

스피커의 주관적 청음 평가치와 라우드니스 측정치 간의 상관 관계

On the Correlation between Subjective Test and Loudness Measurement of the Loudspeaker

신성환*, 이정권*, 정혁**, 유동구***

(Sung-Hwan Shin*, Jeong-Guon Ih*, Hyuk Jeong**, Dong-Gu Yu***)

*한국과학기술원 기계공학과, **한국전자통신연구소, *** ㈜삼미음향기술
(접수일자: 2000년 9월 5일; 채택일자: 2000년 9월 25일)

스피커의 음질에 대한 재생 성능을 판단하기 위하여 개발 및 최종 평가 단계에서 수행되는 청음실험은 시간과 경제적 문제 뿐만 아니라, 평가자의 평가 능력에 따른 재현성 및 신뢰성에 많은 문제를 나타낸다. 이와 같은 이유로 스피커의 성능은 주관적 평가치 뿐만 아니라 객관적인 지표로서도 나타내는 시도가 함께 이루어져야 한다. 이를 위한 시도의 하나로써, 본 연구에서는 청음실험 및 통계 처리를 실행하여 얻은 주관적 평가 결과와 스피커 재생음에 대한 음질인자 해석을 통하여 얻은 객관적 측정 사이의 관계를 밝히고자 한다. 청음실험 결과는 분산분석법 및 Tukey의 방법을 이용한 네 단계의 통계 처리를 통하여 주관적 평가 결과를 얻고, 객관적 평가를 위해서는 라우드니스 측정을 기본으로 한 충실도 평가치 (Fidelity Rating)와 부드러움 평가치 (Softness Rating)를 제안한 후, 두 평가 결과의 상관도를 계산함으로써 타당성을 검증하였다. 본 연구에서의 방법론을 이용하면, 개발품 및 완성품의 스피커 음질 평가에 있어서, 시간이 많이 들고 비싼 주관 평가 작업이 없이도 통계적으로 안정되게 예측 평가할 수 있다고 본다.

핵심용어: 스피커의 주관평가, 스피커의 객관평가, 라우드니스, 분산분석, 충실도, 부드러움

투고분야: 전기음향 분야 (3.2), 음악음향 및 음향심리 분야 (8.2)

Acoustic performance of loudspeakers for sound reproduction has been qualitatively evaluated by using the listening test by juries in the development and final evaluation stages. However, the subjective evaluation method has many problems in the viewpoint of reliability and repeatability that are mainly related to the jury performance, as well as time and economy. In this reason, objective techniques should be tried to evaluate the acoustic performance of loudspeakers as well as the conventional subjective test. The object of this study is to find if there is any correlation between the statistically treated jury test results and the measured results based on the loudness of reproduced sound signals. For the four-step statistical analysis, the analysis of variance (ANOVA) and Tukeys method are employed for dealing with the data from the listening test. For the objective evaluation, Zwickers loudness considering the human hearing characteristics is calculated for the measured sound signal emitted from each loudspeaker and the objective ratings such as fidelity rating (FR) and softness rating (SR) is suggested. The correlation between two ratings has been demonstrated for an actual set of loudspeakers using FR, SR and correlation coefficient. The method in this study can be useful in statistically evaluating commercial or prototype loudspeakers without using very time-consuming and expensive subjective testing.

Key words: Subjective rating of loudspeaker, Objective rating of loudspeaker, Loudness, Analysis of variance, FR, SR

Subject classification: Electro-acoustics (3.2), Musical acoustics and Psychoacoustics (8.2)

I. 서 론

소리를 재생하는 시스템은 크게 플레이어, 증폭기, 스피

커로 구성되는데, 이중 특히 고음질의 재생을 위해 가장 많은 기여를 하는 부분은 소리 재생의 최종 단계인 스피커라고 알려져 있다. 스피커의 재생 음질에 대한 평가는 지금까지 주로 청음실험을 통해서 이루어지고 있는데, 이는 스피커 설계 완성도에 대한 확신을 얻고 음질을 검증하는 수단이 된다. 그러나 청음실험은 시간 및 경제적인

책임저자: 이정권 (ihih@sorak.kaist.ac.kr)
305-701 대전광역시 유성구 구성동 373-1 한국과학기술원 기계공학과
(전화: 042-869-3035; 팩스: 042-869-8220)

면에서의 제약뿐만 아니라, 체계적인 방법과 기준을 적용하지 않을 경우 신뢰성과 재현성에 많은 문제를 나타내기가 쉽다. 이와 같은 이유로, 스피커의 성능을 주관적 평가치뿐만 아니라 기술적인 측정에 의해서 나타낼 수 있는 방법에 대한 요구가 제기 되어왔다.

실제 사람들 대부분의 성향이 비슷하다는 것은 기대할 수 없지만, 과학적으로 제어되는 청음실험을 통해서 얻은 결과를 통계적으로 신뢰할 수 있다면, 사람이 인지하는 단계에서 소리에 대해 공통적으로 호감을 느끼게 하는 인자들을 찾을 수 있을 것이다. 또한 이 인자들에 영향을 미치는 기술적, 물리적인 면이 스피커의 주관적인 선호도에 어떤 영향을 미치는가를 평가할 수 있는 단계까지 발전할 수 있다.

본 연구에서는 개발 단계 및 완성된 스피커의 음질을 분석하는 것을 목적으로, 스피커 음질을 대한 주관적 평가 결과를 얻기 위하여 청음실험을 하고, 스피커에 대한 물리적 특성을 얻기 위해서 재생음에 대한 객관적 측정을 수행하였다. 이 과정에서 청음실험을 통하여 얻어진 평가 결과에 대해서 분산분석법 (analysis of variance: ANOVA) 및 Tukey 기법을 이용한 통계처리 방법을 제안하고 적용하여, 각 스피커의 음질에 대한 주관적인 평가 결과를 분석하였다. 또, 인간의 청각 기관에 대한 특성이 고려된 음질 해석 (sound quality analysis) 지표 중 하나인 라우드니스(loudness)를 이용하여 각 스피커의 객관적인 차이를 밝히도록 하였다. 결과적으로 전자의 주관적인 평가 결과와 후자의 객관적인 측정 결과간의 상호관계를 조사하여, 최종적으로는 사람이 직접 스피커를 통해서 재생되는 소리를 평가 하는 것과 동등하게 스피커의 음질을 파악하고 표현 할 수 있는 객관적 지수를 얻도록 하였다.

II. 주관적 평가

스피커에 대한 주관적인 평가는 실제 사람을 대상으로 하는 청음실험을 사용한다. 그러나 청음실험의 결과는 음악을 듣는 청음실, 청음실험에 참여하는 청음자 등 많은 변수에 의해서 영향을 받기 때문에 이런 변수에 대한 제어가 필요하게 된다. Toole[1]은 청음실험에 나타나는 개인적인 의견에 영향을 주는 요소, 즉 기여변수들에 대한 내용을 정리하고 이를 제어하기 위한 연구를 하였다. 청음실험에 영향을 주는 기여변수 (nuisance variable) 는 음악을 듣는 조건과 관련된 물리적인 환경에 의한 변수, 실험에 입하는 청음자와 관련된 심리 및 생리학적 조건에 관련된 변수, 또 실험 과정에서 고려되는 기여변수와 같이 크게 셋으로 나눌 수 있다.

본 연구에서는 이와 같은 주관적 평가에 관련된 변수를 일관적으로 제어할 수 있는 환경을 검토하고, 관련된 변수들을 고려하며 재현성 있는 결과를 얻기 위한 통계적 방법론을 제시한 후, 이를 객관적 측정 지수와 비교하고자 한다. 이를 위해서 실험 대상이 되는 스피커는 6.2인치 스피커와 3인치 스피커로 구성된 투웨이 시스템 (two-way system)의 스피커 7개를 사용하였고, 스피커를 제외한 다른 구성 부품은 모두 동일하였다.

2.1. 실험 환경

청음실험에 의한 주관적 평가에 영향을 주는 기여변수를 정확하게 제어하지 못 할 경우, 그 결과에 대한 재현성과 신뢰성은 매우 떨어져서 확신할 수 없는 결과를 얻게 된다. 지금까지 발표된 청음실험과 관련된 내용들을 정리하면, 주관적 평가에 영향을 줄 수 있는 기여변수들에 대한 실험 조건이 다음과 같이 고려될 수 있다.

첫째, 물리적인 환경과 관련된 기여변수에 대한 제어가 필요하다. 이 부분에서 가장 중요한 청음실의 크기에 있어서, 본 연구에서는 길이(L) 10.5m, 너비(W) 8.9m, 높이(H) 3.7m인 공간을 선택하였다. 이는 IEC [2]에서 추천하고 있는 크기, 7.0m 5.3m 2.7m (L×W×H)보다는 크지만, IEC에서 제시된 청음실 내부에서 청음자와 스피커 사이의 음향학적 연성에 가장 큰 영향을 주는 내부 공진 모드의 영향을 줄이기 위한 조건은 만족시킨다. 위 공간에서 잔향시간을 측정한 결과는 그림 1과 같다. 잔향시간의 경우 1kHz 이상 부분에서는 IEC에서 제안하는 범위 내에 있지만, 200Hz~1kHz 사이의 범위에서는 최대 0.2초 정도가 크게 나타나고 있는데 이는 청음실 규모에 의존하기 때문이다. 여기서, EDT (early decay time)는 음압의 초기 레벨에서 10dB 감소하는데 걸리는 시간에 6을 곱해서 얻은 잔향시간이고, T20은 초기 음압 레벨에 대해서 5dB 감소한 지점부터 20dB 감소한 지점, T30은 5dB에서 30dB 까지 감소하는데 걸리는 시간에 각각 3과 2를 곱해서 얻은 잔향시간이다.

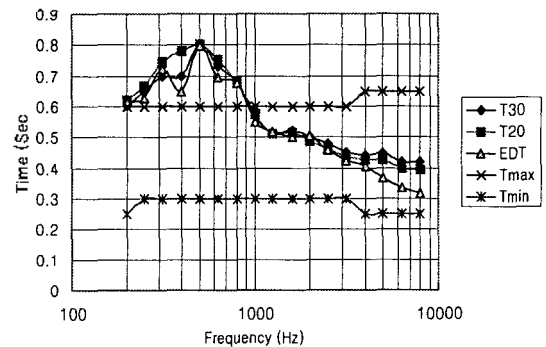


그림 1. 청음실의 잔향 시간
Fig. 1. Reverberation time of the listening room.

음을 재생 시킬 경우, 재생 음의 절대적 크기와 상대적 차이 역시 주관적 평가에 영향을 주는 변수로서 작용한다. 재생되는 음의 절대적 크기는 청음 위치에서 약 75~90dBA를 유지하고 [3], 각 스피커 사이의 상대적 차이는 1dB 이내에서 있도록 한다 [4]. 그림 2는 청음실 내부의 스피커 위치와 각 청음 위치를 나타낸 것이고, 표 1은 각 청음 위치에서 모든 스피커의 음압을 측정된 것으로 스피커 재생음의 절대적인 크기는 약 77.5dBA, 상대적인 음의 차이는 1dB 이내에 있어서 적절한 상태에서 실험이 수행되었음을 보여준다.

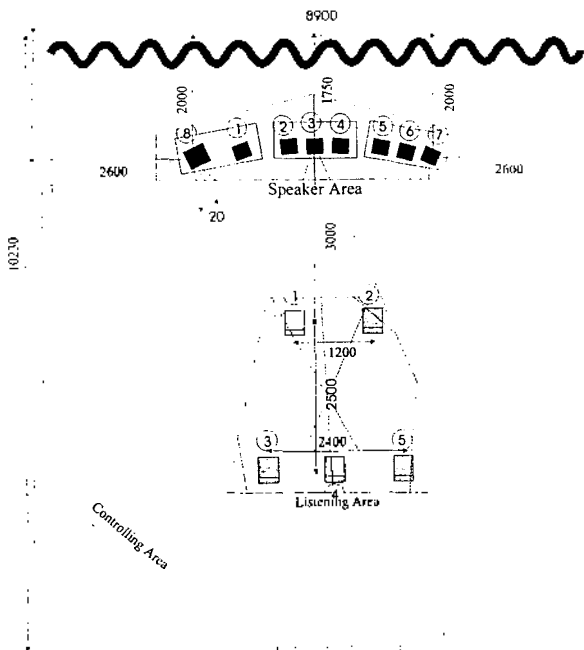


그림 2. 청음실 내부에서 스피커와 청음 위치 (단위:mm)
Fig. 2. Arrangement of the listening condition (unit: mm).

표 1. 각 청음 위치에서 측정된 스피커의 음압 레벨 (dBA)
Table 1. Sound pressure levels of loudspeakers measured at listening positions (dBA).

Seat ID. Speak ID.	1	2	3	4	5
1	78.6	78.4	76.4	76.8	77.2
2	78.8	79.2	76.7	77.6	77.8
3	79	78	76.7	77.3	76.7
4	79	77.9	76.8	77	76.1
5	78.3	78	76.2	76	76.5
6	77.8	78.2	76.1	77.1	77.7
7	77.7	78.2	76.1	76	77.2
8(Anchor sp.)	79.8	78.7	77.9	77.7	77.6
평균	78.46	78.27	76.43	76.83	77.03
편차	0.54	0.44	0.30	0.62	0.63

둘째, 청음자와 관련된 기여변수에 대한 제어가 필요하다. 청음실험의 결과에는 청음자의 청음 능력이 매우 중요한 역할을 한다. 청음 능력은 나이가 들어감에 따라, 또는 각종 질병에 따라 나타나는 생리학적인 청력 손실과 소리에 대한 인지 경험에 따른 청감에 의존한다. 이를 고려하기 위하여 각 청음자에 대한 청력 검사를 수행한 후, 최저가청영역 (MAF: minimum audible field) [5]을 기본으로 각 나이에 따른 청음 역치 (hearing threshold level) [6]에 대해서 10dB 이상의 청력 손실을 갖지 않은 청음자들만을 평가에 참여하도록 하였다. 청음 역치에 대

해서 15dB 이상의 청력 손실이 나타나는 평가자의 평가는 정상적인 청음 능력을 소유한 평가자의 평가와 거의 관계가 없음이 보고 된 바 있다 [3]. 평가자들은 음악에 관련된 분야에 종사하거나, 음악 교육을 받고 있는 사람들, 또는 적어도 한가지의 악기를 다룰 수 있는 사람들을 대상으로 하였다. 표 2는 청음실험에 참가한 평가자들의 구성을 나타낸다.

표 2. 청음 실험에 참가한 평가자 구성
Table 2. Juries participated in the listening test.

Occup	Undergraduate Students		Graduate Students	Professionals	Total
	Others	Related to music			
No.	8	9	10	5	32
Age	18~23	24~29	30~43		
No.	17	9	6		32

셋째, 실험 과정 및 평가방법과 관련된 기여변수의 제어가 필요하다. 청음 실험을 위한 실험 방법은 재생 음의 비교 방법, 음 재생 방법에 따라서 구분 할 수 있다. 본 연구에서 재생 음의 비교 방법은 SDM (semantic difference method)를 기본으로 여러 소리를 차례로 듣고 평가하는 복수비교 방법 (multiple comparison rating method)을 이용하여 재생 음을 비교하였다. 이 방법은 각 스피커의 상대적인 차이를 평가하고, 청음실험에 대한 경험이 부족한 사람과 평가 대상이 되는 스피커의 음질의 차이가 크지 않을 경우에 용이하게 사용되는 방법이다. 상대적인 차이라는 단점을 보완하기 위해서 모니터 스피커 JBL Model 4425를 기준스피커 (anchor speaker)로 사용하였다. 이러한 실험을 위해서 청음실험을 수행하는 청음자를 대상으로 실제 실험 과정과 같은 방법으로 사전 교육을 수행하였고, 재현성 및 신뢰도를 높이기 위해서 4회 이상 실험을 반복하였다. 또한 청음자의 귀의 피로도를 감안하여 각 실험에 걸리는 시간이 30분이 넘지 않도록 휴식 시간을 제공하였다.

음 재생 방법으로는 방의 구조나 배치, 주위 음원에 의한 영향이 작으면서, 청음실험에서 상대적으로 좋은 재현성을 갖는 모노 재생 (monophonic reproduction)을 사용하였다. 모노 재생 방법은 스피커의 잠음을 포함한 기계적인 문제, 그리고 음색 변화를 스테레오 재생에 비해 정확히 표현하는 장점을 갖는다[1]. 스피커의 음질을 평가하는 자각인자 [1]는 음의 전체적인 평가를 담당하는 충실도와 부분적인 평가를 나타내는 부드러움으로 하고, 평가는 0~10까지의 척도를 그림 3과 같은 기준에 의해서 이루어 지도록 하였다. 충실도란 실제 음이 어느 정도 정확히 재생되는지를 나타내는 척도이고, 부드러움은 고음 부분의 재생 정도를 평가하는 것이다.

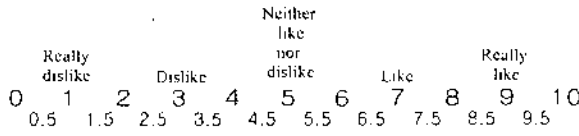


그림 3. 청음실험에 사용된 평가 표
Fig. 3. Subjective rating table for the listening test.

청음실험에서 음의 재생을 위해서 사용된 프로그램은 DDD (digital recording, digital mixing, digital editing) 방식으로 녹음 된 Michael Franks의 Antonios Song 일부 (프로그램 1), ADD (analog recording, digital mixing, digital editing) 방식으로 녹음 된 바흐의 브란덴부르크 협주곡 제 4번 Allegro 일부 (프로그램 2), 차이코프스키의 피아노 협주곡 1번 3악장 Allegro con fuoco 일부 (프로그램 3), DDD 방식으로 녹음 된 사람의 목소리 (프로그램 4)이다.

2.2. 통계 처리 방법

음질을 평가한다는 것은 사람의 청감과 두뇌에 남아있는 잔상이 매우 중요한 역할을 하므로, 그 측정이 매우 어려울 뿐만 아니라 항상 정확한 결과를 얻는다는 것은 거의 불가능 하다. 이런 문제를 제거하면서 보다 정확한 결론을 얻기 위한 방법은 실험이 추구하는 목적에 알맞은 통계 처리 방법을 적용하는 것이나 아직까지 통계 처리 과정은 거의 체계화 되지 않은 상태이다. 따라서 본 연구에서는 복수비교방법에 대한 통계처리 방법을 분산분석법과 Tukey의 방법을 기본으로, 재현성 및 신뢰성을 확보 할 수 있는 네 단계의 통계 처리 방법을 제안하고 적용한다.

첫 단계는 복수비교방법에 평가 대상이 되는 각 스피커의 음질을 청음자가 구분했는지를 파악하기 위하여 일원 배치법 (one-way ANOVA)을 적용한다. 이 때 일원 배치법의 인지는 스피커가 되며, 귀무가설은 스피커 음질을 나타내는 대표값 사이의 차이가 없다이다. 복수비교방법은 임의의 순서에 따라 스피커를 재생시키고 평가하면서 이 과정을 여러 번 반복하여 수행하는 것이다. 따라서 각 청음자가 평가한 스피커에 대해서 여러 번 평가하게 되는데, 각 평가 값들이 일원배치법의 적용 대상이 된다. 청음 능력, 평가 능력 및 학습 정도에 따라서 반복되는 실험에서 정확한 재현성을 나타내지 못하는 청음자가 존재하는데, 이 단계에서는 재현성이 없는 평가를 제거하게 된다. 일반적으로 각 스피커에 대하여 5번 이상 측정이 이루어 질 경우 일원배치법의 가설검정에 사용되는 유의수준은 0.1 이하로 할 수 있다[7].

두 번째 단계는 각 청음자가 스피커를 평가할 때 사용한 평가 범위 차이를 제거하기 위한 단계이다. 일반적으로 평가 척도에 대한 1 / 3 / 5 / 7 / 9점의 매우 추상적인 정의만을 청음자에게 제공하기 때문에, 각 청음자의 음에 대한 경험이나 기준에 따라서 평가에 사용한 범위가 다를 수 있다. 평가상에 사용된 범위가 달라지면 통계 처리

의 최종적인 결론은 스피커의 음질 차이에 의한 구분보다는 각 청음자들 사이의 차이에 의해서 더 큰 영향을 받는다. 이런 결과는 모수인자 (fixed factor)를 스피커로 하고 변량인자 (random factor)를 청음자로 한 이원배치법의 평가 결과를 통해서 쉽게 증명할 수 있다. 따라서 각 청음자에 대한 평가 차이를 보정하는 문제가 제기되고, 이 문제를 해결하기 위해서 다음과 같은 정규화(normalization) 과정을 거치도록 한다.

첫번째 단계에서 귀무가설이 기각된, 즉 일원배치법 결과 스피커의 차이를 인지한 청음자의 평가 결과를 표 3a와 같이 배열하고, 다음과 같은 범위 보정 값 d_i 를 계산한다.

$$d_i = \frac{\sum_{j=1}^J (x_{ij} - \mu_j)}{J} \tag{1}$$

여기서 d_i 는 i 번째 청음자의 범위 보정 값, J 는 청음실험에 사용된 스피커의 개수, x_{ij} 는 i 번째 청음자의 j 번째 스피커에 대한 평가치, μ_j 는 j 번째 스피커의 평균치이다. 이와 같은 과정으로 구한 범위 보정 값은 표 3b에서와 같이 이용된다. 이 과정은 스피커의 평균값을 기준으로 청음자들의 평가 편차 및 평가 값에 대한 차이를 제거한 측정 결과를 얻을 수 있고, 정규화 과정을 거친 후에도 각 스피커에 대한 평균값은 일정하게 유지된다.

표 3. (a) 청음자에 의한 청음 평가 결과
(b) 정규화 과정 후의 결과
Table 3. Normalization process of the raw data.
(a) Arrangement of raw data, (b) normalized data.

Speaker No. / Test Subject	1	2	...	J	
Subject 1	x_{11}	x_{12}	...	x_{1j}	$d_1 = \left(\sum_{j=1}^J (x_{1j} - \mu_j) \right) / J$
Subject 2	x_{21}	x_{22}	...	x_{2j}	$d_2 = \left(\sum_{j=1}^J (x_{2j} - \mu_j) \right) / J$
⋮	⋮	⋮	⋮	⋮	⋮
Subject I	x_{i1}	x_{i2}	...	x_{ij}	$d_i = \left(\sum_{j=1}^J (x_{ij} - \mu_j) \right) / J$
Average	μ_1	μ_2	...	μ_j	$\sum d_i = 0$

(a)

Speaker No. / Test Subject	1	2	...	J
Subject 1	x_{11}	x_{12}	...	$x_{1j} - d_1$
Subject 2	x_{21}	x_{22}	...	$x_{2j} - d_2$
⋮	⋮	⋮	⋮	⋮
Subject I	x_{i1}	x_{i2}	...	$x_{ij} - d_i$
Average	μ_1	μ_2	...	μ_j

(b)

세 번째 단계는 모수인자를 스피커로 하고 변량인자를 청음자로 한 이원배치법을 두 번째 단계에서 정규화 과정을 거친 데이터에 적용하는 것이다. 이 단계를 통해서 현재 다루고 있는 지각인자 (perceptual dimension)에 대해서 스피커 사이의 최종적인 차이를 파악할 수 있다. 이원배치법의 처리 과정에서 두 번째 단계의 정규화 과정을 거쳤기 때문에 청음자에 의한 차이는 절대적으로 무시할 수 있다. 각 스피커에 대한 청음자의 평가치는, 일원배치법에서 사용한 청음자에 따른 각 스피커 평가치의 평균값을 사용한다. 이원배치법의 가설검증에 사용한 유의수준은 0.1이하로 하고, 가능하면 0.05를 사용한다 [7].

분산분석법에서 사용되는 가설검증의 대립가설은 앞에서 언급했듯이 적어도 한 스피커의 음질을 나타내는 대표 값의 차이가 있다는 것으로, 귀무가설이 기각되어 대립가설이 인정된다고 하더라도 스피커 사이에 등급을 나타내지는 못한다. 이는 분산분석법은 최종 결과가 단지 인자수준의 대표 값에 차이가 있다는 것만을 평가하는 것이기 때문이고, 각 인자수준, 즉 스피커 사이의 등급을 평가 하지는 못하는 단점을 갖는다.

청음 실험 결과에 대한 통계 처리 과정의 마지막 단계는 분산분석법에서 다루지 못한 각 스피커 사이의 등급을 판단하기 위한 Tukey 방법의 적용이다. Tukey 방법은 분산분석법의 에러에 대한 평균 제곱 (MSE: mean squares of error)과 통계 분포의 하나인 Studentized Range Distribution ($Q_{\alpha, M, N}$)[7]을 이용하여, 평가 대상이 되는 스피커 평균값 차이를 구분할 수 있는 유효범위를 구한다. 각 스피커의 지각인자에 대한 평균값이 유효범위 안에 위치하면 그 평균값들은 같은 수준의 음질을 가지고 있음을 의미하고, 유효범위에서 벗어나 있으면 각 평균값 사이에 차이가 있음을 뜻한다. Tukey 방법에서의 유효범위 w 는 다음과 같다:

$$w = Q_{\alpha, I, (I-1)(J-1)} \sqrt{\frac{MSE}{J}} \quad (2)$$

여기서, α 는 유의수준, I 는 청음자의 수, J 는 스피커의 수이다.

2.3. 주관적 평가 결과

본 연구에서는 스피커 평가에 사용되는 지각인자 중 충실도 (fidelity)와 부드러움 (softness)에 대한 청음실험을 수행하였다. 충실도는 전체적인 음질을 평가하면서 재생된 음이 어느 정도 원음과 비슷한가를 평가하는 지각인자이고, 부드러움은 고음역의 재생 음질을 중심으로 저음역과의 조화를 평가하는 지각인자이다.

표 4는 각 프로그램에 의한 충실도를 통계 처리한 결과이다. A~D는 Tukey방법에 의한 각 스피커의 등급을 구분한 것이고, 충실도가 뛰어난 스피커 순으로 정리하였다. 프로그램 1에 대해서 보면 스피커 1이 가장 좋은 충실도를 나타내고, 스피커 2 / 4, 스피커 6 / 3이 중간 단계, 스피커 5 / 7이 비교적 좋지 못한 충실도를 가지고 있음을

알 수 있다. 프로그램 2와 프로그램 3에서도 비슷한 결과를 얻을 수 있지만 프로그램 4에서는 중복된 등급의 스피커가 많아 확실한 구분이 불가능 하였다. 전체적으로 보면 주관적으로 평가된 충실도 면에서 스피커 1이 가장 좋은 성능을 가지고 있고, 스피커 7이 가장 나쁜 성능을 갖는 것으로 나타났다.

표 4. 충실도의 각 프로그램에 대한 청음실험 결과
Table 4. Subjective rating of fidelity for given input programs.

Program #1			Program #2		
Speaker No	Rating	Class	Speaker No	Rating	Class
1	7.70	A	1	6.94	A
2	6.39	B	6	6.78	A
4	6.14	B	2	6.49	A B
6	5.50	C	4	5.95	B
3	5.43	C	3	4.88	C
5	5.06	D	7	4.57	C
7	4.70	D	5	4.44	C
Program #3			Program #4		
Speaker No	Rating	Class	Speaker No	Rating	Class
1	7.09	A	6	7.25	A
6	6.87	A B	1	6.85	A B
2	6.50	B	2	6.70	A B
4	5.15	C	7	6.33	B C
5	5.11	C	4	6.25	B C
3	4.95	C	3	5.77	C
7	3.59	D	5	5.75	C

표 5. 부드러움의 각 프로그램에 대한 청음실험 결과
Table 5. Subjective rating of softness for given input programs.

Program #1			Program #2		
Speaker No	Rating	Class	Speaker No	Rating	Class
4	7.02	A	1	7.19	A
1	6.84	A	4	6.47	A
2	6.82	A	2	6.46	A
3	5.92	B	3	5.68	B
5	4.82	C	6	4.96	B
6	4.68	C	5	3.72	C
7	3.52	D	7	3.69	C
Program #3			Program #4		
Speaker No	Rating	Class	Speaker No	Rating	Class
2	7.44	A	1	7.44	A
1	6.64	B	2	7.00	A
6	6.34	B C	3	6.83	A
3	6.28	B C	6	5.97	B
4	5.67	C	7	5.24	C
5	5.47	C	4	5.14	C
7	3.44	D	5	4.94	C

표 5는 각 프로그램에 대한 부드러움을 통계 처리한 결과이다. 주관적으로 평가된 부드러움 면에서, 프로그램 1에서는 스피커 4 / 1이 좋은 성능을, 스피커 7이 나쁜 성능을 나타내고, 프로그램 2에서는 스피커 1 / 4 / 2가 높은 등급을 얻고 있으며, 스피커 5 / 7이 낮은 등급을

가지고 있다. 이와 함께 프로그램 3 / 4의 결과를 모두 종합해 보면, 부드러운 면에서 스피커 1 / 2가 스피커 5 / 7 보다 현저히 좋은 성능을 가지고 있다고 판단할 수 있다.

III. 라우드니스에 의한 객관적 평가

스피커 음질을 객관적으로 평가하기 위해서 이루어진 주파수 응답곡선과 지향성(directionality pattern), 고조파 왜곡(harmonic distortion), 충격응답(impulse response), 음색 변조(difference tone distortion)등을 기본으로 한 기존의 연구들을 되돌아 보면, 인간의 청각 기관에 대한 특성, 즉 인간의 귀는 단순히 주파수를 인식하지 않고 피치를 지각하고, 음의 레벨을 인지하지 않고 라우드니스(loudness) 지각하며, 스펙트럼의 모양, 변조 깊이, 변조 주파수를 지각하기 보다는, 날카로운 느낌(샤프니스: sharpness), 거친 느낌(러프니스: roughness), 변동감(변동 강도: fluctuation strength) 등을 지각한다 [8]. 따라서 스피커 음질 평가 시 인간 청감의 실제 특성을 고려하지 않고, 단순히 기계적인 측정 결과에만 의존하였다는 점에 문제가 있어 왔다. 기계적인 측정에 대해서는 스피커에 대한 입력 신호, 즉 원음이 스피커의 특수한 동적 성능에 의해서 어떠한 형태로 변환하는 것을 측정할 수는 있겠지만, 이런 변화가 소리를 실제 인지하는 특성을 반영하는 인간의 청감에 어떤 작용을 하는지에 대해서는 문제를 밝히지 못하기 때문이다.

본 연구에서는 스피커의 음질에 대한 객관적인 평가 수단으로 사용되어 온 기존의 측정 방법에 인간의 청감 특성이 고려된 신호처리 기법을 적용한 라우드니스라는 음질 평가 요소를 선택하여 적용한다.

3.1. 라우드니스 (loudness)

주관적으로 느끼는 소리의 크기로 정의되는 라우드니스는 그 자체로서 음질을 결정하는 중요한 인자이며, 샤프니스, 러프니스, 변동 강도와 같은 다른 음질 인자 발생의 원인을 설명하는데 필요한 기본적인 청감이다.

본 연구에서는 Zwicker [9]에 의하여 제안된 라우드니스 계산 방법이 Stevens [9]에 의해 제안된 방법보다 복잡하긴 하지만 확산 음장(diffuse sound field) 및 자유 음장(free sound field)에 모두 적용 가능하고, 순음 성분이 있거나 불연속적인 형태의 주파수 특성을 갖는 음의 해석에 유용하기 때문에, 그림 4와 같은 방법으로 라우드니스를 계산하였다. 라우드니스(N: 단위 sone)는 각 임계대역(critical bandwidth)에 대한 비라우드니스(specific loudness: 단위 sone/Bark)를 모두 더한 값이 되는데, 라우드니스는 소리의 전체 크기를 나타내는 반면, 비라우드니스는 각 임계대역에 대한 크기를 나타내기 때문에 소리의 특성을 주파수 대역과 대응시켜 분석할 수 있는 장점이 있다. 임계대역 [8]은 협대역 소음의 라우드니스를 측정하는 과정에서 협대역 소음의 폭이 어느 정점이 되기까지는 일정하다가 그 이상이 되면 갑자기 증가하게

되는 현상에 근거한 청각 필터의 유효 대역이다. 만약 임계 대역을 겹치지 않고 순차적으로 배열하면, 가청 주파수 대역에서 24개의 임계 대역을 구성할 수 있다. 소리의 절대적인 크기는 라우드니스 값 자체로서 의미를 가지지만, 두 소리의 상대적인 차이를 분석할 때는 각 소리의 라우드니스 비를 비교하도록 한다. 즉, 10sone에서의 1sone 차이는 2sone에서의 1sone 차이보다 작다.

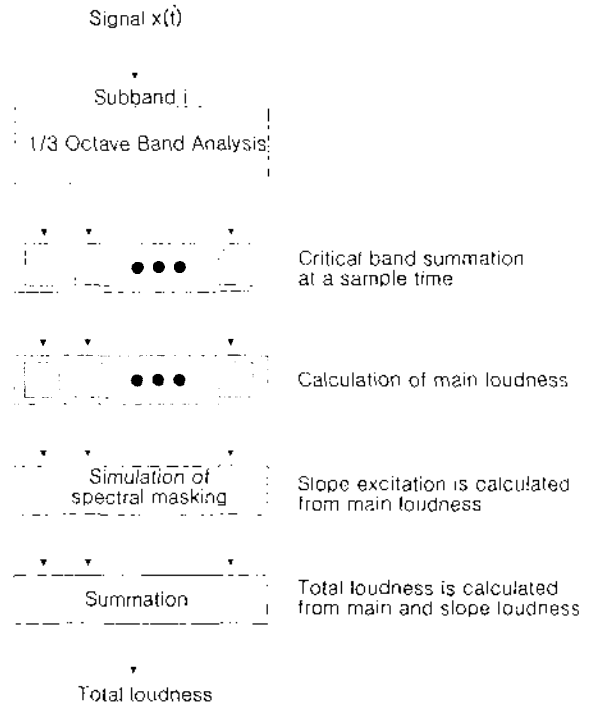


그림 4. Zwicker에 의한 라우드니스 계산 방법
Fig. 4. Flow chart for the calculation of Zwicker's loudness.

3.2. 실험 환경 및 라우드니스 해석 결과

스피커의 객관적 음질 평가를 위해서 스피커에 영향을 줄 수 있는 주위 요소의 영향을 배제하고 스피커 자체의 성능을 파악하기 위하여, 63Hz의 차단주파수를 갖는 무향실에서 pink noise 입력에 따른 각 스피커의 재생음을 측정하였다. 재생음은 스피커 정측 1m 앞에서 측정되었고, 44.1kHz의 샘플링 주파수로 DAT에 녹음하였다.

그림 5는 입력 신호로 사용된 pink noise와 이 신호에 의한 각 스피커의 재생음을 라우드니스 해석한 결과이다. 먼저 전체적인 비라우드니스 패턴을 살펴보면 스피커 1의 경우 입력 신호의 분석 결과와 비교할 때 2Bark~3Bark의 저주파 영역의 비라우드니스 크기와 16Bark~17Bark의 고주파 영역의 비라우드니스 크기가 거의 같다. 특히 16Bark~17Bark 영역에서의 재생 결과가 입력 신호와 매우 유사함을 볼 수 있다. 스피커 2의 경우도 15Bark과 17Bark을 제외하고는 입력 신호와 거의 비슷하다. 이에 반해 스피커 7은 저주파 영역에서는 그리 큰 차이가 나 타나지 않지만, 고주파 영역 15Bark~18Bark의

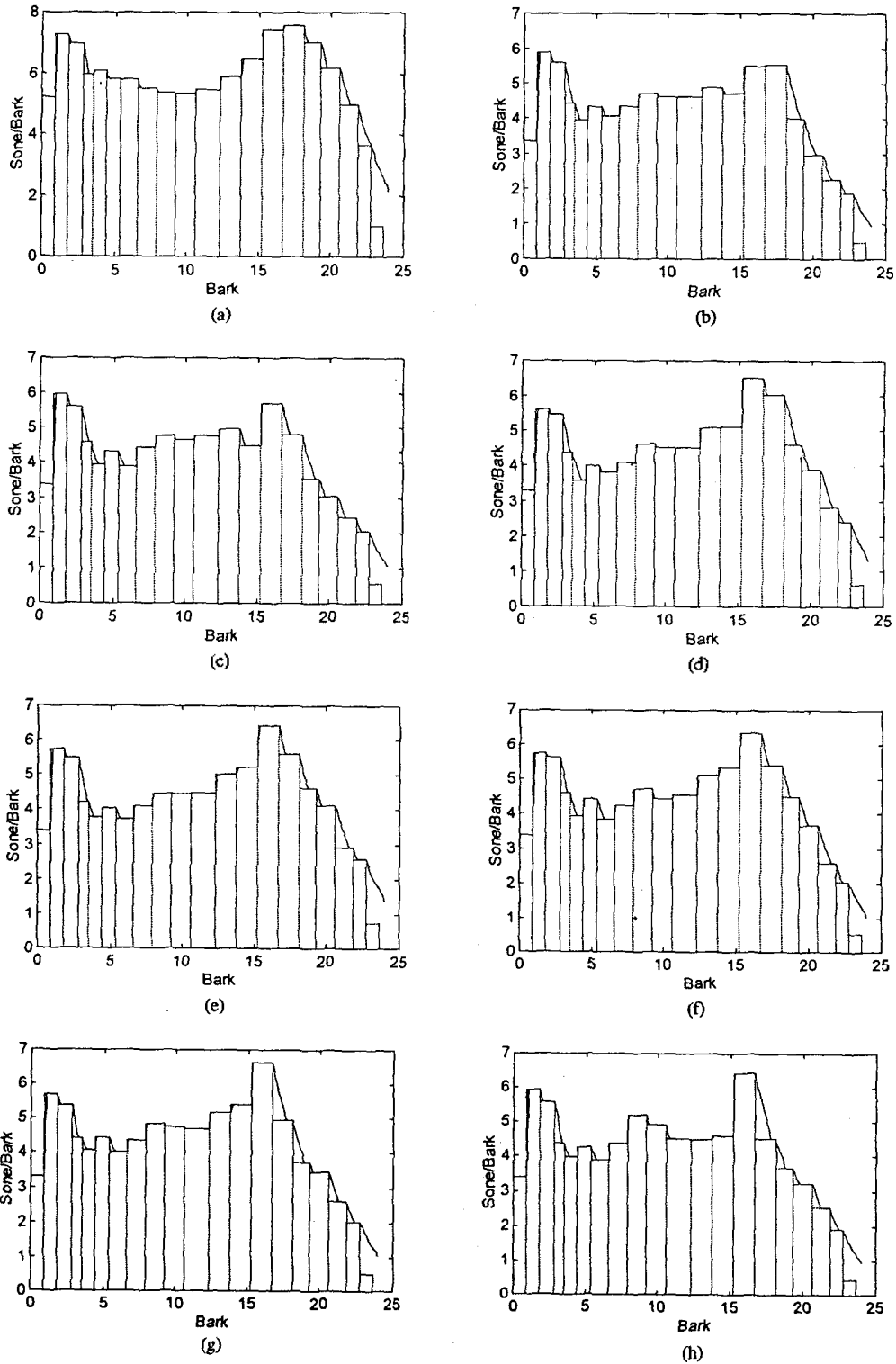


그림 5. 입력 신호 및 재생 음의 라우드니스 해석 결과

(a) 원음, (b) 스피커 1, (c) 스피커 2, (d) 스피커 3, (e) 스피커 4, (f) 스피커 5, (g) 스피커 6, (h) 스피커 7

Fig. 5. Specific loudness analysis of the reproduced sound and original one,

(a) Original signal, (b) speaker 1, (c) speaker 2, (d) speaker 3, (e) speaker 4, (f) speaker 5, (g) speaker 6, (h) speaker 7.

재생 결과가 입력 신호의 분석 결과와 비교할 경우 전혀 다른 모양을 가지고 있고, 특히 16Bark에서의 비라우드니스 값은 실제 입력에 비해서 상대적으로 크게 나타난다. 나머지 스피커 3 / 4 / 5 / 6 의 라우드니스 해석 결과는 각 Bark에 대해서 서로 비슷하다. 그러나 입력 신호와 비교할 경우 15Bark~18Bark 에 해당하는 비라우드니스 분포가 조금씩 차이가 있음을 볼 수 있다.

13Bark 이상의 비라우드니스 분포만을 고려해 볼 때, 세부분, 즉 스피커 1 / 2 / 4, 스피커 3 / 5 / 6, 스피커 7 로 나눌 수 있다. 스피커 1 / 2 는 중주파수가 거의 일정한 비라우드니스를 가지면서 16Bark~17Bark에 해당하는 고주파수 부분이 중주파수 영역보다 약간 높게 나타나지만 실제 입력 신호보다는 낮은 결과를 나타내고 있다. 반면 스피커 7의 경우는 16Bark에서의 비라우드니스가 중주파수 영역에 비해서 현저히 크게 나타나고 있는데, 이 영역의 소리는 다른 주파수 영역에 비해서 상대적으로 크게 재생되는 경향이 있다. 따라서 소리의 음질이 매우 날카롭게 되는 현상이 발생한다. 스피커 3 / 5 / 6의 경우는 14Bark 이상의 비라우드니스 값의 상대적 차이가 실제 입력 신호에 비해서 크게 나타나고 있음을 볼 수 있다. 그러나 단순히 이러한 결과만으로 스피커의 순위를 나타내기는 어렵고 라우드니스 분석을 기본으로 한 새로운 계산 모델의 적용이 필요하다.

IV. 주관적 평가와 객관적 지수의 상관관계

4.1. 충실도와 FR (fidelity rating)

주관적 평가를 위한 지각인자 중, 충실도는 스피커를 통해서 재생되는 음이 원음과 어느 정도 비슷한가를 판단하는 것으로, 이에 대한 객관적 지수를 얻기 위하여,

라우드니스 분석 결과를 이용한 그림 6과 같은 FR (fidelity rating) 계산 방법을 제시한다.

먼저 FR은 입력 신호와 재생된 신호의 라우드니스 해석 후, 각 임계 대역에 대해서 입력 신호에 대한 재생 신호의 비라우드니스 값의 비율 구한다. 이상적인 스피커, 즉 원음을 그대로 재생하는 경우에 임계 대역에 따른 비라우드니스의 비는 서로 같은 값이 되지만 실제 스피커의 경우 이런 경향을 얻는 것은 스피커 구성을 위한 여러가지 원인에 의해서 거의 불가능 하다. 비라우드니스의 비율 비교한 후, 스피커가 사용되는 용도를 고려하기 위해서 전체 비라우드니스 곡선을 세 부분으로 나눈다. 의사 소통을 위한 인간의 일반적인 음성 주파수 대인 600Hz 를 기준으로 그 이하를 저주파 대역(1Bark~7Bark)으로 하고 인간의 귀가 가장 민감하게 반응하는 2kHz~4kHz까지의 영역이 포함되는 부분을 고주파 대역(14Bark~)으로 구분한 후, 두 대역 사이 7Bark~14Bark를 중주파수 대역으로 정의하였다 [8]. 이렇게 세 부분으로 나누어진 입력 신호에 대한 재생 신호의 비라우드니스 비에 각 대역 편차를 더하여 충실도를 객관적으로 파악할 수 있는 지수인 FR을 다음과 같이 정의한다.

$$FR = \sum_{i=1}^3 \left(\alpha_i \times \sum_j \sqrt{(N_{ij} - \mu_i)^2} \right) \tag{3}$$

여기서, N_{ij} 는 i 번째 대역의 j 번째 비라우드니스, μ_i 는 i 번째 대역의 비라우드니스 평균, α 는 각 대역에 대한 가중치로 스피커의 사용 목적이나 주로 재생되는 영역을 고려하여 정하도록 한다. 본 연구에서는 각 대역에 대한 가중치로 동일한 값을 사용하였다.

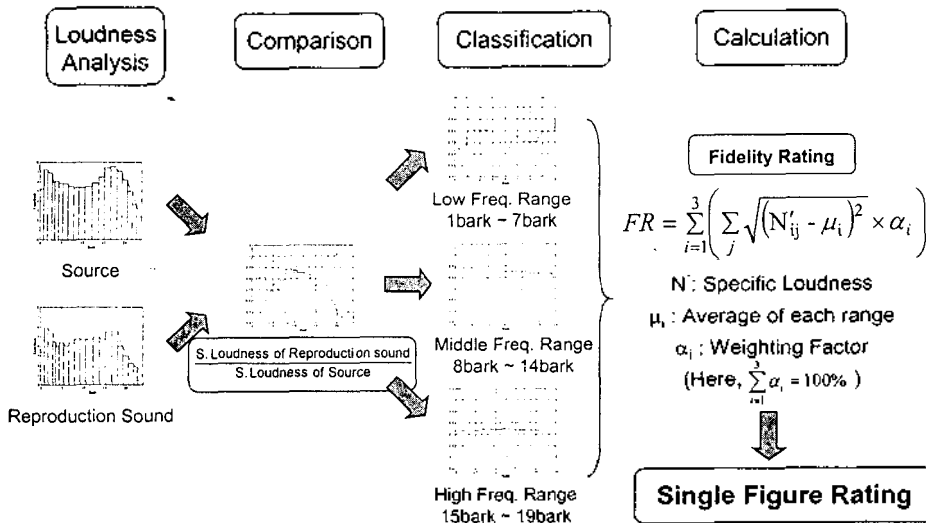


그림 6. FR 값 계산
Fig. 6. Calculation of fidelity rating.

각 스피커 재생음에 대한 FR의 계산 결과는 표 6과 같다. FR 계산은 각 비라우드니스 사이의 편차를 고려한 값이기 때문에 FR이 작을수록 원음을 재생하는 능력이 좋고, FR이 클수록 원음과 다른 음을 재생하는 것을 의미한다. 따라서 스피커 1의 경우가 실험 대상 스피커 중 원음과 가장 비슷한 음을 재생하고, 반대로 스피커 7이 원음과 가장 다른 음을 재생하고 있음을 나타낸다.

표 6. 각 스피커에 대한 재생 음의 FR
Table 6. Fidelity rating of each loudspeaker.

Speaker No.	Ranking	FR	Fidelity
1	1	12.56	Good
2	2	13.04	
3	5	16.50	
4	3	13.37	
5	6	19.42	
6	4	14.73	
7	7	23.25	Poor
Source		0.00	

청음실험을 통한 주관적 평가 결과와 객관적 평가 결과 FR 사이의 상관 계수 (correlation coefficient) 는 표 7과 같고, 상관 계수($\rho_{x,y}$)는 다음의 식으로 구한다.

$$\rho_{x,y} = \frac{Cov(X, Y)}{\sigma_x \cdot \sigma_y} \quad (4)$$

여기서, $Cov(X, Y)$ 는 X와 Y에 대한 공분산 (covariance),

σ_x, σ_y 는 X, Y에 대한 표준편차 (standard deviation)를 의미한다. FR에서는 값이 작을수록, 주관적 평가에서는 값이 클수록 원음 재생 능력이 좋음을 의미하기 때문에 상관계수는 음수로 나타난다. 통계적으로 상관계수는 그 절대값이 0.8 이상일 경우 아주 강한 상관 관계가 있고, 0.6~0.8 사이의 값은 비교적 좋은 상관 관계, 0.5 이하의 값은 갖는 경우는 거의 상관 관계를 가지지 않는다는 것을 의미한다.

표 7. 충실도에 대한 주관적 평가와 FR 사이의 상관 분석 결과
Table 7. Correlation between subjective fidelity and objective fidelity rating.

Program No.	Correlation Coefficient
1	-0.821
2	-0.847
3	-0.830
4	-0.468

이를 근거로 프로그램 1 / 2 / 3에 대한 상관 계수가 0.8 이상이므로 주관적인 충실도를 대신하기 위해서 제안된 객관적 지수 FR이 매우 타당하다고 보여진다. 그러나 프로그램 4의 경우는 0.5 이하의 상관 계수를 갖는데, II.3 장에서 언급했듯이 프로그램 4에 대한 주관적 평가에서 각 스피커 사이의 구분이 불분명했고, 프로그램 4가 인간의 음성을 재생한 것이어서 특정 부분, 즉 인간의 언어 소통에 사용되는 중주파수 대역에서의 재생이 뚜렷하기 때문에 각 대역의 가중치에 대한 고려가 필요하였으나, FR 계산 시 적용되지 않은 것에서 나타나는 문제로 판단된다.

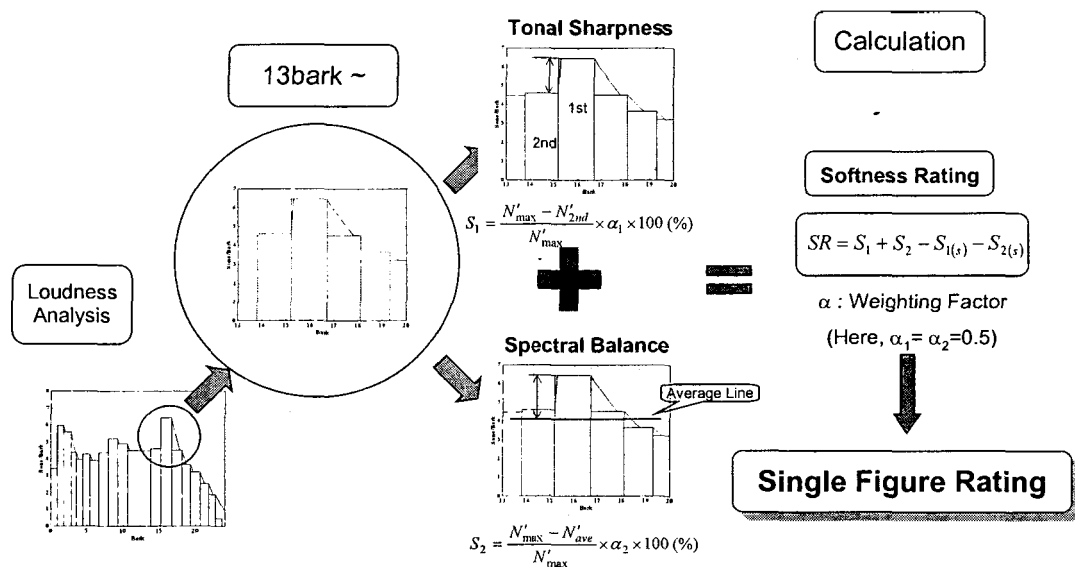


그림 7. SR값 계산
Fig. 7. Calculation of softness rating.

4.2. 부드러움과 SR (softness rating)

주관적 평가를 위한 지각인자 중, 부드러움은 스피커의 고음역 재생을 평가하기 위한 부분적인 평가 요소로서, 이를 위한 객관적 지수를 얻기 위하여 라우드니스 분석 결과를 이용한 그림 7과 같은 SR (softness rating) 계산 방법을 제안한다. SR은 사람의 청각 기관을 고려 해 볼 때 특정 부분, 특히 높은 고주파수를 갖는 소리가 지속될 경우 짜증을 느낄 수 있고, 이 때문에 고 음역에서의 일정 부분이 크게 재생 될 경우에는 소리가 날카롭게 인지되어 부드러움에 대해서 낮은 평가를 받게 되는 점을 고려 하였다.

부드러움이 고음역의 특성을 나타내는 것이므로 SR은 라우드니스 분석 결과에서 13Bark 이상의 비라우드니스 만을 고려한다. 이 부분에 대해서 음 재생 시 특정 대역에서의 날카로움을 측정하기 위한 토널 샤프니스 (tonal sharpness: S_1)와 고주파수 대역의 재생 밸런스를 측정하는 스펙트럴 밸런스 (spectral balance: S_2)를 다음과 같은 식에 의해서 정의한다:

$$S_1 = \frac{N'_{max} - N'_{2nd}}{N'_{max}} \times \alpha_1 \times 100(\%), \quad (5)$$

$$S_2 = \frac{N'_{max} - N'_{ave}}{N'_{max}} \times \alpha_2 \times 100(\%). \quad (6)$$

여기서 N'_{max} , N'_{2nd} , N'_{ave} 는 각각 13Bark 이상의 비라우드니스 최대값, 두번째로 큰 값, 평균값이고, α_1 은 날카로움에 대한, α_2 는 재생 밸런스에 대한 가중치이다. 이때 각 가중치는 0.5가 되도록 한다. 한편, 원음 자체의 날카로운 특성을 고려하기 위하여 원음에 대한 토널 샤프니스 ($S_{1(s)}$)와 스펙트럴 밸런스 ($S_{2(s)}$)를 구하여 다음의 식에 의해서 SR을 계산한다:

$$SR = S_1 + S_2 - S_{1(s)} - S_{2(s)}. \quad (7)$$

표 8. 각 스피커에 대한 재생 음의 SR
Table 8. Softness rating of each loudspeaker.

Speaker No.	Ranking	SR(%)	Softness
1	1	2.60	Good
2	4	12.4	
3	2	8.81	
4	3	12.1	
5	5	15.0	
6	6	20.2	
7	7	28.7	Poor
Source		0.00	

각 스피커에 대한 SR의 계산 결과는 표 8과 같다. SR 계산은 비라우드니스 차이에 의해서 계산되기 때문에, 그 값이 클수록 날카롭게 거슬리며 들리는 경향을 보여, SR

값이 작을수록 재생되는 소리가 부드럽게 들리는 것을 의미한다. 따라서, 스피커 1의 경우가 실험 대상 스피커 중 소리의 재생이 가장 부드럽고, 반대로 스피커 7이 가장 날카롭게 거슬린 느낌을 갖도록 재생되고 있음을 나타낸다.

청음실험을 통한 주관적 평가 결과와 객관적 평가 결과 SR 사이의 상관 계수는 표 9와 같다. FR과 같이 SR도 계산치가 작을수록 부드럽게 재생됨을 의미하는 반면, 주관적 평가에서는 값이 클수록 부드럽게 재생되고 있음을 의미하기 때문에, 상관계수는 음수로 나타난다. 각 프로그램에 대한 상관 계수가 모두 0.65이상이므로 주관적인 부드러움을 나타내기 위해서 제안된 객관적 지수 SR이 적합한 지수임을 보인다.

표 9. 부드러움에 대한 주관적 평가와 SR 사이의 상관 분석 결과

Table 9. Correlation between subjective softness and objective softness rating.

Program No.	Correlation Coefficient
1	-0.853
2	-0.797
3	-0.733
4	-0.655

V. 결 론

스피커에 대한 청음실험은 스피커의 음질을 파악하기 위한 스피커 제작의 최종 단계로서 오랫동안 지속되어 왔지만, 그 과정이 매우 복잡하고, 소비되는 시간과 비용에 비해 신뢰성 및 재현성을 갖는 결과를 얻는 것이 쉽지 않았다. 따라서 본 연구에서는 청음실험을 수행하지 않고도 스피커의 성능을 예측하기 위한 통계적으로 안정된 객관적인 평가 방법을 제안하고자 하였다. 이를 위해서, 신뢰성과 재현성이 인정된 청음실험을 통하여 얻은 주관적 평가 결과와 스피커의 성능을 객관적으로 평가한 물리적인 측정 사이의 상관 관계를 찾기 위한 시도를 하였다. 정확한 청음실험 결과를 얻기 위해서, 청음실험 결과에 영향을 줄 수 있는 기어변수를 ISO, IEC의 규격에 따라 제어하고, 7개의 스피커에 4개의 프로그램을 사용하여 청음실험을 수행 하였다. 청음 실험을 통해서 얻은 평가 결과를 분산분석법과 Tukey방법을 사용한 통계 처리 과정에 적용하여, 청음실험에서 사용되는 음질에 대한 인지 요소 중 충실도 및 부드러움에 대해, 신뢰성과 재현성이 보장된 주관적인 평가 결과를 얻었다. 또, 과거에 수행된 바와 같이 음향 물리량을 직접 측정하는 방법들을 사용하는 것 보다는, 청감 특성을 표현하는 객관적인 음질 요소 중 가장 중요한 라우드니스를 사용한 객관적 지수 FR 및 SR을 제안하였다. 이 과정을 통하여 계산된 결과가 충실도와 부드러움에 대한 주관적 평가 결과와 매우 높은 상관 관계가 있음을 확인하였다.

본 연구 수행에 있어서 객관적인 결과를 얻기 위하여 사용한 신호가 실제 청음실험에서 사용한 음악과 그 성질 자체가 다르다는 문제점과, 스피커 자체의 음 재생 능력에 중점을 두고 연구를 진행하는 과정에서 청음실의 환경에 의한 영향을 배제 했다는 것, 청음실험 시 선택한 프로그램의 음향 특성이 실험 결과에 어떤 영향을 주었는가에 대한 해석이 부족하다는 점, 또 양쪽 귀에서의 청감 특성이(binaural characteristics) 고려되지 못한 점 등이 아직 해결되지 못한 과제로 남아 있다. 또한 충실도와 부드러움을 객관화하기 위한 지수 FR 및 SR 계산에 사용된 가중치에 대한 연구가 필요한데, 이에는 앞으로 인간의 청각 기관과 스피커의 재생 목적을 충분히 고려하는 것이 필요할 것이다.

참 고 문 헌

1. F. E. Toole, "Subjective measurement of loudspeaker sound quality and listener performance," *J. Audio Eng. Soc.* **33**, 2~32, 1985.
2. IEC/TR3 60268-13 (1998-03): Sound System Equipment Part 13-Listening Test on Loudspeaker, Geneva: International Electro-technical Commission.
3. S. Bech, "Selection and training of subjects for listening tests on sound-reproducing equipment," *J. Audio Eng. Soc.* **40**, 590~610, 1992.
4. AES 1996, "AES recommended practice for professional audio-subjective evaluation of loudspeakers," *J. Audio Eng. Soc.* **44**, 383~401, 1996.
5. ISO 226: 1987 Acoustics Normal Equal-Loudness Level Contour, Geneva: International Organization for Standardization.
6. ISO 7029: 1984 Acoustics-Threshold of Hearing by Air Condition as a Function of Age and Sex for Ontologically Normal Persons, Geneva: International Organization for Standardization.
7. W. H. Beyer, *CRC Handbook of Tables for Probability and Statistics* (CRC Press, New York, 1968), Chap.8, pp.361~367.
8. E. Zwicker and H. Fastl, *Psychoacoustics* (Springer-Verlag, Berlin, 1990), Chaps.6~8.
9. ISO 532: 1975 Acoustics Method for Calculating Loudness Level. Geneva: International Organization for Standardization.
10. H. Jcong and J.-G. Ih, "Implementation of a new algorithm using the STFT with variable frequency resolution for the time-frequency auditory model," *J. Audio Eng Soc.* **47**, 240-251. 1999.

▲ 신 성 환(Sung-Hwan Shin) 1975년 9월 12일생
 1997년 2월 : 한국과학기술원 기계 공학과(학사)
 1999년 2월 : 한국과학기술원 기계 공학과(석사; 음향학)
 2000년 3월~현재 : 한국과학기술원 기계공학과 박사과정
 ※ 주관심분야: 실내음향, 심리 음향, 음질 해석, 음향 기기, 소음 및 진동

▲ 이 정 권(Jeong-Guon Ih) 한국음향학회지 제15권 제5호 참조

▲ 정 혁(Hyuk Jcong) 한국음향학회지 제19권 제5호 참조

▲ 유 동 구(Dong-Gu Yu) 1964년 5월 5일생
 1992년 : 성균관대 물리학과(석사; 음향학)
 1998년 : ㈜ 엔케이 텔레콤 음향공학 연구소 책임연구원
 2000년~현재 ㈜ 삼마음향기술 대표
 2000년~현재 주성대학 음향전자기 기학과 겸임교수
 ※ 주관심분야: 스피커 음향 설계 및 제작, 보청기 설계 및 제작