1993년도 한국음향학회 학술논문발표회 논문집(제 12권 1(e)호)

*청각 특성을 이용한 피아노 음원 압축 알고리듬

°김기수, 황덕동, 윤대희, 차일환

연세대학교 전자공학과 음향 음성 및 신호처리 연구실

Piano Sound Compression Algorithm Using Auditory Characteristics

Ki-Soo Kim, Duck-Dong Hwang, Dae-Hee Youn, Il-Whan Cha

A.S.S.P. Lab. Dept. of Electronic Engineering Yonsei University, Seoul, Korea

요약

디지틀 피아노에서 PCM 방식의 음원 합성을 위한 압축 알고리듬을 제안하였다. 디지틀 피아노는 매우 높은 음질을 필요로 한다. 따라서 FM 방식 보다는 PCM 방식의 음원 합성 알고리듬이 주로 사용되어져 왔다. 그러나 PCM 방식은 많은 메모리가 필요한 단점이 있다.

본 논문에서는 피아노 음원을 청각적으로 왜곡이 없도록 압축, 저장하여 음원 데이타를 줄였다. 또한 피아노 음의 시간 영역 특성에 따라 비트 할당을 달리하여 음질을 향상시킬 수 있었다. 이와 같이 부호화할 경우 약 15 : 1 ~ 20 : 1 로 원음을 압축하여도 지각적으로 원음과 동일한 복원음을 얻을 수 있었다.

1. 서론

디지틀 신호처리 기술과 대용량 메모리 및 이식 기술의 발전으로 디지틀 방식의 전자 악기가 등장하였다. 디지틀 전자 악기는 음질 면에서도 기존의 어쿠스틱 악기에 떨어지지 않고 다양한 기능을 제공하므로 새로운 악기의 한 영역으로 발전하고 있다.

아날로그 신호인 악기음을 FM 방식에서와 같이 파라미터를 사용하지 않고 직접 디지틀 신호로 변환하여 저장한 데이타를 가지고 합성할 경우, 음질 면에서 뛰어난 성능을 얻을 수 있으나 메모리가 많이 필요한 단점이 있다. 악기음의 경우에 대역폭이 약 20 kHz 이고, 동적 영역도 96 dB 이상으로 디지틀 악기음이 원음에 가까운 음질을 얻기 위해서는 약 700 kbit/sec의 데이타 양이 필요하다. 따라서 디지틀 피아노와 같이 많은 건반의 음원을 저장하고 있어야 하고 다른 악기의 음원까지 저장해야 할 때는 PCM 방식의 적용에 많은 비용이 드는 단점이 있다.

본 연구는 (주)대우전자 악기 연구소의 연구비 지원에 의한 결과임

따라서 디지틀 피아노와 같이 고음질을 필요로 하는 경우에는 PCM 방식으로 음을 저장하고 합성하여 원음에 비해 음질의 저하를 줄이면서 메모리를 줄이기 위해 음원 데이타를 압축, 저장하여 합성하는 기술이 필요하다.

본 연구에서는 디지틀 피아노의 음원 압축 알고리듬 개발을 수행하였다. 디지틀 피아노는 매우 높은 음질을 요구하므로 음성에서와 같이 모델링에 의해 큰 압축율을 얻을 경우 음질이 떨어지므로 적용할 수 없다. 따라서 파형 부호화 방식이면서도 높은 비상관화 효과를 갖는 변환 부호화나 서브밴드 부호화와 같은 주파수 영역 부호화 방식이 적합하다[1][2][3][4]. 또한 주파수 영역에서의 부호화는 인간의 청각 특성을 이용할 수 있는 장점이 있다. 즉 어떤 주파수 대역에 큰 에너지를 갖는 신호가 있을 때 주변 대역의 약한 신호를 듣지 못하게 되는 마스킹 현상을 적용할 수 있다[5]. 따라서 원음에 의해 양자화 잡음이 마스킹 되어 듣지 못하게 되는 임계값 이하에 양자화 잡음이 분포하도록 비트 할당을 하면 큰 압축율을 얻으면서 원음에 비해 주관적인 음질이 떨어지지 않는 복원 신호를 얻을 수 있다[2][6][7].

본 논문의 구성은 다음과 같다. 2장에서는 피아노 음의 특성과 심리음향에 있어서 마스킹과 같은 여러 특성에 관해 설명하고, 3장에서는 서브밴드 부호화나 변환 부호화와 같은 주파수 영역 부호화와 심리 음향 모델이 결합된 피아노 음원 압축 알고리듬에 관하여 설명하였다. 4장에서 모의 실험을 통한 음질 평가 결과에 대해 설명하고, 5장은 결론으로 피아노음에 적합한 압축 및 합성 방식을 제안하였다

2. 피아노 음의 심리 음향적 특성

2-1. 심리음향에서의 마스킹 현상

우리 귀에서 음을 지각하는 경로는 음압이 고막에서 기계적 진동으로 변환한 후, 내이의 기저막에서 시간 영역 신호를 주파수 정보로 바꾸어 청신경에 전달한다. 이때 기저막의 청신경들은 필터뱅크와 같이 주파수 영역을 각

대역으로 나누어서 음을 인식하고, 그 대역은 임계 대역이라 불리워 진다. 임계 대역의 단위는 Bark이며 비선형적인 로그 함수적 특성을 갖는다. 이러한 임계 대역에 따라 음의 선택성 및 가청 한계가 달라진다[5].

음압이 큰 음원이 더 작은 음원을 들리지 않게 하는 현상을 마스킹이라 한다. 마스킹 현상은 내이의 기저막에서 청신경의 자극 정도에 의해 일어나며 시간 영역에서의 마스킹과 주파수 영역의 마스킹으로 구분할 수 있다.

시간의 차를 두고 일어나는 시간 영역의 마스킹을 순시 마스킹이라 한다. 동시 마스킹이란 어떤 주파수 대역에 순음이 존재하거나 잡음이 존재할 때 주위 대역의 작은 음압 레벨을 갖는 신호들이 들리지 않게 되는 현상을 말한다. 즉 주파수 영역에서 일어나는 마스킹을 동시 마스킹이라 하며 일반적인 마스킹은 동시 마스킹을 가리킨다.

그림 1은 1 kHz에 순음이 존재할 때 각 레벨에 따른 마스킹 곡선을 나타낸다. 일반적으로 마스킹 곡선은 고주파 쪽에서 더 넓은 마스킹 곡선을 가지며, 순음이 여러개 존재할 때는 각각의 순음에 대한 마스킹 곡선이 더해지는 형태의 마스킹 곡선을 갖는다.

그림 1의 아래 곡선은 조용한 환경에서도 음을 인식할 수 있는 임계값을 보여준다. 이 값을 절대 가청 한계라고 하는데 저주파와 고주파 대역에서는 큰 값을 갖고 약 2 kHz에서 5 kHz의 중간 주파수 대역에서는 아주 작은 값을 갖는 것을 알 수 있다. 이 것은 외이의 공명 주파수가 약 3 kHz이고 이 부근에 음성의 대부분의 정보가 모여 있는 것과 일치한다.

어떤 음원이 존재할 때, 그 음원의 마스킹 곡선과 절대 가청 한계를 더한 값을 전체적인 마스킹 임계값이라 하는데 양자화 잡음을 마스킹 임계값 아래에 분포하도록 부호화하면 주관적으로 원음과 거의 동일한 복원음을 얻을 수 있다.

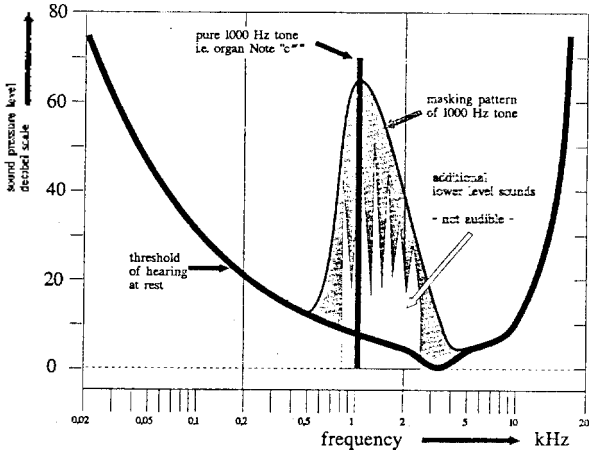

그림 1. 마스킹 곡선

2-2. 피아노음의 심리음향적 특성

피아노의 표준 건반수는 88개이며 음역은 A_1에서 C_7까지 7¼ 옥타아브로 이루어져 있다. 피아노 음의 발생 원리는 쉽이가 다른 현을 해머에 의해 타격을 가하여 발생하는 현진동으로 인한 공기 밀도의 변화에 의해 일어난다. 피아노음은 27.5 ~ 4186 Hz의 기본 주파수(피치)의 정수배로 이루어진 하모닉스 성분들로 구성된다.

피아노 음의 특성은 그림 2와 같다. (a)와 (c)의 그림은 4 번째 옥타아브의 C_4 음의 시간에 따른 그래프와 에너지 감쇄 곡선이다. 그림에서와 같이 피아노음은 크게 attack, steady state, decay 부로 나눌 수 있다. attack 부는 건반이 눌러지고 난 후의 수십 msec 정도로 급격하게 음압이 커지며 고주파 성분도 강하게 나타난다. steady state는 음이 어느 정도 안정된 이후의 상태로 주기적인 성분이 강하게 나타나며 시간에 따라 거의 일정한 스펙트럼을 갖는다. decay부는 음압 레벨이 크게 감쇄하는 부분으로 약 수백 msec 정도이다[10].

(b)는 C_4 음의 steady state에서 주파수 스펙트럼을 살펴본 결과이다. 그림에서 보면 정확하게 기본 주파수(261.6 Hz)와 하모닉스 성분들이 나타남을 알 수 있다. 피아노 음은 그림 2. (b)에서 보면 대역 폭이 비교적 좁고 순음 성분이 매우 강하게 나타난다. 따라서 순음에 의한 마스킹 모델을 적용하면 마스킹 현상을 효과적으로 이용할 수 있으므로 적은 비트율 가지고도 주관적으로 매우 높은 음질의 소리를 얻을 수 있다.

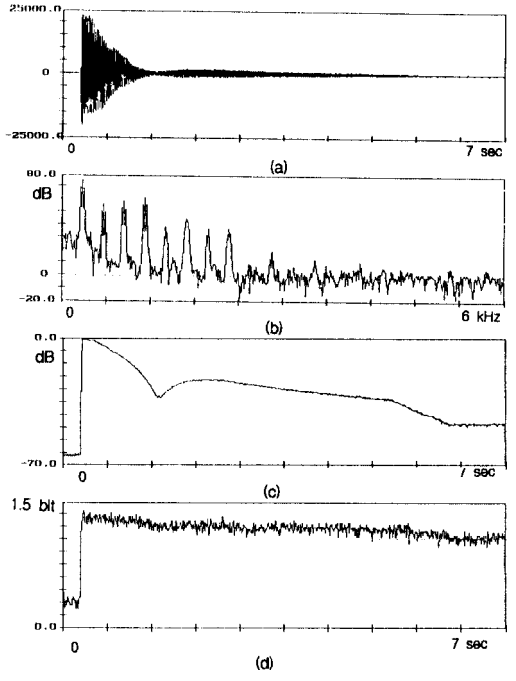

그림 2. 피아노음의 특성

(d)는 C_4 음에 대한 지각 엔트로피를 나타낸 것이다. 지각 엔트로피란 어떤 음을 청각적으로 동일한 음질을 갖도록 부호화하는데 최소 비트 수를 나타낸다. attack 부에서만 약간 많은 비트가 할당되고 steadystate 와 decay 부에서는 샘플당 약 1.2 비트(약 15 : 1) 이내로 부호화하면 원음과 동일한 음질을 얻을 수 있음을 알 수 있다. 또한 steady state 부에서는 거의 일정한 비트가 사용됨을 보여준다.

3. 피아노 음원 압축 알고리듬의 설계

마스킹 현상을 비롯한 여러 청각 특성을 결합한 주파수 영역 부호화 방식으로 일반적인 오디오 신호를 부호화하면 96 ~ 128 kbit/s에서 주관적으로 원음과 동일한 복원음을 얻을 수 있다[7]. 피아노 음은 주기적인 성분이 강하게 나타나고 전체 대역의 크기가 비교적 좁게 나타나기 때문에 약 64 kbit/s 정도에서도 좋은 음질을 얻을 수 있다. 피아노 음원 부호화기에서는 입력된 음원 신호의 주관적인 중복성과 통계적인 중복성을 제거하여 압축된 비트열을 만드는 역할을 한다. 입력 신호는 32개의 필터뱅크를 통과하여 서브밴드 샘플로 바뀌어 진다. 이때 심리음향 모델에서는 부가적인 FFT을 이용하여 마스킹 임계값을 얻어 양자화에 쓰이는 비트 할당 정보를 주게된다. 즉, 필터뱅크의 출력값과 마스킹 임계값을 가지고 신호대 마스크 비를 구하여 주관적으로 양자화 잡음이 신호에 의해 마스킹될 수 있도록 비트 할당을 한다. 양자화된 서브밴드 샘플과 비트 할당 정보 등의 부가 정보를 가지고 비트열을 만든다.

복호화기에서는 압축된 비트열을 풀어 각 서브밴드 샘플들이 복원되고 합성 필터를 통과하여 부호화된 신호의 PCM 샘플(복원 신호)을 얻는다. 복호화기는 부호화기에 비해 간단하며 심리음향에 관한 정보가 필요하지 않으므로 실시간 시스템에 적합하다.

피아노 음원 부호화가 갖는 특징은 다음과 같다.

- 부호화 시스템은 전송이 아닌 저장을 목적으로 하므로 실시간 시스템을 필요로 하지 않는다.
- 복호화 시스템은 계산량과 지연이 적은 실시간 시스템을 필요로 한다.
- 원음의 중복성이 매우 큰 반면, 복원음의 음질은 원음과 거의 같아야 한다.

그림 3은 제안된 피아노음 부호화 알고리듬의 기본 구조이다. 피아노음의 경우, 기본 하모닉스의 정수배에 해당하는 하모닉스 성분들을 충분히 살려주어야 음질이 보존되므로 사용자가 특별히 강조하여 비트를 할당할 밴드가 발생한다. 따라서 심리음향 모델에서의 마스킹 현상을 이용하여 얻어진 최적 비트 할당 값에 사용자가 각 밴드와 프레임에 할당된 비트를 가변할 수 있게 하여 원하는 음질을

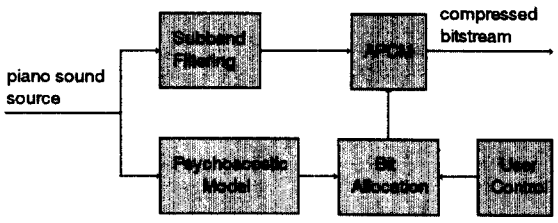

그림 3. 제안된 피아노음 부호화 시스템

얻을 수 있다. 이와같은 방법으로 각 건반음에 적합한 비트 할당이 결정될 수 있다.

본 논문에서 제안한 피아노 음원의 압축 및 합성 방법은 피아노음의 음질에 중요한 역할을 하는 attack 부분은 PCM 데이타로 모든 건반에 대하여 가지고 있고 나머지 부분에 대해서는 일부분의 데이타만을 가지고 음을 합성하는 것이다. 다음 절에서 음원 압축 알고리듬의 각 부분에 대해 상세하게 설명한다.

3-1. 서브밴드 분석과 합성

시간 영역에서의 입력 신호를 주파수 영역과 시간 영역에서 원하는 형상으로 변환시켜주는 역할을 하는 필터뱅크는 서브밴드 부호화기에서 가장 중요한 역할을 한다. 디지탈 피아노 음원 압축에 사용된 필터뱅크는 32개의 동일 크기를 갖는 가중 중첩 가산(Weighted Overlap-Add) 방법의 SSB 필터뱅크이다. 가중 중첩 가산 방법의 필터뱅크는 블럭 단위로 데이타를 처리하여 효율적인 계산이 가능하다. 서브밴드 분석에 사용되는 필터는 512-tap 저역 통과 필터가 기본이 되며 행렬에 의해서 주파수 천이되어 32개의 동일 크기 서브밴드가 된다. 가중 중첩 가산 방식으로 다음과 같은 특징을 갖는다.

- 분석 필터와 합성 필터를 통과할 때 완전 복원 조건을 만족한다.
- 분석과 합성 과정에서 임계적으로 표본화된다.
- 효과적인 서브밴드 부호화를 위해 32개의 필터뱅크를 사용하였다.

서브밴드 합성 과정은 부호화된 32 서브밴드 샘플로 출력 신호를 복원하는 과정으로 분석 과정의 역으로 한다.

3-2. 심리 음향 모델

심리 음향 모델을 사용하면 각 서브밴드에서 원음에 의해 마스킹 되어 들을 수 없는 최대 잡음 레벨을 결정할 수 있다. 이 잡음 레벨(마스킹 임계값)을 사용해서 각 밴드의 실제 양자화기를 결정하는 비트 수를 알 수 있다.

그림 4에 C_4 음에 대한 심리음향 모델 적용의 결과가 나타나 있다. 그림에서 알 수 있듯이 각 밴드에서의 음압 레벨에서 각 밴드에서의 최소 마스킹 임계값을 빼주면 신

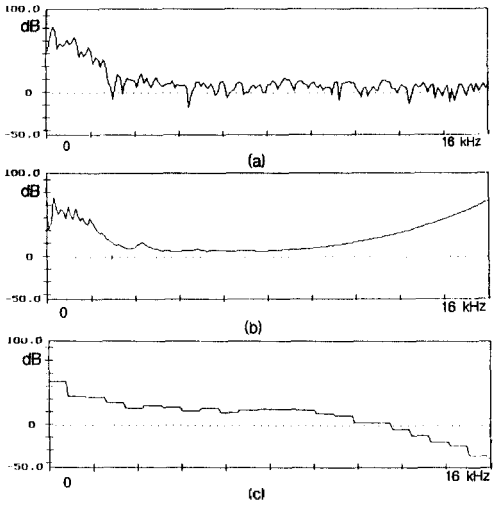

그림 4. 심리음향 모델의 결과
a) FFT 스펙트럼 b) 마스킹 임계값 c) 신호대마스크비

호대 마스크비를 얻을 수 있다. 결국 신호대 마스크비가 작다면 신호의 음압 레벨이 작거나 마스킹이 많이 되는 것이므로 신호대 마스크비가 작을수록 적은 비트를 가지고 효과적인 양자화를 할 수 있다.

3-3. 비트 할당

비트 할당은 그림 2. (d)의 지각 엔트로피에서 알 수 있듯이 시간 영역에 따라 독립적으로 수행하였다. 비트 할당 테이블은 심리 음향 모델에 의한 최적 비트 할당으로부터 구해진다. 심리음향 모델의 결과에 의하면 attack 부는 고주파 성분이 많고 음질을 좌우하는 경우가 많다. 따라서 마스크 대 잡음비(MNR)가 모든 밴드에서 0 dB 이상인 최적의 비트 할당 보다 많은 비트를 할당하여 음질을 향상시켰다. attack 부는 약 5 프레임(약 120 msec) 정도로 최적 비트 할당에서 각 밴드에 2 비트를 더 할당하여 마스크대 잡음비를 약 12 dB 이상이 되도록 하였다. 이러한 이유는 마스킹 임계값이 실험에 의한 값을 평균한 값이므로 각 개인에 따른 가청력을 고려한 것으로 음에 매우 예민한 사람도 원음과 동일한 음질을 얻을 수 있기 때문이다.

안정 상태의 한 구간에서 비트할당은 최적 비트 할당에서 약 0 ~ 2 비트 정도의 일정한 비트를 대역에 따라 달라더하여 비트 할당을 하였다. steady state 부는 비슷한 신호가 반복되는 주기적 함수로 볼 수 있으므로 마스킹 특성도 프레임 간에 유사한 형태로 나타난다. 따라서 최적의 비트 할당을 해도 원음과 거의 동일한 음질을 얻을 수 있지만 약 10 ~ 18 kHz의 대역에 약 2 비트 정도를 할당함으로 고주파 성분에 따라서 음색의 변화가 생기는 것을 막을 수 있다.

위와 같은 방법으로 얻어진 비트 할당 정보를 가지고 서브밴드 샘플을 양자화 할 때는 mid_tread 선형 양자화기를 사용하였다. 각 서브밴드 샘플들은 최대값에 의해 나누어져 정규화된 후 양자화되어 진다.

4. 음질 평가

디지털 피아노 음원 부호화 알고리듬의 음질을 평가하기 위해 객관적 평가 방법으로 구간 신호대 잡음비와 잡음대 마스크비를 사용하였다[8][9]. 압축 비는 12 : 1 이며 음질 평가의 세부 사항은 다음과 같다.

1) 먼저 구간 신호대 잡음비는 512 샘플을 한 구간하여 구한 신호 대 잡음비를 전 구간에 걸쳐 구한 후에 평균값을 취해 얻어졌다. 구간 신호대 잡음비를 구하는 식은 다음과 같다.

$$SNR_{seg} = \frac{1}{L}\sum_{l=0}^{N_f} 10\log\left|\frac{\sum_{k}^{512} x(k)^2}{\sum_{k}^{512}(x(k)-y(k))^2}\right| \quad (4\text{-}1)$$

2) 잡음대 마스크비는 512 샘플마다 신호의 스펙트럼과 잡음의 스펙트럼을 계산한 후 신호의 스펙트럼으로부터 마스킹 임계값을 구한 후에 각 임계 대역별로 최저값을 구한다. 여기서 구해진 최저 마스킹 임계값과 잡음 스펙트럼의 임계 대역별 함과의 차를 구하였다. 즉, 마스킹 임계값과 잡음 스펙트럼의 평균 거리라고 생각할 수 있다.

사용된 피아노음은 Sonic Images Sample Library의 Vol 6.에 있는 그랜드 피아노 음 모음 중에서 Steinway 피아노의 메조 포르테(Mezzo forte) 음이 사용되었다. 사용된 음계는 4 번째 옥타아브의 C_4, D_4, F_4, A_4 이다. 사용된 각 음의 선구간에 대한 지각 엔트로피는 표 1과 같다.

표 1. 지각 엔트로피

	C_4	D_4	F_4	A_4
지각 엔트로피 (bit/sample)	0.89	0.73	0.64	0.67

표 2는 각 음에 대해 구간 신호대 잡음비와 잡음대 마스크비를 구한 결과이다. 구간 신호대 잡음비는 약 30 dB, 잡음대 마스크비는 -11.5 dB 정도의 결과를 얻어 원음과 주관적으로는 거의 동일한 음질을 얻을 수 있음을 알 수 있다.

표 2. 구간 신호대 잡음비와 잡음대 마스크비

	C_4	D_4	F_4	A_4
SNRseg (dB)	32.2	32.2	29.7	28.7
NMR (dB)	-11.23	-11.49	-11.68	-12.05

5. 결론

본 연구에서는 서브밴드 부호화와 청각 특성을 이용한 디지틀 피아노음의 데이타 압축 방법을 제안하였다. 이 방법은 심리음향 모델에서의 마스크대 잡음비를 0 dB 이상이 되도록 하여 원음과 주관적으로 동일한 음질을 갖도록 하였다.

비트 할당 방식으로는 심리음향 모델에 맞게 비트 할당을 하여 낮은 데이타율에서 높은 음질을 보존함을 확인하였다. 심리 음향 모델에 의한 마스킹을 이용하는 비트 할당으로 건반에따라 약 15 : 1에서 20 : 1 정도의 데이타 압축이 가능함을 알 수 있다. 따라서 40 ~ 60 Kbit/s 정도의 데이타율에서 건반음을 부호화 할 수 있다. 필터뱅크로는 SSB 변조 필터뱅크가 가산 중첩 방식으로 구현되어 48 kHz의 표본화율을 가지는 피아노음을 32개의 대역으로 분리하여 압축하였다.

또한 제안된 압축 알고리즘은 건반음마다 적합한 비트 할당을 음의 매 구간마다 테이블화하여, 사용자가 하모닉스의 분포를 고려하여 구간 마다의 비트 할당을 추가적으로 결정하도록 하였다. 피아노음은 일반적인 악기 신호들의 한 예에 속하므로 다른 악기음의 압축에도 동일한 알고리즘이 사용될 수 있다. 더 높은 압축율이 필요한 때에는 엔트로피 부호화, 혹은 벡터 양자화 등이 결합될 수 있다. 이러한 방식은 각 서브밴드 출력의 통계치를 구하거나 패턴을 구분하는 작업이 필요하다.

참 고 문 헌

[1] K. Brandenburg, "OCF - a new coding algorithm for high quality sound signals." *Proc. ICASSP* pp. 141-144, 1987

[2] Y. F. Dehery, *et al.* "A MUSICAM source codec for digital audio broadcasting and storage." *Proc. ICASSP* pp.3605-3608, 1991

[3] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria." *IEEE J. Selected Areas Comm.* pp. 314-323, 1988

[4] M. Iwadare, *et al.* "A 128 kbit/s hi-fi audio codec based on adaptive transform coding with adaptive block size MDCT." *IEEE J. Selected Areas Comm.* pp.138-144, 1992

[5] E. Zwicker, *Psychoacoustics.* Springer-Verlag, New . York, 1982

[6] G. Stoll and Y. F. Dehery, "High quality audio bit-rate reduction system family for different applications" *Proc. ICC* pp.937-941, 1990

[7] ISO/IEC JTC1/SC29/WG11 MPEG-Audio

[8] R. N. Veldhuis, "Bit rates in audio source coding." *IEEE J. Selected Areas Comm.* pp.86-96, 1992

[9] U. Halka, et al. "A new approach to objective quality measures based on attribute matching" Speech Comm. pp. 15-30, 1992

[10] T. D. Rossing, *The Science of Sound.* Addison Wesley, 1990