

프린터 음질평가의 신뢰성을 위한 결정오차 모델설계 및 실험적 검증 Design of Decision Error Model for Reliability of Sound Quality Analysis and Its Experimental Verification

김 의 열*. 이 영 준*. 이 상 권†

Eui-Youl Kim, Young-Jun Lee and Sang-Kwon Lee

(2011년 11월 21일 접수 ; 2012년 6월 11일 심사완료)

Key Words : Sound Quality Analysis(음질분석), Jury Evaluation(청음평가), Reliability(신뢰도), Decision Error(결정오차), Shannon Entropy(새넨엔트로피)

ABSTRACT

In this study, the possibility of decision error is investigated to identify and improve the reliability of participants in the process of conducting the sound quality analysis for laser printers. So far, there is not a way to identify and express the possibility of individual participant quantitatively. Thus, the decision error model is proposed which is based on the expectation value between the perceived sounds. Through the experimental verification on the laser printers, it was found that the possibility of decision error is affected according to the normalized difference. The possibility of decision error has inversely proportional to the normalized difference between the perceived sounds. When the normalized difference becomes small value, the uncertainty between decisions is inversely increase, and then it is difficult to obtain the proper result in the process of the jury evaluation for laser printers. For this reason, in this study, the proposed decision error model is added in the previous step of the correlation verification. Comparing to the conventional process only using the correlation based method, after the reliability of each participant is verified, the correlation with the mean response of participants is verified. It was found that the participants who were recognized as having unusual preferences are actually identified as having the reliability problem. Based on the results of this study, the proposed decision error model will be helpful to identify and improve the reliability of participants in the following study for the sound quality analysis.

1. 서 론

앞서 레이저 프린터 작동시 발생하는 소음의 음질 연구를 수행하는 과정에서 모터, 기어, 쿨링팬 등의 회전체 기계요소로 인한 토널(tonal) 소음 성분이 다수의 청음 평가자들에 의해 부정적인 음질

인자로 고려되고 있음을 실험적으로 확인하였다^(1,2). 하지만, Aures tonality 모델과 같은 음질평가 모델(sound metric)의 청음평가 결과에 대한 상관관계의 부재로 인해 선행 연구에서 제안된 음질 인덱스는 토널 소음 요소의 영향성을 적절히 표현하지 못하는 문제점을 가지고 있다. 그래서 기존 연구들은 낮은 상관관계 문제의 개선을 위해 주로 음질평가 모델과 음질 인덱스 모델링 기법의 개선에 중점을 두고 연구를 수행하였다⁽³⁻⁶⁾. 반면에, Takanashi, A.⁽⁷⁾는 프린터 작동 중에 토널 소음 성분이 다수의 청

† 교신저자; 정회원, 인하대학교 기계공학과
E-mail : sangkwon@inha.ac.kr
Tel : (032)860-7305, Fax : (032)868-1716
* 정회원, 인하대학교 대학원 기계공학과

음 평가자들에 의해 인지되는 것을 확인하였으나 관련 음질 요소 모델들이 다수의 실험 결과에 일관된 상관관계를 가지고 있지 않은 까닭에 프린터 작동 소음의 음질 연구 중에 토널 소음 성분이 적절히 고려되고 있지 못함을 언급하였고 여기에서 일관되지 못한 상관관계 문제의 주요 원인을 비음향 요인(non-acoustic factor)에 의한 주관 평가 결과의 불확실성(uncertainty)에 있다고 고려하였다. 그래서 이 연구는 Takanashi, A.의 논문에서 제시된 의견과 공통적으로 비음향 요인에 의한 청음평가 결과의 불확실성이 문제가 될 수 있음을 고려하고 음질 연구의 원점으로 돌아가 청음 평가자들의 주관적 감성 평가 결과를 검증하는 기존 평가 절차와 방법이 과연 적절한 것인가에 의구심을 가지고 문제점을 분석 및 개선안을 제시를 위해 이 연구를 시작하였다.

레이저 프린터 작동시 발생된 토널 소음 요소의 평가 모델 개발에 관한 선행 연구 결과⁽³⁾와 함께 향후 진행되는 음질 연구를 통해 도출된 모든 결과들의 신뢰성과 정확성에 대한 비교 판단 기준이 되는 청음 평가자들의 주관적 감성 평가결과는 그 역할이 매우 중요하기 때문에 이 연구에서 다루고자 하는 비음향요인들에 의한 청음평가 결과의 불확실성 문제는 충분히 연구할 가치가 있는 연구주제라고 생각한다.

구체적으로 이 연구는 평가 대상 소음 사이에 존재하는 막연한 기대값에 대한 청음 평가자의 반응 특성을 통계적으로 분석하기 위한 결정 오차 모델(decision error model)을 제안하였다. 이를 통해 청음 평가자들의 보편적 결정 오류 특성을 도출하는 것이 가능하며 통계적 결과를 기초로 비정상적인 결정 오류 특성을 가진 평가자들을 찾아내는 것이 가능하다.

즉, 평가자 군집 내에서 상관관계 분석법(correlation based method)의 활용이 평가자 군집의 일반성(generality) 개선을 위해 특이 성향의 평가자들을 찾아내는 것이 목적이었다면 이 연구에서 새롭게 제안하는 결정 오차 모델은 평가자 개별의 선호도 차이가 아닌 평가자 군집의 신뢰성(reliability) 개선을 위해 불성실한 청음평가 참여, 좋지 않은 건강 상태 등의 비음향 요인들의 영향성을 정량적으로 분석하고 음질 인덱스 개발 과정에서 문제 평가자

들을 배제하는데 목적을 두고 있다. 하지만, 기존 평가 절차와 방법이 잘못되었음을 지적하고 수정하려는 것은 아니며 기존 상관관계 분석법 적용 이전 단계에 결정 오차 모델을 추가 적용함으로써 평가자 군집의 일반성과 함께 신뢰성까지 검증 및 개선하는 것이 이 연구의 궁극적인 목적이다.

2. 결정오차 모델의 설계

음질 분석을 위한 주관적 감성 평가 과정 중에 널리 활용되고 있는 의미분별법(SDM, semantic differential method)을 기반으로 랜덤하게 배열된 10개의 프린터 작동 소음을 주어진 임의의 질의어에 대해 청음평가하는 상황을 가정하였다. 우선, 주관 평가 과정의 첫 번째 평가 사이클에서는 평가자에 의해 임의로 선택된 두 개의 작동 소음만이 고려된다. 그리고 이어서 두 번째 사이클에서는 신규 작동 소음과 함께 첫 번째 사이클의 주관 판단 결과가 새로운 주관 평가를 위한 배경 지식으로 고려된다. 이와 같이 반복되는 평가 사이클 중에 평가자의 선택은 재귀시스템(recursive system)처럼 이전 사이클에서 누적된 판단 결과와 작동 소음들에 영향을 받게 된다. 물론, 사람의 기억력 한계와 왜곡 특성으로 인해 다수의 작동 소음과 판단 결과들을 모두 정확하게 기억하고 인지된 작동 소음에 대한 감성을 주어진 임의의 질의어에 맞추어 정량적으로 적절히 평가한다는 것은 쉽지 않은 일이다. 게다가 평가자 개별 특성에 영향을 받기 때문에 평가 오류의 경향성을 예측하는 것 또한 쉽지 않다.

따라서, 청음평가 과정 중에 인위적으로 쉽게 통제 가능한 변수들과 비정상적인 환경 조건들의 부정적 영향을 제외한다면 주관 평가 과정 중에 평가 오류 발생의 원인으로 비음향 요인들에 의한 판단의 모호성을 고려할 수 있다. 여기에서 판단의 모호성이란 평가자에게 주어진 다수의 정보와 선택 사이에 지나치게 너무 많은 경우의 수가 존재하고 모든 경우의 수가 동일한 확률을 가지고 불확실성이 최대가 되는 상태로 정의될 수 있다. 이러한 개념을 기초로 예를 들면, 청음평가에 사용된 다수의 작동 소음들 사이에 존재하는 소리의 다양한 특징들의 차이가 사람의 청각 인지 특성 측면에서 구분하는 것이 명확하지 않다면 판단의 모호성에 대한

정의에 따라 평가 오류 발생의 가능성이 증가하는 것으로 볼 수 있다. 하지만, 지금까지 음질 연구와 관련하여 판단의 모호성을 정량적으로 평가하기 위한 연구 또는 평가용 모델이 구체적으로 제시된 바가 없다. 그래서 이 연구에서는 평가 소음들 사이에 존재하는 막연한 기대값 개념에 기초하여 결정 오차 모델을 제안하고자 한다.

주어진 임의의 질의어를 기준으로 임의의 평가 소음 A와 B사이에는 어느 소음이 크다 작다 또는 같다 라는 등의 막연한 기대값이 항상 존재한다. 예를 들어 프린터 작동 소음 중 성가심의 첫 번째 인자로서 고려된 라우드니스의 크기에서 소음 A가 소음 B보다 크다면 소음 A는 소음 B보다 낮은 주관 평가 값을 가질 것이라는 막연한 기대값이 대부분의 평가자들의 마음속에 존재 할 것이다. 여기에서 두 작동 소음 상에 음질 차이가 정량적으로 어느 정도인지 정확하게 평가하는 것은 부차적인 문제이며 경향성 측면에서 기대값을 기준으로 평가자의 주관적 평가가 동일한 경향을 가지는지 또는 반대 경향을 가지고 있는지 확인 할 수 있다. 물론, 전제 조건은 사용되는 음질 요소가 사람의 주관적 평가 특성과 높은 상관도를 가져야 한다. 결론적으로 이러한 관계에 기초하여 평가 소음의 객관적 분석을 위해 사용되는 음질 요소에 기반하여 두 소음 사이의 기대값에 주관적 평가 값이 부합하는지 여부를 확인 할 수 있다. Fig. 1은 두 작동 소음 사이에 기대값에 평가자의 주관 평가값이 부합하는지 여부를 확인하는 과정을 보여준다. 여기에서 x_i 는 음질 요소 계산 결과를 의미하고 y_i 는 주관적 평가 결과를 의미한다. 음질 요소의 계산 결과는 확률 밀도 함수로 변환되어 x_i 대신 z_i 로 표현된다. 상대적인 크기 차이는 지시 함수를 사용하여 다음과 같이 표현될 수 있다.

$$I_A = \begin{cases} +1 & \text{if } z_n - z_{n-1} > 0 \\ -1 & \text{if } z_n - z_{n-1} < 0 \end{cases} \quad (1)$$

$$I_B = \begin{cases} +1 & \text{if } y_n - y_{n-1} > 0 \\ -1 & \text{if } y_n - y_{n-1} < 0 \end{cases} \quad (2)$$

그리고 기대값은 상호 부합 여부에 따라 +1과 -1으로 표현되며 Table 1과 같이 총 4가지 경우의 수를 가진다.

Table 1 Expected value between the subjective decision making and the perceived sound attribute

I_A	I_B	$E[A B]$
+1	+1	+1
+1	-1	-1
-1	+1	-1
-1	-1	+1

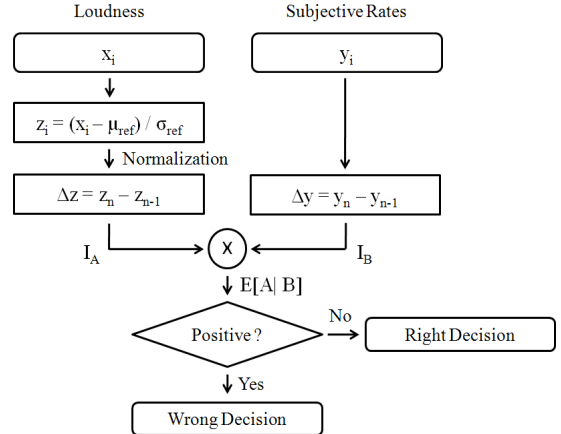


Fig. 1 The flowchart of the decision error based method

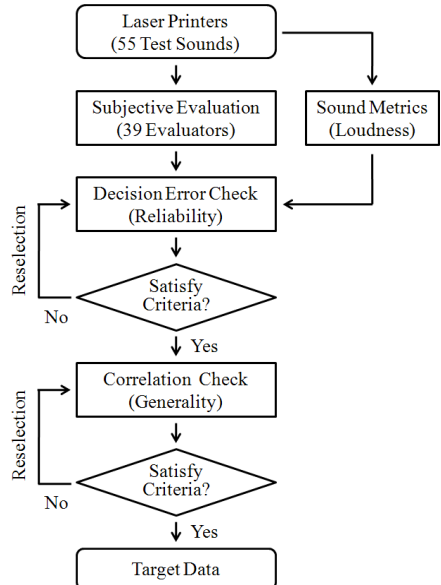


Fig. 2 The proposed post-processing procedure for the subjective evaluation of sound quality

$$E[A|B] = E[I_A | I_B] = P[A|B] = P[A \cap B] = I_A \times I_B \quad (3)$$

이와 같이 단순한 기대값 개념에 기초한 판단 오류 모델은 Fig. 2와 같이 프린터 작동 소음의 청음 평가 결과를 검증하는 과정에 적용되어 주어진 두 작동 소음 사이에 상대적인 차이 Δz 에 따른 평가 오류의 가능성 및 확률 분포 특성을 분석하는데 활용되었다.

3. 합성음(synthesis sound)과 청음평가

3.1 프린터 소음 측정

레이저 프린터의 음질 연구를 위해 Table 2에 정리되어 있듯이 5개 제조사의 9개 프린터 모델이 사용되었다. 일부 모델은 스캐너, 팩스 등의 기능을 포함하고 있는 다기능 프린터(MFP, multi-function printer) 타입의 모델이며 흑백 출력만 가능한 제조사 E의 모델 9번을 제외하고 모두 컬러 및 흑백 출력이 가능하다. 제조사 및 모델에 따라 분당 출력 속도(ppm)가 다르기 때문에 하나의 인쇄 사이클에 소요되는 시간 또한 모두 다르다. 음질 연구에 사용된 프린터 작동 소음은 이 연구실의 무향실에서 ISO 11201⁽⁸⁾에 따라 관찰자(bystander) 위치에서 HEAD acoustics사의 HMS III Artificial Head를 사용하여 측정하였다. 대기 상태에서 첫 인쇄를 준비 하는 과정에서 퓨저 롤러(fuser roller) 가열에 소요되는 시간이 모델에 따라 다르기 때문에 여러 차례 인쇄를 반복한 후에 작동 소음을 측정하였다. 이러한 과정을 통해 총 17개의 프린터 작동 소음이 측정되었으며 제조사 B의 모델 4번이 이 연구의 비교 분석 기준으로 고려하였다.

3.2 합성음 제작

프린터 작동 소음에 대한 청음 평가자들의 주관적 반응 특성을 다중회귀분석(MLR, multiple linear regression) 또는 인공신경회로망(ANN, artificial neural network)을 사용하여 새로운 음질 인덱스를 모델링하는 과정에서 의미분별법의 평가 범위 내에 대응하는 다양한 경우의 작동 소음들을 적절히 고려하는 것은 개발된 음질 인덱스 모델의 신뢰성과 오차를 최소화하는 것과 연관되어 중요하다. 하지

만, 현실적으로 음질 연구에 필요한 모든 작동 소음들을 실험을 통해 직접 측정하는 것은 쉽지 않은 일이다. 그래서 앞서 선행되었던 유사 연구들과^(1,6) 같이 음질 연구에 필요한 프린터 작동 소음들을 주요 관심 대상 모델인 제조사 B의 모델 4번의 컬러 출력 시 작동 소음을 기준으로 시간과 주파수 축에서 관심 대상에 따라 네 가지 항목으로 나누고 총 40개의 합성 소음 신호를 제작하였으며 이 과정은 선행된 연구 논문에 자세히 소개되었다⁽¹⁾.

3.3 청음평가

이 연구에서 청음평가(jury evaluation)는 주관적 감성 평가 결과의 신뢰성과 정확성을 위해 Otto, N. et al.에 의해 제시된 올바른 청음평가 절차와 방법을 참고하였다⁽⁹⁾. 청력 평가를 통해 정상 청력을 가지고 있는 것으로 확인된 39명(남자: 33명, 여자 6명, 연령: 20~45세)의 평가자들이 참여를 하였으며, 청음평가를 수행하기에 앞서 평가자들을 대상으로 평가 방법과 평가용 소음에 대한 사전 교육과 함께 참조 기준 소음의 충분한 청취 기회를 제공하였다. 총 57개(측정소음신호 17개, 합성소음신호 40개)의 소음 신호를 HEAD acoustics사의 음향재생장치 장비와 헤드폰을 사용하여 비교 청취한 후 정숙감(quietness)에 대한 질의어^(1,2)에 대해 의미분별법(SDM, semantic differential method)에 기초하여 4점과 9점 사이의 점수를 평가자가 선택하는 형태로 청음평가가 수행되었다.

Table 2 Print mode and maximum speed of the laser printer models used for sound quality

Model No.	Manufacturer	Print mode	Max print speed
1	A	Color / mono	20 ppm
2	A	Color / mono	20 ppm
3	B	Color / mono	17 ppm
4	B	Color / mono	24 ppm
5	B	Color / mono	20 ppm
6	C	Color / mono	24 ppm
7	C	Color / mono	22 ppm
8	D	Color / mono	20/32 ppm
9	E	Only mono	20 ppm

4. 실험적 검증

청음평가 과정 중에 평가자의 마음속에 존재하는 판단의 모호성을 정량적으로 평가하기 위해 제안된 판단 오류 모델의 유용성을 실험적으로 검증하기 위해 57개 작동 소음에 대한 총 39명의 주관 평가 결과에 제안된 모델을 적용하고 기존 방법의 결과와 비교 분석을 수행하였다. 우선, Fig. 3에서 볼 수 있듯이 57개 프린터 작동 소음 내에서 랜덤 함수를 통해 임의로 선택된 두 개의 작동 소음 사이에 존재하는 막연한 기대값은 ISO 532B⁽¹⁰⁾에서 제시하는 규정에 따라 계산된 라우드니스의 상대적인 크기 차이를 기준으로 계산이 되었고 평가자 개별의 주관 평가 결과에 경향성이 부합되는지 여부가 확인되었다. 여기에서 판단 오류 모델의 검증을 위한 기준 음질 요소로서 라우드니스가 고려된 이유는 프린터 또는 복사기 작동 소음에 대한 기존 음질 연구의 결과에서 공통적으로 라우드니스 또는 dBA가 주관 평가 결과와 높은 상관관계를 가지고 인덱스 개발 과정에서 성가심의 첫 번째 요인으로 고려되었기 때문이다⁽¹¹⁾. 확률 측면에서 모호성의 가능성에 접근하기 위해 랜덤 함수를 통해 임의로 선택된 작동 소음 사이에 존재하는 기대값에 대한 주관 평가의 부합 여부를 확인하는 과정을 N회 반복하였고 x축 상에 0~5σ 범위 내에서 0.5σ 간격으로 나누어진 정규 분포 상에서의 i번째 위치에서 올바른 결정(right decision)의 횟수 a_i 와 잘못된 결정(wrong decision)의 횟수 b_i 는 다음 관계식을 통해 Fig. 3(a)와 Fig. 3(b)에서 볼 수 있듯이 정규 분포 상에서 올바른 결정의 확률, n_r 과 잘못된 결정, n_w 로 연계될 수 있다.

$$n_r = \frac{1}{N} \sum_{i=1}^k a_i \tag{4}$$

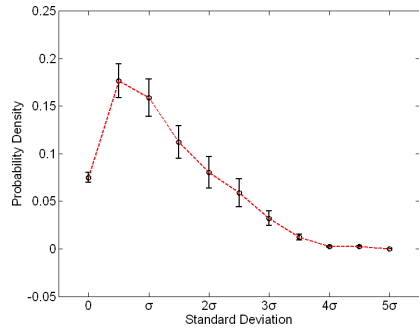
$$n_w = \frac{1}{N} \sum_{i=1}^k b_i \tag{5}$$

그리고 두 확률 분포의 합은 다음 계산 수식과 같이 항상 1의 값을 가진다.

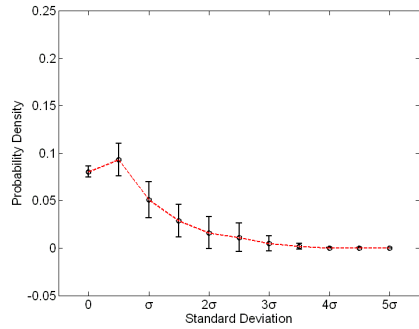
$$n_r + n_w = 1.0 \tag{6}$$

여기에서 올바른 결정은 평가자가 작동 소음 사이에 존재하는 막연한 기대값에 부합하는 감성적 느낌이 평가자 마음속에 존재함을 의미한다. 이러한 관계를 기초로 판단 오류의 가능성이란 다음 관계식과 같이 평가자의 모든 선택의 가능성 중에 잘못된 결정의 비중으로 정의되었다.

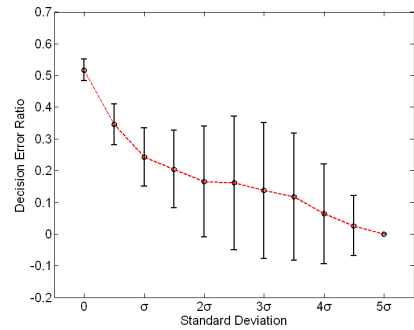
$$e = \sum_{i=1}^k \frac{b_i}{a_i + b_i} \tag{7}$$



(a)



(b)



(c)

Fig. 3 Decision error characteristics of the original group of 39 participants : (a) right decision, (b) wrong decision, (c) decision error ratio

Fig. 3(c)는 이 연구에서 제안된 판단 오류 모델을 프린터 작동 소음에 대한 청음평가 결과에 적용하여 도출된 판단 오류 가능성의 확률 분포 특성을 보여주며 청각 인지 특성과 연계하여 다음과 같이 평균 특성과 표준 오차 특성으로 나누어 살펴 볼 수 있다. 그리고 Table 3과 4는 0~5 σ 범위 내에서 0.5 σ 간격으로 평가자 39명에 대한 판단 오류 가능성의 평균과 표준 편차의 확률 분포 특성을 보여준다. 첫 번째는 판단 오류 가능성의 평균 특성에 관한 것이다. 우선, 정규 분포 상에서 x축 상에 상대적인 차이가 0에 가까운 경우 Table 3에서 볼 수 있듯이 약 50%의 판단 오류 가능성을 가지며 판단의 모호성이 극대화된 상태를 의미한다. 그리고 상대적인 차이가 증가함에 따라 판단 오류의 가능성은 지수 함수 형태로 점차 감소하며 결국 0%의 판단 오류 가능성을 가지게 되며 판단의 모호성은 사라지게 된다. 이처럼 사람의 청각계 인지 특성의 한계에 기반하여 평가용 소음 사이에 존재하는 모호성과 판단 오류 가능성 사이에 관계는 문제의 대상이라고 하기보다 자연스러운 결과이다. 두 번째는 판단 오류 가능성의 표준 오차 특성에 관한 것이다.

정규 분포 상에서 x축 상에 상대적인 차이가 0 또는 5 σ 인 경우 표준 오차 특성은 Table 4에서 볼 수 있듯이 0%에 가까운 값을 가진다. 반면에 상대적인 차이가 3 σ 인 경우에는 정규 분포 형태로 약 21%의 표준 오차 특성이 존재한다. 평균 특성의 경우 사람의 청각계 인지 특성의 한계와 연관되어 있지만 표준 오차 특성은 불성실한 청음평가 참여, 건강 상태 등의 비음향 요인과 밀접한 연관을 가지는 것으로 고려되었다.

즉, 청각인지 측면에서 명확하게 차이가 존재하는 평가용 소음에 대해 일부 평가자들이 비음향 요인에 의해 군집 특성에 부합하지 않으며 정규분포 형태의 표준오차 특성은 신뢰성 문제를 가지고 있는 평가자 검출을 위한 판단 기준으로 활용이 가능하다. 그리고 평가자 군집 전체의 청음평가 결과에 기초하여 특이 성향의 평가자 검출에 용이한 상관관계 분석법과 달리 개별 평가자의 청음평가 결과를 기초로 독립적으로 평가자 신뢰성의 문제 유무 확인이 가능하다. 정규 분포 상에서 x축 상에 상대적인 차이가 명확한 경우에 존재하는 판단 오류 가능성은 식 (7)과 같이 선형 가중치를 적용하여 강조

Table 3 Average possibility of decision error on the normalized difference between sounds in each group

Normalized difference	Original group	Modified group #1	Modified group #2
0	0.5176	0.5175 (-0.02 %)	0.5092 (-1.62 %)
0.5 σ	0.3457	0.3469 (+0.35 %)	0.3250 (-5.99 %)
σ	0.2427	0.2423 (-0.16 %)	0.2030 (-16.36 %)
1.5 σ	0.2048	0.2069 (+1.03 %)	0.1492 (-27.15 %)
2 σ	0.1660	0.1634 (-1.57 %)	0.0801 (-51.75 %)
2.5 σ	0.1609	0.1593 (-0.99 %)	0.0532 (-66.94 %)
3 σ	0.1371	0.1346 (-1.82 %)	0.0301 (-78.05 %)
3.5 σ	0.1168	0.1131 (-3.71 %)	0.0235 (-79.88 %)
4 σ	0.0635	0.0622 (-2.05 %)	0.0000 (-100 %)
4.5 σ	0.0258	0.0271 (+5.04 %)	0.0000 (-100 %)
5 σ	0.0000	0.0000	0.0000

Table 4 Standard deviation of decision error on the normalized difference between sounds in each group

Normalized difference	Original group	Modified group #1	Modified group #2
0	0.0333	0.0338 (+1.50 %)	0.0290 (-12.91 %)
0.5 σ	0.0641	0.0650 (+1.40 %)	0.0534 (-16.69 %)
σ	0.0915	0.0910 (-0.55 %)	0.0455 (-50.27 %)
1.5 σ	0.1218	0.1239 (+1.72 %)	0.0544 (-55.34 %)
2 σ	0.1747	0.1690 (-3.26 %)	0.0497 (-71.55 %)
2.5 σ	0.2111	0.2081 (-1.42 %)	0.0481 (-77.21 %)
3 σ	0.2141	0.2083 (-2.71 %)	0.0400 (-81.32 %)
3.5 σ	0.2008	0.2044 (+1.79 %)	0.0573 (-71.46 %)
4 σ	0.1575	0.1541 (-2.16 %)	0.0000 (-100 %)
4.5 σ	0.0949	0.1007 (+6.11 %)	0.0000 (-100 %)
5 σ	0.0000	0.0000	0.0000

되어 표현되었다.

$$e_w = \sum_{i=1}^k \frac{b_i}{a_i + b_i} \times \Delta z_i \quad (8)$$

Fig. 4는 선형 가중치가 적용된 판단 오류의 컬러 맵을 보여준다. 여기에서 x축은 평가자 번호, y축은 평가용 소음 사이에 상대적인 크기 차이, z축은 선형 가중치가 적용된 판단 오류 가능성 확률값을 의미한다.

특히, 평가자 17번은 가장 비정상적인 결정오류의 가능성을 가지고 있다. 정상적인 청각 인지 특성을 가진 것으로 확인되었으나 건강상의 문제를 청음평가 진행자에게 알리지 않고 장시간 청음평가에 참여하면서 위와 같은 부적절한 결과가 도출되었다. 선행 연구⁽¹⁾에서는 상관관계분석법의 결과에 기초하여 특이성향을 가진 것으로 판단되어 음질 인덱스 개발과정에 배제되었으나 이 논문에서 제안한 결정오차 모델에 기초하여 신뢰성 문제를 가진 것으로 확인이 되었다. 이와 같은 결과는 제안된 결정오차 모델의 유용성을 잘 보여 준다.

5. 모델의 응용

5.1 청음 평가자 군집의 비교 분석

기존 음질 연구에서는 상관관계 분석을 기초로 청음평가를 통해 얻은 평가자 군집의 평균 특성 또는 일반성에 비교하여 낮은 상관관계를 가지는 평가자들은 비음향 요인 중에 하나인 특이 성향을

가지고 있는 것으로 고려되었다^(12,13). 반면에, 이 연구에서 제안된 Fig. 1에서 볼 수 있는 판단 오류 모델은 앞서 제시한 Fig. 4의 결과와 같이 평가자 군집이 아닌 평가자 개별의 청음평가 결과를 토대로 비정상적인 판단 오류의 가능성을 가지고 있는 평가자들은 비음향 요인에 의한 신뢰성 문제를 가지고 있는 것으로 고려되었다. 여기에서 언급된 두 가지 검증 방법은 평가자 군집의 검증 및 개선 과정에 중첩되지 않는 서로 다른 목적을 가지고 있다. 그래서 판단 오류 모델의 활용에 관한 이 연구는 어느 방법이 좋고 나쁨을 확인하기 위한 것이 아니라 이 연구에서 제안된 판단 오류 모델이 기존 음질 연구 과정에 추가 적용됨으로써 긍정적인 기여가 가능한지 여부를 확인하는 것에 목적을 두고 수행되었다.

이와 관련하여 평가자 39명의 원본 군집은 앞서 언급한 두 가지 검증 방법에 기초하여 서로 다른 조합의 수정된 두 개의 평가자 군집들이 생성되었다. 우선, 첫 번째 군집은 상관관계 분석을 기초로 평가자 군집의 평균 특성에 약 70% 이하의 상관도를 가지는 경우 특이 성향을 가지고 있는 문제 평가자로 고려하였다. 그리고 두 번째 군집은 Fig. 2에서 볼 수 있듯이 판단 오류 모델을 상관관계 분석의 앞 단계에 적용을 하였다. 이유는 검증 과정 중에 우선 평가자 개별의 신뢰성이 보장되지 않으면 특이 성향 검증의 기준으로 사용되는 평가자 군집의 평균 특성 또한 신뢰성 문제를 가지기 때문이다. 그리고 각 평가자 군집의 세부적인 조합 근거는 다음의 객관적 결과에 근거하여 결정되었다.

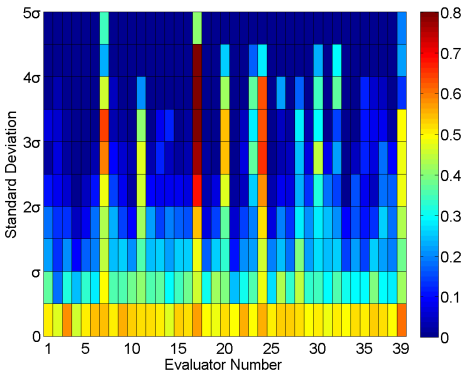


Fig. 4 Comparison of the linear weighted decision error distributions on the color segmentation map

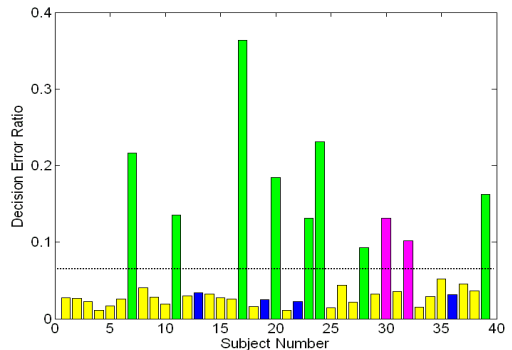


Fig. 5 Comparison of the linear weighted decision error ratios on the bar graph(■ case #1, ■ case #2, ■ case #3)

Fig. 4는 Fig. 5의 결과에서 평가자 중심으로 평균화된 평가자 개별의 판단 오류의 가능성을 보여준다.

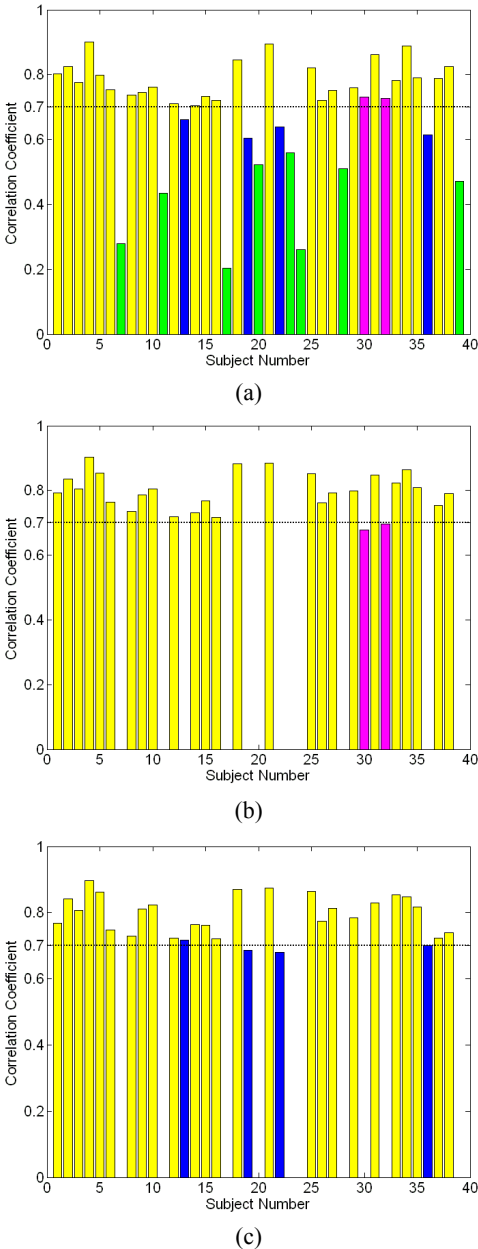


Fig. 6 Correlation coefficients between the objective rate and the mean response of participants according to the methods used for the post-processing of subjective evaluation : (a) original group, (b) correlation based method, (c) decision error based method(■ case #1, ■ case #2, ■ case #3)

여기에서 검은색 점선은 평가자 39명의 평균값을 의미하며 평균값 이상의 판단 오류의 가능성을 가진 평가자 10명을 비정상적인 판단 오류 특성을 가진 것으로 고려하였다. Fig. 6은 세 가지 평가자 군집별로 평가자 개별의 상관관계 분석 결과를 보여준다. Fig. 6(a)는 평가자 원본 군집의 상관도를 보여주며 70% 이하의 상관도를 가진 평가자 12명이 특이 성향을 가진 것으로 고려되어 Fig. 6(b)와 같이 제거되었다.

일부 평가자의 상관도는 원본 군집에 비해 증가되었으나 일부는 70% 이하의 상관도로 감소되었다.

Fig. 6(c)는 앞서 Fig. 5에서 평균값 이상의 판단 오류를 가진 평가자 10명이 제외된 경우의 상관도를 보여준다. 일부 평가자의 상관도가 모호하게 70%에 근접한 값을 가지고 있는 까닭에 허용 오차 범위 내에 포함되는 것으로 고려하고 제거하지 않았다. 결과적으로 평가자 39명의 원본 군집에서 두 가지 형태의 수정된 평가자 군집이 생성되었다. 첫 번째 군집은 12명이 제거되고 27명의 청음평가 결과가 활용되었으며, 두 번째 군집은 10명이 제거되고 29명의

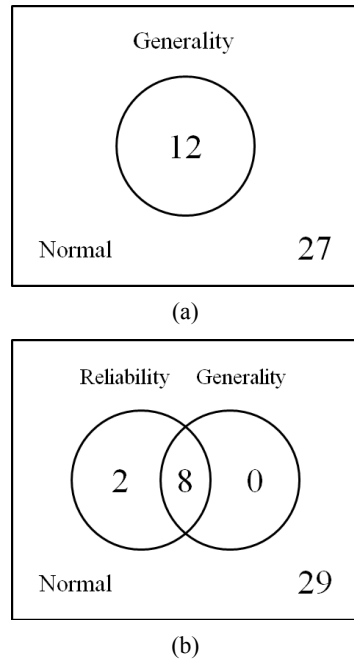


Fig. 7 Comparison of Venn diagram for the reliability and the generality according to the verification methods : (a) correlation based method, (b) decision error based method

청음평가 결과가 활용되었다. 그리고 각 군집에서 제거된 평가자들은 Fig. 5와 6의 결과를 토대로 다음과 같이 세 가지 유형으로 나누어 그 특성을 설명할 수 있다.

첫 번째 유형은 평가자 신뢰성 문제가 특이 성향으로 오인된 경우이며 제거된 평가자 8명이 여기에 포함된다. 두 번째 유형은 평가자 군집의 평균 특성과 충분한 상관관계를 가지고 있으나 평가자 신뢰성 문제를 가진 경우이며 제거된 평가자 2명이 여기에 포함된다.

세 번째 유형은 충분한 신뢰도 및 상관관계를 가진 평가자들이 부적절한 기준에 의해 특이 성향으로 오인되는 경우이며 제거된 평가자 4명이 여기에 포함된다. 여기에서 부적절한 기준이란 상관관계 분석의 기준이 되는 평균 특성이 일부 평가자들의 신뢰성 문제로 인해 잘못된 값으로 도출 될 수 있는 가능성이 존재함을 의미한다. Fig. 7은 청음평가 결과의 검증 및 개선 과정에서 제거된 세 가지 유형의 평가자들의 관계를 벤 다이어그램(Venn diagram)으로 보여준다. Fig. 7(a)는 첫 번째 수정된 평가자 군집에 관한 것이며 첫 번째 유형과 세 번째 유형의 평가자 총 12명이 특이 성향으로 고려되어 제거되었다. 그리고 Fig. 7(b)는 두 번째 수정된 평가자 군집에 관한 것이며 첫 번째 유형과 두 번째 유형의 평가자 총 10명이 신뢰성 문제로 인해 제거되었다.

5.2 결과 분석

평가자 군집 내에 평균 특성에 대한 상관관계와 평가자 개별의 판단 오류 특성을 고려하여 총 세 가지 조합의 평가자 군집을 제시하고 4절에서 평가자 오류 특성을 도출하면서 적용하였던 방법들을 동일하게 적용을 하였다. Fig. 8은 상관관계 분석을 통해 도출된 첫 번째 수정된 평가자 27명 군집의 판단 오류 특성을 보여준다. Fig. 8(a)와 (b)는 각각 올바른 결정과 잘못된 결정의 확률 분포를 보여주며 Fig. 8(c)는 판단 오류 특성을 보여준다. Fig. 9는 판단 오류 모델의 추가 적용을 통해 도출된 두 번째 수정된 평가자 29명 군집의 판단 오류 특성을 보여준다. Fig. 9(a)와 (b)는 각각 올바른 결정과 잘못된 결정의 확률 분포를 보여주며 Fig. 8(c)는 판단 오류 특성을 보여준다. 판단 오류 가능성의 평균 특성은 사람의 청각계 인지 특성의 한계와 연관되어

있기 때문에 Table 3에서 볼 수 있듯이 평가자 군집 사이에 차이는 크지 않다. 반면에 판단 오류 가능성의 표준 오차 값은 비정상적인 판단 오류 특성을 가진 평가자들의 존재를 의미하며 Table 4에서 볼 수 있듯이 평가자 군집 사이에 차이가 명확하게 존재한다. Fig. 3(c)에서 3σ 부근의 표준 오차 값에 비교하여 Fig. 8(c)와 Fig. 9(c)의 표준 오차 값은 상당히

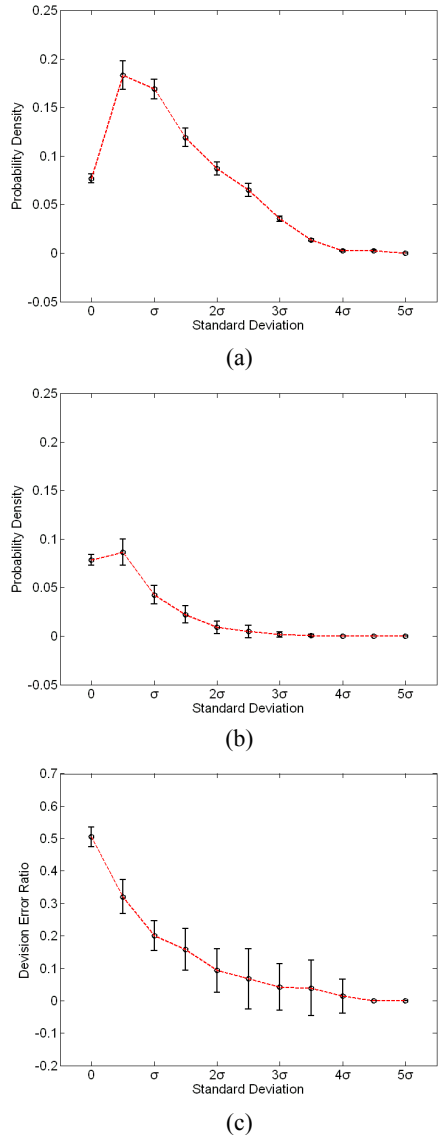
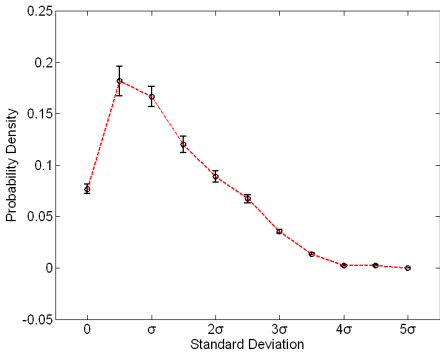


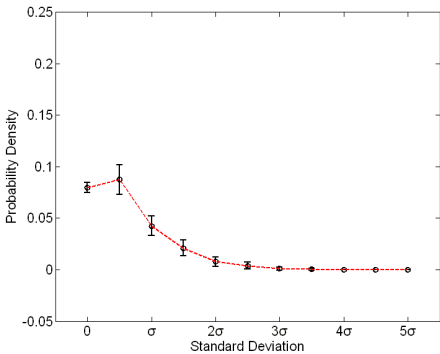
Fig. 8 Decision error characteristics of the modified group of 27 participants selected by using the correlation based method: (a) right decision, (b) wrong decision, (c) decision error ratio

감소되었다. 다르게 표현하면 사람의 청각계를 통해 명확하게 구분이 가능한 정도의 차이가 프린터 작동 소음 사이에 명확하게 존재함에도 불구하고 발생하는 잘못된 판단의 가능성들이 배제하는데 앞서 언급한 두 가지 접근 방법이 공통적으로 효과적임을 알 수 있다. 그리고 첫 번째 수정된 평가자 군집에

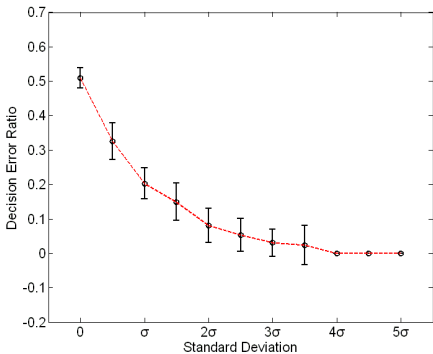
비교하여 신뢰성 개념을 기초로 특이 성향으로 오인된 정상 평가자 4명의 불필요한 손실을 막고 반대로 정상 평가자로 오인된 신뢰성 문제를 가진 평가자 2명을 제거함으로써 결과적으로 평가자 2명의 청음 평가 결과의 손실을 줄이고 판단 오류 가능성의 표준 오차 값을 개선하는데 도움이 되었다. Fig. 10은



(a)

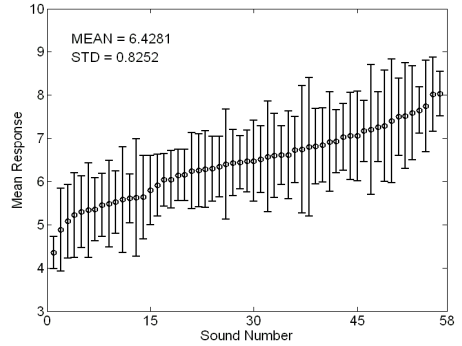


(b)

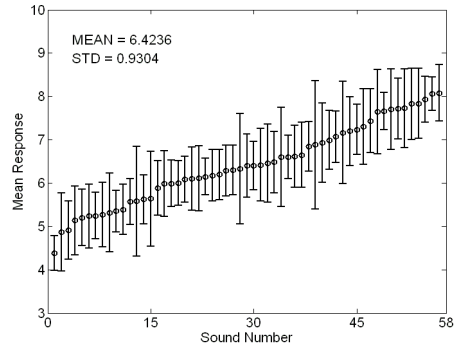


(c)

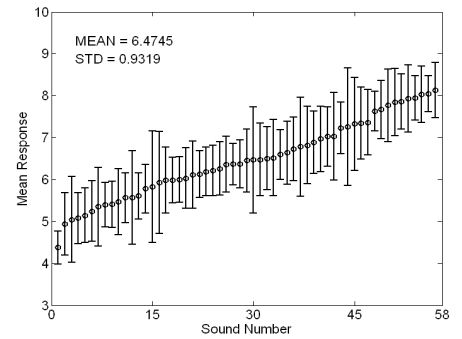
Fig. 9 Decision error characteristics of the modified group of 29 participants selected by using the correlation based method : (a) right decision, (b) wrong decision, (c) decision error ratio



(a)



(b)



(c)

Fig. 10 Mean response and \pm standard deviation with 95 % confidence in an ascending order for 57 printer sounds : (a) original group, (b) correlation based method, (c) decision error based method

세 가지 형태의 평가자 군집 내에서 각 평가 소음의 평균값 기준으로 오름차순 정렬된 청음평가 결과의 통계 결과를 보여준다. 그리고 Table 5는 각 평가자 군집 내에서의 통계 특성을 보여준다.

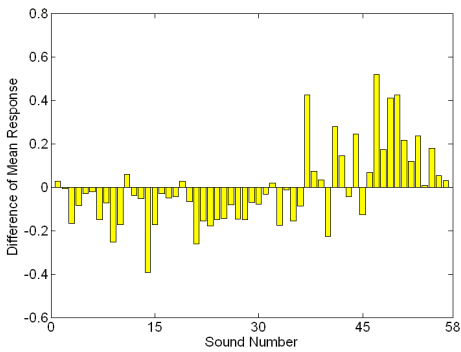
Fig. 11은 원본 군집을 기준으로 두 가지 수정된 평가자 군집에서 오름차순으로 청음평가 결과의

Table 5 Statistical information of the subjective evaluation in each group

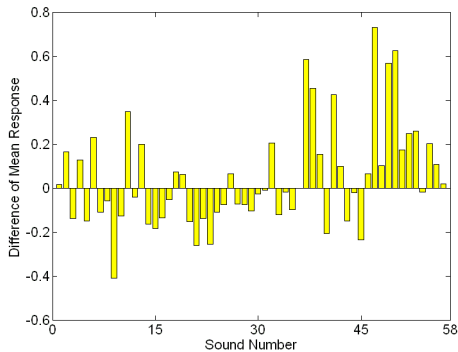
Property	Original group	Modified group #1	Modified group #2
N	39	27	29
Correlation	0.6902	0.7943	0.7872
Max	8.0288	8.0768	8.1302
Mean	6.4281	6.4236	6.4745
Min	4.3565	4.3836	4.3747
STD	0.8252	0.9304	0.9319

평균값 변화를 보여준다. 공통적으로 저평가된 작동 소음의 경우 청음평가 결과의 검증 및 개선 과정을 통해 더 낮은 값으로 보정되었고 고평가된 작동 소음의 경우 보다 높은 값으로 보정되었다.

그리고 Fig. 12는 57개 프린터 작동 소음 중에 Table 2에서 볼 수 있는 9개 모델의 17개 원본 작동 소음 신호에 해당하는 청음평가 결과의 평가자 군집 별 변화를 보여준다. Fig. 12(a)와 (b)는 각각 컬러 및 흑백 출력의 결과이다. 그리고 Table 6과 7은 인쇄 모드 별 청음평가 결과의 변화를 보여준다. 흑백과 컬러 출력 모두 공통적으로 6번과 7번 모델에서 원본 군집의 결과를 기준으로 상대적으로 큰 변화가 존재하는데 동일 회사의 유사한 모델들의 작동 소음이다. 전반적으로 4개 제조사 9개 모델 중에 상위에 속하는 것으로 주관 평가된 모델이며 작동 소음의 전반적인 크기를 의미하는 dBA와 라우드니스는

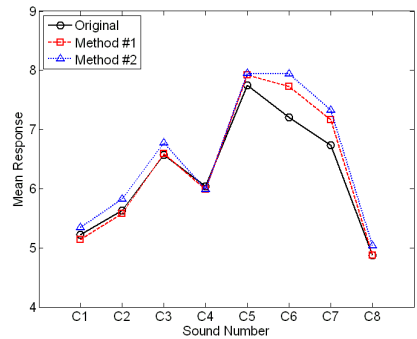


(a)

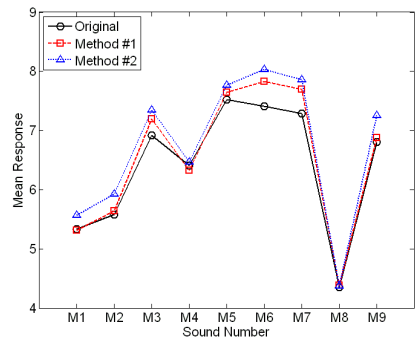


(b)

Fig. 11 Differences of mean response between the original group and the modified groups : (a) correlation based method, (b) decision error based method



(a)



(b)

Fig. 12 Comparison of mean responses for 17 original printer sounds according to the methods used for the post-processing of subjective evaluation : (a) color print (b) black and white print

다른 모델들에 비교하여 상대적으로 작은 값을 가지고 있지만 모터 및 기어 등의 회전체 소음은 토널 소음 요소들의 실제 절대값 보다 배경 소음에 대한 상대적인 관계에 따라 평균 이상으로 명확하게 인지되는 특성을 가지고 있다. 청음평가 수행 과정 중에 평가 모델 중에 상대적으로 작은 크기의 작동 소음과 함께 회전체 소음의 존재를 단지 인식만 하는지 아니면 성가심 유발의 요인으로 인식하는지 여부는 평가자 개별의 선호도 및 비음향 요인에 의해 영향을 받기 때문에 6번, 7번 모델의 작동 소음에 대한 평가의 모호성이 발생하였다고 볼 수 있다.

그 결과 Table 5에서 볼 수 있듯이 원본 군집에

Table 6 Variance of mean responses for the color print mode of laser printers in each group

Property	Original group	Modified group #1	Modified group #2
C1	5.2201	5.1367	5.3496
C2	5.6303	5.5780	5.8291
C3	6.5761	6.5963	6.7810
C4	6.0385	5.9892	5.9865
C5	7.7439	7.9254	7.9474
C6	7.2038	7.7258	7.9369
C7	6.7394	7.1655	7.3261
C8	4.8762	4.8701	5.0409

Table 7 Variance of mean responses for the mono print mode of laser printers in each group

Model no.	Original group	Modified group #1	Modified group #2
M1	5.3371	5.3178	5.5693
M2	5.5770	5.6366	5.9270
M3	6.9203	7.2002	7.3457
M4	6.4043	6.3235	6.4706
M5	7.5214	7.6421	7.7699
M6	7.4074	7.8333	8.0335
M7	7.2872	7.6993	7.8577

Table 8 Correlation coefficients between the sound metrics and the mean response of each group

Sound metric	Original group	Modified group #1	Modified group #2
dBA	-0.8341	-0.8585	-0.8671
Loudness	-0.8214	-0.8666	-0.8960

비교하여 청음평가 결과의 표준 오차값이 증가되었으며 의미분화법의 평가 범위 4-9점 내에서 평가용 작동 소음 사이에 차이를 보다 명확하게 하는데 작게나마 도움이 되는 부분으로 고려된다.

Table 8은 판단 오류 모델의 기대값 계산 과정에 기준 음질 요소로 고려되었던 라우드니스와 함께 dBA에 대한 각 평가자 군집 별 상관도를 보여준다. 두 번째 수정된 평가자 군집의 결과가 가장 큰 상관도를 가지며 원본 군집에 비해 상대적으로 개선이 되었음을 알 수 있다.

세 가지 형태의 평가자 군집 사이에 비교 분석을 통해 얻을 수 있는 가장 중요한 결론은 다음과 같다. 이 결과에서 가장 중요한 부분은 기존 음질 연구 과정 중에 특이 성향으로 고려되어 제거되었던 12명 중에 8명은 평가자 개별의 특이성 보다 신뢰성이 근본적인 문제이며 나머지 4명은 부적절한 평가 기준에 의해 불필요하게 제거되었다. 평가자 1명의 청음평가 결과를 얻기 위해 소요되는 시간과 비용들을 고려한다면 평가자 39명 중에 12명의 데이터를 배제한다는 것은 막대한 손실이다. 음질 연구 과정에 새로운 검증 모델의 추가 적용 및 유용성 검증에 관한 연구의 궁극적인 목적은 제약된 시간과 비용 속에 충분한 신뢰성이 기반이 되는 정확한 청음평가 결과를 얻기 위한 것이다.

상관관계 분석의 경우 평가자 군집의 청음평가 결과가 필요하지만 다수의 평가자들이 청음평가를 수행하기에 앞서 신뢰성 문제로 일관된 결과를 얻는데 어려움이 존재한다. 반면에, 판단 오류 모델은 평가자 개별의 평가 결과 내에서 판단 오류의 가능성에 접근하기 때문에 평가 대상의 이러한 결과는 기존 검증 과정에서 사용되는 상관관계 분석 단계의 앞에 이 연구에서 제안된 판단 오류 모델이 적용되는 경우 보다 신뢰성 측면에서 개선된 상관관계 결과를 얻을 수 있음을 보여주며 이러한 결과들을 기초로 제작되는 음질 인덱스 또한 기존 결과에 대비하여 정확성과 신뢰성 측면의 문제들이 개선하는데 도움이 될 것이라고 예상한다.

6. 선형결정오차 모델

여기에서는 앞서 실험 결과를 토대로 도출된 사람의 청각계 인지 특성의 한계에 의한 판단 오류

가능성을 불확실성 개념 측면에서 정의를 하고자 한다. 청음평가 과정 중에 판단 오류의 가능성은 정규 분포 상에서 평가용 소음들 사이에 상대적인 위치 차이와 연관되어 있음을 확인하였고 Fig. 9(c)의 확률 분포 특성은 다음과 같이 지수 함수 형태로 회귀 분석을 통해 표현 할 수 있다.

$$y = \alpha_1 \cdot \exp(\beta_1 \cdot x) \text{ for } 0 < x < 5\sigma \quad (9)$$

여기에서 식 (9)의 α_1 와 β_1 는 Table 9에 명시되어 있으며 불확실성 개념의 정량적 표현을 위해 다음의 관계식을 통해 y축의 물리량을 확률에 새넨 엔트로피(shannon entropy)로 변환하였다.

$$E_s = \sum_{i=1}^k -p_i \cdot \log_2(p_i) \quad (10)$$

정규 분포 상에서 평가용 소음들 사이에 차이가 0에 가까운 경우 Fig. 9(c)에서 볼 수 있듯이 0%와

Table 9 Parameters of a curve fitting model for the possibility of refined decision error

α_1	β_1	R
0.512	-0.9053	0.9974

Table 10 Possibility of refined decision error based on the Shannon entropy

Normalized difference	Shannon entropy
0	0.9997
0.5 σ	0.9093
σ	0.7297
1.5 σ	0.6098
2 σ	0.4046
2.5 σ	0.2978
3 σ	0.1904
3.5 σ	0.1118
4 σ	0.0000
4.5 σ	0.0000
5 σ	0.0000

Table 11 Parameters of a curve fitting model for the possibility of refined decision error based on the Shannon entropy

α_1	β_1	R
-0.2693	1.0029	0.9941

50%의 판단 오류 가능성은 불확실성 측면에서 Fig. 13과 Table 10에서 볼 수 있듯이 각각 0과 1의 값을 가진다. 그리고 이러한 선형 관계는 회귀분석을 통해 다음과 같이 표현 할 수 있다.

$$E_s(x) = \begin{cases} \alpha_2 \cdot x + \beta_2 & \text{for } x < 3.724\sigma \\ 0 & \text{for } x > 3.724\sigma \end{cases} \quad (11)$$

여기에서 α_2 와 β_2 는 Table 11에 명시되어 있으며 Fig. 14에서 판단 모호성은 프린터 작동 소음의 경우 평가용 소음들 사이에 상대적인 차이가 증가함에 따라 -0.269 Es/ σ 의 기울기를 가지고 선형적으로 감소하며 x축 상에서 3.724 σ 이상 차이가 나는 경우 판단 오류의 가능성 없이 명확한 판단이 가능하다는 것을 의미한다. 그리고 이러한 결과는 비음향 요인에 의한 판단 오류의 가능성은 개별 평가자 특성에 의존하여 예측하는 것이 어렵지만 사람의 청각계 인지 특성의 한계에 근거한 판단의 모호성은 불확

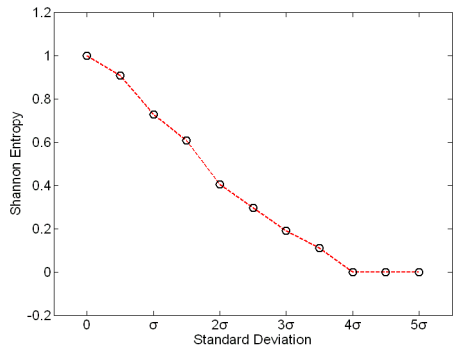


Fig. 13 The linear characteristics of decision error based on Shannon entropy

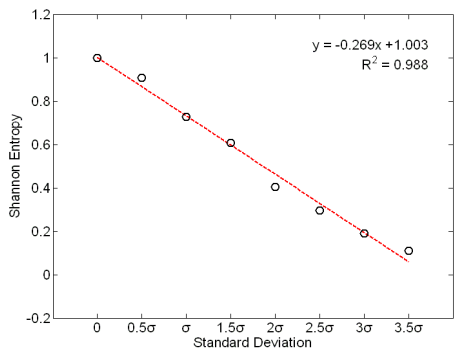


Fig. 14 The linear regression line for the decision error based on Shannon entropy

실성 측면에서 선형 관계식으로 모델링 하는 것이 가능함을 보여준다.

7. 결 론

이 연구는 음질평가 과정 중에 비음향 요인들에 의한 청음평가자들의 결정 오차 가능성을 정량적으로 분석하고 음질 인덱스의 일반성과 신뢰성을 개선하기 위해 제안된 결정 오차 모델의 유용성 검증 및 응용 연구를 수행하였다. 주관적 감성평가 결과의 검증 과정 중에 특히 성향과 신뢰성 문제를 구별하는데 용이한 기준과 이에 부합하는 결과가 도출되었다. 향후 연구에서는 토널 소음 성분 평가용 모델 개발에 관한 결과와 함께 개선된 음질 인덱스를 도출하기 위한 기초 연구 자료로 활용 될 예정이다.

후 기

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 MEST(No. 2010-0014260).

참 고 문 헌

(1) Park, S. W., Lee, H. H., Na, E. W., Lee, S. K., Park, Y. J. and Kim, J. W., 2010, Development of Sound Quality Evaluation System for a Printer Noise Based on Human Sensibility, Transactions of the Korean Society for Noise and Vibration Engineering, Vol. 20, No.5, pp. 427-436.

(2) Park, S. W., Yang, H. J., Na, E. W., Lee, S. K., Park, Y. J. and Kim, J. W., 2010, Identification of Printer Noise Source and Its Sound Quality Evaluation System Development, Transactions of the Korean Society for Noise and Vibration Engineering, Vol. 20, No.11, pp. 1018-1024.

(3) Kim, E. Y., Lee, Y. J. and Lee, S. K., 2012, Tonality Design for Sound Quality Evaluation in Printer, Transactions of the Korean Society for Noise and Vibration Engineering, Vol. 22, No. 4, pp. 318~327.

(4) Hansen, H. and Weber, R. 2009, Semantic Evaluations of Noise with Tonal Components in Japan, France, and Germany: a Cross-cultural Comparison, J. Acoust. Soc. Am., Vol. 125, No. 2, pp. 850~862.

(5) Hellman, R. P., 1982, Loudness, Annoyance, and Noisiness Produced by Single Tone Noise Complexes, J. Acoust. Soc. Am., Vol. 72, No. 1, pp. 62~73.

(6) Lee, K. H., 2006, Perception of Tones in Machinery Noise and Its Influence on Annoyance, Purdue University, Ph.D Thesis.

(7) Takanashi, A., 2008, Study of Problems with Noise and Sound Quality Evaluations of Copying Machines, Printers, and MFD, SQS 2008.

(8) ISO 11201:1995, Acoustics - Noise Emitted by Machinery and Equipment - Measurement of Emission Sound Pressure Levels at a Work Station and at Other Specified Positions - Engineering Method in an Essentially Free Field Over a Reflecting Plane.

(9) Otto, N., Amman, S., Eaton, C. and Lake, S., 2001, Guidelines for Jury Evaluations of Automotive Sounds, Sound and Vibration, Vol. 35, pp. 24-47.

(10) ISO 532B:1975, Acoustics - Method for Calculating Loudness Level.

(11) Fastl, H., Patsouras, C. H., Kuwano, S. and Namba, S., 2001, Loudness, Noisiness, and Annoyance of Printer Sounds. In: Wiss. Hrsg.: Otto v. Estorff, Editor, Tagungsband Fortschritte der Akustik - DAGA 2001, Hamburg, pp. 388-389.

(12) Kim, T. G., Lee, S. K. and Lee, H. H., 2009, Characterization and Quantification of Luxury Sound Quality in Premium-class Passenger Cars, Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, Vol. 223, No. 3, pp. 343-353.

(13) Lee, S. K., Kim, B. S. and Park, D. C., 2005, Objective Evaluation of the Rumbling Sound in Passenger Cars Based on an Artificial Neural Network, Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, Vol. 219, No. 4, pp. 457-469.