

소리 선호도 조사에 사용하는 청음평가법의 비교 연구

A Comparative Study on Methods of Jury Test used for Sound Preferences

조현호*, 성원찬*, 김성현**, 박동철**, 강연준†

Hyeonho Jo, Wonchan Seong, Seonghyeon Kim, Dongchul Park, Yeonjune Kang

Key Words : Psychoacoustics(음향심리학), Sound Quality(음질), Jury Test(청음 평가), Rating Methods(평점 평가법), PCM(쌍대 비교법)

ABSTRACT

Quantifying the perception of human is of great interest for the optimal characterization of the interaction of persons with their environment. The direct approach to such measurements is based on jury-test methods. In this study we considered which jury-test method is more efficient and exact way to measure the preferences for operating sound of car. Moreover we needed to research selection criteria of jury-test methods in accordance with the properties of sound samples. We choose two sample sounds; door locks which generate simple impact sounds and window lifts which are consisted of long and complex sound. For each sound sample we conducted two jury-test methods respectively and discussed about the results

기 호 설 명

The subscript D : 도어 락 작동음

The subscript W : 윈도우 리프트 작동음

1. 서 론

기술의 발전과 생활 수준의 향상은 NVH 연구에 다양한 변화를 불러왔다. 그 중 가장 괄목할 만큼 성장한 분야는 단연 음질 연구로 대표될 수 있는 안락함에 관한 연구이다. 기존의 NVH 연구의 목표가 단순히 소리를 더욱 조용하게 만드는 것에 초점을 맞췄던 것과 달리 점차 사람이 들었을 때 좋은 소리를 만들어 내는 방향으로 변모해 가고 있는 것이다. 이러한 연구 흐름은 소리에만 국한 되는 것이

아니다. 진동 저감 기술 역시 사람이 느끼기에 편안함을 느낄 수 있으려면 어떻게 해야 하는가로 변화해 가고 있고, 나아가서는 전체적인 소리와 진동의 브랜드화 마저 진행되고 있다.

이런 사람의 안락함을 연구할 때 필수적으로 선행되어야 하는 것이 실제 사람이 느끼는 감정, 인식들을 수치화하여 나타내는 작업이다. 과거부터 사람의 느낌을 객관화하려는 노력은 계속 이루어져 왔으며 그 방법은 청음 평가(jury test), 생리적 변화 추적 physiological reactivity), 행동 변화 추적 (overt behavioral acts)의 세가지로 구분될 수 있다.⁽¹⁾ 그 중 가장 직접적이고 분명하여 주로 사용되는 방법은 청음평가이다.

청음 평가는 그 기법에 있어 매우 다양한데, 연구자는 연구의 목적, 자극(stimulus)의 성질에 따라 그 선택을 달리하곤 한다. 가령 도심 교통 소음의 노출 시간에 따라 사람이 보이는 주관적 반응을 연구하는 데에 쌍대 비교법을 사용하여 사람의 주관적 특성을 객관화한 것이 있다.⁽²⁾ 또한 직접적으로 냉장고의 소리가 사람에게 어떻게 들릴지에 대한

† 교신저자; 정회원, 서울대학교 기계항공공학부
E-mail:yeonjune@snu.ac.kr
Tel:(02)880-1691, Fax:(02)888-5950

* 정회원, 서울대학교 대학원 기계항공공학부

** 현대자동차(주)

연구에서는 평점 평가법(Rating Methods)을 사용하였다.⁽³⁾ 뿐만 아니라 의미 분별 척도법(Semantic differential method)를 사용하여 제품 디자인의 목표 설정을 활용하기도 한다.⁽⁴⁾

본 연구는 자동차 차실 내에서 동작하는 다양한 부품들의 작동음을 가지고 음질을 평가할 때 연구자가 선택할 수 있는 가장 효율적이고 정확한 방법은 무엇일까 하는 궁금증으로 시작되었다. 과연 사람의 선호도를 조사하기 위해 어떤 방법을 취하였을 때 시간, 비용적 부담은 최소화 하면서도 결과에 대한 신뢰성을 확보할 수 있는지, 또한 각 방법에 이 가지고 있는 장단점은 무엇인지에 대해 고찰해보고자 하였다. 그를 위하여 같은 음원 샘플을 평점 평가법(Rating Methods)과 쌍대 비교법(Paired Comparison Methods)으로 각각 평가하여 그 결과를 비교하였다. 추가로 본 연구를 통하여 음원의 특성에 따른 청음 평가 기법 선택 기준 역시 고찰하고자 하였다.

2. 도어 락과 윈도우 리프트의 작동음 특성

본 연구는 도어 락, 윈도우 리프트 두 가지 종류의 부품 작동음을 대상으로 진행되었다. 두 부품 모두 차실 내에서 작동음을 들을 수 있는 부품으로 모터로 작동한다는 공통점이 있지만 소리의 시간이나 특성에 있어 큰 차이를 보인다. 소리의 특성에 따른 청음 평가 기법 선정의 유의점을 알아보기 위해 두 부품을 선정하여 연구를 진행하였다.

두 부품의 음원 샘플은 모두 실차에 장착된 상태에서 운전석에 앉아서 녹음하여 얻었으며, 음원 녹음은 사람의 쌍이 특성(Binaural properties)을 반영하기 위해 마이크로폰 헤드셋 BHS II(HEAD Acoustics 社)을 이용하여 진행하였다. 청음 평가를 위한 음원 샘플의 수는 도어 락이 13 개(AD~MD), 윈도우 리프트가 11 개(AW~LW)이다. 차량은 차급, 브랜드 등을 고려하여 선정되었다.

Fig. 1과 Fig. 2는 각각 도어 락, 윈도우 리프트의 작동음을 시간에 따른 음압 레벨 변화를 통해 나타낸 그래프이다.

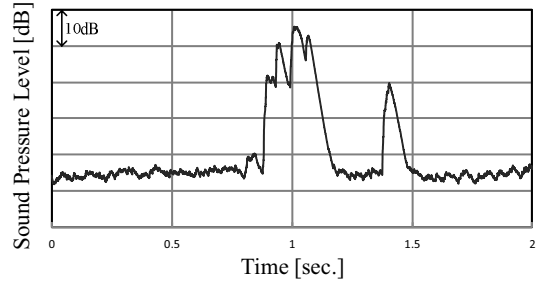


Fig. 1 An operation sound of door lock

도어락은 1 초 미만의 짧은 시간동안 발생하는 충격음으로 부품의 모델마다 세부적 사항으로 차이가 있긴 하지만 공회 작동 시작 시 발생하는 일차 충격음과 0.3~5 초 후에 발생하는 후음으로 구성된다⁽⁵⁾. 일차 충격음과 후음은 평균적으로 15dB 정도의 차이가 난다. 차량 모델 별로 소리의 특성은 차이가 있지만 청감적인 느낌의 차이는 크지 않으며 선호도를 평가하는데 있어 어려움은 그리 크지 않은 소리에 속한다.

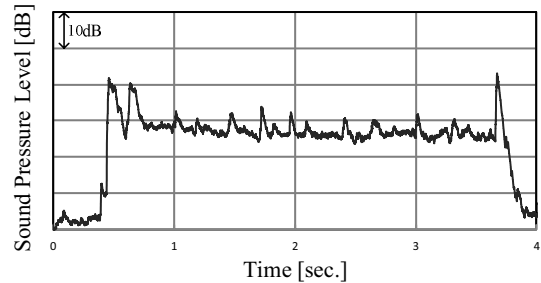


Fig. 2 An operation sound of window lift

윈도우 리프트는 비교적 긴 시간이라 할 수 있는 3~5 초 간 발생하는 소리로서, 시작 충격음, 글래스 이송음, 끝 충격음으로 대략적인 구분이 가능하다⁽⁶⁾. 시작과 끝 충격음은 이송음에 비해 약 10dB 정도 큰 소리이다. 또한 도어 락과는 달리 모델 간 소리의 차이가 존재하는데, 그 차이는 작동 시간, 모듈레이션 여부, 충격음의 크기, 주파수 등으로 다양한 범위에서 나타난다.

3. 청음 평가

청음 평가는 정상적인 청각 기능을 가지고 있는 20~30 대 남성 17 명을 대상으로 진행하였다. 나이가 다양하지 않지만 20 대 후반에서 30 대 초반인 사람의 청각 능력이 평균에 가깝다는 점을 바탕으로 적합한 실험군이라 판단하였다⁽⁷⁾.

음원 재생(Play-back)은 HEAD acoustics 社의 PEQ V 와 고성능 헤드폰을 사용하였으며, 실험에 익숙하지 않은 평가자를 위하여 충분한 사전 교육을 실시한 후에 청음 평가를 진행하였다. 또한 음원은 차급 혹은 브랜드에 따른 영향을 없애기 위해 블라인드 상태로, 무순(無順)으로 들려 주었다.

3.1 평점 평가법

본 연구에 사용된 청음 평가 기법 중 하나인 평점 평가법(Rating Methods)는 샘플에 대해 직접적인 점수를 부과함으로써 평가를 하는 방법으로 5 점, 7 점, 11 점 척도가 가장 많이 사용된다. 각각의 기준에 따라 결과에 있어 어떠한 차이가 존재하는지, 그를 보정하기 위해서 어떠한 노력을 할 수 있는지는 Kim 등의 연구⁽⁸⁾를 참고하였다. 본 연구는 Fig. 3 에 나타낸 것처럼 간단한 표를 이용하여 11 점 척도(1 점~11 점) 평점 평가법으로 진행되었다.

WINDOW LIFT											
	1	2	3	4	5	6	7	8	9	10	11
A			◎								
B					◎						
C							◎				
D									◎		
E		◎									
F											
G	◎										
H									◎		
I											◎
Comments											

Fig. 3 Test sheet used for jury test (rating methods)

평점 평가법으로 청음 평가를 진행하면서 동시에 평가자들로부터 평가의 이유를 간단히 수집할 수 있는 인터뷰 표(Comments)를 제공하였다. Table 1 과 2 는 각각 도어 락과 윈도우 리프트 작동음의 청음평가에서 얻은 인터뷰 결과를 간략하게 나타낸 표이다.

Table 1 Results of interview of jurors for door lock

Engineering expressions	Subjective expressions	Total number of mentions
Loudness related	'quiet', 'loud', etc.	42
Sharpness related	'too high', 'light like a sound of toys', etc.	58
Roughness related	'luxurious soft', 'scratching sound', etc.	5
Time related	'too fast', 'too slow', etc.	3

Table 2 Results of interview of jurors for window lift

Engineering expressions	Subjective expressions	Total number of mentions
Loudness related	'quiet', 'loud', etc.	49
Sharpness related	'too high', 'light like a sound of toys', etc.	36
Roughness related	'scratching sound', etc	28
Time related	'too fast', 'too long', etc	6
Modulation related	'changing sound',	11
Etc.	'powerful', 'shut tightly', 'faintful'	3

이를 통해 고득점과 저득점을 한 샘플들 간에 어떤 차이가 있는지, 평가자들이 어떤 기준을 가지고 평가에 임했는지를 확인할 수 있었다. 이 데이터는 향후 선형 회귀 분석(Linear regression analysis)를 위한 수치 선택이나 청음 평가 결과 비교 분석에 사용되었다. 형용사를 통해 소리를

묘사한다는 점에서 의미 분별 척도법(Semantic differential methods)와 유사하지만 본 인터뷰의 결과는 단순히 회귀식 구성을 위한 아이디어를 얻는데 사용하였다.

도어 락의 경우 인터뷰 결과 나온 주요 음질 인자들이 소리의 크기(Loudness related)나 날카로움(Sharpness related)에 관련된 것들에 집중돼 언급된 것에 반해 윈도우 리프트의 인터뷰 결과는 크기나 날카로움 이외에도 거칠기(Roughness related), 흔들림(Modulation related)에 관련된 인자 역시 소리의 고급감 선호도에 큰 영향을 미친 것으로 나타났다. 이는 앞서 살펴보았듯이 도어 락의 작동음은 모델 별로 비슷한 소리 특성을 갖는 충격음(impact sound)인 것에 비해 윈도우 리프트는 비교적 긴 시간 동안 발생하고 여러 가지 다양한 특성을 갖기 때문인 것으로 볼 수 있다.

3.2 쌍대 비교법

본 연구에 사용된 또 다른 청음 평가 기법인 쌍대 비교법(Paired Comparison Methods)는 음원 샘플을 일대일 비교를 통해 선호도나 적합도를 판단하여 최종적으로 모든 비교 결과를 종합하여 순위를 정하는 방법이다.

WINDOW LIFT		
Pair 01	▶ PLAY Sound A	▶ PLAY Sound B
	Like	Like
Pair 02	▶ PLAY Sound A	▶ PLAY Sound B
	Like	Like
Pair 03	▶ PLAY Sound A	▶ PLAY Sound B
	Like	Like

Fig. 4 Program used for jury test (PCM)

평가는 Fig. 4에 나타난 것처럼 직접 제작한 평가표를 이용하여 진행했으며 평가자가 자유롭게 음원을 재생할 수 있고, 선택을 자유롭게 수정할 수 있도록 하였다. 쌍대 비교법은 샘플 수가 적을 경우에는 비교적 평가가 간단하여 사용하기 적합하지만 샘플의 수가 많아짐에 따라 비교해야 하는 조합이 늘어날 경우 평가자의 집중력이 흐려져 정확한 결과를 얻기 힘들어질 수 있다는 단점이 있다⁽⁹⁾. 본 연구는 그러한 오류의 최소화를 위하여 각각의 평가자의 일관성 지수(Coefficient

of consistency)를 확인하여 일관성을 유지하고 있는 평가자의 데이터만을 최종 결과에 반영하였다.

도어 락 작동음을 대상으로 한 청음 평가에서는 평가자들의 일관성 지수가 전반적으로 높은 값을 보여 0.8을 넘기는 평가자가 총 17명 중 15명이었다. 하지만 윈도우 리프트를 대상으로 한 청음 평가에서는 평가자들의 일관성 지수가 전체적으로 낮은 것을 확인할 수 있었다. 모델 별로 특성 서로 달라 세분화된 기준이 존재할 수 있기 때문에 판단할 수 있다.

4. 청음 평가 결과 및 고찰

4.1 도어 락의 청음 평가 결과 비교

도어 락의 작동음 샘플을 이용하여 평점 평가법과 쌍대 비교법을 모두 이용하여 청음 평가를 수행하였고 그를 통해 얻은 결과를 Fig. 5에 나타내었다.

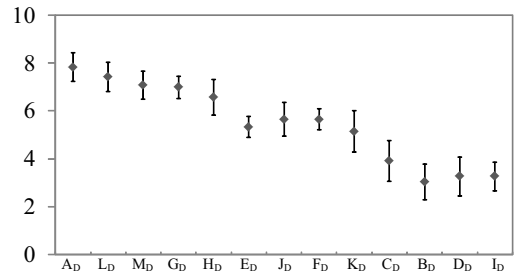


Fig. 5 Comparison of the results from the two test methods, for sounds of door lock

그래프의 가로축은 쌍대 비교법을 통해 얻은 고급감 선호도 순위를 이용하여 순위가 높은 순으로 음원 샘플들을 나열한 것이고 세로축은 평점 평가법을 통해 얻은 점수를 나타낸다. 그래프 상에 표시된 표준 편차는 평점 평가법에서 평가자들 간의 편차를 나타낸다.

도어 락 작동음의 고급감 청음 평가 결과는 그래프를 통해 확인할 수 있듯이 평점 평가법과 쌍

대 비교법 양 기법 간에서 큰 차이를 나타내지 않고 있는 것으로 나타났다. 실제로 평점 평가법을 통해 얻은 점수와 쌍대 비교법을 통해 얻은 순위의 상관관계(Correlation Index)는 약 97%로 두 결과가 매우 유사하다는 것을 알 수 있었다. 또한 각 모델들이 갖는 표준 편차는 약 1.5를 넘지 않는 정도로 평가자들 간의 상관 관계(Correlation)도 상당히 높은 수준인 것을 확인할 수 있었다.

쌍대 비교법이 평점 평가법에 비해 오랜 시간이 걸려 훈련이 잘 되지 않은 평가자들에게 적합하지 않다는 점을 고려하였을 때 이러한 결과는 상당히 유의미한 것이라 할 수 있다.

두 방법의 결과가 완벽하게 일치하지는 않는 현상에 대해 Giovanni 등⁽¹⁰⁾은 샘플 간 상관 관계(coherence)가 있거나 유사성이 존재하는 경우에 나타날 수 있는 현상으로 본 바 있다. 그러한 오차를 해결하기 위해 특성이 유사하여 오류를 불러 일으킬 만한 샘플을 제거하거나 전체적으로 유사성을 가지는 경우에는 차이를 나타낼 수 있는 샘플을 제작하여 평가에 끼워 넣을 것을 제안하고 있다.

4.2 윈도우 리프트의 청음 평가 결과 비교

Fig. 6은 윈도우 리프트 작동음 샘플을 대상으로 실시한 두 가지 방법의 청음평가 결과를 비교할 수 있도록 나타낸 그래프 이다.

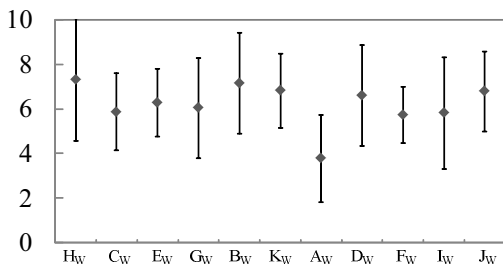


Fig. 6 Comparison of the results from the two test methods, for sounds of window lift

쌍대 비교법을 통해 얻은 순위와 평점 평가법을 통해 얻은 점수에서 큰 상관 관계가 나타나지 않는 결과를 나타내고 있다. 쌍대 비교법을 통해

얻은 결과에서는 7 등으로 고급스러운 것으로 평가됐던 AW 차량이 평점 평가법으로 평가했을 때는 최저점을 받는 등 전체적으로 큰 차이를 보이고 있는 것을 확인할 수 있다. 두 결과 간의 상관 관계를 객관적 수치로 나타내어도 약 27%로 매우 낮음을 알 수 있다. 또한 표준편차 역시 도어 락과는 달리 매우 큰 값을 보인다. 이는 평가자마다 윈도우 리프트 작동음의 고급감 선호에 차이가 존재함을 나타내는 지표로 볼 수 있다. 앞서 살펴본 바와 같이 윈도우 리프트는 작동음이 다양한 소리로 구성되어 있고, 모델에 따른 소리의 특성도 매우 다른 부품으로서 평가자마다 주요하게 생각하는 음질 인자가 다르기 때문인 것으로 판단된다.

4.3 윈도우 리프트의 청음 평가 클러스터링 결과

평가자의 성향에 따라 고급감 선호도 기준이 달라지는 것을 확인하고 심한 이상점 제거를 실시하였다. 평점 평가법 청음 평가 결과 중 평균과 상관 관계(Correlation)가 70% 이하인 평가자 7명을 제거하여 다시 한번 쌍대 비교법 결과와 비교를 실시하였다. Fig. 7은 수정된 윈도우 리프트 청음 평가 결과를 다시 쌍대 비교법과 비교한 그래프이다.

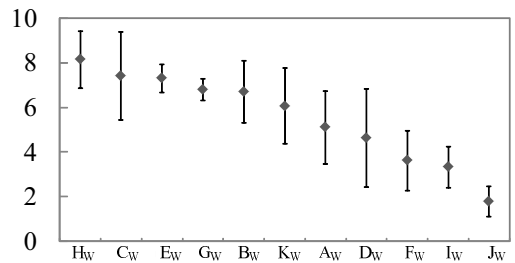


Fig. 7 Comparison of the modified results from the two test methods, for sounds of window lift

앞서 살펴본 결과와는 달리 쌍대 비교법을 통해 얻은 순위와 평점 평가법을 기준으로 매긴 순위가 완벽히 일치함을 확인할 수 있다. 하지만 표준 편차는 역시 도어 락과 비교했을 때 매우 큰 값을 보이고 있고, 지나치게 평가자의 수가 줄어들었다는 약점을 안고 있다.

윈도우 리프트와 같이 작동음의 구성 요소가 다양하여 평가자들마다 상이한 기준이 존재할 수 있는 부품의 경우에는 단순히 쌍대 비교법이나 평점 평가법을 통한 선호도 줄 세우기보다는 보다 세분화된 클러스터링을 통해 목표와 타겟에 맞는 결과를 얻어낼 수 있는 실험을 설계하는 것이 매우 중요할 것이다.

5. 결 론

본 연구를 통해 음질 연구에 사용되는 청음 평가 기법 중 가장 대표적으로 사용되는 평점 평가법(Rating Methods)와 쌍대 비교법(Paired Comparison Methods)은 음원의 특성에 따라 그 결과에 있어 차이를 보일 수 있다는 사실을 알 수 있었다. 평가자가 충분히 잘 교육되었으며, 음원의 종류가 임팩트 소음과 같이 그 특성이 단순하여 평가가 용이하다는 전제 하에 둘은 매우 유사한 결과를 보였다. 하지만 음원이 길고 그 구성 성분이 다양하여 한 두가지의 기준으로 평가가 어려운 경우에는 두 방법 간에 다소 차이가 존재 할 수 있으며, 그러한 문제점을 해결하기 위해서는 AHP나 의미 분별 척도법(Semantic differential)과 같은 세분화된 기준을 갖는 청음 평가 기법이 필요함을 확인할 수 있었다.

후 기

본 연구는 현대자동차(주)와 한국연구재단(BK21 사업)의 지원으로 이루어 졌습니다.

참 고 문 헌

(1) Lang, P. J. 1969. The mechanics of desensitization and the laboratory study of human fear. In C. M. Franks (Ed.), Assessment and status of the behavior therapies. New York: McGraw Hill.
 (2) Hur Deog-Jae, Jo Kyoung-Sook, Choe Byong-Ho. 2008. Study on the Subjective Responses of Loudness and Annoyance according to Exposed Time for the Traffic Noise. Transactions of the Korean Society for Noise and Vibration Engineering. Vol. 17, No. 3, pp. 241~248.
 (3) Jaeman Joo, Jaewon Lee, Jinkyung Lee,

Sangkyoun Oh. 2004. Evaluation of the Sound Quality for the Refrigerator by Consumer's Sense of Hearing. Transactions of the Korean Society for Noise and Vibration Engineering's Confference. pp. 145~148.
 (4) Thomas Hempel, Nicolas Chouard. 1999. Evaluation of interior car sound with a new specific semantic differential design. The Journal of the Acoustical Society of America.
 (5) Hyeonho Jo, Weonchan Sung, Seonghyeon Kim, Dong Chul Park, Yeon June Kang. 2013. A Study on Developing Sound Quality Index of Car Door Latch and Improving Sound Quality by Changing Door Latch Assembly Design. Conference Book of the Korean Society for Noise and Vibration Engineering. Vol.2013. No.2. pp.519~524.
 (6) Seonghyeon Kim, Dong Chul Park, Hyeonho Jo, Weonchan Sung, Yeon June Kang. 2014. The Development and Application of Sound Quality Index for the Improving Luxury Sound Quality of Road Vehicle Power Window System. Transactions of the Korean Society for Noise and Vibration Engineering. Vol. 24, No.23, pp. 108~116.
 (7) Hugo Fastl, Eberhard Zwicker, 2006, Psychoacoustics : Facts and Models, , Springer, Berlin.
 (8) Deuk Sung Kim, Seo Il Chang, Joon Hee Ko, Jin Mo Kim. 2009. On the comparison of 5-point, 7-point and 11-point scale by Jury Evaluation Test.
 (9) Norm Otto, Scott Amman, 1999. Chris Eaton, Scott Lake. Guidelines for Jury Evaluations of Automotive Sounds. SAE TECHNICAL PAPER. 1999-01-1822.
 (10) Giovanni B. Rossi, Francesco Crenna, Michele Codda. 2003. Measurement of quantities depending upon perception by jury-test methods. Measurement. Vol. 34. pp.57~66.