가 더 커진다고 가정하고, 이를 나타내기 위한 *Peak_factor*를 k번째 주파수 인덱스에 대해 다음과 같이 정의하였다.

$$\begin{aligned}
 Peak_factor(k) &= Energy_{dB}(k) - avg_Energy_{dB}(k) \\
 Energy_{dB}(k) &= 10 \times \log_{10} Energy(k) \\
 avg_Energy_{dB}(k) &= \frac{1}{width_k} \sum_{j \neq k} Energy_{dB}(j) \\
 & \quad (|k-j| \leq width_k/2)
 \end{aligned}
 \tag{2}$$

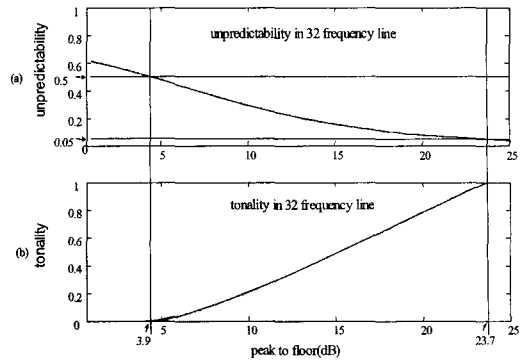


그림 1. 에너지 차이에 따른 비예측성과 톤 지수

평균 데시벨 에너지 $avg_Energy_{dB}(k)$ 는 k번째 주파수 인덱스를 중심으로 일정한 범위를 가지고 평균을 낸 값이다. 평균값을 구하는 범위 $width_k$ 는 k의 주파수에 따라 달라지며 청각의 주파수 특성을 고려하여 고주파 쪽으로 갈수록 넓어지는 임계 밴드(Critical Band) 형태를 따른다. $Peak_factor(k)$ 는 위의 식과 같이 스펙트럼의 데시벨 에너지에서 주변 주파수의 평균 에너지를 뺀 값으로 정의하였다.

식 (2)의 평균 에너지 값은 데시벨 값에 대한 산술 평균이므로 주변 스펙트럼 에너지에 대한 기하 평균과 같다. 따라서 $Peak_factor(k)$ 는 k번째 주파수 스펙트럼 에너지를 주변 평균 주파수의 기하 평균 값으로 나눈 것을 의미한다. 이 값은 산술 평균과 기하 평균 사이의 비를 나타내는 SFM과 유사한 형태이지만, 산술 평균 대신 스펙트럼 에너지와의 비를 구하기 때문에 각 스펙트럼에 대한 톤 성분을 추정할 수 있다. SFM 값이 스펙트럼의 편평도를 나타내는 수치였다면, $Peak_factor(k)$ 는 특정 k번째 주파수 스펙트럼 에너지가 주변 에너지에 비해 얼마나 큰지를 나타낸다.

위의 식 (2)를 통해서 얻은 $Peak_factor(k)$ 값은 톤 성분의 양을 나타내는 [0,1] 사이의 톤 지수 (tonality index) 값으로 변환된다. 이 과정은 $Peak_factor(k)$ 값을 MPEG 심리음향 모델-II의 비예측성 파라미터에 대응시키는 방법을 사용하였다. $Peak_factor(k)$ 값과 비예측성의 관계를 실험적으로 살펴보기 위해 44.1kHz 샘플링 주파수를 가정하여 1.3kHz 주파수의 사인파를 사용하였다. 1.3kHz 주파수는 192시간 샘플에 대해 정확히 6주기가 들어가게 되어 예측이 잘 되며 256 샘플 FFT에 의한 성능 저하도 없다.

그림 1은 1.3kHz에 대해 MPEG 심리음향 모델-II를 사용하여 현재 주파수의 데시벨 에너지와 평

균 데시벨 에너지의 차이에 따른 비예측성과 톤 지수를 측정하는 그림이다. 그림 1(a)와 같이 데시벨 에너지 차이가 커질수록 비예측성은 톤 성분을 나타내는 0에 가까운 값을 가짐을 알 수 있다. 톤 지수는 데시벨 에너지 차이가 3.9dB부터 23.7dB까지 거의 선형적인 증가를 하게 되며 데시벨 에너지 차이가 3.9dB이하에 대해서는 톤 지수가 잡음 성분을 의미하는 0을 나타내게 된다.

MPEG 심리음향 모델-II의 비예측성(c_{ω})을 근사화 방법으로는 정밀한 근사화를 위하여 고차 함수를 이용하는 방법과 선형적으로 근사화하는 방법을 들 수 있다. 2차 함수를 이용한 근사화는 식 (3)을 이용하여 보다 정밀한 근사화가 가능하다. 선형적인 근사화는 식 (4)을 이용하여 근사화 하였다.

$$\hat{c}_{\omega} = 7.5 \times 10^{-4} \times (\text{Peak_Factor} - 30)^2 + 0.001 \quad (3)$$

$$\hat{c}_{\omega} = -\frac{\text{Peak_Factor}}{27} + 0.667 \quad (4)$$

선형적인 근사화 방법이 사용 가능한 이유는 톤 지수가 마스킹 임계값에 미치는 영향의 차이 때문이다. 참고 문헌 [7]에 의하면 톤 지수가 작은 값일 때는 톤 지수의 변화에 따른 마스킹 임계값의 영향이 많지만 톤 지수 값이 커질수록 마스킹 임계값에 미치는 영향은 줄어들게 된다. 따라서 선형적인 근사화의 오차가 많이 있는 톤 지수가 큰 부분이 마스킹 임계값에 미치는 영향이 줄어들게 된다.

제안한 데시벨 에너지 차이를 이용한 톤 지수 결정 방법은 위에서 설명한 바와 같이 데시벨 에너지 차이를 사용하여 근사화된 비예측성(c_{ω})을 얻는다. 비예측성 근사화 후 MPEG 심리음향 모델-II와 같이 인접 밴드 간의 마스킹 영향을 고려하기 위해 스프레딩 함수와의 컨볼루션 과정을 거치고, 톤 지수는 스프레딩된 에너지에 스프레딩된 에너지와 신호의 비예측성이 반영된 에너지의 비를 사용하여 결정된다⁷⁾.

제안한 스펙트럼 에너지를 이용한 톤 성분 검출 방법은 MPEG 심리음향 모델-I과 달리 연속되는 값의 톤 지수와 에너지 스프레딩이 사용되어 세밀한 청각특성 반영이 가능하며, MPEG 심리음향 모델-II에 비해 연산량이 적으므로, 비교적 적은 연산량으로 전대역의 주파수에 대한 톤 지수 측정도 가능하다. 또한 프레임 간의 예측이 아닌 하나의 프레임 내에서 톤 성분을 찾을 수 있기 때문에 MPEG 심리음향 모델-II의 단점이었던, 음성신호와 같은 아주

짧은 시간동안 존재하는 톤 성분도 간과하지 않고 찾아낼 수 있다⁶⁾.

그러나 위에서 제안한 모델은 톤 성분이 아닌 특정 주파수 대역의 일시적인 에너지 증가를 톤으로 오인할 수 있는 한계가 있다. 이것은 현재 프레임의 스펙트럼만을 가지고 톤 성분을 찾기 때문이며, 이러한 경우 특정 주파수 대역의 에너지가 증가할 때 이것을 톤인지 일시적인 에너지 증가인지 구별할 수 있는 기준이 없다. MPEG 심리음향 모델-II의 경우에는 과거의 스펙트럼 데이터로부터 예측을 하여 톤 성분을 찾아내기 때문에 일시적인 에너지 증가를 톤 성분으로 결정하지 않는다. 본 논문에서 제안한 모델을 개선하기 위해서는 MPEG 심리음향 모델-II에서와 같이 과거의 데이터를 사용해야 할 필요가 있다. 본 논문에서는 연산량을 고려하여 식 (5)와 같은 간단한 평탄화(smoothing) 알고리즘을 적용하였다.

$$tb_{b,n} = \mu \times tb_{b,current} + (1 - \mu) \times tb_{b,n-1} \quad (5)$$

$tb_{b,current}$ 는 n번째 프레임에서 계산하여 얻어진 톤 지수이며, $tb_{b,n}$ 은 위의 식 (5)의 평탄화 알고리즘을 거쳐서 최종적으로 얻어진 톤 지수이다. μ 는 [0,1] 사이의 값을 갖는 평탄화 계수이며, 이 값이 1에 가까울수록 과거 프레임보다 현재 프레임에서 계산된 값을 좀더 많이 반영한다. μ 값은 톤 지수가 몇 프레임의 지연을 가지고 천천히 변하게 하지만, 지각 부호화의 특성상 통계적으로 정적인 신호 성분에 대해 더 많은 비트를 할당하는 것이 중요하기 때문에⁶⁾, 이렇게 톤 성분이 갑자기 크게 발생하는 동적인 신호 변화에 대한 톤 지수의 지연 현상은 부호화에 심각한 영향을 미치지 못한다. 마지막으로 이 평탄화 알고리즘은 일시적인 에너지 증가를 고려하지 않기 위한 것이므로 톤 지수가 과거 프레임보다 증가하는 경우에만 적용되어야 한다.

그림 2는 위의 평탄화 알고리즘을 적용시키기 전과 후의 톤 지수의 차이를 보여준다. 톤 성분이 아닌 일시적인 에너지 증가의 예를 위해 톤 성분이 존재하지 않는 백색 가우시안 잡음을 입력 신호로 사용하였다. 평탄화를 적용하기 전인 그림 2(b)의 경우 잡음의 일부 주파수 성분의 큰 에너지가 톤 성분 검출에 영향을 끼치고 있다. 그러나 0.5의 μ 값으로 평탄화를 적용한 후의 그림은 그림 2(a)에서는 그러한 영향이 상당부분 줄어들었음을 확인할 수 있다. 제안한 알고리즘은 평탄화를 사용하여

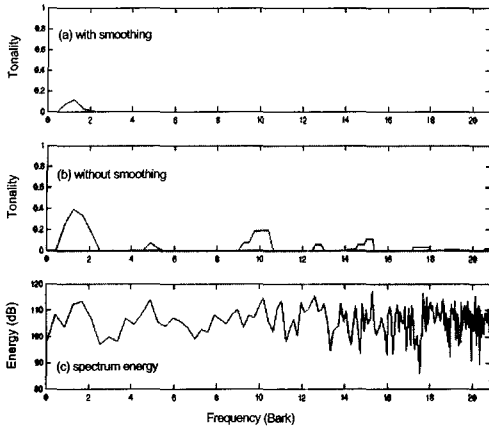


그림 2. 평탄화를 적용시키기 전과 후의 톤 지수 변화

MPEG 심리음향 모델-II의 위상 정보 예측을 대신 하기 때문에, 예측에 사용되는 초월 함수들에 의한 많은 연산량을 크게 줄일 수 있다.

IV. 실험 결과

1. 톤 성분 검출 성능 평가

제안한 알고리즘의 성능을 측정하기 위하여 사인파 신호와 여성의 음성 신호에 대해 톤 성분을 검출하였다. 먼저 실험에 사용된 사인파는 44.1 kHz의 표본화 주파수를 가지며, 1024 샘플의 FFT 주파수 해상도인 43.07Hz의 정수배가 되는 주파수들로 선택하였다. 이렇게 선택된 주파수는 86.133 Hz, 1.292 kHz, 2.584 kHz, 7.752 kHz 이다. 그림 3은 이 사인파 신호에 대해 제안한 알고리즘을 SFM, MPEG 심리음향 모델-I/II 와 비교한 그림이다. MPEG 심리음향 모델-II의 톤 지수 값은 임계치 계산 밴드 단위로 구해지므로, 각 주파수 성분 별 톤 성분을 보기 위하여 (1-c) 값을 표시하였다.

실험 결과 제안한 알고리즘은 4개의 사인파를 모두 찾아내고 있으며, MPEG 심리음향 모델-II와 유사한 성능을 보이고 있다. MPEG 심리음향 모델-II는 1.3 kHz 부근에서 톤 성분을 검출한 위치가 조금 다른데, 이것은 1.292 kHz가 심리음향 모델-II에서 사용되는 256샘플 FFT의 주파수 해상도인 172.27 Hz의 정수배가 아니기 때문이다. 실험 결과에서 나타났듯이 MPEG 심리음향 모델-II는 예측을 통한 톤 성분 검출의 정확도는 높지만 고주파 대역의 낮은 해상도로 인하여 톤 성분의 위치를 넓은 범위에서 대략적으로 얻게 되며, 경우에 따라 그

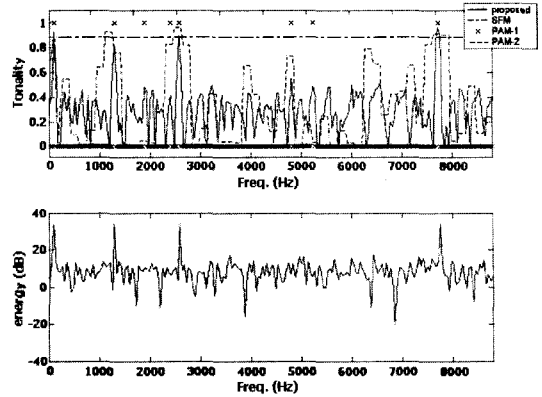


그림 3. 4개의 사인파에 대한 톤 성분 검출 실험 결과

림 3과 같이 부정확한 추정으로 톤 성분이 없는 인접 밴드에 영향을 끼쳐서 마스킹 임계치의 주파수 해상도를 낮아지게 만드는 단점도 있다. 제안된 알고리즘은 위상 정보 대신 평탄화를 사용하여 MPEG 심리음향 모델-II와 동등한 수준의 톤 성분 검출 성능을 갖으며, 또한 발견한 톤 성분에 대해서는 높은 주파수 해상도로 톤 성분의 위치를 비교적 정확하게 찾을 수 있다.

제안한 알고리즘이 실제의 오디오 신호에 대해 톤 성분을 잘 검출하는지를 확인하기 위해 사람의 음성 신호를 사용하였다. 사람의 음성은 유성음과 무성음으로 나뉘며 각각 톤과 잡음의 전형적인 특징을 갖기 때문에 톤 성분 검출 결과를 관찰하기에 용이하다. 그림 4는 여성의 목소리 신호에 대해 실험한 결과이다. 테스트 신호는 수잔 베가의 Tom's Diner 란 곡이며, 반주 없이 목소리만으로 된 음악이다. 이 곡의 처음 "I'm sitting" 하는 부분의 "m

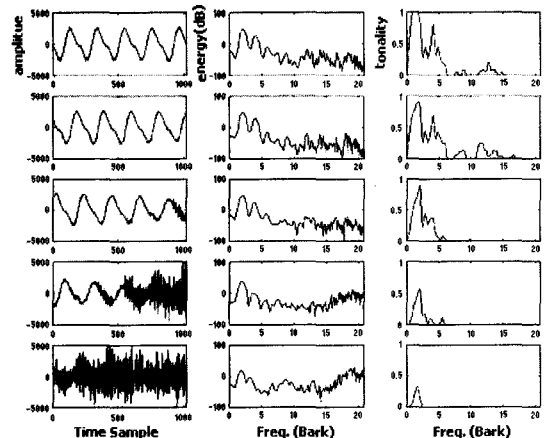


그림 4. 여성 목소리에 대한 톤 성분 검출 실험 결과

표 1. 알고리즘 별 연산량 비교 (MIPS)

심리음향모델	MPEG	Proposed
FFT	21.65	21.65
톤 지수 계산	> 32	2.88
마스킹 임계치	3.59	3.59
기타	0.07	0.07
합계	> 57.31	28.19

s" 부분을 연속된 프레임으로 나타내었다. 그림 4는 유성음에서 무성음으로 바뀌면서 잡음 성분이 많아지고 톤 지수가 점점 낮아지는 것을 보여준다. 따라서 톤 성분이 많은 대표적인 신호인 사람 목소리 신호에 대해서도 톤 성분을 잘 선택하고 있으며, 그 밖의 다른 많은 테스트 신호에 대한 성능 평가에서도 동일한 좋은 성능을 보임을 확인하였다.

2. 연산량 비교

본 절에서는 제안한 알고리즘의 연산량을 기존의 알고리즘들 중 유사한 성능의 MPEG 심리음향 모델-II와 비교하였다. MPEG 심리음향모델-II는 여러 알고리즘들 중 가장 좋은 성능을 가지고 있지만 실시간 구현에 어려운 많은 초월함수를 사용하고 있어 구현 시에 많은 연산량이 필요하다⁷⁾.

실제로 DSP를 사용하여 2채널 44.1 kHz, 128 kbps 부호화에 대한 연산량을 측정한 결과 제안한 알고리즘을 사용한 심리음향 모델은 표 1과 같이 기존의 MPEG 심리음향모델-II에 비해서 절반 이상 연산량이 줄어드는 것으로 나타났다. 연산량은 24비트의 고정소수점 DSP를 기준으로 하였으며⁸⁾, 초월함수 연산은 테이블에서 얻은 값을 선형 보간법으로 근사화하는 알고리즘을 사용하였다⁹⁾.

V. 결론

본 논문에서는 낮은 복잡도로 톤 성분을 검출할 수 있는 알고리즘을 제안하였다. 신호의 각 스펙트럼 성분이 톤과 잡음 중 어떤 것으로 이루어져있는가는 신호의 마스킹 임계치를 결정하는 중요한 요소이며, 결과적으로 지각 부호화기의 비트 할당에도 영향을 미친다. 따라서 톤 성분의 양을 정확하게 추정하는 것은 부호화의 효율을 높이는데 필수적이다.

본 논문에서는 적은 연산량으로 구현 가능한 새로운 톤 성분 검출 알고리즘을 제안하였다. 제안한 알고리즘은 MPEG 심리음향 모델-II에 비해 훨씬

적은 연산량으로 톤 성분의 양을 MPEG 심리음향 모델-II와 동등한 성능으로 나타내며, 심리음향 모델-II에 비해 높은 주파수 해상도로 톤 성분이 위치한 곳을 보다 정확하게 찾을 수 있는 장점이 있다. 복잡도를 측정하기 위한 연산량 평가에서는 제안한 알고리즘을 사용한 심리음향 모델이 MPEG 심리음향 모델-II에 비해 절반 이상 복잡도가 낮은 것으로 나타났다. 따라서 제안된 톤 성분 검출 알고리즘은 적은 하드웨어 자원으로 부호화기를 구현해야 하는 휴대용 기기와 같은 응용 분야에서 유용하게 사용될 수 있다.

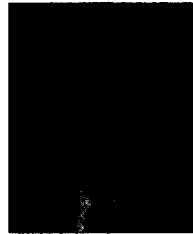
참고 문헌

- [1] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, 1990.
- [2] Nikil Jayant et al., "Signal Compression Based on Models of Human Perceptual Noise Criteria," *IEEE Journal of Selected Areas on Communication*, 6(2), Feb. 1988
- [3] R. Hellman, "Asymmetry of Masking Between Noise and Tone," *Perception and Psychoacoustics*, 11(3):241-246, 1972
- [4] James D. Johnston, "Estimation of Perceptual Entropy Using Noise Masking Criteria," *proceedings of ICASSP 1988*, pp 2524-2527
- [5] K. Brandenburg and J. D. Johnston, "Second Generation Perceptual Audio Coding: The Hybrid Coder," *88th Convention of AES*, Mar. 1990, Preprint No. 2937
- [6] Anibal J.S.Ferreira, "Tonality Detection in Perceptual Coding of Audio," *98th Convention of AES*, Feb. 1995, Preprint No. 3947
- [7] MPEG/IEC JTC1/SC29/WG11 No.71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - CD 11172-3(Part3. MPEG-Audio)"
- [8] Motorola, *DSP56300 Family Manual: 24-Bit Digital signal Processor*, Nov.2000
- [9] Keun-Sup Lee, Hyen-Oh Oh, Young-Cheol Park, Dae Hee Youn, "High Quality

MPEG-Audio Layer III Algorithm for a 16-bit DSP," *IEEE International Symposium on Circuit and Systems*, vol. II, pp.205-208, May 2001

윤 대 희(Dae Hee Youn)

정회원



1977년 2월 : 연세대학교
전자공학과 졸업
1979년 8월 : Kansas State Univ. 석사
1982년 8월 : Kansas State Univ. 박사

현재 : 연세대학교 전기전자공학과 교수
<관심분야> 디지털 신호처리, 적응 신호처리, 음성 신호처리, 오디오 신호처리

이 근 섭(Keun-Sup Lee)

정회원



1997년 2월 : 연세대학교
전자공학과 졸업
1999년 8월 : 연세대학교
전기컴퓨터공학과 석사
1999년 9월~현재 : 연세대학교
전기전자공학과 박사과정

<관심분야> 디지털 신호처리, 오디오 신호처리

연 규 철(Kyu-Chel Yeon)

정회원

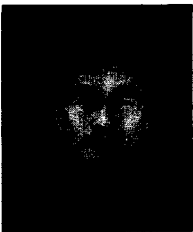


1999년 2월 : 건국대학교
전기공학과 졸업
2002년 8월 : 연세대학교
전기전자공학과 석사
2002년 9월~현재 : LG전자
UMTS 단말연구소
주임연구원

<관심분야> 디지털 신호처리, 오디오 신호처리, 통신 신호처리

박 영 철(Young-Cheol Park)

정회원



1986년 2월 : 연세대학교
전자공학과 졸업
1988년 2월 : 연세대학교
전자공학과 석사
1993년 2월 : 연세대학교
전자공학과 박사

현재 : 연세대학교 정보기술학부 교수
<관심분야> 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리