

디지털 오디오 주관적 음향레벨 계측기 구현을 위한 ITU-R BS. 1387-1의 알고리즘 특성 분석

Performance analysis of subjective Loudness meter with ITU-R BS. 1387-1 algorithm for digital audio

뉴엔 보 바오 느안*, 박 성 균** , 노 승 환* , 한 찬 규*

Nguyen Vo Bao Ngan*, Seonggyoon Park** , Soonghwan Ro, Chankyu Han

Abstract

In this paper, the perceived loudness metering algorithm based on ITU-R BS.1387-1 was investigated and implemented, and its performance was evaluated by applying to 23 pure tones and 9 digital audio samples. Error of the tone test results compared with ISO226:2003 was below 5%, and sample test results, in comparison with Moore's algorithm, showed deviation of less than 4.7% and correlation of 0.96. On the other hand, it was investigated how the implemented algorithm's performance was subject to auditory pitch scale. Its result showed that the algorithm with 37 auditory filters, through correcting a bias effect, has a good performance of less than 2% in comparison with the one with 109 auditory filters.

요 약

본 논문에서는 객관적 오디오 품질 분석에 대한 권고안인 ITU-R BS.1387-1의 귀의 모델을 토대로 라우드니스를 객관적으로 측정할 수 있는 알고리즘을 구현하며, 그 성능을 23개의 순음과 9개의 샘플 디지털 사운드에 적용하여 평가하였다. 순음의 경우 ISO226:2003에서 제공하는 실험 데이터와 비교한 결과 5%이하의 오차를 보이며, 디지털 사운드 실험 결과도 Moore 모델의 측정결과와 비교할 때 4.7%이하의 편차와 0.96이상의 코릴레이션을 나타내어 좋은 성능을 보였다. 한편 구현 알고리즘에 적용한 청각필터 수에 따른 성능 변화를 분석하기 위해 21개의 별도의 샘플 디지털 사운드에 적용하였다. 그 결과를 분석하여 보면, 37개의 청각필터를 갖는 구현 알고리즘에 바이어스 보상값을 적용함으로써 109개 필터를 사용하는 경우에 대해 2%미만의 오차를 갖는 양호한 성능을 갖도록 할 수 있었다.

Key words : perceived loudness metering, ITU-R BS.1387-1, auditory filter band

1. 서론

현재 방송국 및 프로그램 제작 스튜디오 장비에

사용되고 있는 오디오 라우드니스 계측기기는 아날로그 방송을 위해 주로 개발된 것이며, 물리적인 음향의 음압레벨을 측정하여 단순한 귀의 주파수 응답에 따른 가중치만 적용한 것이다.

* School of Information and Communication Engineering, Kongju National University)

Corresponding author

Manuscript received Dec. 14, 2012; revised Dec. 19, 2012; accepted Dec. 21, 2012

psk@kongju.ac.kr, 041-521-9198

디지털 방송에서는 전송되는 콘텐츠에 따라 음향레벨의 동적범위가 매우 넓을 수 있다. 예를 들어 뉴스와 같은 음성 위주의 콘텐츠들은 좁은 동적범위의 음향레벨을 갖지만, 영화와 같은 음성과 다양한 음향이 섞여있는 오디오 콘텐츠들은 매우 넓은 동적범위를 가지게 된다. 따라서 방송을 수신하는 청취자

입장에서는 채널 간 또는 방송 프로그램간 천이 시에 본인이 원하는 음향 레벨이 계속 유지되기를 원할 것이다.

이를 위해 최근의 오디오 코딩 시스템을 사용한 수신기는 메타데이터를 이용하여 음향 레벨을 정규화함으로써 보다 균일한 재생 오디오 레벨을 얻을 수 있도록 하고 있다. 이러한 방식이 제대로 동작하기 위해서는 올바른 메타데이터를 생성할 수 있도록 정확한 방송 프로그램의 음향 레벨을 측정할 수 있어야 하며, 측정되는 음향 레벨은 단순히 물리적인 음향 세기가 아니라 인간이 주관적으로 실제 느끼고 인지하는 세기(라우드니스, loudness)이어야 한다.

음향 레벨의 주관적 인식에 관한 연구는 독일의 Zwicker의 연구[1]가 근간을 이루며 1975년도에 국제 표준안 ISO532B로 채택되었고[3][4], 이를 토대로 영국의 Moore[2]가 개선하고 보완하여 1997년도에 모델 [5][6]을 완성하여 제시하였다.

Moore의 모델은 이전의 방대한 임상실험 연구결과를 토대로 외이 및 중이의 전달특성, 청각필터의 임계대역과 중심주파수의 관계 함수를 구하고 이를 반영하고 있다. 그러나 이 모델은 대략 적어도 150개 이상의 청각필터를 요하는 것으로 FFT기반으로 알고리즘을 구현할 경우 계산 시간이 너무 오래 걸린다는 단점이 있다.[5][6]

이에 따라 본 연구에서는 오디오 음질의 객관적 분석을 위한 표준권고안인 ITU-R BS.1387-1[14]의 분석 결과를 토대로 라우드니스 계측 알고리즘을 도출하고, 그 청각필터의 수에 따른 특성을 조사 분석함으로써 경제적이면서 실시간적으로 동작 가능한 주관적 라우드니스를 객관적으로 측정할 수 있는 계측기 구현에 활용 가능성을 점검하고자 하며, 이를 토대로 향후 보완된 보다 나은 음향 레벨 미터, 즉 라우드니스(loudness) 미터를 구현할 수 있는 토대를 마련하고자 한다.

II. 본론

1. ITU-R BS.1387-1의 FFT 기반 청각 모델 개요

권고안 ITU-R BS.1387-1은 오디오 품질을 객관적으로 평가하기 위한 방법을 규정한 것이다. 오디오 품질을 평가하기 위해 계산되어야 하는 가장 기본적인 파라미터로서 여기패턴과 비라우드니스(specific loudness)가 있다. 따라서 이 권고안을 토대로 사람의

인지 라우드니스 계측 알고리즘을 구성할 수 있다.

권고안에서 채용한 귀의 모델은 FFT기반 모델과 필터뱅크기반 모델의 2가지 모델을 모두 사용하고 있다. 필터뱅크 모델은 직류성분(DC) 제거, 두 번의 힐버트 변환과 다운 샘플링의 과정 등을 담고 있어 FFT기반에 비해 상대적으로 좀 더 복잡하다고 판단하여, 본 연구에서는 FFT기반의 모델[14]을 사용하였으며, 그림1에 제시하였다. 이 모델의 자세한 수식과 특징은 참고문헌[15]에 제시되어 있다.

FFT기반 모델의 입력은 48kHz로 샘플된 신호를 0.042초 간격으로 프레임화하고 50%가 중첩되도록 처리한다. 각 프레임은 Hann 프레임 창을 사용하고, 단 푸리에변환(Short Term Fourier Transform)으로 스펙트럼을 구하여 재생레벨로 크기가 조절된다. 이 스펙트럼에 외이와 중이의 주파수 전달 특성을 갖는 주파수 영역의 가중치를 곱한다.

다음으로 내이로 전달되는 스펙트럼을 피치(pitch) 형태로 변환하기 위하여 외이/중이의 가중치가 적용된 스펙트럼을 임계대역(critical bandwidth)으로 그룹화하게 된다. 그리고 주파수의존 오프셋 값이 귀의 내부잡음으로 첨가되고, 스펙트럴 마스크와 시간영역 마스크 효과가 적용된 여기패턴과 비라우드니스 패턴

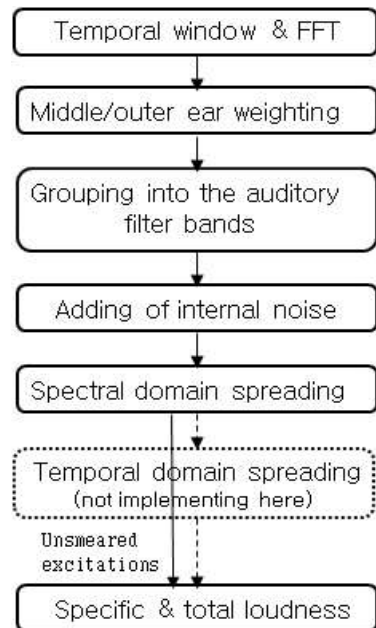


Fig 1. FFT-based ear model of ITU-R BS.1387-1
그림1. ITU-R BS.1387-1 FFT기반 귀의 모델

이 얻어지게 된다.

외이와 중이의 주파수 응답 특성은 주파수의 함수로 나타나는 가중치로 나타나며, FFT가 적용되는 선 스펙트럼에 대한 주파수 응답으로 계산된다. 다음으로 가청주파수 범위를 80Hz에서 18000Hz까지로 규정하고, 청각필터의 수 및 대역을 정의한 후, 앞 단계에서 구한 가중치가 곱해진 선 스펙트럼 에너지로부터 각 청각필터 그룹별 에너지를 산출하는 과정을 거치게 된다.

각 청각필터의 주파수 그룹을 피치라고 하는데, 청각 피치 스케일(auditory pitch scale)은 다음 식(1)으로 표현하고 단위는 'Bark'라고 하지만 Zwicker가 정의한 것과 정확히 같은 스케일은 아니며 근사식이다.[7][10]

$$Bark = 7 \cdot \operatorname{arcsinh}\left(\frac{f/Hz}{650}\right) \quad (1)$$

식 (1)으로부터 각 청각필터의 영역은 각 청각필터의 임계대역 해상도에 의존하며, 이를 함수로 나타내면 다음 식 (2)과 같다.

$$f(i)[Hz] = 80 + 650 \cdot \sinh((i-1) * res/7) \quad (2)$$

여기서 res는 임계대역 해상도를 의미하는 것으로 단위는 Bark이며, i번째 청각필터의 주파수 범위는 하한이 f(i)이고 상한이 f(i+1)이 되고, 중심주파수는 (f(i)+f(i+1))/2이다. 본 연구에서는 0.25, 0.5, 0.75, 1 등 총 4가지 해상도에 대해 청각필터를 구성하고, 각 경우의 음향 레벨 측정 특성을 분석하고자 한다.

따라서 해상도를 선택하고 해당되는 주파수 영역을 참조하여, 외이/중이 주파수 응답 가중치를 적용한 FFT 선 스펙트럼을 각 피치, 즉 각 청각필터 주파수 그룹별 에너지를 구한다.

청각의 임계치 레벨은 주파수의 함수로서 오프셋 값으로 구해져 각 피치 그룹의 에너지에 더해진다.[12] 이로부터 얻어지는 결과를 'Pitch patterns'라 부른다.

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 27dB/Bark이고, 높은 주파수 영역에서는 주파수와 청각필터 그룹 에너지에 따

라 달라진다. 즉, 기울기는 다음 식에 의해 계산된다.

$$S[k, L[k, n]] = -24 - \frac{230Hz}{f_c[k]} + 0.2 \cdot L[k, n]/dB \quad (3)$$

$$S_f[k, L[k, n]] = 27dB/Bark \quad (4)$$

여기서 $L[k, n] = 10 \log_{10}(P_p[k, n])$ 이고, 각 청각필터 그룹의 에너지 레벨이다.

주파수 영역의 확산을 각 주파수(피치) 그룹 k별로 적용하여 얻어지는 여기패턴은 다음 식(5)와 같이 표현된다.

$$E[k, n] = \frac{1}{Norm_{sp}[k]} \left(\sum_{j=0}^{-1} (E_j[j, k, n])^{0.4} \right)^{0.4} \quad (5)$$

이러한 주파수 영역의 확산을 각 주파수(피치) 그룹별로 적용하여 여기패턴이 얻어지게 된다.

한편, 시간영역의 매스킹 효과도 있는데, 계측 알고리즘의 성능을 검증하기 위해 비교할 대상인 기존 라우드니스 계측 모델 중의 하나인 Moore 모델은 정상음을 가정하고 계측하는 모델이어서 시간영역 매스킹 효과를 적용하지 않으므로 본 연구에서 이 효과는 적용하지 않을 것이다.

또한 디지털 방송용 음향 레벨 계측은 2ms 이하의 지속시간을 갖는 급격하게 변하는 과도음 레벨 측정을 주로 하는 것이 아니라 정상음 계측에 가깝다고 보는 것이 타당하므로 이 효과의 배제가 음향 레벨 계측의 정확성에 큰 문제가 되지는 않는다.

마지막으로 각 프레임의 총 라우드니스는 0보다 큰 모든 피치 그룹의 비라우드니스 값을 합하여 구하게 된다.

$$N_{total}[n] = \frac{24}{Z} \cdot \sum_{k=0}^{Z-1} \max(N[k, n], 0) \quad (6)$$

여기서 Z는 필터뱅크의 수이고, $\max(N(k, n), 0)$ 는 본 알고리즘에서 비라우드니스 N(k,n)값이 계산상에서 음의 값을 가질 수가 있는데, 실제로는 음의 값이 물리적 의미가 없으므로 영의 값으로 만들기 위한 것이다.

2. 알고리즘 적용 시뮬레이션

본 연구에서는 음성 파일은 모노, 16bit, 48kHz로 샘플링된 .wav 포맷 데이터 신호를 사용한다. 이 신호를 약 40msec 크기로 프레임을 만들기 위해 한 프

레이의 데이터 포인트 수는 2048로 한다. 그리고 중첩율이 50%이므로 1024개 포인트가 중복된다. 이를 수식으로 표현하면 다음과 같다.

$$[k_i, n] = t[1024 \times n + k_i] \quad (7)$$

여기서 n 은 프레임 번호이고, k_i 는 한 프레임내의 카운터이다. 그리고 재생 레벨 조정(Playback level scaling)을 위해 적절한 가중치를 적용하여야 하는데 다음 식과 같이 계산되는 가중치를 FFT 처리된 프레임 데이터에 곱하면 된다.

$$fac = \frac{10^{20}}{orm} \quad (8)$$

여기서 L_p 는 재생될 음압 레벨(SPL: Sound Pressure Level)이고, 'Norm'의 값은 입력 신호를 주파수 1019.5Hz이고, 0dB 플 스케일인 사인 파형을 적용하여 10프레임에 걸친 FFT의 주파수 계수의 최대값으로 사용한다. 시뮬레이션 결과 Norm은 0.151로 결정되었다.

본 연구에서 구현한 음향 레벨 계측 알고리즘의 유용성을 검증하기 위하여 2가지 실험을 수행하였다.

먼저 등음압레벨을 갖는 순음(pure tone)의 라우드니스 계측 정확도를 검증하기 위하여 ISO226:2003의 데이터와 비교하는 것이고, 다음으로 실제 디지털 샘플 오디오 데이터에 대한 음향 레벨 측정결과를 기존 Moore 모델을 본 연구자가 ISO226:2003에 적합하게 개선한 알고리즘의 측정결과와 비교하여 봄으로써 그 유용성을 검증하고자 한다.

한편, 구현 알고리즘에서 귀의 임계대역 해상도에 따른 청각필터 수와 복잡도 및 성능에의 영향을 분석할 필요가 있다. 이에 따라 임계대역 해상도 및 청각필터 수를 달리하여 구현된 알고리즘에 대해 순음과 샘플 오디오 사운드를 적용한 실험도 수행되었다.

가. 등음압레벨 순음 적용 실험

순음 적용 실험을 하기 위해 음향관련 소프트웨어인 Cooledit를 이용하여 ISO226에서 실험한 23개(80hz~12500hz)의 주파수에 대한 순음을 만든다. 이 순음은 모노 형태로 48kHz 표본 주파수, 16bit, 디지털레벨 0dBFS를 갖는 0.2초의 디지털 사운드 파일로

만들어져 구현된 음향 레벨 알고리즘에 입력된다.

구현 알고리즘의 정확도를 검증해보기 위해 20dB부터 80dB까지 10dB간격으로 총 7개의 등음압레벨 그룹에 대해 라우드니스를 계측하여 Moore 모델 계측 결과와 ISO226:2003 데이터와 비교하였으며, 그 결과가 그림 2에 제시되었다. 구현 알고리즘은 임계대역 해상도를 0.25로 하여 청각필터 수를 109개로 하여 구현된 것이다.

결과를 살펴보면 Moore 모델이나 구현 알고리즘의 순음 측정 결과가 음압레벨에 대해 ISO226:2003 데이터 곡선과 비슷한 추이를 보인다. 구체적으로 살펴보면 순음 응답 곡선의 315hz이하의 낮은 주파수 영역과 10khz근처의 높은 주파수 영역에서는 오차가 크게 나타난다.

Moore 알고리즘은 음압레벨이 40dB일 때 가장 정확하고, 20과 30dB에서는 음압레벨이 낮을수록 순음 라우드니스가 크게 측정되고, 50dB이상의 높은 음압레벨에서는 측정치가 약간 낮게 나타나고 있다. 구현 알고리즘은 모든 음압레벨에서 대체로 일정한 순음 라우드니스 곡선을 보여준다. 1kHz와 2kHz 사이의 주파수 영역에서 ISO226와 비교할 때 선명한 골짜기 모양의 특성을 보이지 않다는 것과 10kHz 주파수 영역에서의 불규칙한 특성 곡선 모양을 따르지 않는다는 점을 제외할 때 비교적 좋은 일치성을 보여준다.

구체적인 수치로 비교해보기 위하여 음압레벨별로 순음 라우드니스 측정결과에 대해 ISO226의 데이터를 기준으로 측정 오차를 구하였으며, 그 결과가 그림 3과 같다.

그림 3은 Moore와 구현 알고리즘을 이용한 측정결과를 23개의 순음 주파수 전체에 대해 음압레벨별로 순음 측정 오차를 구한 것과 실제 귀의 주파수 응답 가중치가 비교적 높아서 디지털 사운드의 라우드니스에 영향을 크게 주는 주파수 영역인 315Hz~6300Hz에서의 측정 오차를 구한 것을 그래프로 제시한 것이다.

전체 주파수 영역에 대한 측정 오차를 살펴보면 Moore보다 구현 알고리즘이 낮은 레벨에서 오차가 더 작으며, 높은 음압레벨일수록 오차가 비슷해지며, 50dB 이상의 음압레벨에서 평균 10%이하의 오차를 보인다. 주요 주파수 영역(315~6300Hz)에서는 구현 알고리즘이 Moore알고리즘 보다 더욱 오차 특성이 좋아서 30dB이상에서는 평균 5%이하의 오차를 보인다.

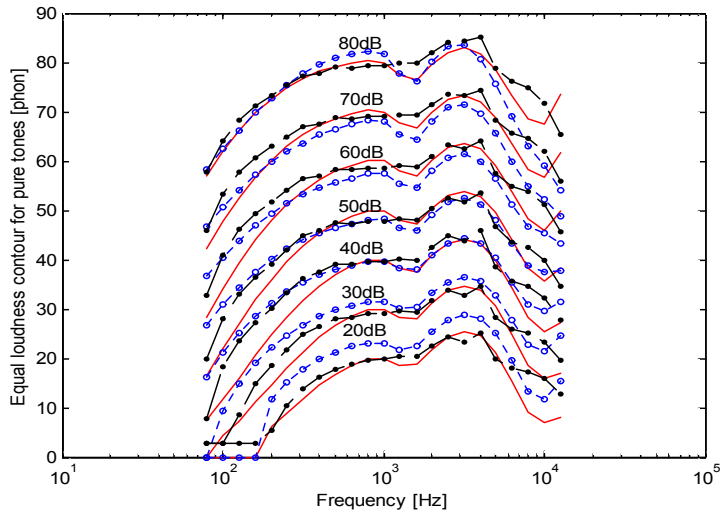


Fig 2. Loudness comparison of pure tones with same sound pressure level

그림 2. 등음압레벨 순음 라우드니스 선도 비교
(실선:ISO226, o:Moore, • : Implemented)

순음 라우드니스의 유사성 분석을 통한 정확도를 검증하기 위하여 통계적으로 코릴레이션을 구해보면 ISO226에 대해 Moore는 0.9887, 구현 알고리즘은

0.9928로서 유사성 분석에서도 구현 알고리즘이 높게 나타나 본 연구에서 구현한 알고리즘의 특성이 좋음을 알 수 있다.

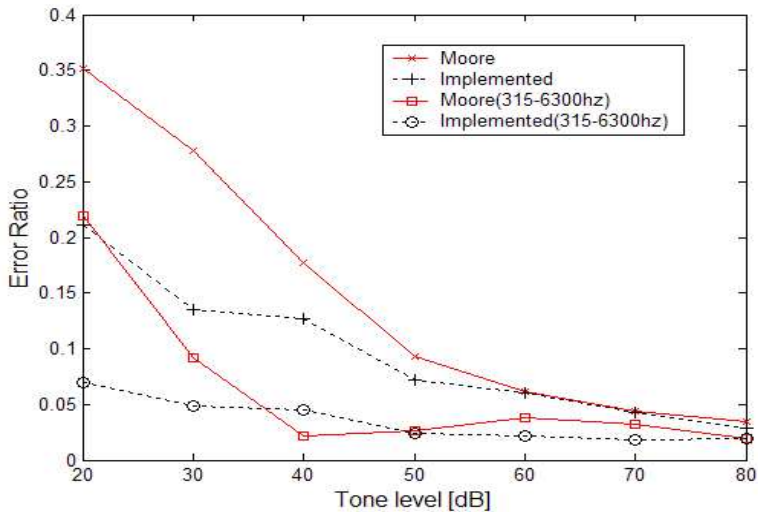


Fig 3. Comparisons of the calculated loudness for pure tones

그림 3. 순음 라우드니스 측정 오차 비교

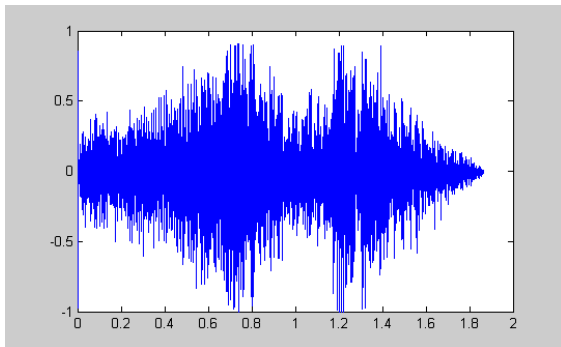
나. 샘플 오디오 사운드 적용 실험

샘플 오디오 사운드는 디지털 오디오 콘텐츠 음원으로부터 상이한 오디오 패턴을 9개 정도(2초~5초 구간)를 골라서 최대 사운드 레벨이 0dBFS이 되도록 Cooledit에서 조정하였다. 선택한 사운드들은 다양한 스펙트럼 패턴을 갖는 인공음(잔디깎기 기계, 폭죽, 종소리 등), 자연음(동물, 벌레), 사람의 육성 등이다.

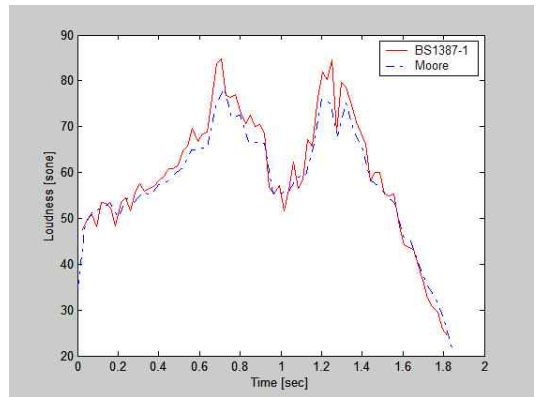
그리고 구현 알고리즘을 샘플 오디오 사운드에 적용한 결과를 비교 검증하기 위하여 순음의 경우와 마찬가지로 Moore 알고리즘을 선택하였다.

그림 4에 실험에 사용된 디지털 사운드 중 하나의 파형과 그 음향 레벨(라우드니스) 측정 결과를 제시하였다. 재생 음압 레벨은 90dB로 설정하였다. 그림 4에서 보는 바와 같이 구현 알고리즘의 측정결과가 기존 Moore 알고리즘의 적용결과와 거의 동일하게 나타난다. 두 알고리즘의 측정결과간 코릴레이션을 구하여 보면 0.96이상의 좋은 일치도를 보이며, 차이도 평균 4.7%정도로 양호하게 나타나고 있다.

즉, 구현 알고리즘의 계측 성능이 디지털 사운드에 충분히 적용할 수 있음을 알 수 있다.



(a) 샘플 파형



(b) 음향 레벨 측정 결과

Fig 4. Comparisons of the calculated levels for sample digital audio sounds

그림 4. 샘플 오디오 사운드 음향 레벨 측정 결과 비교

3. 청각 필터 수에 따른 계측 성능 분석

앞에서 언급한 바와 같이 본 연구에서 구현한 알고리즘은 청각필터 수를 다르게 할 수 있다. 직관적으로 생각할 때 청각필터의 수가 증가하면 당연히 정확성이 더욱 좋아질 것이라 생각할 수 있다. 그러나 알고리즘의 복잡도와 실행시간은 증가할 것이다.

실제로 matlab의 스톱워치 타이머(stopwatch timer) 기능을 이용하여 구현 알고리즘의 주요 블럭별 실행시간 점유도를 분석하여 보면 디지털 사운드의 한 프레임에 대해 계산을 수행할 경우, temporal 윈도우 및 FFT계산 블럭은 0.38%, 청각필터 그룹별 에너지 계산 블럭은 85.23%, spectral spreading 적용 여기패턴 계산 블럭은 14.2%이다. 이 결과를 통해서도 알 수 있듯이 청각필터의 수는 구현 알고리즘의 실행시간에 가장

큰 영향을 주고 있다. 따라서 성능과 실행시간은 상호 상충되는 설계 목표가 되므로 이를 trade-off하기 위해서는 청각필터의 수에 따른 성능 변화를 측정하여 불필요가 있다.

청각필터의 수는 임계대역 해상도를 0.25, 0.5, 0.75, 1 등 총 4가지로 하여 구하여 보면 109개, 55개, 37개, 28개가 각각 얻어진다. 앞 절에서 109개인 경우는 살펴보고 나머지 3가지 경우에 대해 순음의 라우드니스 측정하고 ISO226에 대한 측정 오차를 구하여 제시한 것이 그림 5이다.

측정오차를 살펴보면 음압레벨이 낮을수록 청각필터의 수가 적을수록 오차가 크게 나타나며, 음압레벨이 50dB이상으로 높아짐에 따라 오차율이 5%이하로 수렴함을 볼 수 있다.

실제 등음압레벨을 따라 순음 라우드니스 선도를

그러보면 청각필터의 수가 작을수록 주요 주파수 대역(315~6300Hz)에서 등고선 모양의 요동(fluctuation)이 커지면서 ISO226의 곡선 모양으로부터 점점 많이 벗어나는 것을 확인할 수 있을 것이다. 이것은 청각필터의 수가 줄어들게 됨으로써 주파수 분해 능력이 떨어지게 된다는 것을 의미하는 것이다.

이러한 현상을 등고선 모양의 패턴 유사도를 측정함으로써 확인할 수 있는데 통계적으로 ISO226에 대한 청각필터 수에 따른 측정결과와 코릴레이션을 구하여 보면 109개, 55개, 37개, 28개의 각각의 경우에 대해 0.9928, 0.9913, 0.9893, 0.9528로 나타나며 28개인 경우 코릴레이션 열화의 폭이 커짐을 알 수 있다.

한편, 청각필터의 수를 달리하여 구현한 알고리즘을 샘플 오디오 사운드에 적용하여 그 결과를 분석함으로써 실제적인 청각필터 수의 영향을 분석하고자 한다.

여기서 측정할 샘플 오디오 사운드는 앞 절에서 사용한 사운드와는 다른 것으로 음악, 라디오 방송, 영화 등에서 다양한 샘플을 얻어 cooledit로 모노, 16bit, 48kHz 형식으로 조정하여 적용하였다. 사용한 디지털 음원은 총 21개이다. 구현 알고리즘으로 측정할 때 재생 레벨은 실제 가정에서의 일반적인 재생 레벨로 받아들여지는 60, 70, 80dB의 3가지의 레벨을 적용하였다.

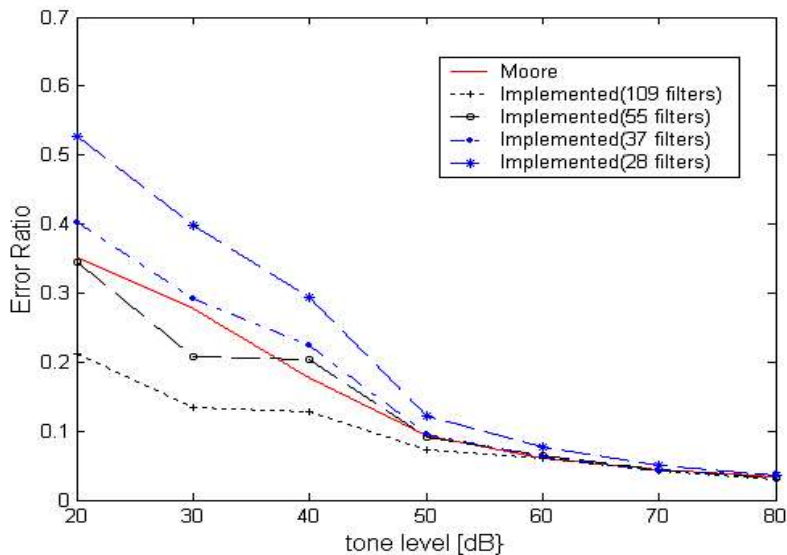


Fig 5. Comparisons of calculation errors for pure tones according to number of auditory filter banks

그림 5. 청각필터 수에 따른 순음 라우드니스 선도 측정오차 비교

60과 80dB의 측정 결과를 토대로 109개 필터를 기준으로 측정 오차를 구하여 그림 6에 제시하였다. 측정 오차를 잘 살펴보면 55개 필터인 경우가 가장 측정 오차가 적고 성능이 우수한 것은 예상대로이지만, 28개 필터인 경우가 37개 필터인 경우보다 측정 오차가 더 우수하게 나타난 것은 예상과 정반대의 결과이다.

이러한 현상이 발생한 이유를 정확히 알아보기 위해서 109개 필터에 대한 나머지 경우의 측정결과간 코릴레이션을 구하여 분석하였다. 그 결과, 코릴레이션

값이 37개 필터의 경우 모두 0.99이상으로 나타나는 반면, 28개 필터의 경우는 0.98인 경우도 나타나서 37개 필터인 경우가 더 좋게 나타난다. 이 분석 결과를 토대로 추정하여 보면 각 샘플 오디오 사운드에 대한 라우드니스 패턴의 일치도는 청각필터의 수가 많을수록 더욱 좋으나, 청각필터 수의 변화에 따라 실제 라우드니스 값에서 일정한 바이어스(bias) 현상이 발생하는 것으로 생각할 수 있다.

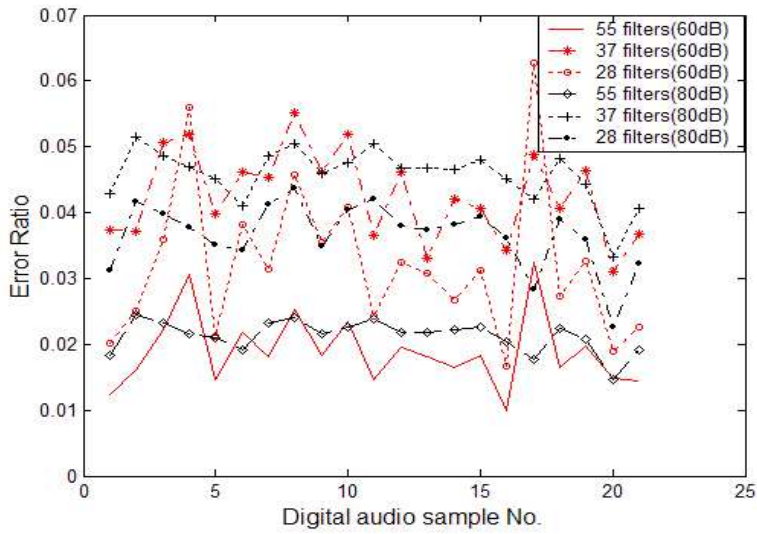


Fig 6. Comparisons of calculation errors for 21 sample digital audio sounds

그림 6. 21개 샘플 오디오 사운드에 대한 측정 오차 비교

이를 확인하기 위하여 각 샘플 오디오 사운드에 대한 라우드니스 패턴과 평균치를 구하여 그려보면 그림 7의 예시와 같이 나타나게 된다. 그림7에서 보는 바와

같이 109개 필터의 경우와 37개 필터인 경우가 라우드니스 패턴은 거의 동일하지만 평균값의 차이만큼 바이어스되어 있음을 알 수 있다.

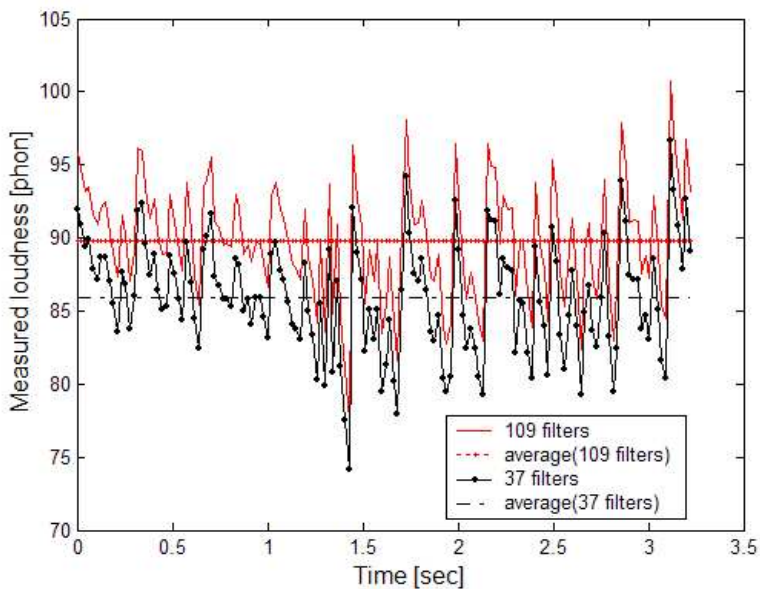


Fig 7. The estimated bias effect of calculated loudness patterns from varying number of auditory filter banks

그림 7. 청각필터 수의 변화에 따른 측정 라우드니스 패턴의 바이어스 현상

그림 7의 경우 109개 필터의 평균치에 대한 차이만큼 바이어스 효과를 보상하여 주면 37개 필터의 경우의 측정 오차가 4.28%에서 0.31%로 줄어든다는 것을 확인할 수 있었다.

그림 6에서 측정 오차율의 패턴을 살펴보아도 80dB 재생레벨에 대해서는 37개 및 28개 필터인 경우 모두 동일한 패턴과 작은 편차를 보이고 있으나, 60dB 재생레벨에서는 28개 필터인 경우는 37개인 경우에 비해 상대적으로 측정 오차가 큰 편차를 보이고 있다. 큰 편차를 보인다는 것은 일률적인 바이어스 보상값을 결정하기 어렵다는 것을 의미한다.

결론적으로 37개 필터인 경우는 구현 알고리즘에 바이어스 보상값을 적용함으로써 55개 필터인 경우와 거의 동일한 성능을 가질 수 있으며 109개 필터에 비해 2% 미만의 오차를 가지는 양호한 성능을 보여준다. 28개 필터인 경우는 모든 재생레벨에서 일정한 바이어스 보상을 적용하기 어려우므로 37개 필터에 비해 상대적으로 부적합한 것으로 판단된다.

III. 결론

현재 디지털 방송의 오디오 부분에서는 방송 채널 또는 방영 프로그램들간에 균일한 재생 주관적 인지 라우드니스를 유지할 수 있는 음향 레벨 조정이 중요하다. 이를 위해서는 주관적 음향레벨, 즉 라우드니스를 객관적으로 정확하게 계측할 수 있는 방법이 필요하며, 또한 가능하면 복잡도가 낮은, 즉 실행시간이 짧은 알고리즘 구현이 요구되고 있다.

본 연구에서는 ITU-R의 표준안 중에서 음질 분석을 위해 권고된 BS1387-1을 분석하고, 이를 토대로 라우드니스 계측 알고리즘을 구현하였으며, 순음 신호와 일반 디지털 오디오 사운드에 적용하여 그 성능을 검증하고, 구현 알고리즘의 실행시간에서 가장 큰 비중을 차지하는 청각필터의 수를 가변시켜 그에 따른 성능 분석을 수행하였다.

23개의 순음에 대한 계측 결과를 ISO226:2003의 데이터와 비교하였을 때 109개 필터를 갖는 구현 알고리즘의 정확도가 Moore 알고리즘보다 다 좋았으며, 또한 9개의 샘플 디지털 오디오 사운드에 적용한 결과를 Moore 알고리즘과 비교하여 보면 평균 4.7%이하의 오차를 갖고 0.96이상의 코릴레이션 값을 갖는 좋은 성능을 보여 디지털 사운드의 라우드니스 계측에 충분히 사용할 수 있음을 확인하였다.

한편 21개의 샘플 오디오 사운드를 준비하고 109개,

55개, 37개, 28개의 청각필터를 갖는 구현 알고리즘을 각각 적용하여 성능 분석한 결과, 청각필터 수가 줄어들면서 라우드니스 바이어스 현상이 발생하므로 바이어스 보상값을 적용하면 적은 필터로도 좋은 성능을 구현할 수 있을 확인하였다. 28개 필터인 경우는 측정 오차의 편차가 크므로 37개 필터를 갖는 구현 알고리즘에 바이어스 보상값을 적용하여 보면 109개 필터를 갖는 경우에 비해 2%미만의 오차를 갖는 양호한 계측 알고리즘을 얻을 수 있다.

References

- [1] Zwicker E. and Fastl H., "Psycho-acoustics, Facts and Models", Springer Verlag, 1990.
- [2] Moore B.C., "An introduction to the psychology of hearing", Academic Press, London, 1989.
- [3] Zwicker E., Fastl H., 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] Zwicker E., Fastl H., 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.
- [5] B.C.J. Moore and B.R. Glasberg, "A Revision of Zwicker's Loudness Model," Acustica, vol.82, pp.335-345, 1996.
- [6] B.C.J. Moore, B.R. Glasberg, 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.
- [7] Schroeder M.R., Atal B.S. and Hall J.L. "Optimizing digital speech coders by exploiting masking properties of the human ear", J. Acoust. Soc. Am., Vol. 66, pp.1647-1652, 1979.
- [8] Beerends J.G. and Stermdink J.A., "A perceptual audio quality measure based on a psychoacoustic sound representation", J. Audio Eng. Soc., Vol. 40, pp.963-978, 1992.
- [9] Beerends J.G. and Stermdink J.A., "A perceptual

speech quality measure based on a psychoacoustic sound representation", J. Audio Eng. Soc., Vol. 42, pp.115-123, 1994.

- [10] Terhardt E., "Calculating Virtual Pitch", Hearing Research, Vol. 1, pp.155-182, 1979.
- [11] Karjalainen J., "A new auditory model for the evaluation of sound quality of audio system", Proceedings of the ICASSP, Tampa, Florida, pp.608-611, 1985.
- [12] Cohen E.A. and Fielder L.D., "Determining noise criteria for recording environments", J. Audio Eng. Soc., Vol. 40, pp.384-402, 1992.
- [13] International Standard ISO226:2003, "Normal equal-loudness-level contours", 2003
- [14] Recommendation ITU-R BS.1387-1, "Method for objective measurements of perceived audio quality", 2001
- [15] Seonggyoon Park, "A Study on Real-Time Loudness Metering Algorithm for Digital Broadcasting", Journal of KEES, vol.16, No.4, pp.427-438, 2005

BIOGRAPHY

Nguyen Vo Bao Ngan (Student Member)



He received a B.S degrees from the Department of Physics and Electronics at University of Science Hochiminh city, Vietnam in 2010. He had worked for Renesas Design Vietnam Co.,Ltd for one year.

Since September 2011, he has started to study a master program of Information and Communication Engineering at Kongju National University, Korea. His research interests include mobile communication and embedded systems.

Park Seonggyoon (Member)



1985 : B.S. degree in Electronics Engineering, Yonsei University.
1987 : M.S. degree in Electronic Engineering., Yonsei University.
1994 : Ph.D. degree in Electronics Engineering, Yonsei University

1994.9 ~Present : Faculty of Department of Information and Communication at Kongju National University

<Research interest> communication signal processing, Spectrum engineering

Soonghwan Ro (Member)



He received B.S., M.S., and Ph.D degrees from the Department of Electronics Engineering at Korea University in 1987, 1989, and 1993, respectively.

He was a research engineer of Electronics and Telecommunications Research Institute and University of Birmingham in 1997 and 2003, respectively. Since March 1994 he has been a professor at Kongju National University, Korea. His research interests include pervasive computing, mobile communication, and embedded systems.

Han Chankyu (member)



1991 : B.S. degree in Electronics Engineering, Kwandong University.
2001 : M.S. degree in Information and communication Engineering., Kongju National University.

2011 : Ph.D. course , Kongju National University

2005.6~Present : STA system corp.

<Research interest> Spectrum engineering, Mobile communication, signal processing