PERFORMANCE AND PREFERENCE OF RESPONSE SCALES FOR SEMANTIC DIFFERENTIALS IN AUDITORY PERCEPTION AMONG UNIVERSITY STUDENTS

Wonyoung Yang* **and Jin Yong Jeon** †
Department of Architectural Engineering, Hanyang University, Seoul, Korea.

For Murray, my dearest mentor

Résumé

On sait que les échelles de réponse sont essentielles à la qualité des réponses. Une échelle numérique unipolaire à 11 points conforme à la norme ISO/TS 15666:2003 a été largement utilisée pour évaluer la perception auditive à l'intérieur et à l'extérieur, ainsi que pour les évaluations de terrain et la recherche psychoacoustique. Toutefois, dans de nombreuses disciplines, une échelle visuelle analogique a été utilisée à des fins académiques plus approfondies. Cette étude vise à comparer la performance et la préférence de deux échelles de réponse, une échelle visuelle analogique bipolaire et une échelle numérique unipolaire, pour les différences sémantiques dans la perception auditive à l'aide d'un dispositif basé sur le Web. Deux échelles de réponse différentes ont été comparées dans cinq stimuli acoustiques (niveau de bruit de fond de 38 dBA, bruits de l'eau et bruit du trafic de 42 et 61 dBA, respectivement) avec deux mesures répétées. Les deux échelles de réponse étaient acceptables pour leur fiabilité et leur sensibilité. Cependant, l'échelle analogique visuelle bipolaire était plus fiable que l'échelle numérique unipolaire à 11 points dans les mesures répétées, et l'échelle numérique unipolaire à 11 points était plus sensible que l'échelle analogique visuelle bipolaire pour distinguer les différences subtiles entre sources sonores. L'échelle analogique visuelle bipolaire était évidemment préférée par les participants. Le choix des adjectifs sémantiques est une condition préalable essentielle pour déterminer les échelles de réponse pour la perception auditive. En résumé, une échelle visuelle analogique unipolaire est proposée pour évaluer la perception auditive à des fins de recherche psychoacoustique chez les jeunes adultes instruits.

Mots clefs : échelles de réponse, échelle unipolaire, échelles bipolaire, échelle visuelle analogique, échelle numérique unipolaire à 11 points, préférence du répondant, intensité, caractère bruyant, agacement

Abstract

It is known that response scales are critical for achieving the quality of the responses. A unipolar 11-point numerical scale in accordance with ISO/TS 15666:2003 has been widely used for assessing auditory perception both indoors and outdoors, as well as for field assessments and psychoacoustic research. However, in many disciplines, a visual analogue scale has been used for more in-depth academic purposes. This study aims to compare the performance and preference of two response scales, a bipolar visual analogue scale, and a unipolar numeric scale, for semantic differentials in auditory perception using a web-based device. Two different response scales were compared in five acoustic stimuli (background noise level of 38 dBA, water sounds and traffic noise of 42 and 61 dBA, respectively) with two repeated measurements. Both response scales were acceptable for their performance of reliability and sensitivity. However, the bipolar visual analogue scale was more reliable than the unipolar 11-point numerical scale in repeated measurements, and the unipolar 11-point numerical scale was more sensitive than the bipolar visual analogue scale in distinguishing subtle differences between sound sources. The bipolar visual analogue scale was obviously preferred by participants. The choice of semantic adjectives is a critical prerequisite for determining response scales for auditory perception. In summary, a unipolar visual analogue scale is proposed for assessing auditory perception for psychoacoustic research purposes for young educated adults.

Keywords: Response scales, unipolar scale, bipolar scale, visual analogue scale, 11-point numerical scale, respondent's preference, loudness, noisiness, annoyance

1 Introduction

The evaluation of the acoustic environment is mainly based on the subjective rating scale responses to questions about acoustic sensation and perception. The quality of the responses depends on the design of the response scales [1]. The 5-point verbal and 11-point numerical scales proposed by ICBEN (International Commission on Biological Effects of Noise) [2] are the two major methods for measuring the response to subjective questions about acoustic sensation and perception.

For better understanding of how humans react to sound, it is necessary to investigate both the negative and positive aspects of sound. The ICBEN recommendation was developed for assessing and comparing environmental noise annoyance, and was later adopted as the international

* wyang@hanyang.ac.kr
† jyjeon@hanyang.ac.kr

standard ISO/TS 15666:2003 [3]. No positive acoustic aspects were taken into account in the ICBEN methods. About a decade later, the soundscape was defined as an acoustic environment perceived or experienced and/or understood by a person or people, in the context of the first ISO standard, ISO/TS 12913-1:2014 [4]. Situational differences between measuring annoyance and measuring soundscape preference were taken into consideration in the methods. The 5-point verbal response scale was adopted as the international standard ISO/TS 12913-2:2014 [5] for soundscape data collection and reporting.

For more in-depth psychological understanding of human sensation, perception, and recognition of sound as well as speech, a visual analogue scale (VAS) was used in previous psychoacoustic studies [6-16]. VAS may be preferred in research due to better sensitivity [17]. VAS is known for its high sensitivity to discriminate subjective feelings [18]. At first glance, it may seem that VAS may have better precision and be more sensitive to detect changes than numerical scales, simply because of the finer gradations of response levels [19].

Comparisons of the visual analogue scale and the numerical scale have been reported in clinical, market research, and psychology [17, 20-31], and these studies have yielded contradictory findings. The use of VAS on a multipoint scale is beneficial with regard to sensitivity,[17, 21, 24] respondent preference [20], accuracy [32], and response time [30]. On the other hand, a few studies have evidence supporting the use of a multipoint scale over VAS regarding response rates [25, 26], respondent preference [33], and response time [25, 28]. However, many studies reported no significant difference in the use of the scales [22, 23, 25-28, 31, 33].

Munson et al. [11] recommended the use of continuous rating scales in their phonetic research, because visual analogue scales are well correlated with acoustic parameters and can be easily implemented both in field research on phonological acquisition and in the clinic. In audiology, VAS loudness and VAS annoyance are valid and effective measurements for capturing the reductions in the severity of tinnitus in patients with chronic tinnitus [13]. In indoor environmental discipline, although visual analogue scale has been used in several laboratory studies [8-10, 15, 16], to date, no study has compared response scales to verify the quality of subjective responses in psychoacoustic research.

Recently, a few comparative studies between 5-point verbal and 11-point numerical scales reported noise annoyance [34-37]. Brink et al. [34] found that standardized average annoyance scores were slightly higher when using the 11-point numerical scale, whereas the percentage of highly annoyed respondents was higher based on the 5-point verbal scale. The frequency distributions of the two upper categories (very and extremely) of 5-point verbal scale in the highest categories out of 10 of 11-point numerical scale are almost the same. Nguyen et al. [35] expanded the annoyance response study in Japan and Vietnam. In Japanese, it was found that the highest category of 11-point and 5-point scales basically corresponds to the top category of 5-point and 11-point scales, respectively. However, in Vietnamese, the highest category of 5-point and 11-point scales corresponded to the two upper categories of 11-point and 5-point scales, respectively. It was found that logistic regression curves with high annoyance, defined by the three upper categories of the 11-point scale, have a good fit to the quadratic curves with high annoyance, defined by a cutoff point of 28%, as recommended by Miedema and Vos. [38]. However, these curves are separated from logistic regression curves with high annoyance, defined by the two upper categories of the 5-point scale in both countries. Bjerre et al. [36] reported on consistency between the 5-point verbal scale and the 11-point numerical scale in their on-site and laboratory evaluations of the urban soundscape. Tristán-Hernándes et al. [37] found no statistically significant differences between the 5-point and 11-point scales when evaluating noise annoyance inside university facilities.

The purpose of this study was to investigate the performance and preference of two response scales, a bipolar visual analogue scale, and a unipolar numerical scale, for semantic differentials in auditory perception using a web-based device. Specific research interest was the impact of polarity and types of the scale, which were the questionnaire related factors in young adults.

2 Method

2.1 Participants

Overall, 50 university students (23 men and 27 women) participated in a 60-minute session. No hearing impaired participants were examined by the interview. Informed consent was obtained from each of the participants, and they received financial support for their participation. The mean age of participants was 22.5 (S.D. 2.0) years.

2.2 Testing laboratory and experimental conditions

The experiment was conducted in a test laboratory (4.0 m × 5.0 m × 2.4 m), which was built for indoor environmental research. The indoor environment was maintained at the air temperature of 24.5 ºC and humidity of 40%. The ventilation system was in operation during the experiment. The local air velocity was measured to be less than 0.1 m/s. The mean illuminance levels along the desk surface during the experiments were 995.0 lx.

A loudspeaker system (Turbosound Milan M10) was used as a sound source and was located on the rear side to minimize the spatial sensitivity of sound sources. The reverberation time in the testing laboratory was measured as 0.3 s at 500 Hz for octave bands (01 dB dB4). The ambient noise level in the laboratory was 38 dBA (01 dB solo) when the thermal and ventilation systems were operated.

Four different sound sources (water sound and traffic noise of 42 and 61 dBA) were reproduced through the loudspeaker, considering the average measured daytime noise exposure levels [39]. Water sounds, representing a positive sound, were acquired from an open website [40], and traffic noises, representing a negative sound, were recorded in the living room of a residential building. The levels of the sound sources were adjusted using an audio controller. The differences in sound level across the positions of the participants were measured at ± 0.3 dBA. Figure 1 shows the octave band frequency spectra of the sound sources, including ambient noise in the chamber.

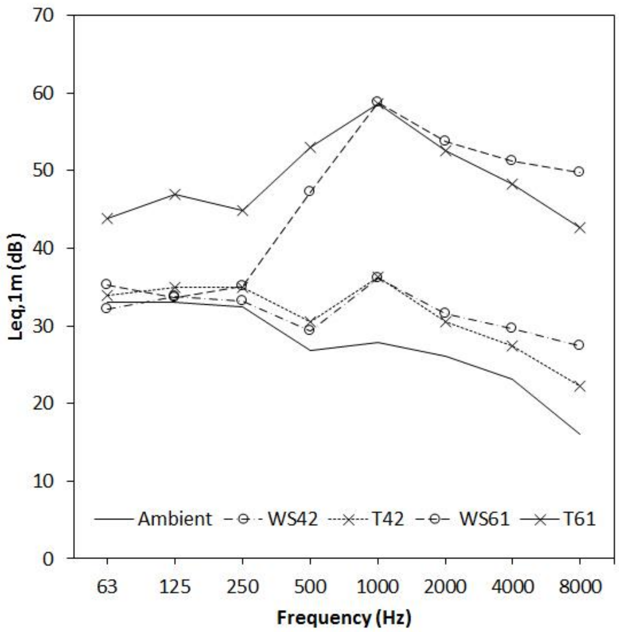

Figure 1: Frequency spectra of sound sources

2.3 Response scales

A web-based tablet interface was used for subjective assessments. Figure 2 shows the two response scales and their tablet interfaces. A unipolar 11-point numerical scale (unipolar11) with endpoint and midpoint labels was adopted based on ISO/TS 15666:2003,[3] which was developed for socio-acoustic noise annoyance surveys. It is assumed that a 0-to-10 scale would be more understandable and manageable than the shorter ones. Most people are familiar with the base-10 numeric systems through currency and other familiar counted materials. Radio buttons were also used to create 11 discrete scales from 0 to 10. Three verbal labels "Not at All," "Neutral," and "Extremely" were placed at the top of "0," "5," and "10." The number of questions has doubled on the bipolar scales, because a unipolar scale could only evaluate to a degree of one attribute.

A bipolar visual analogue scale (bipolar VAS) was introduced in the study. The questionnaire content was identical to the unipolar 11-point scale, except for the polarity. VAS consists of a plain, mostly horizontal line with a length of 100 mm and mostly verbal end labels. Respondents give a rating by placing a mark on the line. In this study, a numerical value from -10.0 to 10.0 was assigned to the responses for statistical analysis. A slider was placed at the left end in the default setting as an indicator of the rating mark. However, respondents were required not to drag, but click on the slider to avoid potential technical problems of dragging with their fingers.

The semantic attributes of the questionnaire were four pairs of adjectives: soft versus loud, quiet versus noisy, pleasant versus annoying, and uncomfortable versus comfortable. For a unipolar scale questionnaire, each semantic attribute was listed one by one. For a bipolar scale, soft, quiet, pleasant, and uncomfortable were positioned on the left end, and noisy, loud, annoying and comfortable were positioned on the right end.

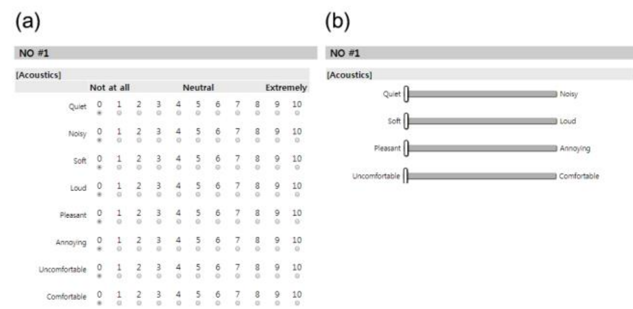

Figure 2: Two types of questionnaire: (a) unipolar 11-point, (b) bipolar VAS

2.4 Experimental design and procedure

A factorial within-subject design with repeated measurements was employed with two independent variables: response scale (unipolar 11 and bipolar VAS) and sound source (ambient, water sound 42 dBA, traffic noise 42 dBA, water sound 61 dBA, and traffic noise 61 dBA).

A maximum of six participants simultaneously assessed the acoustical conditions in a test laboratory. The response data provided by the participants were automatically saved on a server. In each session of 60 min, a 20-min adaptation period was implemented at the beginning of the session for relaxation and environmental adaptation, as shown in Figure 3.

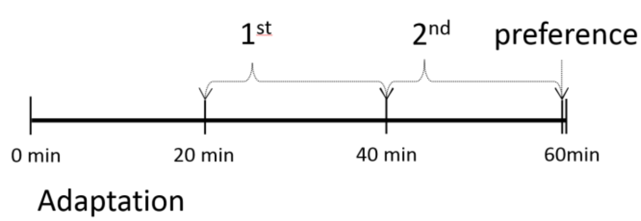

Figure 3: Experimental procedure for each session

Participants were seated during the adaptation period. Each sound stimulus was presented for 50 s, and a response time was provided until all participants in the test group submitted their responses. Ambient sounds for four different response scales were assessed at the beginning and at the end. Four sound sources combined with four response scales were randomly presented in each test session, and their replicas were also presented in random order.

At the end of the session, a paired comparison was conducted to investigate participants' preferences for the scales between the unipolar 11-point scale and the bipolar visual analogue scale.

2.5 Statistics

Statistical analyses were performed using two different approaches: original and normalized data analyses. The original data from respondents were used to analyze the correlation performance for reliability. Fisher's Z transformation was applied to compare the correlation coefficients of repeated measurements on each response scale. The original data were also applied to a factorial analysis of variance (ANOVA) to validate the effects of sound sources on each response scale. ANOVA is a powerful statistical test and it was used in this case, although normality cannot be guaranteed for subjective ratings [41, 42].

A repeated-measurement ANOVA was also used to test the scale factor for two repeated measurements. The original data were converted to unipolar 0.0-to-10.0 scales to perform ANOVA on two response scales with different numerical ranges. If a response value was greater than zero, it was treated as a right-end semantic attribute, and if a response value was less than zero, it was treated as a left-end semantic attribute. Three corrections (Greenhouse-Geisser, Huynh-Feldt, and the lower boundary) for violations of sphericity were used to test the sphericity. The Mauchly sphericity test requires more than three repeated measurements, but only two measurements were performed in this study. An epsilon (ε) value of 1 was found for the three corrections across all subjective attributes, which indicates that the condition of sphericity was exactly met. A Bonferroni post-hoc test was applied.

3 Results

3.1 Original data analysis

Correlations were assessed for a pair of the first and second measurements for each response scale. The bipolar VAS had higher correlation strength than the unipolar 11-point scale using the Fisher's Z transformation ($P < 0.05$) for all subjective attributes in Table 1. In the unipolar 11-point scale, the reliabilities of loudness, noisiness, and annoyance were significantly higher than those for softness, quietness, and pleasantness. Loudness and noisiness showed higher correlation coefficients than any other subjective attributes. In the bipolar VAS, a pair of quietness and noisiness showed the highest reliability, and the pairs of softness/loudness and pleasantness/annoyance followed. The attributes associated with acoustic comfort, both for the unipolar 11-point scale, and for the bipolar VAS, were observed as the least reliable attributes.

The bipolar soft/loud pair correlates better with the unipolar loudness than the unipolar softness. The unipolar noisiness correlates better with the bipolar quiet/noisy pair than with unipolar quietness.

The bipolar acoustic uncomfortable/ comfortable pair also correlates better with unipolar acoustic comfort. The unipolar 11-point scales were, in general, correlated better with the right-end attributes of the bipolar VAS than the left-end attributes of the bipolar VAS. The bipolar quiet/noisy pair was observed as the most reliable measure in repeated measurements, and it showed the best reliability in response scale comparisons.

The bipolar soft/loud pair correlates better with the unipolar loudness than the unipolar softness. The unipolar noisiness correlates better with the bipolar quiet/noisy pair than with unipolar quietness. The bipolar acoustic uncomfortable/ comfortable pair also correlates better with unipolar acoustic comfort. The unipolar 11-point scales were, in general, correlated better with the right-end attributes of the bipolar VAS than the left-end attributes of the bipolar VAS as listed in Table 2. The bipolar quiet/noisy pair was observed as the most reliable measure in repeated measurements, and it showed the best reliability in response scale comparisons.

Table 3 lists the results of the Bonferroni post hoc test for each subjective attribute according to the sound sources. Mean values that do not share the letters in each attribute are significantly different. The unipolar scale could differentiate between quietness, noisiness, pleasantness, annoyance, acoustic discomfort, and acoustic comfort between the water sounds and the traffic noise, even at the same sound levels. However, the bipolar scale cannot differentiate any subjective attributes between water sounds and traffic noise at 42 dBA, except for the pair of acoustic uncomfortable/ comfortable. The unipolar quietness could distinguish between background noise and 42 dBA. The unipolar discomfort and the bipolar discomfort/comfort pair also could differentiate between background noise and 42 dBA sounds.

Table 1: Pearson's correlation coefficients between repeated measures ($P < 0.0005$) and Fisher's Z transformation ($P < 0.05$) results (coefficients that do not share a letter are significantly different, A > B > C > D)

N=250	Pearson's CC (P<.0005)	Fisher's Z transformation (P<.05)
Unipolar 11		
Soft	0.766	D
Loud	0.897	AB
Quiet	0.856	C
Noisy	0.905	AB
Pleasant	0.772	D
Annoying	0.845	C
Uncomfortable	0.735	D
Comfortable	0.715	D
Bipolar VAS		
Soft-Loud	0.875	B
Quiet-Noisy	0.913	A
Pleasant-Annoying	0.871	B
Uncomfortable-Comfortable	0.813	C

Table 2: Pearson's correlation coefficients between the unipolar 11-point scale and the bipolar VAS ($P < 0.0005$) and Fisher's Z transformation ($P < 0.05$) results (coefficients that do not share a letter are significantly different, A > B > C > D)

Bipolar VAS	Unipolar 11			
	Left-end		Right-end	
Soft-Loud	-0.772	C	0.868	B
Quiet-Noisy	-0.870	B	0.905	A
Pleasant-Annoying	-0.767	C	0.791	C
Uncomfortable-Comfortable	-0.702	D	0.756	C

Table 3: Results of Bonferroni pairwise comparisons according to sound sources (Mean values that do not share a letter are significantly different, A > B > C > D. $P < 0.05$)

	BN	W42	T42	W61	T61
Soft	A	A	A	B	B
Loud	B	B	B	A	A
Soft-Loud	C	C	C	**B**	**A**
Quiet	**A**	**B**	B	C	C
Noisy	C	C	C	**B**	**A**
Quiet-Noisy	C	C	C	**B**	**A**
Pleasant	A	**A**	**B**	C	D
Annoying	C	C	C	**B**	**A**
Pleasant-Annoying	C	C	C	**B**	**A**
Uncomfortable	C	C	C	**B**	**A**
Comfortable	**D**	**C**	**B**	AB	A
Uncomfortable-Comfortable	**D**	**C**	**B**	AB	A

3.2 Normalized data analysis

The original data of the bipolar VAS from -10.0 to 10.0 were normalized to unipolar 0.0-to-10.0 scales to perform ANOVA with repeated measurements on the two response scales with different numerical ranges.

Table 4 lists the significance levels and size of the effect of the repeated-measurement ANOVA results for normalized subjective responses. The effects of repetition were found only in acoustic discomfort. The effects of the response scales were found in softness, loudness, noisiness, pleasantness, annoyance, and acoustic comfort. The right-end attributes, loudness, noisiness, annoyance, and acoustic comfort showed higher values with the unipolar 11-point numerical scale than with the bipolar VAS. Softness and pleasantness among the left-end attributes had higher values with the bipolar VAS than with the unipolar 11-point scale. No effects of the response scales were found in quietness and acoustic discomfort. The effects of sound sources were found in all subjective attributes, as expected. The positive attributes, namely, quietness, pleasantness, and acoustic comfort can distinguish differences between background noise, water sounds of 42 dBA and traffic noise of 42 dBA. However, these positive attributes can not differentiate between the sounds of 61 dBA water sound and traffic noise. On the other hand, the negative attributes, namely, loudness, noisiness, annoyance, and acoustic discomfort can distinguish between the sounds of a 61 dBA water and traffic noise, but can not distinguish sounds of lower levels. Figure 4 shows normalized mean values with two different response scales according to sound sources.

3.3 Preference results

The 86% of participants voted for the bipolar VAS as shown in Figure 5. Only two options of choice were provided to participants. Non-response did not occur in this question.

Table 4: Results of significance level ($P < 0.05$) and effect size (η^2) of repeated-measurement ANOVA using normalized data (D: discomfort, C: comfort) uncomfortable

		Softness		Loudness		Quietness		Noisiness		Pleasant-ness		Annoyance		Discomfort		Comfort	
Within subjects																	
Repeat	P													0.027			
	η^2													0.015			
R1	Mean													4.454	A		
R2	Mean													3.894	B		
Between subjects																	
Scale	P	<.0005		<.0005				<.0005		0.008		0.010				0.004	
	η^2	0.029		0.071				0.129		.0270		0.036				0.030	
Unipolar 11	Mean	4.392	B	4.022	A	4.712	A	4.066	A	4.126	B	3.694	A	3.932	A	4.978	A
Bipolar VAS	Mean	5.339	A	3.133	B	4.551	A	2.811	B	4.726	A	2.786	B	3.370	A	4.156	B
Sound	P	<.0005		<.0005		<.0005		<.0005		<.0005		<.0005		<.0005		<.0005	
	η^2	0.117		.0.349		0.443		0.223		0.130		0.307		0.218		0.188	
BN	Mean	6.865	A	0.841	C	7.671	A	0.813	C	6.470	A	1.087	C	1.826	C	6.802	A
W42	Mean	6.620	A	1.301	C	6.749	B	1.370	C	6.053	A	1.398	C	1.901	C	6.271	A
T42	Mean	6.255	A	1.434	C	6.528	B	1.462	C	5.182	B	1.790	C	2.276	C	5.403	B
W61	Mean	2.675	B	6.848	B	1.445	C	6.176	B	2.693	C	5.304	B	5.316	B	2.619	C
T61	Mean	1.913	B	7.465	A	0.763	C	7.373	A	1.731	C	6.622	A	6.684	A	1.739	C

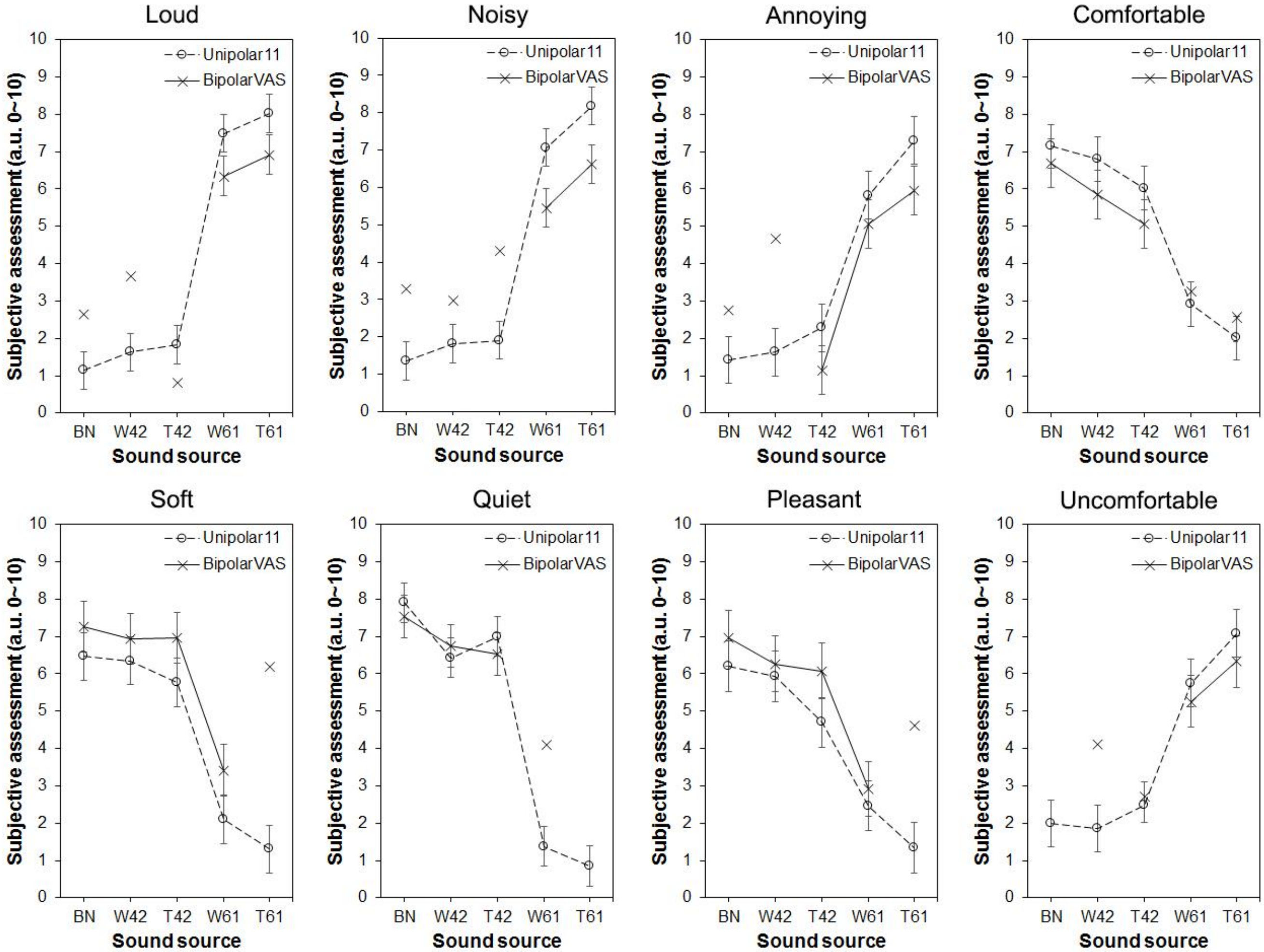

Figure 4: Normalized mean subjective judgment with a significant difference on the response scale (black: bipolar VAS, red: bipolar 7, blue: unipolar 11, purple: combined

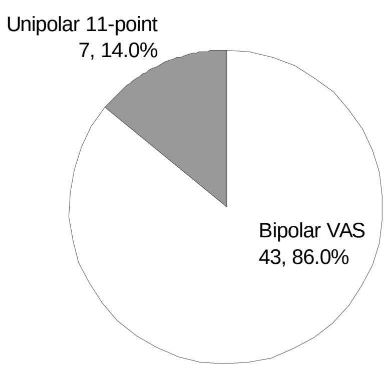

Figure 5: Paired comparison test results between a bipolar VAS and a unipolar 11-point scale

4 Discussion

4.1 Numeric scale vs. visual analogue scale

Reliability over repeated measurements of the two response scales was acceptable [43]; however, it was higher with the bipolar VAS than with the unipolar 11-point scale within each pair of subjective attributes. This is consistent with Rausch and Zehetleitner [28], although they could not have statistical significance in terms of reliability in comparison between the VAS and the 4-point scale as measures of a conscious experience of motion. However, Lewis and Erdinç [31] reported that the reliability of 7- and 11-point Likert-type scales and the VAS had no obvious advantage over each other in the context of user experience research. Clear evidence regarding the highest reliability has not yet been found.

4.2 Unipolar vs. bipolar

The scale sensitivity, the degree of differentiation by sounds, was higher with the unipolar 11-point scale than with the bipolar VAS. The bipolar VAS had two adjectives on each side of the scale. As listed in Table 2, the correlation coefficients between the unipolar 11-point scale and the bipolar VAS were higher than 0.7 on all subjective attributes, which meant a strong correlation [43]. However, the bipolar VAS of a pair of subjective attributes was highly correlated and yielded a similar precision in discriminating sound sources with the unipolar 11-point scale of the right-end attributes, rather than the left-end ones with statistical significance (Table 3). This means that the left-end subjective attributes of the bipolar VAS cannot be assessed as reliably as the right-end subjective attributes. The polarity of the response scale affected the sensitivity of subjective attributes. However, it is not clear whether the position of the right-end is simply preferable to the left-end, or the subjective attributes of loudness, noisiness, annoyance, and acoustic comfort are more impressive than the softness, quietness, pleasantness, and acoustic discomfort for the participants.

The impact of scale polarity on data quality has not yet been clearly investigated. Alwin [44] reported that unipolar scales are somewhat more reliable than bipolar scales.

However, in this study, the polarity of the scale was related to sensitivity, not reliability.

4.3 Participants preferences

The bipolar visual analogue scale was preferable to the unipolar 11-point scale among young adults in this study. User preferences on rating scales were studied mainly in medicine or psychology. The effects of socio-economic educational factors were significant on the user preferences of response scales according to studies in medicine [27]. It has been observed that VAS is not a priority of preferences for response scales in pain scale studies [27, 45, 46]. In psychology, Preston and Colman [47] reported respondents' preference on the response scales for 149 undergraduate students. For young adults, scales with 6, 7, and 10 response categories were the most preferable for ease of use, but the 101-point scale was the most favorable rating for adequate expression of feelings.

User preference may be a factor in choosing a rating scale, given the positive association between user performance and their subjectively expressed preferences [48].Understanding the socio-economic and educational status of respondents would be the basis for considering user preferences for response scales.

4.4 Semantic adjective attributes

The differences between the background noise, the water sound of 42 dBA, and the traffic noise of 42 dBA could be distinguished by quietness, pleasantness, and acoustic comfort, which are all positive attributes, except for softness. The differences between the water sound and traffic noise of 61 dBA showed up in all the negative attributes, without exception. The bipolar VAS could not differentiate the left-end adjectives. Furthermore, there were no adjective attributes in the unipolar 11-point scale that could distinguish all five different sound sources. Choosing the right semantic attributes is a prerequisite for defining response scales. Participants tended to focus on semantics. The response of the participants was not so sensitive to the response to sounds belonging to the semantic category, if the sounds that they heard were not in the semantic category. For example, as soon as the sounds were not heard loudly subjectively, their subtle differences of sounds were not evaluated in loudness assessment, and vice versa. Therefore, in the case of a broad range of sound levels to be assessed, semantic attributes should be chosen more carefully based on the purpose of the study.

Research information on semantic differentials in auditory perception is still scarce. More research is needed in this area.

4.5 Limitations

First, the test configurations were combined with numerical versus analogue and unipolar versus bipolar to reduce the number of comparisons. If a two-by-two matrix (2 response types x 2 polarities) was used for comparisons, more direct results could be obtained. Secondly, the test participants were limited to young, educated participants. However, the non-randomized sample was justified by the purpose of the study. Thirdly, the preference question was a simple paired comparison, therefore it was impossible to analyze the cause and the effect on the participant preference on the response scales.

5 Conclusion

The bipolar visual analogue scale and the unipolar 11-point numerical scale were compared to assess their performance and the preference of auditory perception among university students. Both response scales were acceptable for their reliability and sensitivity. However, the bipolar visual analogue scale was more reliable than the unipolar 11-point numerical scale in repeated measurements, and the unipolar 11-point numerical scale was more sensitive than the bipolar visual analogue scale to distinguish subtle differences between sound sources. Participants obviously preferred the bipolar visual analogue scale. The choice of semantic adjectives is a critical prerequisite for determining response scales for auditory perception. In summary, a unipolar visual analogue scale is proposed for assessing auditory perception for psychoacoustic research purposes for young educated adults.

Acknowledgments

This work was supported by the Basic Science Research Program of the National Research Foundation (NRF) [grant number 2018R1D1A1B07048157] funded by the Ministry of Education, Republic of Korea.

References

[1] DeCastellarnau A. A classification of response scale characteristics that affect data quality: A literature review. Quality & Quantity. 2018;52:1523-59.

[2] Fields J, De Jong R, Gjestland T, Flindell I, Job R, Kurra S, et al. Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. Journal of Sound and Vibration. 2001;242(4):641-79.

[3] Standardization IOf. ISO/TS 15666:2003. Acoustics -- Assessment of noise annoyance by means of social and socio-acoustic surveys. Geneva, Switzerland2003.

[4] Standardization IOf. ISO/TS 12913-1:2014. Acoustics -- Soundscape -- Part1: Definition and conceptual framework. Geneva, Switzerland2014.

[5] Standardization IOf. ISO/TS 12913-2:2014. Acoustics -- Soundscape -- Part2: Data collection and reporting requirements. Geneva, Switzerland2014.

[6] Mcallister A, Sundberg J, Hibi SR. Acoustic measurements and perceptual evaluation of hoarseness in children's voices Logopedics Phoniatrics Vocology. 1998;23(1):27-38

[7] Bradley MM, Lang PJ. Affective reactions to acoustic stimuli. Psychophysiology. 2000;37(2):204-15.

[8] Pellerin N, Candas V. Effects of steady-state noise and temperature conditions on environmental perception and acceptability. Indoor Air. 2004;14(2):129-36.

[9] Nagano K, Horikoshi T. New comfort index during combined conditions of moderate low ambient temperature and traffic noise. Energy and Buildings. 2005;37(3):287-94.

[10] Tiller DK, Wang LM, Musser A, Radik MJ. Combined effects of noise and temperature on human comfort and performance. ASHRAE Transactions. 2010;116:522-40.

[11] Munson B, Edwards J, Schellinger SK, Beckman ME, Meyer MK. Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana Clinical Linguistics and Phonetics. 2010;24(4-5):245-60.

[12] Kong E, Edwards J, editors. Individual Differences in Speech Perception: Evidence from Visual Analogue Scaling and Eye-Tracking. 17th International Congress of Phonetic Sciences; 2011 August 17-21, 2011; Hong Kong, China.

[13] Adamchic I, Langguth B, Hauptmann C, Tass PA. Psychometric Evaluation of Visual Analog Scale for the Assessment of Chronic Tinnitus. American Journal of Audiology. 2012;21(2):215-25.

[14] Hall DA, Irwin A, Edmondson-Jones M, Phillips S, Poxon JEW. An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. Applied Acoustics. 2013;74(2):248-54.

[15] Yang W, Moon HJ, Kim M-J. Combined effects of short-term noise exposure and hygrothermal conditions on indoor environmental perceptions. Indoor and Built Environment. 2018;27(8):1119-33.

[16] Yang W, Moon HJ. Cross-modal effects of noise and thermal conditions on indoor environmental perception and speech recognition. Applied Acoustics. 2018;141:1-8.

[17] Herr KA, Mobily PR. Comparison of selected pain assessment tools for use with the elderly. Applied Nursing Research. 1993;6(1):39-46.

[18] Bond A, Lader M. The use of analogue scales in rating subjective feelings. British Journal of Medical Psychology. 1974;47(3):211-8.

[19] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 5th edition ed: Oxford University Press, USA; 2015.

[20] Joyce CRB, Zutshi DW, Hrubes V, Mason RM. Comparison of fixed interval and visual analogue scales for rating chronic pain. European Journal of Clinical Pharmacology. 1975;8(6):415-20.

[21] Seymour RA. The use of pain scales in assessing the efficacy of analgesics in post-operative dental pain. European Journal of Clinical Pharmacology. 1982;23(5):441-4.

[22] Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. Journal of Chronic Diseases. 1987;40(12):1129-33.

[23] Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthritis and Cartilage. 2003;11(7):499-507.

[24] Skovlund E, Bretthauer M, Grotmol T, Larsen IK, Hoff G. Sensitivity of pain rating scales in an endoscopy trial. The Clinical Journal of Pain. 2005;21(4):292-6.

[25] Couper MP, Tourangeau R, Conrad FG, Singer E. Evaluating the Effectiveness of Visual Analog Scales:A Web Experiment. Social Science Computer Review. 2006;24(2):227-45.

[26] Davey HM, Barratt AL, Butow PN, Deeks JJ. A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. Journal of Clinical Epidemiology. 2007;60(4):356-60.

[27] Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review. Journal of Pain and Symptom Management. 2011;41(6):1073-93.

[28] Rausch M, Zehetleitner M. A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. Consciousness and Cognition. 2014;28:126-40.

[29] Funke F. A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. Social Science Computer Review. 2016;34(2):244-54.

[30] Voutilainen A, Pitkäaho T, Kvist T, Vehviläinen-Julkunen K. How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. Journal of Advanced Nursing. 2016;72(4):946-57.

[31] Lewis JR, Erdinç O. User experience rating scales with 7, 11, or 101 points: does it matter? Journal of Usability Studies. 2017;12(2):73-91.

[32] Funke F, Reips U-D. Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales. Field Methods. 2012;24(3):310-27.

[33] Schaik PV, Ling J. Design parameters of rating scales for web sites ACM Transactions on Computer-Human Interaction. 2007;14(1):4.

[34] Brink M, Schreckenberg D, Vienneau D, Cajochen C, Wunderli J-M, Probst-Hensch N, et al. Effects of Scale, Question Location, Order of Response Alternatives, and Season on Self-Reported Noise Annoyance Using ICBEN Scales: A Field Experiment. International Journal of Environmental Research and Public Health. 2016;13(11):1163.

[35] Nguyen T, Yano T, Morihara T, Yokoshima S, Makoto M, editors. Comparison of annoyance response measured with ICBEN 5-point verbal and 11-point numerical scales in Japanese and Vietnamese. 12th ICBEN Congress on Noise as a Public Health Problem; 2017 June 18-22; Zurich, Switzerland.

[36] Bjerre LC, Larsen TM, Sørensen AJ, Santurette S, Jeong C-H. On-site and laboratory evaluations of soundscape quality in recreational urban spaces. Noise & Health. 2017;19(89):183-92.

[37] Tristán-Hernández E, Garcıa IP, Navarro JML, Campos-Cantón I, Kolosovas-Machuca ES. Evaluation of Psychoacoustic Annoyance and Perception of Noise Annoyance Inside University

Facilities. International Journal of Acoustics and Vibration. 2018;23(1):3-8.

[38] Miedema HME, Vos H. Exposure-response relationships for transportation noise. Journal of Acoustical Society of America. 1998;104(6):3432-45.

[39] Kraus U, Breitner S, Hampel R, Wolf K, Cyrys J, Geruschkat U, et al. Individual daytime noise exposure in different microenvironments. Environmental Research. 2015;140:479-87.

[40] Pitkanen P. Strem small water flow 2018 [cited 2018 September 17]. Available from: https://www.soundsnap.com.

[41] Budescu DV, Appelbaum MI. Variance Stabilizing Transformations and the Power of the F Test. Journal of Educational Statistics. 1981;6(1):55-74.

[42] Box GEP. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effects of Inequality of Variance and of Correlation Between Errors in the Two-Way Classification. The Annals of Mathematical Statistics. 1954;25(3):484-98.

[43] Evans JD. Straightforward statistics for the behavioral sciences: Thomson Brooks/Cole Publishing Co; 1996.

[44] Alwin DF. Margins of error: A study of reliability in survey measurement: John Wiley & Sons; 2007.

[45] Akad K, Solmaz D, Sari I, Onen F, Akkoc N, Akar S. Performance of response scales of activity and functional measures of ankylosing spondylitis: numerical rating scale versus visual analog scale. Rheumatology International. 2013;33(10):2617-23.

[46] Yazici Sayin Y, Akyolcu N. Comparison of Pain Scale Preferences and Pain Intensity According to Pain Scales among Turkish Patients: A Descriptive Study. Pain Management Nursing. 2014;15(1):156-64.

[47] Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica. 2000;104(1):1-15.

[48] Nielsen J, Levy J. Measuring usability: preference vs. performance Communication of the ACM. 1994;37(4):66-75