

디지털 방송용 오디오 레벨 계측 알고리즘의 실시간화 연구

A Study on Real-Time Loudness Metering Algorithm for Digital Broadcasting

박 성 균

Seong-Gyoon Park

요 약

본 논문에서는 디지털 방송에서 요구되는 디지털 오디오 레벨에 대한 주관적 음향 레벨, 즉 라우드니스의 계측 알고리즘을 실시간화 하는 방법을 제안한다. 이를 위해 객관적 오디오 품질 분석을 위한 권고안인 ITU-R BS1387-1의 분석 결과를 토대로 FFT 기반의 라우드니스 계측 알고리즘을 구현하고, 이 알고리즘을 참조테이블(look-up table)을 이용하여 실시간화 하는 방법을 제시하였다. 그리고 이 알고리즘의 성능과 계측 시간 분석을 위해 23개의 순음과 30개의 디지털사운드 샘플을 적용하고, 그 결과를 분석하였다.

실험결과를 살펴보면, 실시간화 된 알고리즘의 계측값은 실시간화 방법이 적용되지 않은 원래 알고리즘과 비교할 때, spectral spreading 관련 참조테이블의 레벨 해상도가 10 dB로 상당히 큰 경우에도 2 % 미만의 차이를 보여 매우 양호한 측정 결과를 보였다. 계측 시간의 감소 효과를 살펴보면, 알고리즘 전체적으로는 1/21 정도로 실행시간이 줄었으며, 가장 계측시간 점유비율이 큰 청각필터 그룹별 에너지 계산 시간은 1/450로 줄어들고, 여기 패턴을 구하는 계산 시간도 약 1/3.57로 좋은 계측 시간 단축 효과를 얻었다. 결론적으로 제안한 알고리즘을 통해 향후 DSP 기반의 완전 실시간 라우드니스 미터를 구현할 가능성을 확보하게 되었다.

Abstract

In this paper, the perceived audio level metering algorithm of digital audio sound to be able to operate in real-time is proposed. Through analyzing a conventional recommendation ITU-R BS1387-1 for objective audio quality analysis, FFT-based loudness metering algorithm is implemented and the real-time method of that algorithm was advised and proved. The proposed method is based on look-up table. In order to prove the proved method, using 23 pure tones and 30 preselected digital audio samples, its performance and operation time is evaluated.

Its performance, compared with an original algorithm's, have a good figure of less than 2 % error even if look-up table related with spectral spreading have large level resolution of 10 dB. The proposed algorithm take only 1/21 of original algorithm's measuring time. Also, in the proposed algorithm, auditory pitch group energy calculation take 1/450 of original algorithm's and excitation calculation take 1/3.57. In conclusion, the proposed algorithm is expected to be implemented into DSP-based real-time loudness meter.

Key words : Loudness Metering Algorithm, ITU-R BS1387-1, Look-Up Table, Real-Time

I. 서 론

디지털 방송에서는 디지털 오디오 콘텐츠에 따라

음향 레벨의 동적 범위가 매우 넓을 수 있다. 예를 들어 뉴스와 같은 음성 위주의 콘텐츠들은 좁은 동적 범위의 음향 레벨을 갖지만, 영화와 같은 음성과

공주대학교 정보통신공학부(School of Information and Communication Engineering, Kongju National University)

· 논문 번호 : 20050222-022

· 수정완료일자 : 2005년 4월 8일

다양한 사운드들이 섞여 있는 콘텐츠들은 매우 넓은 동적 범위를 가지게 된다. 따라서 방송을 수신하는 청취자 입장에서는 채널간 또는 방송 프로그램간 천이 시에 본인이 원하는 음향 레벨이 계속 유지되기를 원할 것이다.

따라서 수신기는 메타데이터를 이용하여 음향 레벨을 정규화함으로써 보다 균일한 재생 오디오 레벨을 얻도록 한다. 그러나 이를 위해서는 올바른 메타데이터를 생성할 수 있도록 정확한 방송 프로그램의 음향 레벨을 측정할 수 있어야 하며, 측정되는 음향 레벨은 인간이 주관적으로 실제 느끼고 인지하는 세기이어야 한다.

주관적 음향 레벨을 계측하기 위한 심리음향적인 연구는 독일의 Zwicker와 영국의 Moore에 의한 연구가 가장 주목된다. 청각계의 임계대역을 근간으로 하는 Zwicker 모델이 청각계의 비선형성을 보다 우수하게 표현하고 있는 것으로 평가되며, 국제표준안인 ISO532B에 반영되었다^{[2],[7],[8],[14]}. 이후 Zwicker는 음향의 노출시간과 크기에 따른 비선형성을 나타낼 수 있는 아날로그 회로를 고안하고, 이를 이용한 과도음 분석 모델도 제안하였다. 영국의 Moore와 Glasberg는 정상신호에 대해 기존의 다양한 청각 특성에 대한 다양한 임상실험 결과를 토대로 임계대역의 새로운 정의와 스펙트럴 마스크 효과를 적용하기 위한 새로운 함수를 개발함으로써 Zwicker의 모델을 개선한 모델을 제안하였다^{[6],[12],[13]}.

한편, 기존 ITU-R의 오디오와 관련된 권고안 중에서 오디오 품질의 객관적 측정 알고리즘에 관한 것으로 BS1387-1이 있는데, 오디오 품질을 측정하기 위해서는 기본적으로 오디오 레벨, 즉 라우드니스(loudness)를 측정하여야 하며 이 권고안에서도 Beerends, Stermdink, Terhardt 등의 1990년대 연구 결과를 참조하여 구성된 비 라우드니스(specific loudness) 및 총 라우드니스(total loudness) 계측 관련 알고리즘이 포함되어 있다^{[2],[9],[11],[16]}.

그리고 최근 ITU-R WP6P에서 완전 디지털 방송에 대비한 디지털 사운드의 계측 방법에 대한 연구가 2000년부터 진행되고 있다^[15].

본 연구의 목적은 1~2개의 DSP로 실시간 계측 시스템 구현이 가능한 방법을 찾고자 하는 것이다. 즉, 적은 비용(low cost)으로 실시간(real-time)적으로

실제 인간이 인지하는 오디오 레벨을 측정하는 시스템을 구현하는 방법을 찾는 것이다.

이러한 라우드니스 미터를 구현하는 방법에는 크게 필터 뱅크(filter bank) 기반과 FFT 기반이 있다. 필터 뱅크로 구현할 경우 실시간화가 가능할 지는 모르지만 일반적으로 병렬로 입력신호를 처리하는 필터뱅크의 개수가 최소한 24개 이상이므로 DSP 프로세서가 너무 많은 개수를 요하게 된다. 따라서 저비용으로 구현하기 위해서는 FFT 기반의 구현 방법이 적합하다.

FFT 기반으로 구현이 가능한 라우드니스 계측 방법으로 Moore 방법을 생각할 수 있는데, Moore의 모델은 청각 필터의 개수가 최소한 150개 이상으로 설정하여 계산하므로 실행시간이 오래 걸리는 문제가 있다^{[5],[12]-[14]}. 다른 방법으로는 객관적 음질 분석 권고안인 ITU-R BS1387-1을 토대로 라우드니스 계측 알고리즘을 구성해볼 수 있다. 이 방법은 청각 필터 개수를 109개로 줄일 수 있으므로 실시간화 대상으로 BS1387-1 기반의 계측 알고리즘이 보다 적합하다^[6].

본 논문에서는 BS1387-1의 FFT 기반의 귀의 모델을 적용한 라우드니스 계측 알고리즘 특성을 구성한 다음, 이의 유용성을 검증하고, 이를 토대로 실시간화 방법을 제시하고 그 타당성을 검증하기 위한 실험 결과를 제시한다. 실시간화를 위한 방법으로는 참조테이블(look-up table)을 이용하여 계산량을 줄이는 방법을 적용하여 그에 따른 성능 변화 및 계산 시간 감축 효과를 고찰하였다.

II. ITU-R BS1387-1를 토대로 한 FFT기반 라우드니스 계측 알고리즘

2-1 개요

여기서 제시할 알고리즘은 주요 과정은 세 부분으로 구성된다. 첫째, 입력신호의 스펙트럼을 구한 후 중이 및 외이의 스펙트럼 특성을 적용한 후 임계대역에 따른 청각 필터 그룹별 에너지를 계산하는 과정과 둘째, 청각 필터 그룹별 에너지가 인근 필터 그룹으로 확산하는 효과를 고려한 여기 패턴의 계산 과정이 있고, 마지막으로 여기 패턴으로부터 비 라우드니스 및 라우드니스를 산출하게 된다.

이러한 과정을 좀 더 자세하게 살펴보면 다음과 같다.

2-1-1 프레임 처리 및 재생 레벨 조정

적용할 음성 파일은 모노, 16 bit, 48 kHz로 샘플링된 .wav 포맷 데이터 신호를 대상으로 하는데 한 프레임의 데이터 포인트 수는 2048(약 40 ms)로 한다. 그리고 중첩율이 50 %이므로 1024개 포인트가 중복된다. 이를 수식으로 표현하면 다음과 같다.

$$t_n[k_f, n] = t[1024 \times n + k_f] \quad (1)$$

여기서 n 은 프레임 번호이고, k_f 는 한 프레임 내의 카운터이다.

프레임화된 신호의 시간영역에서 주파수 영역으로의 매핑은 Hann 윈도우를 적용한 후, 푸리에변환(FFT)을 사용하여 수행된다. 그리고 FFT 처리된 2048 포인트 프레임 데이터 $F[k_f, n]$ 는 재생 레벨 조정(playback level scaling)을 위해 적절한 상수(fac)를 곱하여 다음 식과 같이 계산된다.

$$fac = \frac{10^{\frac{L_p}{20}}}{Norm},$$

$$F[k_f, n] = fac \cdot F_f[k_f, n] \quad (2)$$

여기서 L_p 는 재생될 음압 레벨(SPL: Sound Pressure Level)이고, k_f 는 라인 스펙트럼 인덱스이다. 또한 'Norm'의 값은 입력 신호를 주파수 1019.5 Hz이며, 0 dB 플 스케일인 사인 파형을 적용하여 10프레임에 걸친 FFT의 주파수 계수의 최대값으로 사용한다. 본 연구에서 시뮬레이션 결과 Norm=0.151로 결정되었다. 여기서의 재생 레벨은 디지털 사운드의 최대 레벨을 기준으로 설정되는 것이다.

2-1-2 외이/중이의 주파수 응답 특성

외이/중이의 주파수 응답 특성(middle and outer ear weighting)은 주파수의 함수로 나타나는 가중치로 나타나며, 음향 스펙트럼 k 번째 라인에서의 주파수 응답으로 계산된다. 즉,

$$W[k]/dB = -0.6 \cdot 3.64 \cdot \left(\frac{f[k]}{kHz}\right)^{0.8} + 6.5 \cdot e^{-0.6 \left(\frac{f[k]}{kHz} - 3.3\right)^2} - 10^{-3} \cdot \left(\frac{f[k]}{kHz}\right)^{3.6} \quad (3)$$

$$\frac{f[k]}{kHz} = \frac{k \cdot f_s}{1000N} \quad (4)$$

$$F_e[k_f, n] = abs(F[k_f, n]) \cdot 10^{\frac{W[k_f]}{20}} \quad (5)$$

식 (4)에서 f_s 는 주어진 신호의 샘플링 율이고 N 은 FFT 계산 포인트 수이며, 식 (5)에서의 'abs'는 절댓값을 의미한다.

2-1-3 청각 필터 그룹별 에너지 계산

이 과정은 가청 주파수 범위를 80 Hz에서 18,000 Hz까지로 규정하고, 임계대역 해상도에 따른 청각 필터 대역을 정의한 후, 앞 단계에서 구한 가중치가 곱해진 스펙트럼 에너지로부터 각 청각필터 그룹별 에너지를 산출하는 것이다.

각 청각 필터의 주파수 그룹을 피치(pitch)라고 하는데, 청각 피치 스케일(auditory pitch scale)은 다음 식 (6)으로 표현하고 단위는 'Bark'라고 하지만 Zwicker가 정의한 것과 정확히 같은 스케일은 아니며 근사식이다^{[1][2]}.

$$z/Bark = 7 \cdot \arcsin h\left(\frac{f/Hz}{650}\right) \quad (6)$$

식 (6)으로부터 각 청각 필터의 영역은 각 청각 필터의 임계대역 해상도에 의존하며, 이를 함수로 나타내면 다음 식 (7)과 같다.

$$f(i)/Hz = 80 + 650 \cdot \sinh((i-1)res/7) \quad (7)$$

여기서 res 는 임계대역 해상도를 의미하는 것으로 단위는 Bark이며, i 번째 청각필터의 주파수 범위는 하한이 $f(i)$ 이고 상한이 $f(i+1)$ 이 되고, 중심주파수는 $(f(i)+f(i+1))/2$ 이다. 본 연구에서는 0.25의 해상도에 대해 청각 필터를 구성하고, 음향 스펙트럼 에너지로부터 청각 필터 그룹별 에너지 $P_e[k, n]$ 를 계산한다.

2-1-4 내부잡음 첨가

청각의 임계치 레벨은 주파수의 함수로서 P_{Thres} 오프셋값으로 구해서 각 피치 그룹의 에너지에 더해진다. 이로부터 얻어지는 결과 $P_p[k, n]$ 을 'Pitch patterns'라 부른다^[10].

$$P_{Thres}[k] = 10^{0.1 - 0.361 \left(\frac{f[k]}{kHz}\right)^{0.3}} \quad (8)$$

$$P_p[k, n] = P_c[k, n] + P_{Thres}[k] \quad (9)$$

2-1-5 Spectral Domain Spreading

주파수 영역에서의 확산은 스펙트럴 매스킹 효과를 나타내는 것으로서 주파수와 에너지 레벨에 의존하는 확산함수로 정의된다.

확산함수는 양측 지수 함수인데, 피치의 중심주파수보다 낮은 주파수 영역에서는 확산 기울기가 항상 27 dB/Bark이고, 높은 주파수 영역에서는 주파수와 피치 그룹 에너지에 따라 달라진다. 즉, 기울기는 다음 식에 의해 계산된다^{[9],[11]}.

$$\frac{S_u[k, L[k, n]]}{dB/Bark} = -24 - \frac{230Hz}{f_c[k]} + 0.2 \cdot L[k, n]/dB \quad (10)$$

$$S_l[k, L[k, n]] = 27 \frac{dB}{Bark} \quad (11)$$

여기서 $L[k, n] = 10 \log_{10}(P_p[k, n])$ 이다.

확산은 각 주파수(피치) 그룹 k 별로 독자적으로 수행되므로 다음 식 (12)와 같이 표현된다.

$$E_z[k, n] = \frac{1}{Norm_{sp}[k]} \left(\sum_{j=0}^{z-1} (E_{ine}[j, k, n])^{0.4} \right)^{0.4} \quad (12)$$

여기서 $Norm_{sp}[k]$ 와 $E_{ine}[j, k, n]$ 은 해당 주파수 그룹에 인근 주파수 그룹으로부터의 확산 효과를 적용한 여기값을 계산한 것이며, 이를 구하는 구체적인 수식은 권고안에 나타나 있다.

2-1-6 Temporal Domain Spreading

시간영역에서의 매스킹 효과를 반영하는 것으로 포워드 매스킹만을 적용하며, 각 청각 필터 그룹에 있는 에너지의 1차 LPF 출력을 구하는 과정이다.

방송용 오디오 레벨 계측은 정상음(steady sound) 계측에 가깝다고 보는 것이 타당하므로 이 과정은 본 연구의 계측 알고리즘에서는 적용하지 않는다.

2-1-7 비 라우드니스 및 총 라우드니스

비 라우드니스(specific loudness) $M[k, n]$ 를 구하는 공식은 다음과 같다^{[9],[11],[16]}.

$$E_{Thres}[k] = 10^{0.364 \cdot \left(\frac{f_c[k]}{1kHz} \right)^{-0.8}}$$

$$s[k] = 10^{\frac{1}{10} \left(-2 - 2.05 \cdot atm \left(\frac{f_c[k]}{4kHz} \right) \right)} \cdot 10^{\frac{1}{10} \left(-0.75 \cdot atm \left(\left(\frac{f_c[k]}{1600Hz} \right)^2 \right) \right)}$$

$$M[k, n] = const \cdot \left(\frac{1}{s[k]} \cdot \frac{E_{Thres}[k]}{10^4} \right)^{0.23} \cdot \left[\left(1 - s[k] + \frac{s[k] \cdot E[k, n]}{E_{Thres}[k]} \right)^{0.23} - 1 \right] \quad (13)$$

식 (13)에서 스케일 상수 const는 1 kHz, 40 dB SPL (음압레벨) 사인 파형이 1sone의 라우드니스 값에 정합하도록 결정하는데 const=1.07664를 사용한다.

마지막으로 전체적인 총 라우드니스는 0보다 큰 모든 피치 그룹의 비 라우드니스 값을 합하여 구하게 된다. 즉,

$$N_{total}[n] = \frac{24}{Z} \cdot \sum_{k=0}^{z-1} \max(M[k, n], 0) \quad (14)$$

2-2 특성 분석

2-2-1 성능 분석

구현한 오디오 레벨 계측 알고리즘의 성능을 분석하기 위하여 2가지 실험을 수행하였다.

등음압레벨을 갖는 순음(pure tone)의 라우드니스 계측 정확도 검증을 위한 ISO226:2003 데이터^[17]와의 비교 분석과 실제 디지털 샘플 오디오 데이터에 대한 음향 레벨 측정 결과 분석이다. 또한 기존 Moore 모델을 적용한 계측 결과도 함께 비교함으로써 구현 알고리즘의 신뢰성을 검증하고자 한다.

순음 적용 실험을 하기 위해 cooledit를 이용하여 ISO226에서 실험한 23개(80 Hz~12,500 Hz)의 주파수에 대한 순음을 모노 형태로 48 kHz 표본 주파수, 16 bit, 0 dBFS을 갖는 0.2초의 디지털 사운드 파일로 만들어 실험한다.

구현 알고리즘의 정확도를 검증해 보기 위해 20 dB부터 80 dB까지 10 dB 간격으로 총 7개의 등음압레벨 그룹에 대해 라우드니스를 계측하여 Moore 모델 계측 결과와 함께 ISO226:2003 데이터와 비교하였으며, 그 결과가 그림 1에 제시되었다.

주요 주파수 영역(315~6,300 Hz)에서 구현 알고리즘이 Moore 알고리즘보다 오차 특성이 좋아서 30 dB 이상에서는 평균 5% 이하의 오차를 보인다. 순음 라우드니스의 유사성 분석을 통한 정확도를 검증

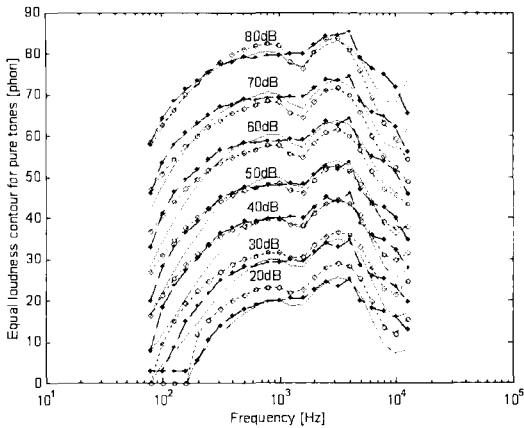


그림 1. 등음압 레벨 순음 라우드니스 선도 비교(-:실선), ISO226, o: Moore, ∴ Implemented)
 Fig. 1. Loudness contour of pure tones for SPLs of 10 dB interval.

하기 위하여 통계적으로 코릴레이션을 구해 보면 ISO-226에 대해 Moore는 0.9887, 구현 알고리즘은 0.9928로서 유사성 분석에서도 구현 알고리즘이 높게 나타나 구현 알고리즘의 특성이 좋음을 알 수 있다.

한편, 샘플 오디오 사운드는 디지털 오디오 컨텐츠 음원으로부터 상이한 오디오 패턴을 12개(2초~5초 구간)를 골라서 최대 사운드 레벨이 0 dBFS이 되도록 cooledit에서 조정하여 적용하였다.

그림 2에 실험에 사용된 디지털 사운드 중 하나의 라우드니스 측정 결과를 제시하였다. 재생 음압 레

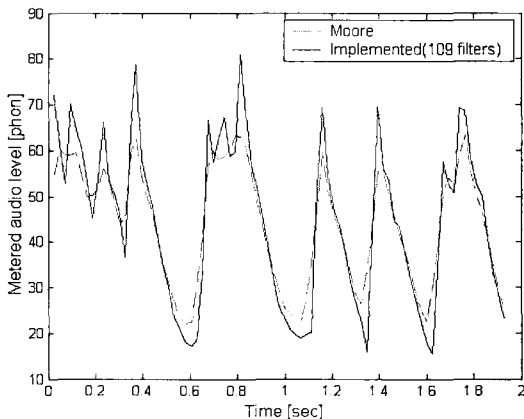


그림 2. 구현 알고리즘 검증을 위한 디지털 사운드의 계측 비교 결과
 Fig. 2. Comparison of Moore algorithm with the implemented one for measuring a sound sample.

벨은 90 dB로 설정하였다. 그림 2에서 보는 바와 같이 구현 알고리즘의 측정 결과가 기존 Moore 알고리즘의 적용 결과와 거의 동일하게 나타난다. 두 알고리즘의 측정 결과간 코릴레이션을 구하여 보면 약 0.95 이상의 좋은 일치도를 보이며, 차이도 평균 5% 이하로 양호하게 나타나고 있다.

2-2-2 계측 시간

구현 알고리즘의 주요 블록별 실행시간을 측정해 보는 것은 실시간화 작업에 상당한 도움이 될 것이다. 이를 위해 디지털 사운드의 한 프레임에 대해 구현 알고리즘이 전체 계측과정을 실행하는 걸리는 시간에 대한 각 주요 블록별 실행시간 점유도를 측정하였다.

주요 블록은 FFT 계산, 청각필터 그룹(auditory pitch)별 에너지 계산, spectral spreading 적용 여기 패턴 계산 등의 세 가지 블록을 대상으로 하였다. 시간 측정 수단은 matlab의 스톱워치 타이머(stopwatch timer) 기능을 이용하였으며, 프로그램 실행 PC의 CPU는 인텔 P4 1.4 G이다. 측정된 결과는 표 1과 같다.

이 결과에서 볼 수 있듯이 청각 필터 그룹별 에너지 계산 블록이 전체 계측 시간에서 차지하는 비중이 절대적이어서 이 블록의 실시간화가 가장 중요하다. 그 이유는 각 그룹에 속하는 선 스펙트럼을 검색하기 위해 매 그룹 에너지를 계산할 때마다 전체 선 스펙트럼을 스캐닝하여야 하기 때문이다.

다음으로 스펙트럴 확산을 적용한 여기 패턴 계산 시간의 비중도 큰 편이어서 이를 실시간화 하는 것도 중요함을 알 수 있다. 여기 패턴의 계산 시간이 큰 이유는 권고안 BS.1387-1의 관련 내용을 자세히 살펴보면 알 수 있겠지만 각 청각 필터 그룹의 에너지가 다른 필터 그룹에 주는 영향을 모두 계산하기 위한 루프 프로그램 비중이 크기 때문이다. 한편, 계측

표 1. 계측 알고리즘 주요 블록별 실행 시간
 Table 1. Calculation time of the key blocks of the implemented algorithm.

블록 항목	블록				
	전체	FFT	Auditory pitch	Spectral spreading	Other blocks
시간(sec)	2.64	0.007	2.25	0.375	0.008
점유도(%)	100	0.27	85.23	14.2	0.3

알고리즘의 나머지 블록들은 실행 시간 점유도가 약 0.3 % 정도로 실시간화에 큰 영향을 주지 않는다.

Ⅲ. 계측 알고리즘의 실시간화

심리음향(psycho-acoustics) 이론상 라우드니스 계측 알고리즘의 틀을 새롭게 바꾸는 것은 거의 불가능하다고 볼 수 있다. 왜냐하면 사람 귀가 소리를 인지하는 메커니즘은 불변하고, 단지 그 과정을 모델링하는 함수 또는 관련 상수값들이 새로운 임상 실험 결과에 따라 바뀔 수 있을 뿐이다. 그러므로 계측 알고리즘을 실시간화 하기 위해서는 알고리즘 내에서의 계산량을 최대한 줄이는 방법을 찾아야 하며, 본 연구에서는 가장 적합한 방법으로서 참조테이블(look-up table)을 이용한 실시간화 방법을 선택하였다.

3-1 참조테이블의 구성

계측 알고리즘의 실시간화를 위해서 가장 먼저 참조테이블을 만들어야 하는 부분은 2-2절에서 보여준 바와 같이 청각 필터 그룹별 에너지와 스펙트럴 확산 적용 여기 패턴 계산 블록이다. 이외에 1-2절의 중이/외이 전달 특성 가중치와 1-4절의 내부 잡음도 참조테이블로 만들어져 적용된다.

3-1-1 청각필터 그룹별 에너지 계산용 참조테이블

청각필터 그룹별 에너지를 계산하려면 각 그룹에 속하는 선 스펙트럼 에너지를 합하면 된다. 그런데 이 과정에서 FFT 계산 포인트 수에 따른 주파수 해상도를 고려하여 청각 필터 그룹의 상하한 경계에 위치한 선 스펙트럼 에너지의 처리를 포함하여 전체적으로 계산하여야 할 경우의 수는 총 4가지로 다음과 같다.

- ① 선 스펙트럼이 청각필터 대역에 완전히 포함
- ② 선 스펙트럼 한 구간(즉, 주파수 해상도)이 청각 필터 대역을 완전히 포함
- ③ 선 스펙트럼 한 구간 내에 청각필터 대역의 하한이 위치
- ④ 선 스펙트럼 한 구간 내에 청각필터 대역의 상한이 위치

표 2. Auditory pitch 에너지 계산 관련 참조테이블
Table 2. Look-up tables related with calculation of auditory pitch energy.

구분	참조테이블 내용
①	- 각 그룹 소속 스펙트럼 라인 번호들(fg1)
②	- 각 그룹 관련 스펙트럼 라인 번호(fg2) - 각 그룹 대역폭 대 주파수 해상도 비율(fgw2)
③	- 각 그룹 관련 스펙트럼 라인 번호(fg3) - 각 그룹대역 하한부분과 해당 선 스펙트럼의 중첩폭 대 주파수 해상도의 비율(fgw3)
④	- 각 그룹 관련 스펙트럼 라인 번호(fg4) - 각 그룹대역 상한부분과 해당 선 스펙트럼의 중첩폭 대 주파수 해상도의 비율(fgw4)

따라서 디지털 사운드의 샘플링 율, FFT 계산 포인트 수, 그리고 임계대역 해상도에 따른 청각필터 대역이 주어질 경우, 4가지 각각의 경우에 대해 참조테이블을 만들어 각 그룹별 에너지를 구할 때 적용한다면 매 그룹별 에너지 계산 시 전체 선 스펙트럼을 스캐닝하지 않아도 되므로 실행 시간을 상당히 단축할 수 있을 것이다. 각 경우에 대한 참조테이블 생성 개요는 표 2와 같다.

3-1-2 Spectral Spreading 적용을 위한 참조테이블

권고안 BS.1387-1에서 spectral spreading을 적용한 여기 패턴 계산 부분이 가장 복잡한 수식으로 표현되어 있으며, 청각필터 그룹별 에너지 계산 다음으로 루프 계산량이 크므로, 이를 참조테이블로 대체하는 것은 계산 시간을 줄이는데 효과적일 것이다.

Spectral spreading은 해당 필터 그룹 에너지가 인근 필터 그룹에 주는 영향을 기울기 함수로 나타낸 것으로, 기울기는 1-5절의 식 (10)에서 보는 바와 같이 필터의 중심 주파수와 에너지 레벨의 함수이다. 따라서 spectral spreading 적용을 위한 참조테이블은 그림 3과 같은 2차원 어레이(array) 구조를 갖게 된다.

여기 패턴의 계산 시에는 기울기 함수의 먹급수가 필요하므로 이를 고려하여 참조테이블을 3차원 구조로 만들 수도 있으나 이는 나중에 DSP 기반의 하드웨어 시스템으로 구현 시 플래시 메모리 용량을 지나치게 크게 요구하게 되므로 바람직하지 않다.

한편, 그림 3에서 보는 바와 같이 참조테이블에서 필터 그룹의 에너지 레벨 해상도는 가변적일 수 있

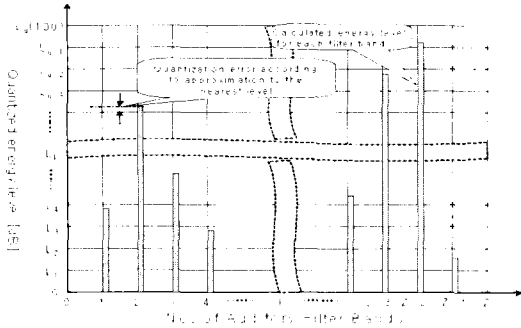


그림 3. Spectral spreading 적용을 위한 참조테이블 구성
Fig. 3. Structure of look-up table constructed by spectral spreading effect.

다. 에너지 레벨 해상도를 너무 작게 할 경우에는 테이블 크기가 너무 커져서 앞서 언급한 바와 같이 하드웨어로 구현 시 부담이 될 수 있으므로 시뮬레이션을 통해 적정 해상도를 찾아야 한다. 그리고 식 (12)의 Norm_sp도 참조테이블로 만들어 적용한다.

이 참조테이블들을 실제 적용하는 방법은 3-2절에서 자세히 언급할 것이다.

3-1-3 기타 참조테이블

계산량을 조금이라도 더 줄이기 위하여 계측 알고리즘의 실행 시간 비중은 작지만 중이/외이 전달 특성 가중치와 내부 잡음도 참조테이블로 만들어져 적용된다.

중이/외이 특성 가중치는 1-2절에서 보는 바와 같이 FFT의 주파수 해상도와 직접적으로 연계되므로 디지털 사운드의 샘플링율이나 FFT 계산 포인트 수가 변하면 참조테이블의 값이나 크기가 변하게 된다.

그리고 내부잡음은 청각필터 그룹의 중심주파수에 대한 함수(1.4절 참조)이므로 109개 청각필터의 중심주파수에 대한 내부잡음 값을 참조테이블로 만들어 적용하게 된다.

3-2 참조테이블을 적용한 계측 알고리즘

계측 알고리즘의 계산과정을 2장의 내용을 따라서 참조테이블과 연계하여 그림 4에 제시하였다.

먼저, 입력 음향 신호를 시간 윈도우 함수를 곱한 후, FFT 처리하는 것은 앞에서의 구현 알고리즘과 동일하며, FFT 스펙트럼에 중이/외이 전달 특성 가

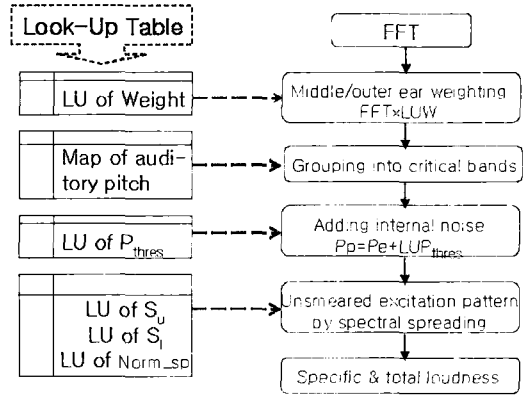


그림 4. 참조테이블을 이용한 계측 알고리즘
Fig. 4. Loudness measuring procedure involved with look-up tables.

중치 참조테이블 값을 읽어 해당 스펙트럼 라인에 곱하게 된다.

두 번째 과정으로 청각 필터 그룹별로 앞에서 설명한 4가지 경우에 대한 참조테이블 내용을 읽어서 관련 선 스펙트럼 에너지에 적용하여 그룹별 에너지 $P_e[k]$ 를 구하게 된다.

이를 좀 더 자세히 설명하면 다음과 같다.

- ①의 경우 :

$$P_e[k] = P_e[k] + F_{sp}[fg1[k]]$$

- ②의 경우 : ③, ④의 경우도 동일

$$P_e[k] = P_e[k] + F_{sp}[fg2[k]] \cdot fgw2[k]$$

위에서 F_{sp} 는 선 스펙트럼 에너지를 의미하며, $fg1$, $fg2$, $fgw2$ 는 Auditory pitch 에너지 계산 관련 참조테이블을 나타내는 표 2의 ①과 ②와 같다.

다음으로 세 번째 과정에서는 앞에서 얻어진 청각 필터 그룹 에너지에 내부 잡음 참조테이블의 값을 각각 더한 후, 네 번째로 spectral spreading을 적용한 여기 패턴을 구하게 된다.

Spectral spreading 관련 참조테이블을 적용할 때, 각 청각 필터 그룹의 에너지 레벨 근처의 참조테이블 값을 읽어서 선형 보간법을 적용하여 정확한 기울기 값을 얻음으로써 정확한 여기 패턴을 구할 수 있을 것이다. 예를 들어 그림 3에서 4번 청각 필터 그룹의 에너지는 참조테이블의 L_2 와 L_3 사이에 위치하므로 4번 청각 필터 그룹 에너지에 상응하는 기울

기 값은 참조테이블의 4번 청각 필터의 L_2 와 L_3 에 해당하는 기울기 값을 읽어서 선형 보간법으로 구할 수 있다. 그러나 보간법은 계산량이 크게 줄이지는 못하므로 계산량을 보다 확실하게 줄이기 위해 그림 3에서 볼 수 있는 것처럼 청각 필터 그룹의 에너지 레벨에 가장 가까운 참조테이블 레벨에 해당하는 값을 읽어서 기울기 값으로 적용하는 일종의 양자화 방법을 택하기로 한다. 결국 이 방법의 성공 여부는 양자화 오차를 줄이는데 있으며, 이를 위해서는 참조테이블의 에너지 레벨 해상도가 어느 정도이냐에 달려 있으며 이 효과를 4장에서 실험을 통해 분석할 것이다.

마지막 과정으로 여기 패턴으로부터 비 라우드니스와 총 라우드니스를 구하게 되는데 이것은 원래 구현한 알고리즘과 동일하다.

IV. 실험결과 및 고찰

4.1 실험의 개요

실시간화 된 계측 알고리즘의 성능과 계측시간 단축 효과를 정량적으로 분석하기 위하여 순음과 디지털 오디오 사운드에의 적용 실험을 수행하였다. 계측시간 비중이 가장 큰 청각 필터 그룹별 에너지 계산과정에서의 참조테이블 결과는 생성된 참조테이블이 실제 오리지널 계측과정을 이용하여 생성되었으므로 당연히 오리지널 알고리즘 적용 결과와 동일하게 나타난다. 따라서 참조테이블을 이용한 청각 필터 그룹별 에너지 계산과정은 계측시간 단축에 얼마나 기여하는가가 훨씬 중요한 의미를 갖는다.

실시간화 된 계측 알고리즘의 실제 성능에 가장 영향을 주는 부분은 spectral spreading 계산과정이다. 왜냐하면 앞서서도 밝혔듯이 참조테이블 적용 시 청각필터 그룹별 에너지 레벨의 근사화로 인한 양자화 오차가 발생할 것이기 때문이다. Spectral spreading 계산량의 감소를 위해 사용될 참조테이블은 109개 청각필터 그룹에 대해 에너지 레벨 해상도가 0.1 dB, 1 dB, 2 dB, 3 dB, 5 dB, 10 dB 등의 총 6가지 경우로 작성하여 각각 적용하고, 그 결과를 근사화 하지 않은 원래 알고리즘의 결과와 비교 분석함으로써 참조테이블 생성 시 적절한 청각필터 그룹 에너지

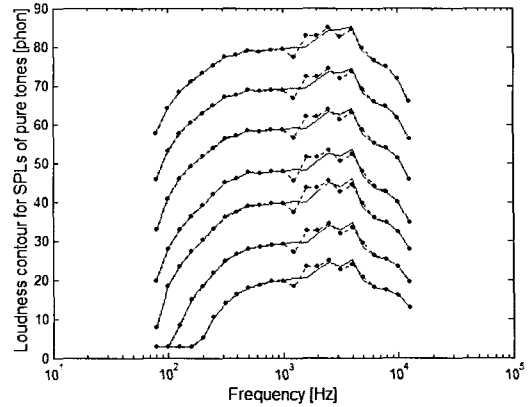


그림 5. 실시간화 된 알고리즘의 순음 테스트 비교 결과(-실선): original, ··: 2 dB 양자화

Fig. 5. Comparison of the proposed algorithm to reduce calculation time with the originally implemented one for testing pure tones.

레벨 해상도를 결정할 근거를 마련할 수 있다.

4.2 순음 및 샘플 디지털사운드 계측 결과 분석

먼저 2장 2절에서처럼 23개의 순음에 대해 실시간화 된 계측 알고리즘을 테스트하고, 그 결과를 실시간화 하지 않은 원래 알고리즘의 계측 결과와 비교하였으며, 그 결과를 그림 5에 제시하였다.

제시한 그래프는 원래 알고리즘과 2 dB 간격으로 양자화 된 spectral spreading 참조테이블을 적용한 알고리즘의 결과를 비교한 것이다. 10 dB 간격으로 양자화된 경우의 결과도 2 dB 경우와 거의 차이가 없게 나타난다.

얻어진 결과로부터 오차와 코릴레이션을 계산하여 보면 원래 알고리즘에 대한 오차가 0.1 dB~10 dB사이의 총 6가지 양자화 된 참조테이블의 모든 경우에서 2% 미만이고, 코릴레이션도 모두 0.98 이상으로 매우 양호하였으며, 특이한 점은 참조테이블의 해상도에 따라 오차나 코릴레이션의 거의 차이가 나지 않는다는 점이다.

이러한 현상이 순음에만 국한된 것인지 실제 디지털 사운드에 대해서도 동일한 결과가 나타나지는지를 확인해 보기 위하여 30개의 샘플 디지털 사운드를 만들어 계측하여 보았다. 30개의 샘플은 방송의 뉴스, 드라마 등, 영화의 특수 음향을 비롯한 다양한 사운드, 음악, 인공음, 자연의 소리 등 다양한 음향

패턴을 갖는 음원을 택하여 cooledit에서 40 KHz, 16 bit, 모노로 조정하여 계측 알고리즘에 적용되었다. 그리고 재생 레벨은 각 샘플당 40 dB, 60 dB, 그리고 80 dB의 세 가지로 설정하여 계측하였다.

30개의 샘플 디지털사운드를 계측한 결과, 원래 계측 알고리즘의 계측 결과와의 코릴레이션을 계산하여 보면 재생 레벨과 6가지 참조테이블 해상도에 관계없이 0.997 이상으로 1에 가까울 정도로 완벽한 라우드니스 패턴의 일치치를 보였으며, 오차도 참조테이블을 적용하지 않은 원래의 알고리즘에 비해 10 dB 레벨 해상도를 갖는 경우까지도 30개 샘플에 대해 오차가 평균 약 2 %에 불과하였다.

그 이유를 알아보기 위해 각 청각 필터 그룹별 spectral spreading을 계산할 때 입력되는 실제 에너지 레벨이 양자화 되므로 샘플 디지털사운드에 대해 참조테이블의 레벨 해상도별로 확산 기울기(spreading slope)와 이에 따른 비 라우드니스(specific loudness) 패턴을 분석하여 보았다. 확산 기울기의 경우 식 (10)과 (11)에 보는 바와 같이 주파수가 낮은 측으로의 확산 기울기는 주파수와 레벨에 무관한 상수이고, 주파수가 높은 측으로의 확산 기울기만 주파수와 레벨의 함수이므로 레벨 해상도에 영향을 받는 upper spreading slopes만 분석한다.

샘플 디지털사운드에 대한 upper spreading slopes와 비 라우드니스 패턴 예시를 그림 6과 7에 제시하였다. 제시된 그래프는 양자화 효과를 확실하게 볼 수 있도록 원래 알고리즘과 2 dB 및 10 dB 레벨 해

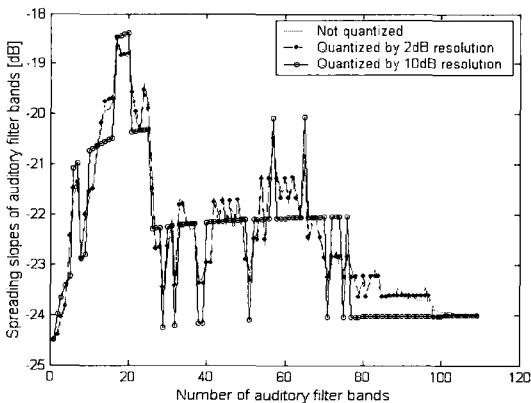


그림 6. 샘플 디지털사운드의 upper spreading slopes
Fig. 6. Upper spreading slopes for digital sound sample.

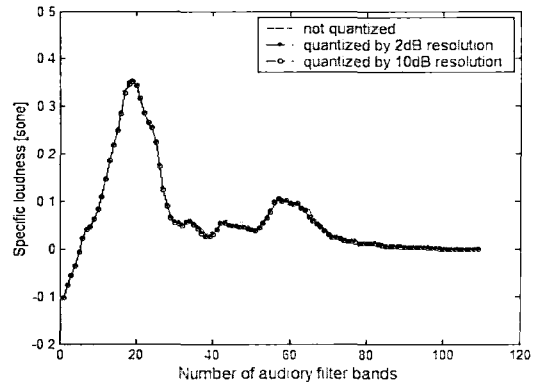


그림 7. 샘플 디지털사운드의 비 라우드니스 패턴
Fig. 7. Specific loudness pattern of digital sound sample.

상도의 참조테이블을 갖는 알고리즘을 비교한 것으로 재생레벨을 60 dB로 설정한 경우이다.

그림 6을 보면 청각 필터 그룹별 에너지의 양자화 효과에 따라 upper spreading slopes의 근사화 현상이 분명하게 보여지고 있다. 그리고 2 dB의 경우에 비해 10 dB의 경우가 보다 양자화 효과가 크게 나타남을 알 수 있다. 그러나 3가지 알고리즘의 결과를 비교하여 보면 최대오차가 -1 dB를 넘지 않는다. 실제 -19 dB와 -20 dB의 1 dB 차이는 선형스케일로 0.0026의 기울기 차이를 의미한다.

비 라우드니스 패턴은 2장에서 언급하였듯이 각 청각필터 그룹 에너지에 각 그룹별 확산 기울기를 곱하고, 식 (12)에 의하여 주어진 지수에 의한 비선형적인 방식으로 확산 효과가 더해져 얻어진 여기 패턴을 이용하여 식 (13)에 의해 구해지게 된다. 이 과정에서 곱해진 확산 기울기의 오차 효과가 상당 부분 감쇠되어 그림 7에서 보는 바와 같이 거의 비 라우드니스 패턴에 영향을 주지 못하는 것으로 파악된다. 비 라우드니스 패턴에 대해 총 6가지의 참조테이블 레벨 해상도에 따른 효과를 살펴보면, 1 dB 이하의 레벨 해상도에 대해서는 비 라우드니스 패턴이 거의 완벽하게 일치하고, 2 dB 이상의 경우에 대해 그림 7에서처럼 부분처럼 약간의 차이를 보여주는 정도였다.

4-3 계측 시간 단축 효과 분석

본 연구의 목적은 사람이 실제 인지하는 오디오

표 3. 실시간화 계측 알고리즘 주요 블록별 실행 시간
Table 3. Calculation time of the key blocks of the proposed algorithm to reduce computation time.

블록 항목	전체	FFT	Auditory pitch	Spectral spreading	Other blocks
시간(sec)	0.125	0.007	0.005	0.105	0.008
점유도(%)	100	5.6	4.0	84.0	6.4
단축효과 (A/B)	21.12	1	450	3.57	1

A: 실시간화 되지 않은 계측 알고리즘(표 1 참조).

B: 실시간화 된 계측 알고리즘.

레벨, 즉 라우드니스를 계측 알고리즘을 실시간화 하는 알고리즘을 개발하는 것이므로 제안한 계측의 복잡도 감소 방법이 계측 시간 단축에 어느 정도 효과가 있는지 검증하는 것은 매우 중요하다.

이를 분석하기 위해 matlab을 이용하여 2장의 2.2 절에서 적용한 것과 동일한 방법으로 실시간화 한 계측 알고리즘의 주요 부분별 실행 시간을 측정하여 보았다. 측정결과를 다음과 같이 표 3에 제시하였다.

표 3에서 보는 바와 같이 청각필터 그룹별 에너지를 구하는 시간이 상당히 감소되어 알고리즘의 실시간화에 큰 기여를 하고 있음을 알 수 있다. 그리고 spectral spreading에 의한 여기패턴을 구하는 부분도 약 3.5배의 실행시간 감소를 보여 참조테이블을 이용한 계측 알고리즘의 실시간화 방법이 상당한 효과를 가짐을 확인할 수 있다.

V. 결 론

디지털 방송의 음향 부문에서는 방송 채널 또는 프로그램간의 오디오 레벨이 시청자가 듣기에 균일하게 느끼도록 하기 위하여 주관적 음향 레벨, 즉 라우드니스를 객관적으로 정확하게 측정할 수 있는 방법이 요구되고 있다. 이러한 라우드니스 계측 알고리즘은 계측 시간이 짧아서 실시간적으로 측정 결과를 보여줄 수 있도록 고안되어야 한다.

본 연구에서는 오디오 품질 분석을 위한 권고안인 ITU-R BS1387-1을 분석하고, 이를 토대로 FFT기반의 라우드니스 계측 알고리즘을 구현하여 그 성능을 검증하였으며, 이 구현된 계측 알고리즘을 실

시간화하기 위한 방법을 제안하였다. 구현한 알고리즘과 제안한 계측 알고리즘은 23개의 순음과 30개의 샘플 디지털사운드에 적용하여 실험함으로써 그 성능과 계측 시간 특성을 분석하였다.

제안한 실시간화 된 라우드니스 계측 알고리즘 방법은 참조테이블(look-up table)을 이용하여 계산량을 줄이고 계산과정을 단순화하는 것이다. 원래 구현된 알고리즘에서 가장 실행시간을 많이 차지하는 부분이 청각 필터 그룹별 에너지를 계산하는 과정인데, 이를 청각 필터 그룹의 주파수 대역 및 관련 파라미터를 참조테이블로 만들어 계산과정을 단순화함으로써 1/450의 시간 단축효과를 가져왔다.

다음으로 실행시간 비중이 높은 부분이 spectral spreading 효과를 고려한 여기패턴 계산과정인데, 이것은 청각필터 그룹에 대한 레벨 양자화를 전제로 한 참조테이블을 구성하여 적용함으로써 약 1/3.57의 시간 감소 효과를 얻었다. 전체적으로는 원래 실행시간의 약 1/21로 계측 시간이 단축되었다.

한편 성능은 실시간화 되지 않은 원래 알고리즘과 비교할 때 제안한 알고리즘의 성능이 거의 차이가 없었다. 특히 여기패턴을 계산하기 위한 spectral spreading 효과 적용 참조테이블의 레벨 해상도에 거의 무관하게 2% 미만의 오차를 갖는 양호한 성능을 보여 제안한 방법이 성능의 열화를 거의 가져오지 않음을 확인하였다.

이제 본 연구 결과를 바탕으로 향후 spectral spreading 부분의 계산량을 좀 더 줄이는 방법을 보완하고, 제안된 라우드니스 계측 알고리즘을 DSP로 직접 구현하여 실험함으로써 완전 실시간 계측이 가능한 라우드니스 미터 구조 설계와 구현이 가능할 것이다.

참 고 문 헌

- [1] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear", *J. Acoust. Soc. Am.*, vol. 66, pp. 1647-1652, 1979.
- [2] E. Terhardt, "Calculating virtual pitch", *Hearing Research*, vol. 1, pp. 155-182, 1979.
- [3] E. Zwicker, H. Fastl, and C. Dallmayr, "Basic-

program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532B", *Acustica*, vol. 55, no. 63, 1984.

- [4] J. Karjalainen, "A new auditory model for the evaluation of sound quality of audio system", *Proceedings of the ICASSP*, Tampa, Florida, pp. 608-611, 1985.
- [5] B. C. J. Moore, B. R. Glasberg, "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns", *Hearing Research*, vol. 28, pp. 209-225, 1987.
- [6] B. C. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, London, 1989.
- [7] E. Zwicker, H. Fastl, *Psycho-Acoustics, Facts and Models*, Springer Verlag, 1990.
- [8] E. Zwicker, H. Fastl, U. Widmann, K. Kurakata, S. Kuwano, and S. Namba, "Program for calculating loudness according to DIN 45631(ISO532B)", *J. Acoust. Jpn(E)*, vol. 12, pp. 39-42, 1991.
- [9] J. G. Beerends, J. A. Stermdink, "A perceptual audio quality measure based on a psychoacoustic sound representation", *J. Audio Eng. Soc.*, vol. 40, pp. 963-978, 1992.
- [10] E. A. Cohen, L. D. Fielder, "Determining noise

criteria for recording environments", *J. Audio Eng. Soc.*, vol. 40, pp. 384-402, 1992.

- [11] J. G. Beerends, J. A. Stermdink, "A perceptual speech quality measure based on a psychoacoustic sound representation", *J. Audio Eng. Soc.*, vol. 42, pp. 115-123, 1994.
- [12] B. C. J. Moore, B. R. Glasberg, "A revision of Zwicker's loudness model", *Acustica*, vol. 82, pp. 335-345, 1996.
- [13] B. C. J. Moore, B. R. Glasberg, and Thomas Baer, "A model for the prediction of thresholds loudness and partial loudness", *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 123-177, 1997.
- [14] Hyuk Jeong, "Sound quality analysis of non-stationary acoustic signals", KAIST Ph.D. Thesis, 1999.
- [15] ITU-R Question 2/6, "Audio metering characteristics suitable for use in digital sound production", 2001.
- [16] Recommendation ITU-R BS.1387-1, "Method for objective measurements of perceived audio quality", 2001.
- [17] International Standard ISO226:2003, "Normal equal-loudness-level contours", 2003.

박 성 균



1985년 2월: 연세대학교 전자공학과 (공학사)
 1987년 2월: 연세대학교 전자공학과 (공학석사)
 1987년 4월 ~ 1989년 8월: 삼성전자연구소
 1994년 2월: 연세대학교 전자공학과 (공학박사)

1994년 9월 ~ 현재: 공주대학교 정보통신공학부 교수
 [주 관심분야] 전파신호처리, 이동통신시스템 분석, 디지털 방송 음향 신호처리