

청각 특성을 고려한 적응적인 오디오 음량 자동 제어 기술 개발

*이영한 **조충상 ***김제우

전자부품연구원

{*yhlee, **ideafisher, ***jwkim}@keti.re.kr

Development of Adaptive Audio Loudness Control Technology based on Human Acoustic Characteristics

*Young Han Lee **Choongsang Cho ***Je Woo Kim

Korea Electronics Technology Institute (KETI)

요약

2016년 이후 방송법 개정을 통해 디지털 방송의 음량을 ITU-R / EBU에서 제안한 측정 방식을 활용하여 채널 및 프로그램 사이의 방송 음량을 맞추어 제공하고 있다. 본 논문에서는 방송 음량을 조절하는 기술 중의 하나로 ITU-R 1770-3 측정 방식을 기반으로 하여 청각적 특성을 고려한 적응적 오디오 음량 자동 제어 기술을 제안한다. 오디오 음량 자동 제어 기술은 프로그램의 음량을 기준치에 맞추는 동시에 오디오 데이터의 왜곡을 최소화해야 한다. 제안한 기술은 음량을 기준인 -24 LKFS에 맞추는 동시에 왜곡의 최소화하면서 명료도를 높이는 것을 목표로 개발되었다. 이를 위해, 가청/비가청 구간에 따라 적응적으로 이득을 조절할 수 있는 구조를 개발하였다. 제안한 방식의 성능을 확인하기 위해 주관적 음질 평가 방식을 실시하였으며 이를 통해 기존 음량 제어 기술과 비교하여 음질이 향상됨을 확인하였다.

1. 서론

디지털 방송으로의 전환 후, 디지털 방송 음량에 대한 기준이 없는 상태로 방송을 하면서 TV 시청자들은 채널 간 또는 프로그램 간의 전환 시에 오디오 음량 레벨의 급격한 변화로 인해 많은 불편을 겪었다. 이를 해결하기 위해서 우리나라에서는 2014년 5월 방송법을 개정하여 '디지털 텔레비전 방송프로그램 음량 등에 관한 고시'를 통해서 2016년 5월부터 국제 권고 수준인 평균 음량 -24 LKFS 수준에 맞춰 방송 프로그램을 송출하도록 규제하고 있다 [1].

우리나라의 방송 프로그램의 음량 기준은 방송 프로그램에서 주관적으로 인지하게 되는 음량을 객관적으로 표시할 수 있는 방송 음량 측정 방법인 BS.1770-3 [2]을 기반으로 미국, 캐나다, 일본, EU 등의 법률과 기준을 참고하여 ITU 권고 기준에 준수하고 있다.

현재 국내를 비롯한 국외에서는 방송 프로그램의 음량 기준을 준수하기 위해 다음의 세 가지 방법으로 음량 기준을 준수하고 있다. 첫 번째는 방송 프로그램의 제작 단계에서 음량 기준에 맞춰서 오디오의 음량을 조절하여 방송 프로그램을 제작하는 방법이고, 두 번째는 방송 송출 바로 전단계인 송출용 오디오 인코더(일반적으로 비디오 인코더 내에 포함) 앞에서 '라우드니스 조절 장비'를 설치하여 실시간으로 송출되는 방송 프로그램의 음량을 실시간으로 자동 조절하는 방법이다. 마지막으로 기존 제작된 방송 프로그램을 '라우드니스 변환 장비'를 사용하여 파일기반으로 제작된 방송 프로그램의 음량을 기준에 맞춰 재생하는 방법이다. 라우드니스 조절 장비를 이용하는 방법은 실시간

으로 입력되는 오디오 데이터를 이전 입력 오디오 데이터로부터 예측하여 음량을 제어하기 때문에 원 음원의 왜곡이 불가피하게 발생하게 된다. 하지만, 라우드니스 변환 장비의 경우에는 일반적으로 프로그램의 전체 오디오 데이터를 입력받아 먼저 전체 오디오의 음량을 측정된 후에 음량 기준(-24 LKFS)에 맞춰 선형적으로 음량 이득(Gain)을 조절하므로 원 음원의 왜곡 없이 음량을 조절할 수 있는 장점을 갖는다.

그러나, 기존 제작된 방송 프로그램들은 음량 기준에 대한 고려 없이 프로그램을 제작한 경우가 많기 때문에 '라우드니스 조절 장비' 또는 '라우드니스 변환 장비'를 적용하여 출력되는 오디오 데이터들은 "사람의 목소리가 작아져서 소리의 명료도 및 의사 전달 기능이 저하"되는 경우가 다수 발생하고 있다. 즉, 최근 방송 프로그램의 오디오 음량에 대한 시청자들의 불만은 규제 이전의 '채널 간 또는 프로그램 간의 음량 불균형'에서 규제 이후 '전체적인 음량의 감소 또는 음성의 상대적 음량 감소'로 인한 명료도 하락'으로 전환되었다.

본 논문에서는 이러한 시청자들의 불만을 해소하기 위해서 상대적으로 음성의 명료도를 강화할 수 있는 사람의 청각 특성을 고려하여 가청음과 비가청음을 분석하여 적응적으로 오디오 음량을 제어하는 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 ITU-R BS.1770-3 권고 표준의 방송 음량 측정 알고리즘 개요에 대해 기술하고, 3절에서는 본 논문에서 제안하는 가청음/비가청음 결정 기법 및 적응적 음량 제어 알고리즘, 그리고 구현된 알고리즘의 성능 결과를 기술하고, 마지막으로 본 논문에 대한 결론을 맺는다.

2. ITU-R BS.1770-3 음량 측정 기법 개요

사람이 실제로 느끼는 소리의 감각적인 크기를 음량이라고 하는데, ITU에서는 디지털 방송 프로그램의 음량 측정을 위해 사람의 청각 특성을 고려하여 사람이 실제로 느끼는 소리의 감각적 크기를 [2]의 음량(loudness)이라는 개념으로 수치화하여 LKFS (Loudness, K-weighted relative to Full Scale) 측정 단위로 측정하였다.

그림 1은 [2]의 방송 음량 측정 알고리즘 블록도를 보인 것으로, 예로서 5.1채널 방송 프로그램 음량 측정 방법을 보인 것이다. 5.1채널의 경우에 저주파효과(LFE)는 음량 측정에서 제외된다. 그림 1에 보이는 것과 같이, 방송 음량 측정은 pre-filter와 RLB filter로 구성된 K-filter, 채널 가중치 모듈, 그리고 Gate 모듈로 구성된다.

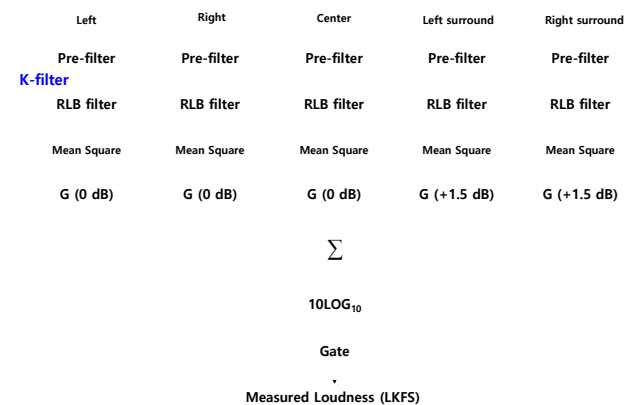
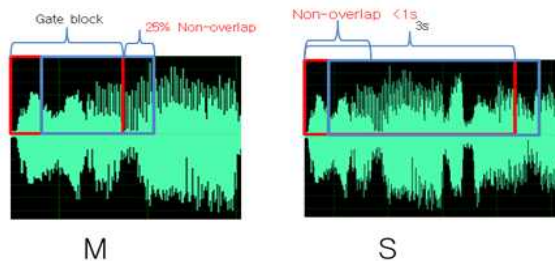


그림 1. ITU-R BS.1770-3의 음량 측정 알고리즘 블록도



(a) M-LKFS

(b) S-LKFS

그림 2. 음량 측정 단위 (Gate block, EBU R128 기준)

[2]의 음량 측정 최소 단위는 gate block으로, gate block으로 400ms 길이의 오디오 PCM 데이터를 의미한다. 이 단위를 유럽방송연합(EBU)에서 오디오 음량의 측정 및 정규화를 위한 권고안 R128 [3]에서는 momentary (M) -LKFS로 정의하고 있다. 먼저 데이터가 입력되면 400ms의 데이터에 대해서 음량을 측정하고, 그림 2-a)와 같이 다음 gate를 위해 75% overlap shift가 이루어진다. 다음 gate는 75% overlap된 데이터와 새로 추가된 25%의 데이터로 구성되어 음량을 측정하게 된다. 즉, 각 M-LKFS는 100 ms 단위로 계산되어 결과 값이 출력된다.

또한, EBU R128에서는 그림 2-b)에 보이는 것과 같이 3초의 오디오 데이터에 대해서 M-LKFS 값을 평균하여 short-term (S) -LKFS라고 정의하고 있으며, 이 또한 M-LKFS와 유사하게 1초의 over-shift

를 통해서 다음 S-LKFS 값을 측정한다. 일반적으로 방송 프로그램의 음량 규제 단위는 R128의 integrated (I) -LKFS로서, 한 프로그램의 평균 음량을 의미한다.

3. 가청/비가청음 기반 적응적 음량 제어 알고리즘

최근 라우드니스 장비들을 통해서 방송 프로그램들이 시청자들에게 제공되면서 '사람의 목소리가 작아져서 소리의 명료도 및 의사 전달 기능이 저하'되는 현상이 다수 발생하고 있다. 이를 해결하기 위해서 본 논문에서는 사람의 청각 특성을 고려한 적응적인 음량 제어 알고리즘을 제안한다.

제안하는 방법은 먼저 원 오디오로부터 사람이 들을 수 있는 가청 음인지, 사람이 들을 수 없는 비가청음인지를 판별하고, 이 데이터들을 분리하여 음량 제어를 하는 기법이다. 이 기법은 상대적으로 작은 부분의 소리를 배제함으로써 사람에게 들리는 부분에 대해 좀 더 많은 음량을 할당하여 음량 제어 성능을 향상시키는 것으로, 가청음/비가청음 분석 모듈과 적응적 음량 제어 모듈로 구성된다.

가청음(유성음)과 비가청음(무성음)을 판단하는 기준은 [3]에서 권고한 절대적 기준과 상대적 기준을 적용하였다. 절대적 기준은 M-LKFS 값이 -70 LKFS 이하인 경우를 의미하고, 일반적으로 -70 LKFS이하의 음량을 갖는 데이터는 과형이 -20 ~ 20 사이의 값으로 구성된다.

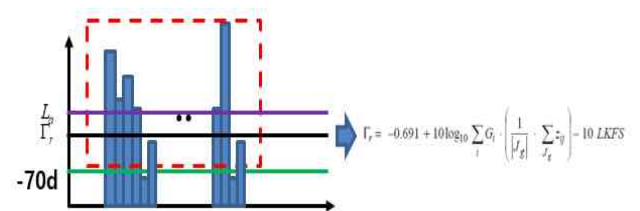


그림 3. 가청음/비가청음 판별을 위한 상대적 기준

상대적 기준은 그림 3에 보이는 것과 같이 절대적 기준에 의해 소리가 있는 신호에 대해서 I-LKFS를 측정하고, 이를 기준으로 일정 bias 만큼 기준을 낮추어 상대적 비가청음 기준을 설정하였다. 일반적으로 상대적 기준은 I-LKFS 값을 기준으로 -10 dB 이하인 경우를 의미한다. 여기서, I-LKFS 값을 획득하기 위해서는 한 개 방송 프로그램의 전체 데이터가 필요하지만 실시간 자동 음량 조절을 하는 경우에는 기 입력된 데이터들을 기준으로 I-LKFS를 구하여 연산한다.



그림 4. 가청음/비가청음 기반 적응적 오디오 음량 제어 알고리즘

제안된 적응적 음량 제어 알고리즘은 그림 4와 같이 ITU-R 기반 음량 측정, target LKFS를 고려한 gain 연산, 유성음/무성음 판단 기반 gain 연산, 그리고 smoothing 필터를 적용한 gain 적용 등의 모듈로 구성된다.

먼저 ITU-R 기반 음량 측정은 실시간으로 입력되는 오디오 데이터를 gate block 단위로 M-LKFS를 연산하고, 계산된 M-LKFS 값과 이전 M-LKFS 값을 이용하여 S-LKFS 값과 temporal I-LKFS 값을 연산한다. Target LKFS를 고려한 Gain 연산에서는 현재까지 컨트롤된 오디오 신호의 LKFS가 측정되어 현재 입력된 오디오 신호의 크기와 앞으로 컨트롤 될 신호의 방향에 따라서 현재 프레임에 대한 이득을 연산한다.

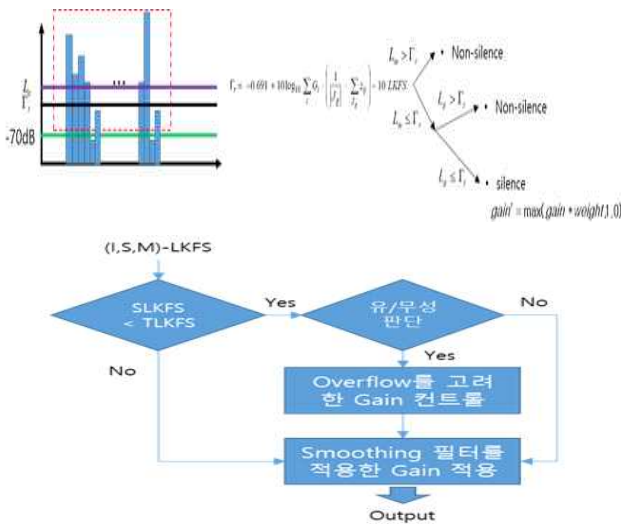


그림 5. 가청음/비가청음 기반 Gain 연산 및 제어 기법

유성음/무성음 판단 기반 Gain 연산은 그림 5에 보이는 것과 같이 가청음인 경우에는 이전 연산에서 계산된 gain에 weight를 더하여 보다 많은 gain을 얻도록 하고, 비가청음인 경우에는 상대적 기준을 적용하여 앞 단의 이득이 감소를 의미하는 경우 현재 gain을 보정할 지를 판단하여 gain을 연산을 한다. 제안된 알고리즘에서는 급격한 신호의 변화에 대처하기 위해 M-LKFS 값을 기준으로 비교하지 않고, S-LKFS 값을 기준으로 가청음과 비가청음을 판단하여 gain을 연산한다.

마지막으로 연산된 gain에 대해서 만약 현재 가청음에 속하는 경우는 바로 Smoothing 필터에 적용하여 이득을 적용하고, 비가청음으로 판단된 경우는 이득이 감쇄되어 오디오 신호가 더 작아지는 경우를 방지하기 위한 이득 조절을 하여 음량이 조절된 데이터를 출력한다.

제안된 알고리즘은 오디오 음질 평가를 위해서 ITU-R BS.1116 [4] 평가 방식을 통해서 주관적인 음질 평가를 수행하였다. 음질 평가를 위해서 원 프로그램의 오디오 데이터, 라우드니스 변환 장비를 적용하여 왜곡 없이 음량 조절되어 음량 기준을 만족하는 오디오 데이터, 그리고 제안된 알고리즘을 적용하여 음량 기준을 만족한 오디오 데이터를 20명 이상의 오디오 전문가들에게 청음 테스트를 수행하였다. 테스트 데이터는 각각 10초 분량으로 10개의 테스트 데이터를 생성하여 비교하였다.

그림 6에 보이듯이 제안된(proposed) 알고리즘과 왜곡 없이 음량 조절된(local) 데이터를 비교한 결과, MOS 5-score를 기준으로 제안된 알

고리즘은 약 4.5 score를 획득하여, local 결과에 비해 0.8 score 정도 우수한 성능 결과를 보였다. 그리고, 두 방식 중 어느 것이 음성의 명료도가 나은 지에 대한 선호도를 비교한 결과, 그림 7에 보이는 것과 같이 75% 정도가 제안된 알고리즘이 우수하다고 선택하였다.



그림 6. 제안된 알고리즘의 성능 결과 (BS.1116 방식 음질 평가)

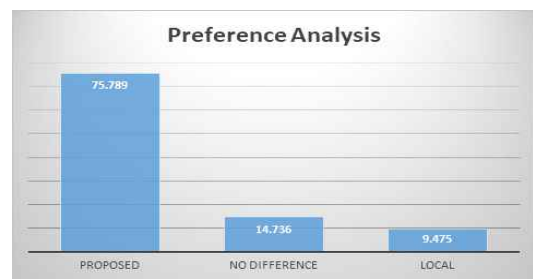


그림 7. 제안된 알고리즘의 성능 결과 (선호도 비교)

4. 결론 및 향후 계획

본 논문에서는 방송 프로그램의 음량 문제를 해결하기 위해서 가청음과 비가청음 기반으로 음량 데이터를 판별하고, 이 데이터들을 분리하여 적응적으로 오디오 음량 제어를 하여 가청음의 명료도를 높이는 음량 제어 알고리즘을 기술하였다. 제안된 알고리즘은 음질의 주관적인 평가를 통하여 기존의 왜곡 없이 음량 조절된 음원에 비해서 음성의 명료도 및 음질이 우수함을 보였다.

향후에는 목음, 배경음, 음성 등 오디오 데이터의 종류를 구별하여 보다 음성의 명료도를 높이기 위한 음량 제어 기술을 개발하고자 한다.

Acknowledgement

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2017-0-00788, 딥러닝 기반 지능형 오디오 분석을 통한 적응적 오디오 콘텐츠 변환 솔루션 개발)

참고문헌

- [1] 미래창조과학부고시 제2014-87호, 디지털 텔레비전 방송프로그램 음량 등에 관한 기준, 2014년 11월.
- [2] ITU-R Rec. BS.1770-3, "Algorithms to measure audio programme loudness and true-peak audio level," Aug, 2012.
- [3] EBU R128, "Loudness Normalization and Permitted Maximum Level of Audio Signals", Jun, 2014.
- [4] ITU-R Rec. BS.1116, "Methods for the subjective assessment of small impairments in audio systems", Feb. 2015.