

지각적으로 가중된 매칭 퍼슈잇을 이용한 오디오 신호의 정현파 모델링

Sinusoidal Modeling of Audio Signals Using Perceptually Weighted Matching Pursuit

김 연 지*, 이 인 성*
(Yeon-Ji Kim*, In-Sung Lee*)

* 충북대학교 전자공학과

(접수일자: 2002년 11월 7일; 채택일자: 2003년 1월 9일)

본 논문은 매칭 퍼슈잇 (Matching Pursuit)에 인지적인 가중치를 사용하여 오디오 신호를 정현파 모델링하는 것에 대해 기술한다. 매칭 퍼슈잇은 입력 신호에서 에너지가 가장 큰 성분들을 반복적으로 추출해 원 신호와 재생 신호의 차이가 영에 도달할 때까지 계속된다. 본 논문에서는 매칭 퍼슈잇에 심리음향모델을 사용한 인지적인 매칭 퍼슈잇으로 입력 신호로부터 인지적으로 중요한 성분을 차례대로 추출하였다. 인지적인 매칭 퍼슈잇의 성능을 평가하기 위해 인지적인 가중치를 주지 않은 정현파 매칭 퍼슈잇과 비교 평가하였다. 여러 가지 입력 신호에 대한 실험 결과 정현파 매칭 퍼슈잇보다 인지적인 매칭 퍼슈잇의 성능이 좋았고, 특히 시간축 변화율이 큰 신호일 경우에 인지적인 매칭 퍼슈잇을 통해 원래 신호의 음질을 더 잘 복원할 수 있었음을 확인하였다.

핵심어: 매칭 퍼슈잇, 인지적인 가중치, 정현파 모델, 심리음향 모델

투고분야: 음향 신호처리 분야 (1.2), 음악음향 및 심리음향 분야 (8.6)

This paper describes a method for sinusoidal modeling of audio signals using perceptually weighted matching pursuit. Matching pursuits extracts iteratively the greatest energy signals from the input signals until the residual between the original and the reconstructed signal is zero. In this paper, perceptual matching pursuits using psychoacoustic model to matching pursuit extracts greatest perceived energy iteratively. To evaluate the performance of the perceptual matching pursuits it is compared with the sinusoidal matching pursuits which is not included perceptual weighting. For various audio signals the result of simulation shows that the perceptual matching pursuit is superior to the sinusoidal matching pursuits, especially for a high change rate in time domain it can synthesized original signal.

Keywords: Matching pursuits, Perceptual weighting, Sinusoidal modeling, Psychoacoustic model

ASK subject classification: Acoustic signal processing (1.2), Musical acoustics and psychoacoustics (8.6)

I. 서론

정현파 모델은 오디오 신호를 주기적인 순음 성분들의 합으로 표현하는 방법으로 초기에 음성 신호용과 컴퓨터 음악 등에 활용되어 왔으나, 배경 잡음과 바음성 신호에 강한 특성으로 인해 오디오 신호의 분석, 합성, 부호화에도 활용되기 시작했다[1]. 정현파 모델링은 신호를 사인파의 합으로 표현할 수 있다는 이론으로 입력 오디오

신호에 STFT (Short-Time Fourier Transform)을 취하여 분석되는 값은 신호에 포함된 실제 정현파 성분의 주파수와 진폭으로 모델링함으로써 시간축 변화나 주파수 변화와 같은 신호 변형이 쉽고, 우수한 음질을 얻을 수 있다는 방법으로 알려져 있다[2]. 일반적으로 오디오 신호 중의 순음 성분은 STFT 스펙트럼의 피크로 해석된다. 입력 신호의 스펙트럼에는 많은 피크들이 있는데, 정현파 모델을 적용하기 위한 순수한 음의 피크만을 검출하는 것은 중요하고도 어려운 일이다. 스펙트럼 피크들 중에는 큰 잡음의 피크도 포함되어 있기 때문이다. 잡음의 피크를 순음의 피크로 잘못 검출하여 정현파 모델링하

책임 저자: 김연지 (yeon-ji@hotmail.com)
361-763 충북 청주시 흥덕구 개신동 산 48
충북대학교 전자공학과
(전화: 043-262-8477; 팩스: 043-271-4647)

면 합성되는 신호는 원음과 달라진다. 이는 전체적인 합성 오디오 신호의 음질을 나빠지게 하는 큰 원인이 된다. 본 논문에서는 정현파 추출을 위한 스펙트럼 분석 도구로 매칭 퍼슈잇을 사용한다. 매칭 퍼슈잇은 입력 신호로부터 에너지가 가장 큰 성분들을 반복적으로 추출하여 보다 적응적으로 입력 신호의 순음을 검출할 수 있다. 매칭 퍼슈잇의 반복 과정은 원 신호와 재생신호의 차인 잔여 신호의 에너지가 영이나 어떤 정지 기준에 가까워질 때까지 수행된다[3]. 본 논문에서는 모의 실험을 통해 매칭 퍼슈잇에 심리음향모델을 적용한 경우와 적용하지 않은 경우 각각을 비교하였다.

II. 심리 음향

오디오 부호화 기술에 있어서 가장 기본이 되는 것은 사람의 청각 특성이다. 사람의 청각은 신호의 진원지가 어느 수준 이하가 되면 듣지 못한다. 개인적인 차이는 있지만 대부분 2~5 kHz 사이에서 청각은 가장 민감하게 반응한다. 심리음향모델은 큰 값을 갖는 주파수 성분에 대하여 상대적으로 작은 인접한 주파수 성분을 감지하지 못하는 청각상의 특성을 고려한 이론이다. 이러한 심리음향적 특성은 현재 오디오 부호화에 널리 활용되고 있다[4].

2.1. 절대음압

절대음압은 잡음이 없는 환경에서 청취자에 의해 검출될 수 있는 에너지의 크기로 나타내며, 이 레벨 아래의 신호음들은 들리지 않는다. 절대음압은 dB SPL (Sound Pressure Level)로 표현된다. 그림 1은 절대음압의 크기

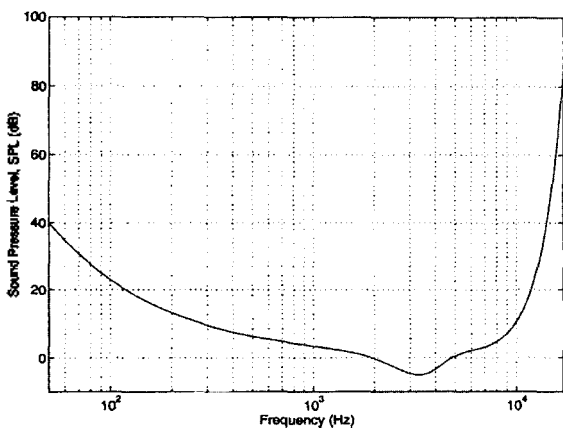


그림 1. 절대음압
Fig. 1. Absolute threshold of hearing in quiet.

를 보여준다. 그림에서 보듯이 절대음압의 크기, SPL이 낮은 부분의 주파수에서 청각은 민감하게 반응한다.

2.2. 주파수 마스킹 효과

낮은 레벨의 신호가 동시에 발생하는 강한 레벨의 신호로 인해 들리지 않게 되는 주파수 상에서의 현상을 마스킹 효과라 한다[5]. 이때 강한 레벨의 신호인 마스킹하는 성분을 마스커(masker)라 하고, 들리지 않게 되는 신호를 마스크(maskee)라 한다.

2.3. 심리음향모델-1의 마스킹 임계치 계산 절차

본 연구에서 사용된 심리음향모델-1의 마스킹 임계치는 다음과 같은 과정을 거쳐 입력 신호의 특성에 따라 프레임별로 계산된다[6].

1. 입력 신호의 주파수 변환 및 음압 산출.
2. 절대 음압 고려.
3. 잡음과 순음의 마스커 위치 판별.
4. 주요 마스커만을 선택.
5. 개별적인 순음/잡음 마스킹 임계치 계산.
6. 전체적인 마스킹 임계치의 계산.

III. 매칭 퍼슈잇

매칭 퍼슈잇은 식 (1)에서처럼 M 개의 성분들로 구성된 사전(D)으로부터 현재 신호와 가장 잘 정합되는 오름을 찾아내고, 그 오름으로 합성한 신호를 현재 신호로부터 빼내어 잔여 신호를 만들어내며 잔여 신호의 에너지가 영에 수렴될 때까지 계속된다[3].

$$D = \{g_m\}; m = 0, 1, \dots, M-1 \quad (1)$$

현재 신호와 가장 잘 정합되는 요소를 사전에서 찾아내는 것은 사전의 모든 요소들과 현재 신호의 상관계수 계산(내적)을 통해 얻어진다. 각각의 반복과정에서는 오름의 인덱스와 그에 따른 신호에 대한 오름의 기여도를 나타내는 값인 상관계수 항목이 계산된다. 예를 들어 반복 과정이 k 번째 일 경우, 현재 신호 r_k 로부터 사전에서의 k 번째 인덱스 m_k 와 그에 따른 상관계수 항목 a_k 가 계산되면 잔여 신호 r_{k+1} 를 구할 수 있다. 이러한 과정은 식 (2)와 같이 표현되고, $k=0$ 일 때 잔여 신호 r_0 는 입력 신호 x 와 같다.

$$r_{k+1} = r_k - \alpha_k g_{m_k} \quad (2)$$

매칭 퍼슈잇에서 이러한 반복 과정을 무한히 할 때 잔여 신호는 영에 근접하게 되어 매칭 퍼슈잇을 통한 합성음은 원래 신호에 근사화하게 된다. 일반적으로 유한한 K 번째까지의 반복 과정으로 추출된 항목들로부터 합성되는 신호는 다음과 같다.

$$\mathbf{x} \approx \hat{\mathbf{x}} = \sum_{k=0}^{K-1} \alpha_k \mathbf{g}_{m_k} \quad (3)$$

IV. 정현파 매칭 퍼슈잇

정현파 매칭 퍼슈잇은 일반적인 매칭 퍼슈잇과 마찬가지로 입력 신호로부터 에너지가 가장 큰 성분들을 차례대로 추출해 내어 잔여 신호를 만들고 그 잔여 신호로부터 계속해서 중요 성분들을 추출해 낸다. 정현파 매칭 퍼슈잇의 사전은 식 (4)와 같이 복소 지수들로 구성된다.

$$\mathbf{g}_m = \{g_m[n] = \frac{1}{N} e^{j2\pi \frac{m}{M} n} ; n=0,1,\dots,N-1\} ; m=0,1,\dots,M-1 \quad (4)$$

사전의 부분 공간은 두 개의 사전 요소들로 구성되는데 사전의 요소와 그것의 복소 공액 요소이다. 정현파 매칭 퍼슈잇의 파라미터들은 현재 신호와 사전과의 내적을 통해 찾아낸다. 이렇게 해서 k번째 반복 과정에서의 잔여 신호는 식 (5)와 같다.

$$\begin{aligned} r_{k+1}[n] &= r_k[n] - \alpha_k g_{m_k} - \alpha_k^* g_{m_k}^* \\ &= r_k[n] - \frac{2\alpha_k}{N} \cos\left[2\pi \frac{m_k}{M} n + \theta_k\right] \end{aligned} \quad (5)$$

이러한 사전의 성분들로 인해 정현파 매칭 퍼슈잇은 잔여 신호를 현재 신호로부터 가장 큰 에너지를 가지는 성분, 즉 정현파들을 추출해 내는 과정으로 나타내어진다.

V. 인지적인 매칭 퍼슈잇

인지적인 매칭 퍼슈잇이란 일반적인 매칭 퍼슈잇처럼 단순히 에너지가 큰 성분을 반복적으로 빼내는 것이 아닌, 정현파 모델링 매칭 퍼슈잇에 심리음향모델을 적용해 사전에 인지적인 가중치를 주어 인지적으로 중요도

가 높은 성분을 차례대로 추출해 내는 과정이다. 인지적인 매칭 퍼슈잇을 만들기 위해서는 매칭 퍼슈잇에 두 가지 항목 γ 와 \mathbf{W} 가 첨가된다[7]. 여기서 $\gamma[m]$ 은 프레임 별로 나누어진 입력 신호에 따라 구해지는 마스킹 임계치 값으로 반드시 실수이어야 하며, 모든 m값에 대해 영의 값을 가져서는 안된다.

$$\mathbf{g}_m = \{g_m[n] = \gamma[m] e^{j2\pi \frac{m}{M} n} ; n=0,1,\dots,N-1\} ; m=0,1,\dots,M-1 \quad (6)$$

또한 식 (7)에서와 같이 내적 계산시 가중된 내적으로 일반화하여 계산하는데, 이때 \mathbf{W} 는 윈도우이다.

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{W}} = \mathbf{y}^* \mathbf{W} \mathbf{x} \quad (7)$$

k번째 인덱스 m_k 와 상관계수 α_k 는 식 (8)과 식 (9)에 의해 계산된다.

$$m_k = \max_m^{-1} \frac{|\langle \mathbf{g}_m, r_k \rangle_{\mathbf{W}}|}{\langle \mathbf{g}_m, \mathbf{g}_m \rangle_{\mathbf{W}}} \quad (8)$$

$$\alpha_k = \frac{\langle \mathbf{g}_{m_k}, r_k \rangle_{\mathbf{W}}}{\langle \mathbf{g}_{m_k}, \mathbf{g}_{m_k} \rangle_{\mathbf{W}}} \quad (9)$$

5.1. 이산 푸리에 변환 구현

각각의 반복 과정에서 파라미터들은 상관계수 계산을 통해 얻어진다. 사전 내에서의 현재 신호와 가장 잘 정합되는 오트는 현재 신호 r_k 와 사전의 모든 요소들과의 상관계수 계산을 통해 그 값이 가장 클 때를 찾아낸다. 이렇게 찾아낸 오트의 위치가 인덱스 m_k 가 되고, 상관계수 값이 기여도 α_k 이다. 인지적인 매칭 퍼슈잇에서 상관계수 계산은 두 항목 γ 와 \mathbf{W} 의 선택으로 인해 이산 푸리에 변환(DFT: Discrete Fourier Transform)의 형태를 가지게 됨으로써 쉽게 구현이 가능하다. 인덱스 m_k 는 식 (10)에서 결과 값이 가장 클 때의 인덱스 m 을 찾는다.

$$\begin{aligned} \frac{|\langle \mathbf{g}_m, r_k \rangle_{\mathbf{W}}|}{\langle \mathbf{g}_m, \mathbf{g}_m \rangle_{\mathbf{W}}} &= \frac{|\mathbf{g}_m^* \mathbf{W} r_k|}{\mathbf{g}_m^* \mathbf{W} \mathbf{g}_m} \\ &= \frac{\left| \sum_{n=0}^{N-1} \gamma^*[m] e^{-j2\pi \frac{m}{M} n} \omega[n] r_k[n] \right|}{\sum_{n=0}^{N-1} \gamma^*[m] e^{-j2\pi \frac{m}{M} n} \omega[n] \gamma[m] e^{j2\pi \frac{m}{M} n}} \\ &= \frac{\left| \sum_{n=0}^{M-1} \omega[n] r_k[n] e^{-j2\pi \frac{m}{M} n} \right|}{\gamma[m] \sum_{n=0}^{M-1} \omega[n]} \\ &= \frac{|R_k^w[\frac{m}{M}]|}{\gamma[m] W[\frac{m}{M}]} \end{aligned} \quad (10)$$

이런 과정을 거쳐 찾은 인덱스 m_k 를 식 (11)에 대입해 상관계수 a_k 를 계산해 낼 수 있다.

$$\alpha_k = \frac{R_k^*[\frac{m_k}{M}]}{\gamma[m_k]W[\frac{0}{M}]} = \frac{a_k e^{j\theta_k}}{\gamma[m_k]W[\frac{0}{M}]} \quad (11)$$

식 (10)에서 분모항목의 $\gamma[m]$ 은 입력 신호의 마스크 임계치이고, 분자항목은 현재 신호 $r_k[n]$ 의 M -point DFT 취해진 값이다. 따라서 $k=0$ 일 때 식 (10)은 입력 신호의 신호 대 마스크 비(SMR; Signal to Mask Ratio)가 된다. SMR의 값이 1보다 작다는 것은 인지적으로 부적절하다는 것을 나타내는 반면, SMR이 최대가 될 때는 신호의 성분이 심리 음향적으로 아주 중요하다는 것을 의미한다. 이런 특성을 이용하여 본 논문에서는 잔여 신호로부터 인지적으로 중요도가 높은 성분들을 차례대로 추출해 내어 SMR의 값이 1보다 작아질 때까지 매칭 퍼슈잇을 수행하였다.

각각의 반복 과정을 통해 계산된 파라미터들로 신호를 합성할 수 있다. 식 (12)는 K 번의 반복 과정으로 합성되는 i 번째 프레임의 재생신호를 나타낸다.

$$\begin{aligned} \hat{x}_i[n] &= \sum_{k=0}^{K-1} (\alpha_k g_m[n] + \alpha_k^* g_m^*[n]) \\ &= \sum_{k=0}^{K-1} \frac{2a_k}{W[\frac{0}{M}]} \cos \left[2\pi \frac{m_k}{M} n + \theta_k \right] \end{aligned} \quad (12)$$

일반적인 매칭 퍼슈잇의 파라미터인 인덱스 m_k 와 가중치 a_k 대신, 인지적인 매칭 퍼슈잇에서는 이들 값으로 유도되는 크기, 주파수, 위상의 세 가지 파라미터를 이용해 정현파 모델링하여 오디오 신호를 재생한다.

$$\{A_k = 2a_k / W[0/M], f_k = m_k / M, \theta_k = \theta_k\} \quad (13)$$

식 (12)에서 신호의 복원을 위해 사용된 사전은 심리음향적인 가중치 요소를 포함시키지 않은 식 (14)와 같다.

$$\tilde{g}_m[n] = e^{j2\pi \frac{m}{M} n} \quad (14)$$

신호를 분석할 때와는 달리 재생시에는 사전에 심리음향적인 가중치를 포함시키지 않음으로써, 프레임마다 계산되는 마스크 임계치로 인한 오버헤드를 없앨 수 있다. k 번째 매칭 퍼슈잇의 결과가 정지조건에 도달되면 반복 과정을 종료하지만, 도달되지 않았을 경우 잔여 신호

r_{k+1} 에서 인지적인 중요도가 높은 성분을 계속해서 추출해 낸다. k 번째 반복 과정에서와 마찬가지로 $k+1$ 번

째 반복 과정에서의 파라미터들은 잔여 신호 r_{k+1} 과 사전의 성분들과의 직접적인 내적으로 계산된다. 그러나 식 (15)에서 보는 바와 같이 전 단계의 반복 과정인 k 번째에서의 항목들을 사용함으로써 $k+1$ 번째 반복 과정에서의 계산량을 줄일 수 있다.

$$\frac{\langle g_m, r_{k+1} \rangle}{\langle g_m, g_m \rangle} = \frac{\langle g_m, r_k \rangle - \alpha_k \langle g_m, g_{m_k} \rangle - (\alpha_k \langle g_m, g_{m_k} \rangle)^*}{\langle g_m, g_m \rangle} \quad (15)$$

식 (15)는 식 (10)과 마찬가지로 DFT의 형태로 풀이되며, 식 (16)과 같다.

$$\frac{R_k^*[\frac{m}{M}] - \frac{A}{2} (e^{j\theta_k} W[\frac{m-m_k}{M}] + e^{-j\theta_k} W[\frac{m+m_k}{M}])}{\gamma[m]W[\frac{0}{M}]} \quad (16)$$

VI. 실험 결과 및 고찰

본 논문에서는 인지적인 가중치 계산을 위해 입력 신호에 MPEG-1, 2 오디오 표준의 심리음향 모델-1을 적용하였다. 입력 신호 분석 시 프레임의 길이는 512샘플로 고정하였고, 해밍 윈도우, 50% 중첩-가산과정으로 분석/합성하였다. 사전의 크기는 1024로 설정하였다. 모의 실험을 통해, 매칭 퍼슈잇에 심리음향모델을 적용한 경우와 적용하지 않은 경우 각각을 비교하였다. 이때 심리음향 모델이 적용되지 않은 매칭 퍼슈잇의 정지 기준은 잔여 신호의 에너지가 영에 가까워질 때까지, 심리음향모델이 적용된 매칭 퍼슈잇에서는 인간의 청각특성이 고려된 신호 대 마스크 비가 1 이상이 될 때를 조건으로 매칭 퍼슈잇을 수행하였다. 입력 신호의 특성에 따라 매칭 퍼슈잇의 정지 기준에 따른 반복 과정이 무한대에 가까워질 수도 있기 때문에 최대 오톨의 개수를 프레임당 200개로 제한하였다.

그림 2는 오보에 입력 신호의 임의의 한 프레임에 대한 인지적인 매칭 퍼슈잇을 수행할 때의 정현파 추출 과정을 나타낸 것이다. 그림에서 각 반복 과정의 사전 내 인덱스 값과 그 때의 SMR 값을 나타내고 있으며, 입력 신호로부터 인지적인 중요도가 가장 큰 것을 찾아가며 SMR의 값이 1보다 작아질 때까지 반복되는 과정을 보이고 있다. 그림에서 나타내는 스펙트럼 값은 입력 신호에 DFT를 취해준 후, 인지적인 가중치로 나누어 준 값이다. 이 때 값이 크면 인지적인 중요도가 높은 것을 의미한다. 스펙트럼 상에서 인지적인 중요도가 높은 것을 차례대로 추출해

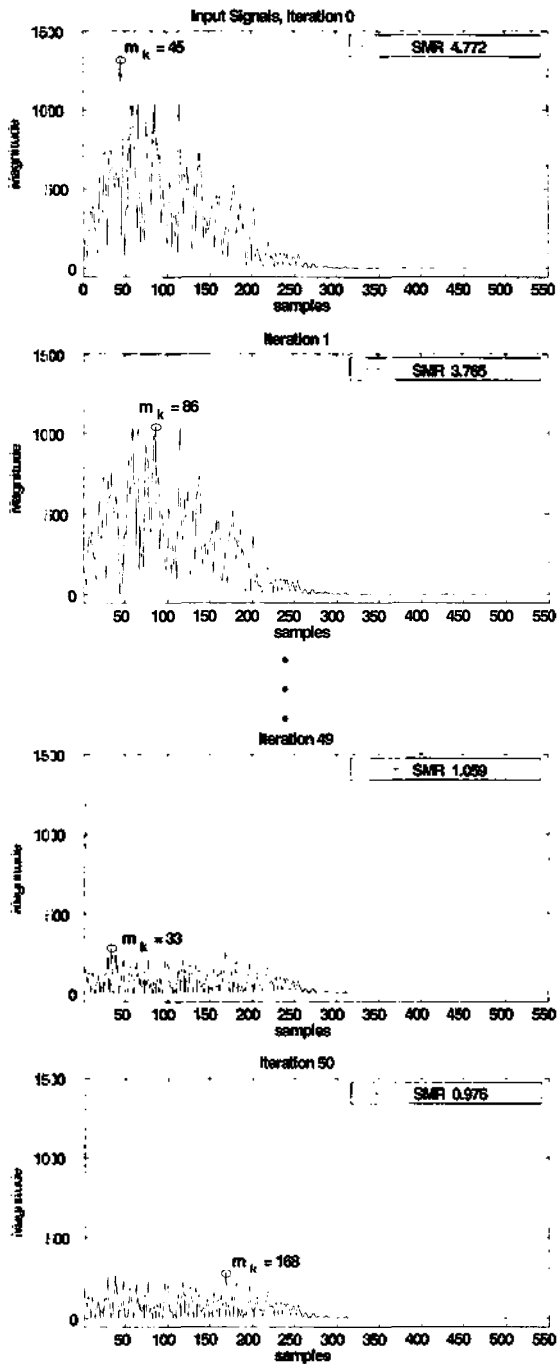
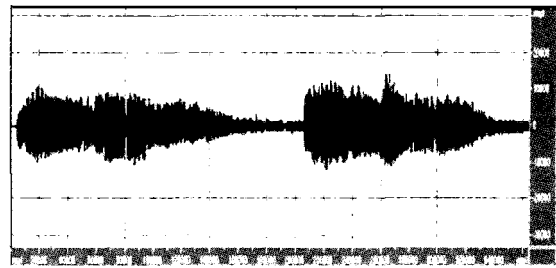


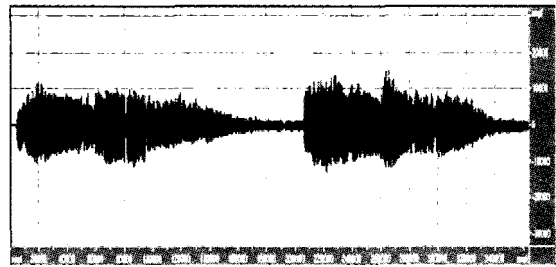
그림 2. 인지적인 매칭 퍼슈잇의 정현파 추출 과정
Fig. 2. Sinusoids extracting of perceptual matching pursuit.

가며 사전 내의 인덱스 m_k 를 찾아내어 그 때의 SMR을 계산하고 값이 1보다 큰 가에 대해 비교한 후 인지적인 매칭 퍼슈잇의 종료 여부를 결정한다. 그림 2에서 보듯이 반복의 수가 늘어감에 따라 스펙트럼이 줄어들고, SMR의 값도 점차 감소해 1보다 작은 값, 0.976일 때 매칭 퍼슈잇이 종료되었다.

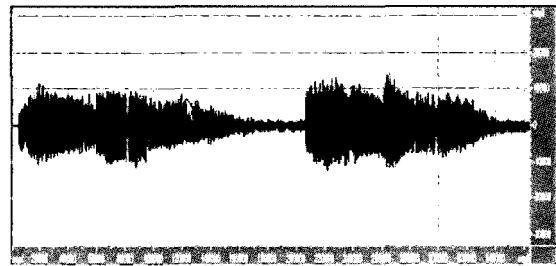
그림 3은 오보에 입력 신호에 대해 정현파 매칭 퍼슈잇, 인지적인 매칭 퍼슈잇을 적용해 추출된 정현파들로 합성



(a) 원음
(a) Original signal



(b) 정현파 매칭 퍼슈잇
(b) Sinusoidal matching pursuit



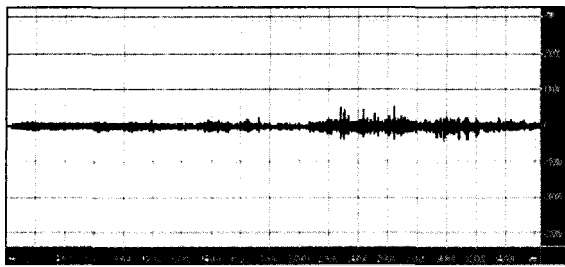
(c) 인지적인 매칭 퍼슈잇
(c) Perceptual Matching Pursuit

그림 3. 오보에 소리의 매칭 퍼슈잇을 통한 합성 신호
Fig. 3. Synthesized signal for 'Oboe'.

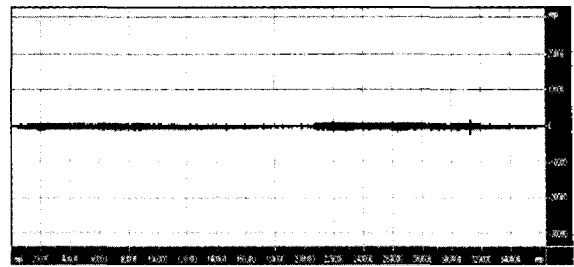
표 1. 오보에 입력 신호에 대한 SNR과 SEGSNR의 비교
Table 1. SNR/SEGSNR for 'Oboe'.

	Sinusoidal Matching Pursuit	Perceptual Matching Pursuit
SNR	16.45	21.62
SEGSNR	19.06	21.63

된 신호의 파형을 나타낸다. 그에 따른 결과는 그림 4의 오차 신호의 파형과 표 1에서 더 구체화된다. 그림 4에서 보듯이 정현파 매칭 퍼슈잇에 대한 오차 신호보다 인지적인 매칭 퍼슈잇에 대한 오차 신호가 영에 더 가깝게 수렴되었으며, 표 1의 신호 대 잡음 비 (SNR: Signal to Noise Ratio)와 구획화된 신호 대 잡음 비 (SEGSNR: Segmented SNR)의 수치도 높았다. 이러한 결과들과 같이 주관적인 음질 실험에서도 인지적인 매칭 퍼슈잇의 합성 음이 원음과 차이가 거의 없었음을 확인하였다. 그러나 정현파 매칭 퍼슈잇의 합성음은 오차 신호가 크게 나타나

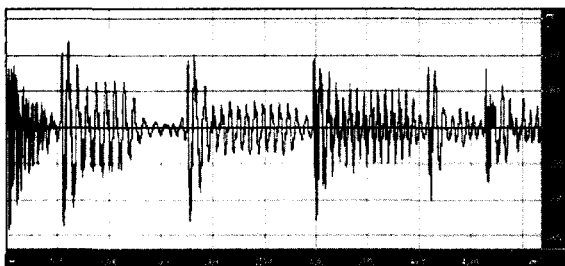


(a) 정현파 매칭 퍼슈잇
(a) Sinusoidal Matching Pursuit

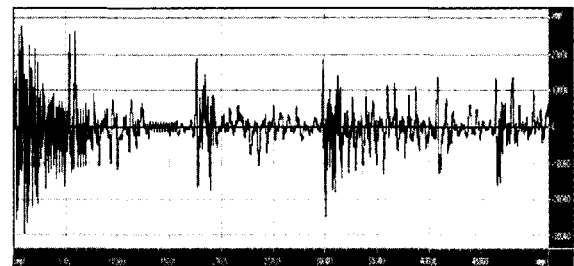


(b) 인지적인 매칭 퍼슈잇
(b) Perceptual Matching Pursuit

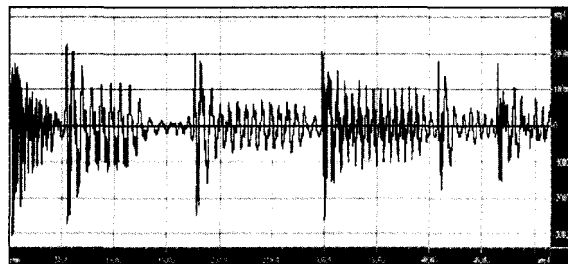
그림 4. 오보에 입력 신호에 대한 오차 신호의 파형
Fig. 4. Error signal for 'Oboe'.



(a) 원음
(a) Original signal



(b) 정현파 매칭 퍼슈잇
(b) Sinusoidal matching pursuit



(c) 인지적인 매칭 퍼슈잇
(c) Perceptual matching pursuit

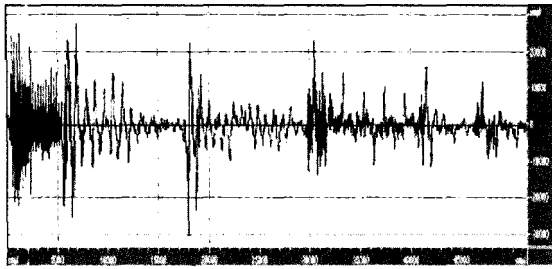
그림 5. 기타 소리의 매칭 퍼슈잇을 통한 합성 신호
Fig. 5. Synthesized signal for 'Guitar'.

는 부분에서 잡음소리가 섞여 청각적으로 원음과 약간의 차이가 났다.

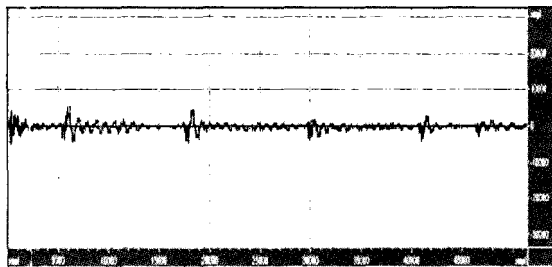
그림 5는 기타 입력 신호에 대한 각 매칭 퍼슈잇을 통한 합성 신호에 대한 것이다. 그림에서 보듯이 기타 소리가 입력으로 주어졌을 경우 정현파 매칭 퍼슈잇, 인지적인 매칭 퍼슈잇을 통한 합성음의 파형은 모두 원음의 파형과 약간 다르게 나타났다. 정현파 매칭 퍼슈잇 음질은 합성음의 파형에서처럼 원음과 많이 달랐지만, 인지적인 매칭 퍼슈잇의 음질은 원음과 별 차이가 없었다. 표 2의 SNR/SEGSNR을 참조해 보면 기타 소리의 정현파 매칭 퍼슈잇의 음질이 좋지 않음을 알 수 있다. 그림 6의 인지적인 매칭 퍼슈잇 오차 신호의 파형에서 에너지가 비교적 큰 성분이 잔여 신호에 남아 있었지만, 이 성분이 제외되

어 합성된 신호와 원래 입력 신호의 소리의 음질이 인지적으로 같았다.

매칭 퍼슈잇의 과정에서 현재 신호와 가장 잘 정합되는 요소를 사전 내에서 찾을 때 에너지와 인지적인 중요도가 모두 큰 경우엔 정현파 매칭 퍼슈잇, 인지적인 매칭 퍼슈잇 모두 같은 인덱스를 찾지만 에너지는 크지만 인지적인 중요도가 낮을 경우엔 각각의 매칭 퍼슈잇이 찾아내는 인덱스가 달랐다. 이로 인해 각각의 매칭 퍼슈잇이 인덱스를 찾아내는 곳과 프레임당의 반복 과정의 개수가 달라진다. 입력 신호의 특성에 따라 다르긴 하지만 정현파 매칭 퍼슈잇에서 인지적인 매칭 퍼슈잇과 같이 원음과 인지적으로 같아지도록 하기 위하여, 프레임당 제한된 오톨의 수를 늘려 주어도 해결이 되지 않는 경우도 있었다.



(a) 정현파 매칭 퍼슈잇
(a) Sinusoidal matching pursuit



(b) 인지적인 매칭 퍼슈잇
(b) Perceptual matching pursuit

그림 6. 기타 입력 신호에 대한 오차 신호의 파형
Fig. 6. Error signal for 'Guitar'.

표 2. 기타 입력 신호에 대한 SNR과 SEGSNR의 비교
Table 2. SNR/SEGSNR for 'Guitar'.

	Sinusoidal Matching Pursuit	Perceptual Matching Pursuit
SNR	2.94	17.58
SEGSNR	2.76	17.33

특히 기타 소리 같이 시간축 변화율이 큰 부분이 포함되어 있는 소리의 경우엔 더욱 그러했다. 정현파 매칭 퍼슈잇에서의 오름 수를 늘려 주는 것이 꼭 필요한 정현파를 더 찾는 것 보다는 인지적인 중요도와는 상관없이 에너지가 큰 정현파들을 추출해 내었기 때문이다.

VII. 결론

본 실험을 통해, 여러가지 소리를 비교해 보았을 때 전반적으로 SNR/SEGSNR, 주관적인 음질 모두 인지적인 매칭 퍼슈잇에서의 실험 결과가 더 좋았다. 인지적인 매칭 퍼슈잇은 정현파 매칭 퍼슈잇보다 더 정확한 정현파들을 추출해 내어 좋은 음질을 얻어낼 수 있었다. 그것은 각각에 대한 오차 신호의 파형과 SNR/SEGSNR에서 확인할 수 있다. 이로써 오디오 신호를 인지적인 가중치의 매칭 퍼슈잇으로 정현파 모델링하는 것이 적절하다는 것을

확인할 수 있었다. 또한 기타 소리 같이 신호에 뛰는 성분(attack)이 포함되어 있을 경우에 인지적인 매칭 퍼슈잇이 정현파 매칭 퍼슈잇보다 성능이 좋음을 확인하였다.

본 실험에서의 인지적인 매칭 퍼슈잇과 정현파 매칭 퍼슈잇의 성능을 비교, 평가하기 위해 사용된 SNR/SEGSNR은 단순히 원음의 에너지와 잔여 신호의 에너지의 비를 나타낸다. 인지적인 매칭 퍼슈잇에서, 성분의 에너지는 크더라도 인지적으로 중요도가 낮은 성분일 경우 잔여 신호에 남게 되어 SNR/SEGSNR의 값이 작아지게 된다. 본 실험에서의 기타 소리의 인지적인 매칭 퍼슈잇의 합성 결과를 보면 잔여 신호에 에너지가 큰 성분이 남아 있어서 전체적으로 SNR/SEGSNR의 값이 낮게 나왔다. 실험의 결과에서 SNR/SEGSNR은 인지적인 매칭 퍼슈잇과 정현파 매칭 퍼슈잇과의 상대적인 비교에는 적절하게 쓰였지만, 인지적인 매칭 퍼슈잇의 성능을 객관적으로 평가하기엔 곤란한 점이 있다. 보다 정확한 음질의 객관적이고 정량적인 측정을 위한 다른 성능평가 방법이 요구된다.

감사의 글

본 연구는 한국과학재단 특정기초 연구지원에 의해 연구되어짐 (과제번호 R01-2000-000-00249-0)

참고 문헌

1. E. B. George and M. J. T. Smith, "Analysis/synthesis overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones," *J. of Audio Eng. Soc.*, 40 (6), 497-516, June, 1992.
2. R. J. McAulay and T. F. Quatieri, "Speech analysis-synthesis based on a sinusoidal representation," *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP-34 (4), 744-754, Aug. 1986.
3. S. Mallat and Z. Zang, "Matching pursuits with time-frequency dictionaries," *IEEE-SP*, 41, 3397-3415, Dec. 1993.
4. K. Brandenburg and M. Bosi, "Overview of MPEG audio: current and future standards for low-bit rate audio coding," *J. Audio Eng. Soc.*, 45 (12), Jan/Feb 1997.
5. E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, 1990.
6. MPEG Committee, "Information technology-coding of moving pictures and associated audio for digital storage media at up to about 51.5Mbit/s part 3: Audio," ISO/IEC 11172-3, 1992.
7. T. S. Verma and T. H. Y. Meng, "Sinusoidal modeling using frame-based perceptually weighted matching pursuit," 1999

IEEE International Conference on, 2, 981-984, 1999.

- 8. X. Serra, "Musical sound modeling with sinusoids plus noise," *Musical Signal Processing*, Swets & Zeitlinger Publisher, 1997.
- 9. M. Goodwin and M. Vetterli, "Atomic decompositions of audio signals," *IEEE-ASSP*, 19-22, Oct, 1997.
- 10. T. Painter, A. Spanias, "Perceptual coding of digital audio," *in Proc. IEEE*, 88, 451-513, April 2000.
- 11. P. Noll, "MPEG digital audio coding," *IEEE Signal Processing Magazine*, 14, 59-81, Sept, 1997.

저자 약력

● 김 연 지 (Yeon-Ji Kim)



1995년~1999년: 충북대학교 전자공학과
 2001년~현재: 충북대학교 대학원 전자공학과 석사과정
 * 주관심분야: 음성 및 오디오 신호 처리

● 이 인 성 (In-Sung Lee)



1983년: 연세대학교 전자공학과 (공학사)
 1985년: 연세대학교 전자공학과 (공학석사)
 1992년: Texas A&M Univ. (공학박사)
 1986년~1987년: 한국통신 연구 개발단 전임연구원
 1993년~1995년: 한국전자통신연구원 이동통신기술 연구단 선임연구원
 1995년~현재: 충북대 전기전자공학부 부교수
 * 주관심분야: 음성 및 영상신호압축, 이동통신, 적응필터