

Article

Usability of Visual Analogue Scales in Assessing Human Perception of Sound with University Students Using a Web-Based Tablet Interface

Wonyoung Yang ¹  and Jin Yong Jeon ^{2,*} ¹ Division of Architecture, Gwangju University, Gwangju 61743, Korea; wyyang@gwangju.ac.kr² Department of Architectural Engineering, Hanyang University, Seoul 04763, Korea

* Correspondence: jyjeon@hanyang.ac.kr

Abstract: Response scales in auditory perception assessment are critical for capturing the true responses of listeners. Despite its impact on data, response scales have received the least attention in auditory perception assessment. In this study, the usability of visual analogue scales for auditory perception assessment was investigated. Five response scales (a unipolar visual analogue scale—negated to regular, a unipolar visual analogue scale—regular to negated, a bipolar visual analogue scale—positive to negative, a bipolar visual analogue scale—negative to positive, and a unipolar 11-point scale (ISO/TS 15666:2021)) for auditory perception assessment are presented. Music and traffic noise were presented to 60 university students at two different levels, i.e., 45 and 65 dBA, respectively. A web-based experimental design was implemented, and tablet pads were provided to the respondents to record their responses. The unipolar 11-point scale required the longest response time, followed by the two unipolar visual analogue scales and two bipolar visual analogue scales with statistical significance. All response scales used in this study achieved statistical reliability and sensitivity for the auditory perception assessment. Among the five response scales, the bipolar visual analogue scale (negative to positive) ranked first in reliability over repeated measures, exhibited sensitivity in differentiating sound sources, and was preferred by the respondents under the conditions of the present study. None of the respondents preferred the unipolar 11-point scale. The visual analogue scale was favoured over the traditional unipolar 11-point scale by young educated adults in a mobile-based testing environment. Moreover, the bipolar visual analogue scale demonstrated the highest reliability and sensitivity, and it was preferred the most by the respondents. The semantic labelling direction from negated to regular, or from negative to positive, is preferred over its opposite counterpart. Further research is necessary to investigate the use of response scales for the general public including children and the elderly, as well as that of semantic adjectives and their counterparts for auditory perception assessment.

Keywords: response scales; visual analogue scales; unipolar scale; bipolar scale; auditory perception



Citation: Yang, W.; Jeon, J.Y. Usability of Visual Analogue Scales in Assessing Human Perception of Sound with University Students Using a Web-Based Tablet Interface. *Sustainability* **2021**, *13*, 9207. <https://doi.org/10.3390/su13169207>

Received: 25 June 2021

Accepted: 12 August 2021

Published: 17 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

Auditory perceptions were measured using a questionnaire survey. The response scales used in the questionnaire survey are critical for obtaining true responses and minimising the impact of the survey design [1]. However, the choice of response scales has not been prioritised, even in broad psychological assessments [2].

In auditory perception, the unipolar five-point verbal and unipolar positive 11-point numerical scales proposed by the International Commission on Biological Effects of Noise [3], adopted as the international standard ISO/TS 15666:2021 [4], are widely used to assess noise. Owing to the negative aspects of noise, the two abovementioned unipolar scales span from the neutral to the extremely negative end. The positive aspects of sound assessment can be measured through a soundscape, which is an acoustic environment

perceived or experienced and/or understood by one or more persons [5]. The unipolar positive five-point verbal response scale and the unipolar negative and positive five-point verbal response scale based on questions were adopted by the ISO/TS 12913-2:2018 [6] for soundscape data collection and reporting.

The visual analogue scale (VAS) is another option for auditory perception experiments. The VAS has been widely used for subjective rating in many disciplines owing to its high sensitivity in rating subjective feelings [7]. In particular, numerous studies pertaining to the use of the VAS in clinical settings have been performed [8–29]. Compared with multipoint discretised scales, the VAS offers a few benefits. The VAS appears to be more satisfactory than the multipoint scale for patient self-rating of chronic pain intensity [8]. It is more sensitive than multipoint verbal or numerical scales, and can distinguish slight changes effectively [9,15,18]. The VAS is a valuable instrument for comparing scores in different groups of subjects, comparing treatments in individuals, and observing the scores of individual subjects over a duration [14]. Furthermore, the VAS avoids the ceiling effect better than the multipoint scale, and the time required to complete the questionnaire was 28% shorter than that of the multipoint scale [22].

Despite the apparent advantages of VAS, it has not been widely used in survey research. Two key features of less popularity of the VAS were analysed by Couper et al. [30]: first, it required self-administration and second, it is visual; that is, it cannot be administered using an aural medium, such as the telephone. These characteristics, along with the extra effort needed to measure and record the answers provided, may limit the use of VAS in surveys.

However, recent developments in graphical user interfaces have raised the possibility of greater use of VAS in web-based survey applications [30–33]. Furthermore, a mobile device or multi-device on a web-based survey has been addressed [34,35]. Recently, head-mounted displays (HMDs) have been used for perception evaluation in various disciplines, such as soundscape [36] and noise annoyance [37]. However, the optimal response scales for capturing true responses utilising newly adopted devices in auditory perception studies are in the early stages of research.

1.2. Literature Review

A few studies [38–40] have investigated the effects of response scales in noise surveys based on ISO/TS 15666:2021 [4]. Brink et al. [38] discovered some disagreements between a five-point verbal scale and an 11-point numerical scale with 2386 Swiss residents. Standardised average annoyance scores were higher using the 11-point numerical scale than the 5-point verbal scale with statistical significance. The percentage of highly annoyed respondents was significantly higher based on the 5-point verbal scale than the 11-point numerical scale. Nguyen et al. [39] investigated the correspondence between 5-point verbal and 11-point numerical scales from 15 social surveys conducted in Japan ($N = 3652$) and Vietnam ($N = 7149$). The exposure-response relationships obtained by the logistic regression function with the high annoyance defined by the top three categories of the 11-point scale located lower than the logistic regression relationships with the high annoyance defined by the two categories of the 5-point scale. However, other studies with relatively small number of participants ($N = 22$ [41], $N = 33$ [40]) found no statistical differences between the two standardised response scales for noise assessment. Tristan-Hernandez et al. [40] evaluated psychoacoustic annoyance and perception of noise annoyance inside university facilities using both a 5-point verbal scale and an 11-point numerical scale [4], and found no significant differences between the answers of both scales. Bjerre et al. [41] also used both a 5-point verbal scale and an 11-point numerical scale [4] for their on-site and laboratory evaluations of soundscape quality in recreational urban spaces, and no significant differences were resulted.

The visual analogue scale (VAS) is not widely used in auditory perception experiments. Previous studies using VASs for auditory perception primarily involved thermal comfort [42–47]. For thermal comfort assessments, a bipolar seven-point scale was stan-

standardised according to ISO 10551:2019 [48]. The choice of visual analogue scale could be a mediated choice, since the standardised response scales for subjective surveys in each research field were different. The effects of the VAS were not the main topic of the previous studies. Recently, Yang and Jeon [49] compared the performance and preference of a bipolar VAS and standardised unipolar 11-point numerical scale for auditory perception. Both response scales were acceptable for reliability and sensitivity; however, subtle differences were observed. The bipolar visual analogue scale was more reliable than the unipolar 11-point numerical scale in repeated measures, whereas the unipolar 11-point numerical scale was more sensitive than the bipolar visual analogue scale in distinguishing differences between sound sources. The respondents preferred the bipolar VAS. Yang et al. [50] expanded the response scale study to assess the combined environmental perception. They compared four different response scales based on ISO/TS 15666:2021 [4] and ISO 10551:2019 [48]. The degree of relative differentiation based on indoor physical factors did not differ significantly across the four response scales, including the VAS score. Respondents subjectively preferred the bipolar VAS. Despite the two previous studies pertaining to the response scale in auditory perception, the effects of the response scale on aural stimuli have yet to be fully investigated.

Chyung et al. [51] reported on the effects of delivery media and its transition in a survey. Traditionally, surveys have been administered on paper and ruler-type devices. However, owing to the popularity of web-based survey systems and respondents' increased access to the web via desktop and mobile devices, practitioners and researchers now administer their surveys frequently via the web. The administrative drawbacks of using a VAS can be mediated via the web. The selection of the response scale should be reconsidered based on the delivery media shift from analogue to digital.

1.3. Research Objectives

The objective of the present study was to investigate the usability of VASs for auditory perception assessments of university students exposed to mobile applications in a web-based testing environment. Four unipolar VASs were compared with the unipolar 11-point numerical scale corresponding to the ISO/TS 15666:2021 [4] for auditory perception experiments. The effects of labelling direction on the VAS were investigated using negated and reversed end labels.

2. Methods

The present study was designed as an experimental study with repeated measures in a laboratory. All participants responded in all conditions of the experimental design. No separate participant groups were designed for comparisons. This study was approved by the Institutional Review Board of Hanyang University.

2.1. Participants

Sixty university students (30 men and 30 women) participated in the response scale comparison testing. No hearing-impaired participants were examined during the interviews. All participants provided written informed consent before the start of the study and received financial support for their participation. The mean age of the participants was 21.1 (SD 2.3) years. All participants were mobile phone users who reported their proficiency in using mobile devices.

2.2. Experimental Conditions

The experiment was conducted in a small laboratory dedicated to indoor environmental experiments. The indoor environment was maintained at an air temperature of 24 °C and humidity level of 45%. The mean illuminance level along the testing desk surface was 860 lx (Konica Minolta T-10A, Tokyo, Japan). The background noise level was measured to be 41 dBA (Rion Rionote, Tokyo, Japan) while the air handling unit was turned on. The

reverberation time in the laboratory was measured to be 0.93 s at mid-frequencies (mean of 500 Hz and 1 kHz).

Four loudspeakers (Genelec 8020C, Iisalmi, Finland) were placed at each corner of the room facing the corner to present sounds with no directional information in order to normalise the sound direction for all participants in the testing room. Four different sound sources (music sound and traffic noise of 45 and 65 dBA, respectively) were presented through the loudspeaker system. The differences in sound level across the room were measured to be ± 0.4 dBA. The first movement of Vivaldi's *The Four Seasons*, "Spring", performed by Amsterdam Sinfonietta in Concertgebouw in 2014 was used as the music sound, i.e., a positive sound source. Meanwhile, traffic noise, representing a negative sound source, was recorded on a street near the university. Figure 1 shows frequency spectra of sound sources.

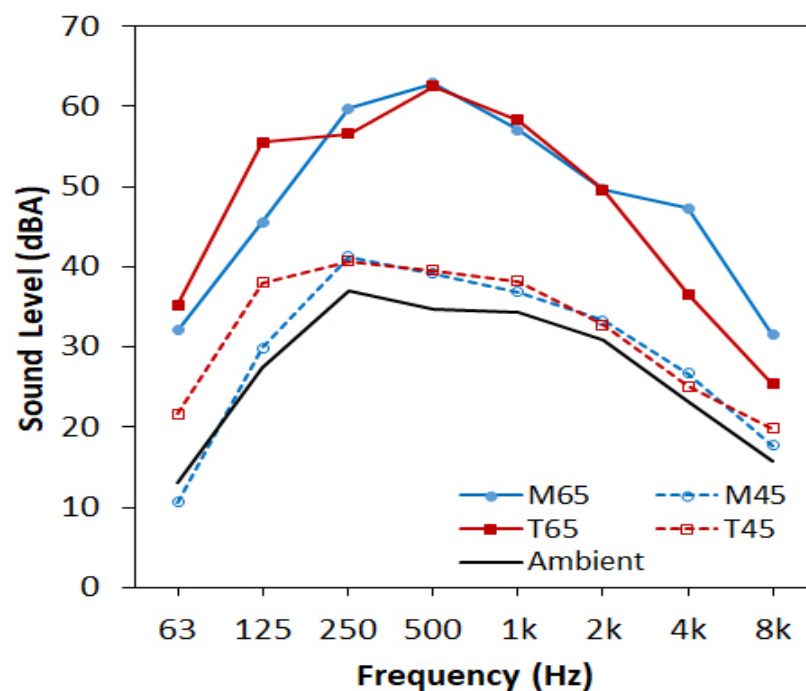


Figure 1. Frequency spectra of sound sources (blue solid line: music 65 dBA; blue dashed line: music 45 dBA; red solid line: traffic 65 dBA; red dashed line: traffic 45 dBA; black solid line: ambient).

2.3. Response Scales and Semantic Adjectives

Four different VASs and a unipolar 11-point numerical scale were used in this study, as shown in Figure 2. The unipolar 11-point numerical scale with endpoint and midpoint labels was adopted based on the ISO/TS 15666:2021 [4], which was developed for socio-acoustic noise annoyance surveys. The four VASs comprised a pair of unipolar VASs and a pair of bipolar analogue scales. Each pair had reversed left and right verbal end labels. The subjective attributes asked were identical in all five response scales; however, the number of questions for the unipolar scales doubled on the bipolar scales.

The semantic attributes of the questionnaire were four pairs of adjectives: soft vs. loud, quiet vs. noisy, pleasant vs. annoying, and uncomfortable vs. comfortable [52]. The loud, noisy, and annoying were widely used for subjective noise evaluation [47,53–57]. Comfortable is an attribute used in many human perception evaluations other than auditory perception [58–61], and was selected as a higher-level semantic attribute inferred by the results of other semantic attributes. The reversed items were then selected for counterbalanced rating scales for validity [62].

For the two unipolar VASs, each semantic attribute was positioned on the left and right ends, and its minimum level was placed using "not" at the other end, which are

called negated items [63]. For the two bipolar VASs, two adjective antonyms, so-called reversed items [63] in auditory perception were used on both sides of the line.

The VAS comprised a plain, horizontal 100-mm-long line mm and primarily verbal end labels. In this study, a web-based tablet interface was developed, and the length of the horizontal line was approximately 120 mm on the tablet screen. Respondents provided rating by tapping a mark on the line. A slider was placed at the left end of the default setting as an indicator of the rating mark.

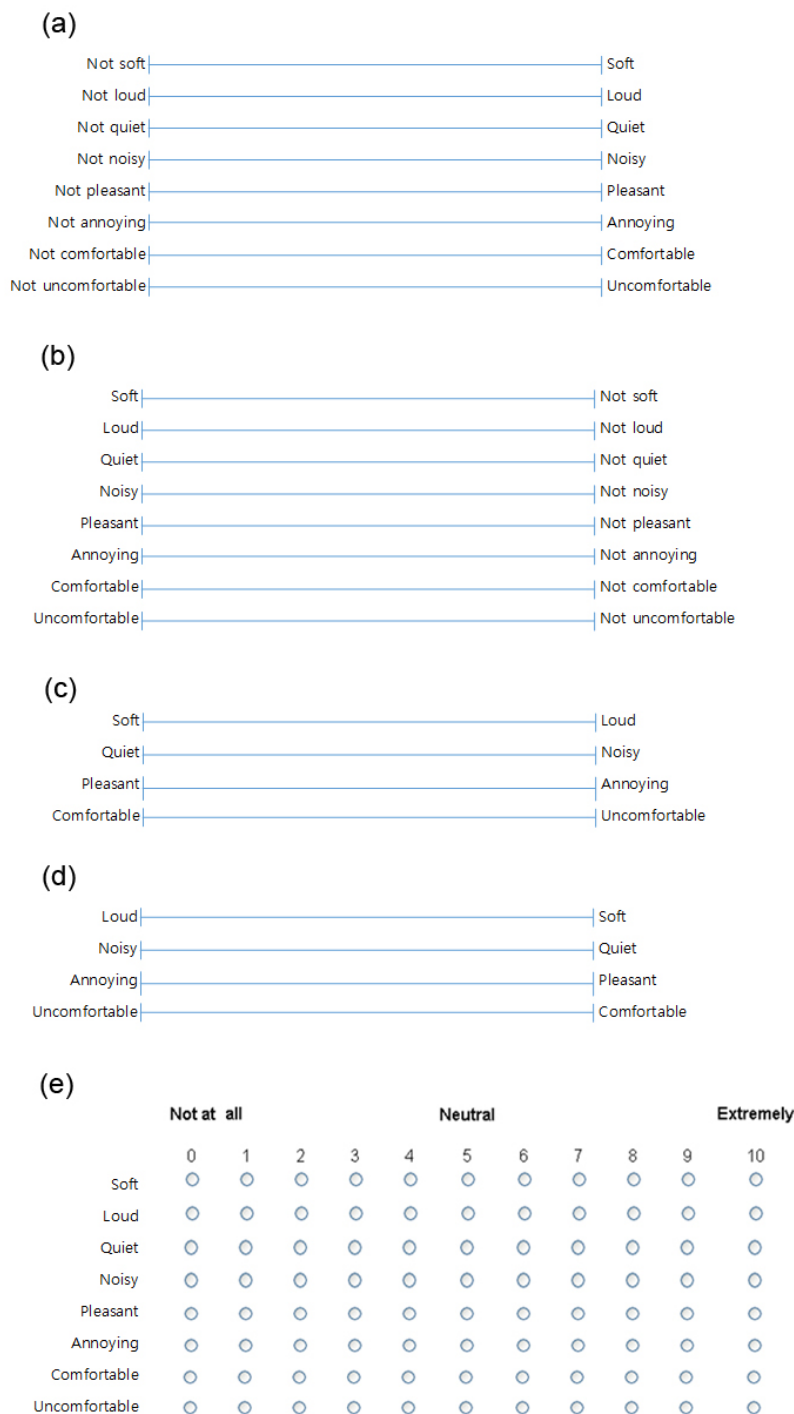


Figure 2. Response scales. (a) Scale 1: unipolar VAS negated to regular; (b) scale 2: unipolar VAS regular to negated; (c) scale 3: bipolar VAS positive to negative; (d) scale 4: bipolar VAS negative to positive; (e) scale 5: unipolar 11-point numerical scale.

2.4. Experimental Design and Procedure

In each 60-min-long testing session, a 15-min adaptation period was implemented at the beginning of the session. A total of 25 cases were created by combining five sound sources and five response scales. Ambient noise perception was asked at the beginning of the session from the first to the fifth to familiarise the participants with the response scales before the music sound and traffic noise were tested, as shown in Figure 3. Subsequently, the response scale was randomised, and the sound source was presented randomly within the same response scale. When 25 cases were completed, the same 25 cases were tested in duplicates.

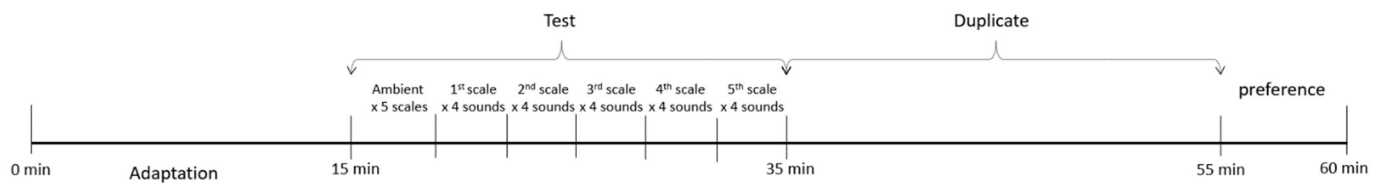


Figure 3. Experimental procedure for each session.

A maximum of eight respondents participated simultaneously during the testing session. A quick demonstration of the method to use the tablet was provided to the respondents. The respondents were required to click on instead of dragging the slider to avoid potential technical problems from dragging using their fingers. Most respondents were familiar with the use of touchscreen devices.

Each sound source was prepared for 50 s. However, the response time was not fixed. Even if 50 s of a sound source playback had or had not been completed, when all respondents submitted their responses in the test group, the next sound source started. The response data were automatically saved on a server.

Repeated measures analysis of variance (ANOVA) was employed with two independent variables, i.e., response scale and sound source, for eight subjective attributes to confirm the effects of the repeated measures and response scales simultaneously. ANOVA is an effective statistical test; however, it cannot guarantee normality for subjective ratings [64,65]. In this study, the response time was analysed using the ANOVA as well. The response time for each case was calculated by subtracting the former submission time from the latter submission time such that the first submission of each respondent had no response duration value in this study. Fisher's Z-transformation was applied to compare the correlation coefficients of response scales.

In this study, a numerical value from 0.00 to 10.00 was assigned to the responses from the VASs for statistical analysis. Furthermore, original responses from the numerical scale were used. The responses from the bipolar VASs (−10.00 to +10.00) were converted to unipolar 0.00 to 10.00 scales to perform an ANOVA on the five response scales for all eight subjective attributes. In the conversion rule, if the response is negative, then the absolute value of the response is assigned to the left-end subjective attribute, and the value of 10 minus the absolute value of the response is automatically assigned to the right-end subjective attribute. If the response is positive, then the response value is assigned to the right-end subjective attribute, and the left-end subjective attribute automatically assumes a value of 10 minus the response value.

3. Results

3.1. Response Times

Response time was dependent on the order of the questionnaire. Table 1 lists the mean response times based on questionnaire order. The response times from numbers 2 to 6 were significantly longer than that of number 25, regardless of the test combinations. To achieve an order-free dataset, numbers 2 to 7 and 25 were excluded, as listed in Table 2. The response scale significantly affected response time. The numerical 11-point response

scale indicated the longest response time. The bipolar VASs required less time than the others, as expected, because the number of responding scales was half that of the unipolar scales. Repeated measures significantly affected response time. The duplicates were shorter than the first response times. Moreover, the sound source affected the response time. The response time for 45 dBA traffic noise was longer than that for other sound sources.

Table 1. Response time based on question number.

Order.	N	Mean (s)	Grouping
2nd	120	62.16	A
6th	120	58.57	AB
5th	120	52.47	BC
3rd	120	50.08	CD
4th	120	43.85	DE
7th	120	41.84	EF
17th	120	40.9	EF
14th	120	40.88	EF
19th	120	40.64	EF
15th	120	40.23	EFG
18th	120	39.31	EFG
22nd	120	39.24	EFG
12th	120	39.04	EFG
16th	120	37.77	EFG
8th	120	37.5	EFG
20th	120	36.82	EFG
10th	120	36.67	EFG
21st	120	36.58	EFG
11th	120	36.41	EFG
9th	120	36.083	FG
23rd	120	35.875	FG
13th	120	35.55	FG
24th	120	35.48	FG
25th	120	32.892	G
F		25.33	
<i>p</i>		<0.0005	

Table 2. ANOVA results of response time for order-free data (from 8th to 24th).

		N	Mean (s)	Grouping	
Repeat	1st	1080	40.99	A	F = 151.31 <i>p</i> < 0.0005
	2nd	1080	34.92	B	
Scale	scale 1	432	41.39	B	F = 72.26 <i>p</i> < 0.0005
	scale 2	432	39.96	B	
	scale 3	420	33.15	C	
	scale 4	423	30.98	C	
	scale 5	453	44.30	A	
Sound sources	T45dB	543	40.93	A	F = 10.46 <i>p</i> < 0.0005
	T65dB	519	37.59	B	
	M45dB	567	37.06	B	
	M65dB	531	36.25	B	

3.2. Data Normalisation

The numerical 11-point scale (scale 5) and unipolar VASs (scales 1 and 2) can be compared directly from the responses; however, the bipolar VASs had different increments on the same line. Although the values allotted for the bipolar VASs ranged from -10.00 to $+10.00$, the actual length of the response scale line was the same as that of the unipolar VASs.

Therefore, the increment was modified using Formula (1), which can result in approximately the same position of the unipolar scales on the line, as shown in Figure 4. The modified values, scale 3' and scale 4', were used for repeated measures ANOVA analysis.

$$\text{scale } N' = \left\lfloor \frac{1}{2} \text{scale } N + 5 \right\rfloor$$

$$-10.00 < \text{scale } N < +10.00 \quad (1)$$

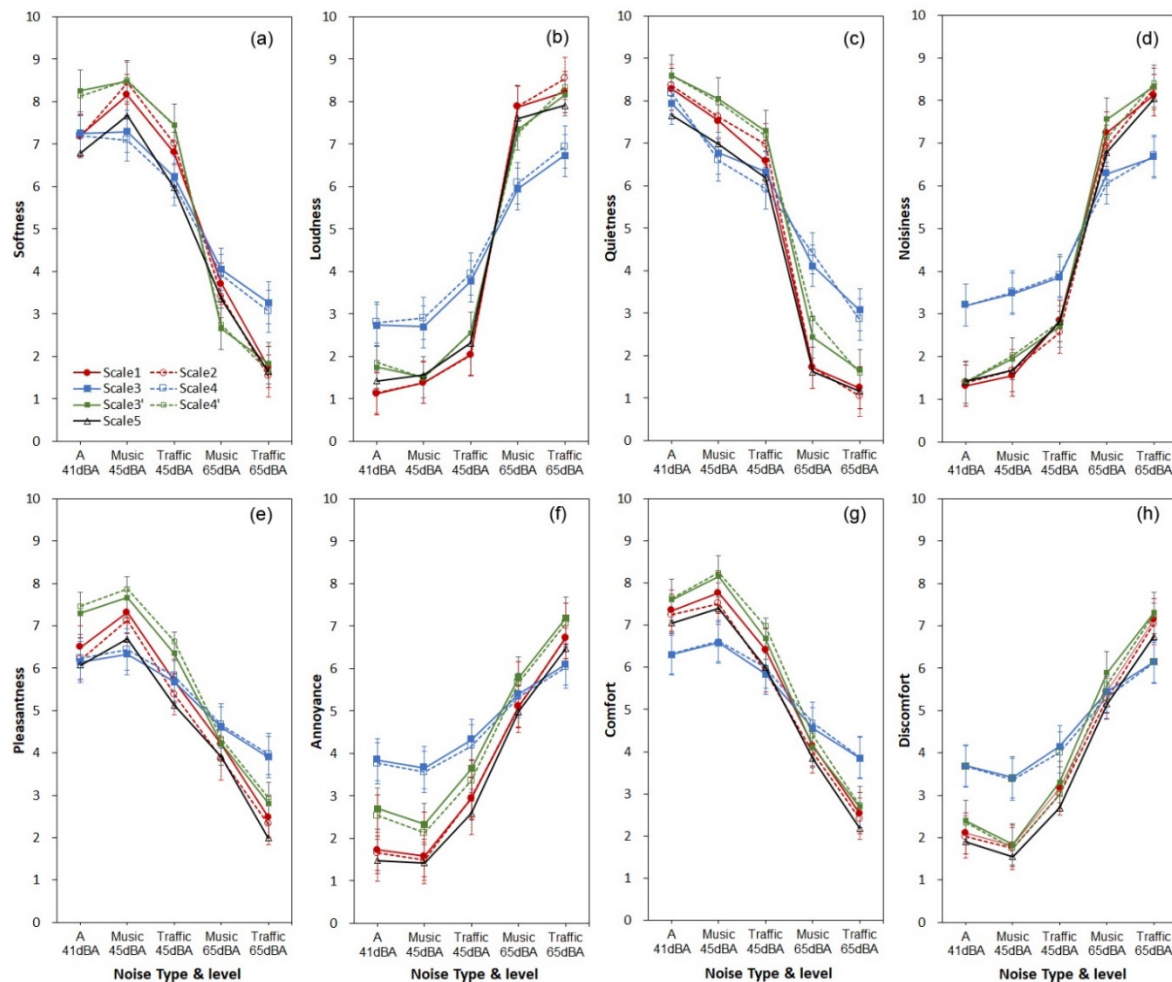


Figure 4. Mean subjective judgment based on sound source. (a) Softness; (b) loudness; (c) quietness; (d) noisiness; (e) pleasantness; (f) annoyance; (g) comfort; (h) discomfort; (red solid line: scale 1—unipolar VAS negated to regular; red dashed line: scale 2—unipolar VAS regular to negated; blue solid line: scale 3—bipolar VAS positive to negative; blue dashed line: scale 4—bipolar VAS negative to positive; green solid line: scale 3'—normalised bipolar VAS positive to negative; green dashed line: scale 4'—normalised bipolar VAS negative to positive; black solid line: scale 5—unipolar 11-point numerical scale).

3.3. Repeated Measures ANOVA

Table 3 lists the *p*-values and effect size (μ) of the repeated measures ANOVA for all subjective attributes. As expected, the sound source was a dominant factor, according to the effect size, whereas the response scale affected the subjective perception with a relatively small effect size.

3.4. Correlation Coefficients and Fisher's Z-Transformation

In this study, Pearson's correlation coefficients between repeated measures were analysed using Fisher's Z-transformation ($p < 0.05$), as listed in Table 5. For softness, the most reliable scale of response was scale 4, followed by scale 3. No statistical difference in reliability was discovered between the unipolar VAS and numerical 11-point scales. For loudness, scales 2 and 1 were more reliable than the other scales. For quietness, scale 5 was rated the worst in terms of reliability; however, for noisiness, scale 5 was rated the most reliable. The scale 4 was the most reliable scale for pleasantness. For annoyance, scale 5 was more reliable than the others. For comfort and discomfort, scale 4 was more reliable than scales 1 and 2. In summary, the bipolar VASs were more reliable than other scales, and scale 4, which is a bipolar VAS from negative to positive semantic attributes, was ranked the most reliable response scale.

Table 5. 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 within a row; $A > B > C$).

	Scale 1		Scale 2		Scale 3		Scale 4		Scale 5	
Softness	0.739	C	0.746	C	0.8	B	0.822	A	0.71	C
Loudness	0.843	B	0.877	A	0.8	C	0.822	BC	0.831	BC
Quietness	0.753	B	0.827	A	0.821	AB	0.802	B	0.754	C
Noisiness	0.782	B	0.757	C	0.821	A	0.802	AB	0.824	A
Pleasantness	0.596	C	0.593	C	0.666	B	0.707	A	0.648	B
Annoyance	0.669	B	0.683	B	0.666	B	0.707	B	0.755	A
Comfort	0.653	B	0.673	B	0.701	AB	0.733	A	0.701	AB
Discomfort	0.632	C	0.671	BC	0.701	AB	0.733	A	0.693	AB

Table 6 lists Pearson's correlation coefficients between paired scales and their Fisher's Z-transformation results for testing the end point label order. The end point label order was more reliable at the VASs than the numerical 11-point scale. Among the VASs, the bipolar VASs were more reliable than the unipolar VASs for end point label ordering in most subjective attributes.

Table 6. Pearson's correlation coefficients between paired scales ($p < 0.0005$) and Fisher's Z-transformation ($p < 0.05$) results (coefficients that do not share a letter are significantly different within a row; $A > B > C$).

	Scale 1 vs. Scale 2		Scale 3 vs. Scale 4		Scale 5 vs. Scale 1	
Softness	0.850	B	0.882	A	0.796	C
Loudness	0.918	A	0.882	B	0.878	B
Quietness	0.856	A	0.875	A	0.800	B
Noisiness	0.869	A	0.875	A	0.866	A
Pleasantness	0.838	A	0.826	A	0.797	B
Annoyance	0.818	A	0.826	A	0.782	B
Comfort	0.836	A	0.857	A	0.835	A
Discomfort	0.799	B	0.857	A	0.796	B

3.5. Respondents' Preference

In general, 37.5% of the respondents preferred the bipolar VAS from negative to positive semantic attributes, as shown in Figure 5. To determine the preferred scale, no responses were collected on a numerical 11-point scale. For the first response, scales 3 and 4 appeared similar, but for the second response, 43.3% of the respondents preferred scale 4, i.e., the bipolar VAS from negative to positive semantic attributes. The bipolar VAS from negative to positive was preferred twice compared with the bipolar VAS from positive to negative.

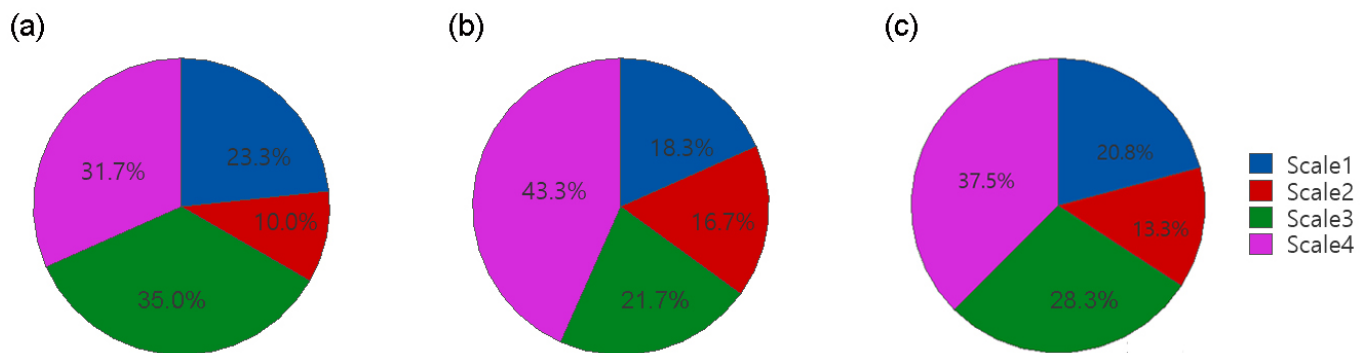


Figure 5. Respondents' preference, (a) first response; (b) second response; (c) mean of first and second responses.

The mean response times of the preferred scale by each respondent were analysed, as shown in Figure 6. The response time appeared to have contributed to the selection of the preferred response scale. To confirm the relationship between response time and response scale preference, the Pearson correlation coefficients between them were calculated, as listed in Table 7. For the first response, the correlation between response scale preference and response time was significant. For the second response, although the correlation strength was not as high as that of the first response, a negative correlation was observed.

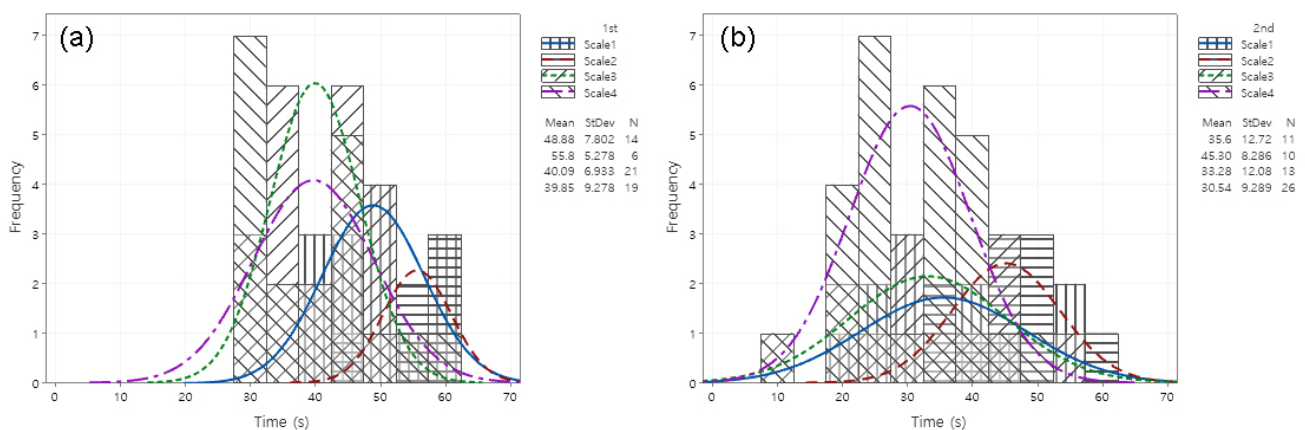


Figure 6. Response time for response scale preferred by respondents: (a) first response; (b) second response.

Table 7. Pearson's correlation coefficients between response time and percentage of preference ($p < 0.05$).

Response	1st	2nd
Pearson's Correlation Coefficient	−0.907	−0.796

4. Discussion

4.1. Reliability over Repeated Measures

In general, the most reliable response scale was the bipolar VAS from negative to positive, followed by that from positive to negative, as listed in Table 5. The unipolar scales, including numerical and analogue scales, did not demonstrate high reliability compared with the bipolar scales, which is consistent with previous studies [49,50]. Yang and Jeon [49] reported that the bipolar visual analogue scale was more reliable than the unipolar 11-point numerical scale for auditory perception. Furthermore, Yang et al. [50] discovered that the bipolar VAS appeared reliable over repeated measures for all subjective attributes, including auditory perception.

It is noteworthy that the reliability over repeated measures was independent of the type of response scale but dependent on the polarity of the response scale. For unipolar scales, the unipolar 11-point scale was more reliable than the unipolar VAS in this study. To the best of our knowledge, the present study is the first to assess the effects of the polarity of response scales.

4.2. Sensitivity for Degree of Differentiation in Auditory Perception

The sensitivity of the response scale was different for the subjective attributes, as listed in Tables 3 and 4. In general, the bipolar VAS from negative to positive indicated the highest sensitivity for all subjective attributes. The unipolar 11-point scale is applicable for loudness, noisiness, and annoyance perception with traffic noise, which shows that the unipolar 11-point scale fulfils the purpose of ISO/TS 15666:2021 [4]; however, its sensitivity in differentiating quietness, pleasantness, comfort, and discomfort was insufficient compared with the other scales. The bipolar VAS from negative to positive could only distinguish loudness and noisiness between music and traffic noise at 65 dBA (Table 4).

In previous studies [49,50] where water sound and traffic noise at 42 and 61 dBA were used, respectively, no statistical difference was exhibited in response scale sensitivity. In this study, music sounds were presented to the participants instead of water sounds. The music sound might be a more explicit positive sound than the water sound, thereby yielding differences in the results.

4.3. Preference by Young Adults as Mobile Users

Two-thirds of the participants preferred bipolar VASs, which is consistent with previous studies [49,50]. None of the respondents selected the unipolar 11-point scale, which was the only numerical scale in the present study. The respondents preferred semantic directivity on the response scales, such as from negated to regular and from negative to positive. The polarity of the response scale (unipolar vs. bipolar) imposed more significant effects on respondents' preferences than the semantic direction on the response scale. The shorter response time of the bipolar scales might have contributed to the significant effects of the polarity of the response scale on the respondents' preferences.

The participants were familiar with the mobile environment. Unfortunately, comparisons between a traditional paper VAS and an electronic VAS were beyond this research scope; however, their preference might be different from that of people who are not familiar with mobile devices.

Socioeconomic educational (SES) factors significantly affect the user preference of response scales in medicine [17]. For the pain scale, the VAS was not a preference as a response scale [17,19,20], since the respondents' socioeconomic educational background was much wider than that in the present study. The mean age of the respondents in this study was 21 years, and all were university students. The respondents' preferences affected the selection of the response scale, owing to the positive association between the respondents' performance and their preferences [66]. Therefore, the socioeconomic educational factors of the respondents should be considered when the response scale was selected for the survey.

4.4. Numerical Scale vs. VAS: Types of Response Scale

To the best of our knowledge, this is the first study to directly compare VASs and numerical scales to assess auditory perception. Only unipolar scales were compared in this study because the experimental design was based on ISO/TS 15666:2021 [4].

The comparison results between a numerically discretised scale and a VAS appeared complicated; hence, a simple conclusion could not be obtained. In the overall analysis, the numerical 11-point scale indicated lower values for softness and pleasantness with statistical significance (i.e., the positive aspects of the sound). The reliability of the numerical 11-point scale for noisiness and annoyance was stronger than that of the unipolar VAS from

negated to regular. None of the respondents preferred a numerical 11-point scale over the VAS.

If the positive attributes are not intentionally assessed in the questionnaire, then the two response scales can be substituted for each other in auditory perception. For comparisons with previous noise assessment studies, the use of standardised response scales may be practical for noise assessment. However, considering the association between respondents' preferences and their performance [66], the VAS may be an alternative to numerical scales for the auditory perception of young adults.

4.5. Unipolar vs. Bipolar: Polarity of Response Scale

In the present study, comparisons between unipolar and bipolar scales were analysed using VAS. In auditory perception, unipolar response scales have been used as international standards, because the sensation and perception of the sounds are considered unipolar. However, bipolar response scales have been discovered to be the most reliable, sensitive, and preferred for auditory perception assessment. Consequently, the preference and performance associations were validated using bipolar VASs.

The normalisation formula was necessary to apply ANOVA on the unipolar and bipolar scales. The normalisation formula developed in this study showed that the respondents selected the visually similar position on the scale line instead of the linearly proportional position between the unipolar and bipolar scales, regardless of whether the response scale was unipolar or bipolar, even though the actual increment of the scale line of the bipolar scale was half that of the unipolar scale for each semantic attribute. For example, once the music sound of 45 dBA was judged as soft, the respondents' selections for softness on the VAS were similar regardless of the polarity of the scale. Therefore, selecting the loudness of the sound was less valuable than selecting the softness of the sound because it was already judged as soft, and vice versa. This might explain the respondents' strong preference for bipolar scales.

In the present study, equivalent sound levels at 45 and 65 dBA, which can be classified as likes and dislikes, were used. Therefore, an increased number of questions with unipolar scales may not be preferred because of the explicit sound level differences. The other inference on bipolar preference is that semantic attributes can be clearly perceived when its semantic counterpart is provided. Therefore, semantic attributes and their semantic counterparts for auditory perception should be further investigated.

4.6. Negation and Reversal Labels

The bipolar VAS with the end labels from negative to positive meaning was preferred by respondents than the bipolar VAS positive to negative meaning. For unipolar VASs, the order of end labels from negation to regular was preferred by the respondents. The participants might associate "negative" meaning or form with left and "positive" meaning or form with right. Spatial-numerical associations have been studied extensively [67–69], and a leftward bias for visual perception [70]. However, references regarding spatial-positive relations have not been found, at least to the knowledge of the authors.

5. Conclusions

VASs can be used to assess auditory perceptions by substituting the traditional numerical 11-point response scale among young educated adults in a mobile environment.

Among the VASs investigated in this study, the bipolar VAS with a negative left end and a positive right end was considered the most reliable, sensitive, and preferred response scale.

The effect of the polarity (unipolar vs. bipolar) was stronger than that of the labelling direction (negated to regular; negative to positive) on the VASs. Further research is necessary to investigate the use of response scales for the general public including children and the elderly, as well as that of semantic adjectives and their counterparts for auditory perception assessment in mobile and virtual reality environments.

Author Contributions: Conceptualisation, W.Y.; methodology, W.Y.; software, W.Y.; validation, W.Y. and J.Y.J.; formal analysis, W.Y.; investigation, W.Y.; resources, W.Y. and J.Y.J.; data curation, W.Y.; writing—original draft preparation, W.Y.; writing—review and editing, J.Y.J.; visualisation, W.Y.; supervision, J.Y.J.; project administration, W.Y. and J.Y.J.; funding acquisition, W.Y. and J.Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: 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. This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (no. 2020R1A2C2009716). This study was also conducted with research funds provided by Gwangju University in 2021.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Hanyang University (HYUIRB-202010-011, 22 October 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We gratefully acknowledge the contributions of all participants in our experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

- DeCastellarnau, A. A classification of response scale characteristics that affect data quality: A literature review. *Qual. Quant.* **2018**, *52*, 1523–1559. [[CrossRef](#)] [[PubMed](#)]
- Wetzel, E.; Greiff, S. The world beyond rating scales. *Eur. J. Psychol. Assess.* **2018**, *34*, 1–5. [[CrossRef](#)]
- Fields, J.M.; De Jong, R.G.; Gjestland, T.; Flindell, I.H.; Job, R.F.S.; Kurra, S.; Lercher, P.; Vallet, M.; Yano, T.; Guski, R.; et al. Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. *J. Sound Vib.* **2001**, *242*, 641–679. [[CrossRef](#)]
- ISO/TS 15666: 2021. *Acoustics—Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys*; International Organization for Standardization: Geneva, Switzerland, 2021; Volume 15666.
- ISO 12913-1: 2014. *Acoustics-Soundscape-Part 1: Definition and Conceptual Framework*; International Organization for Standardization: Geneva, Switzerland, 2014.
- ISO/TS 12913-2: 2018. *Acoustics-Soundscape-Part 2: Methods and Measurements*; International Organization for Standardization: Geneva, Switzerland, 2014.
- Bond, A.; Lader, M. The use of analogue scales in rating subjective feelings. *Br. J. Med. Psychol.* **1974**, *47*, 211–218. [[CrossRef](#)]
- Joyce, C.; Zutshi, D.; Hrubes, V.; Mason, R. Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur. J. Clin. Pharmacol.* **1975**, *8*, 415–420. [[CrossRef](#)] [[PubMed](#)]
- Seymour, R.A. The use of pain scales in assessing the efficacy of analgesics in post-operative dental pain. *Eur. J. Clin. Pharmacol.* **1982**, *23*, 441–444. [[CrossRef](#)] [[PubMed](#)]
- Guyatt, G.H.; Townsend, M.; Berman, L.B.; Keller, J.L. A comparison of Likert and visual analogue scales for measuring change in function. *J. Chronic Dis.* **1987**, *40*, 1129–1133. [[CrossRef](#)]
- Herr, K.A.; Mobily, P.R. Comparison of selected pain assessment tools for use with the elderly. *Appl. Nurs. Res.* **1993**, *6*, 39–46. [[CrossRef](#)]
- Bolognese, J.A.; Schnitzer, T.J.; Ehrich, E.W. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. *Osteoarthr. Cartil.* **2003**, *11*, 499–507. [[CrossRef](#)]
- Herr, K.A.; Spratt, K.; Mobily, P.R.; Richardson, G. Pain intensity assessment in older adults: Use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin. J. Pain* **2004**, *20*, 207–219. [[CrossRef](#)]
- Cox, J.M.; Davison, A. The visual analogue scale as a tool for self-reporting of subjective phenomena in the medical radiation sciences. *Radiographer* **2005**, *52*, 22–24. [[CrossRef](#)]
- Skovlund, E.; Bretthauer, M.; Grotmol, T.; Larsen, I.K.; Hoff, G. Sensitivity of pain rating scales in an endoscopy trial. *Clin. J. Pain* **2005**, *21*, 292–296. [[CrossRef](#)] [[PubMed](#)]
- Davey, H.M.; Barratt, A.L.; Butow, P.N.; Deeks, J.J. A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *J. Clin. Epidemiol.* **2007**, *60*, 356–360. [[CrossRef](#)]
- Hjermstad, M.J.; Fayers, P.M.; Haugen, D.F.; Caraceni, A.; Hanks, G.W.; Loge, J.H.; Fainsinger, R.; Aass, N.; Kaasa, S. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: A systematic literature review. *J. Pain Symptom Manag.* **2011**, *41*, 1073–1093. [[CrossRef](#)]
- Demoly, P.; Bousquet, P.; Mesbah, K.; Bousquet, J.; Devillier, P. Visual analogue scale in patients treated for allergic rhinitis: An observational prospective study in primary care: Asthma and rhinitis. *Clin. Exp. Allergy* **2013**, *43*, 881–888. [[CrossRef](#)]

19. 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. *Rheumatol. Int.* **2013**, *33*, 2617–2623. [[CrossRef](#)] [[PubMed](#)]
20. Sayin, Y.Y.; Akyolcu, N. Comparison of pain scale preferences and pain intensity according to pain scales among turkish patients: A descriptive study. *Pain Manag. Nurs.* **2014**, *15*, 156–164. [[CrossRef](#)] [[PubMed](#)]
21. Bahreini, M.; Jalili, M.; Moradi-Lakeh, M. A comparison of three self-report pain scales in adults with acute pain. *J. Emerg. Med.* **2015**, *48*, 10–18. [[CrossRef](#)]
22. Voutilainen, A.; Pitkääho, 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. *J. Adv. Nurs.* **2016**, *72*, 946–957. [[CrossRef](#)]
23. Alghadir, A.H.; Anwer, S.; Iqbal, A.; Iqbal, Z.A. Test–retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. *J. Pain Res.* **2018**, *11*, 851. [[CrossRef](#)]
24. Le May, S.; Ballard, A.; Khadra, C.; Gouin, S.; Plint, A.C.; Villeneuve, E.; Mâsse, B.; Tsze, D.S.; Neto, G.; Drendel, A.L. Comparison of the psychometric properties of 3 pain scales used in the pediatric emergency department: Visual analogue scale, faces pain scale-revised, and colour analogue scale. *Pain* **2018**, *159*, 1508–1517. [[CrossRef](#)]
25. Karcioğlu, O.; Topacoglu, H.; Dikme, O.; Dikme, O. A systematic review of the pain scales in adults: Which to use? *Am. J. Emerg. Med.* **2018**, *36*, 707–714. [[CrossRef](#)] [[PubMed](#)]
26. Safikhani, S.; Gries, K.S.; Trudeau, J.J.; Reasner, D.; Rüdell, K.; Coons, S.J.; Bush, E.N.; Hanlon, J.; Abraham, L.; Vernon, M. Response scale selection in adult pain measures: Results from a literature review. *J. Patient Rep. Outcomes* **2018**, *2*, 1–9. [[CrossRef](#)] [[PubMed](#)]
27. Elera-Fitzcarrald, C.; Vega, K.; Gamboa-Cárdenas, R.V.; Zúñiga, K.; Zevallos, F.; Reátegui-Sokolova, C.; Pastor-Asurza, C.; Perich-Campos, R.; Bellido, Z.R.; Aranow, C. Reliability of visual analog scale and numeric rating scale for the assessment of disease activity in systemic lupus erythematosus. *JCR J. Clin. Rheumatol.* **2020**, *26*, S170–S173. [[CrossRef](#)]
28. Shafshak, T.S.; Elnemr, R. The visual analogue scale versus numerical rating scale in measuring pain severity and predicting disability in low back pain. *J. Clin. Rheumatol. Pract. Rep. Rheum. Musculoskelet. Dis.* **2020**. [[CrossRef](#)]
29. Turnbull, A.; Sculley, D.; Escalona-Marfil, C.; Riu-Gispert, L.; Ruiz-Moreno, J.; Gironès, X.; Coda, A. Comparison of a mobile health electronic visual analog scale app with a traditional paper visual analog scale for pain evaluation: Cross-sectional observational study. *J. Med. Internet Res.* **2020**, *22*, e18284. [[CrossRef](#)] [[PubMed](#)]
30. Couper, M.P.; Tourangeau, R.; Conrad, F.G.; Singer, E. Evaluating the effectiveness of visual analog scales: A web experiment. *Soc. Sci. Comput. Rev.* **2006**, *24*, 227–245. [[CrossRef](#)]
31. Reips, U.-D.; Funke, F. Interval-level measurement with visual analogue scales in Internet-based research: VAS generator. *J. Behav. Res. Methods* **2008**, *40*, 699–704. [[CrossRef](#)]
32. Toepoel, V.; Funke, F. Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Math. Popul. Stud.* **2008**, *40*, 699–704. [[CrossRef](#)]
33. Gummer, T.; Kunz, T. Using only numeric labels instead of verbal labels: Stripping rating scales to their bare minimum in web surveys. *Soc. Sci. Comput. Rev.* **2020**. [[CrossRef](#)]
34. De Bruijne, M.; Wijnant, A. Can mobile web surveys be taken on computers? A discussion on a multi-device survey design. *Surv. Pract.* **2013**, *6*, 2886. [[CrossRef](#)]
35. Toepoel, V.; Lugtig, P. Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. *Methods Data Anal.* **2015**, *9*, 8.
36. Hong, J.Y.; Lam, B.; Ong, Z.-T.; Ooi, K.; Gan, W.-S.; Kang, J.; Feng, J.; Tan, S.-T. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Build. Environ.* **2019**, *149*, 1–14. [[CrossRef](#)]
37. Jo, H.I.; Jeon, J.Y. Downstairs resident classification characteristics for upstairs walking vibration noise in an apartment building under virtual reality environment. *Build. Environ.* **2019**, *150*, 21–32. [[CrossRef](#)]
38. Brink, M.; Schreckenber, D.; Vienneau, D.; Cajochen, C.; Wunderli, J.-M.; Probst-Hensch, N.; Rösli, M. Effects of scale, question location, order of response alternatives, and season on self-reported noise annoyance using ICBEN scales: A field experiment. *Int. J. Environ. Res. Public Health* **2016**, *13*, 1163. [[CrossRef](#)] [[PubMed](#)]
39. Nguyen, T.; Yano, T.; Morihara, T.; Yokoshima, S.; Morinaga, M. Comparison of annoyance response measured with ICBEN 5-point verbal and 11-point numerical scales in Japanese and Vietnamese. In Proceedings of the 12th ICBEN Congress on Noise as a Public Health Problem, Zurich, Switzerland, 18–22 June 2017.
40. Tristán-Hernández, E.; Pavón García, I.; López Navarro, J.M.; Campos-Cantón, I.; Kolosovas-Machuca, E.S. Evaluation of psychoacoustic annoyance and perception of noise annoyance inside University facilities. *Int. J. Acoust. Vib* **2018**, *23*, 3–8.
41. Bjerre, L.C.; Larsen, T.M.; Sørensen, A.J.; Santurette, S.; Jeong, C.-H. On-site and laboratory evaluations of soundscape quality in recreational urban spaces. *Noise Health* **2017**, *19*, 183. [[PubMed](#)]
42. Pellerin, N.; Candau, V. Effects of steady-state noise and temperature conditions on environmental perception and acceptability. *Indoor Air* **2004**, *14*, 129–136. [[CrossRef](#)] [[PubMed](#)]
43. Nagano, K.; Horikoshi, T. New comfort index during combined conditions of moderate low ambient temperature and traffic noise. *Energy Build.* **2005**, *37*, 287–294. [[CrossRef](#)]

44. Tiller, D.K.; Wang, L.M.; Musser, A.; Radik, M. Combined effects of noise and temperature on human comfort and performance. *ASHRAE Trans.* **2010**, *116*, 522–540.
45. Hall, D.A.; Irwin, A.; Edmondson-Jones, M.; Phillips, S.; Poxon, J.E. An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. *Appl. Acoust.* **2013**, *74*, 248–254. [[CrossRef](#)]
46. Yang, W.; Moon, H.J.; Kim, M.-J. Combined effects of short-term noise exposure and hygrothermal conditions on indoor environmental perceptions. *Indoor Built Environ.* **2018**, *27*, 1119–1133. [[CrossRef](#)]
47. Yang, W.; Moon, H.J. Cross-modal effects of noise and thermal conditions on indoor environmental perception and speech recognition. *Appl. Acoust.* **2018**, *141*, 1–8. [[CrossRef](#)]
48. IOS 10551: 2019. *Ergonomics of the Thermal Environment—Assessment of the Influence of the Thermal Environment Using Subjective Judgement Scales*; International Organization for Standardization: Geneva, Switzerland, 2019.
49. Yang, W.; Jeon, J.Y. Performance and preference of response scales for semantic differentials in auditory perception among university students. *Can. Acoust.* **2019**, *47*, 41–50.
50. Yang, W.; Moon, H.J.; Jeon, J.Y. Comparison of response scales as measures of indoor environmental perception in combined thermal and acoustic conditions. *Sustainability* **2019**, *11*, 3975. [[CrossRef](#)]
51. Chyung, S.Y.; Swanson, I.; Roberts, K.; Hankinson, A. Evidence-based survey design: The use of continuous rating scales in surveys. *Perform. Improv.* **2018**, *57*, 38–48. [[CrossRef](#)]
52. Ma, K.W.; Wong, H.M.; Mak, C.M. A systematic review of human perceptual dimensions of sound: Meta-analysis of semantic differential method applications to indoor and outdoor sounds. *Build. Environ.* **2018**, *133*, 123–150. [[CrossRef](#)]
53. Berglund, B.; Berglund, U.; Lindvall, T. Scaling loudness, noisiness, and annoyance of aircraft noise. *J. Acoust. Soc. Am.* **1975**, *57*, 930–934. [[CrossRef](#)]
54. Berglund, B.; Berglund, U.; Lindvall, T. Scaling loudness, noisiness, and annoyance of community noises. *J. Acoust. Soc. Am.* **1976**, *60*, 1119–1125. [[CrossRef](#)]
55. Berglund, B.; Lindvall, T. *Community Noise*; Center for Sensory Research: Stockholm, Sweden, 1995.
56. Marquis-Favre, C.; Premat, E.; Aubrée, D. Noise and its effects—A review on qualitative aspects of sound. Part II: Noise and annoyance. *Acta Acust. United Acust.* **2005**, *91*, 626–642.
57. Yang, W.; Moon, H.J. Combined effects of sound and illuminance on indoor environmental perception. *Appl. Acoust.* **2018**, *141*, 136–143. [[CrossRef](#)]
58. Fanger, P.O. *Thermal Comfort. Analysis and Applications in Environmental Engineering*; Danish Technical Press: Copenhagen, Denmark, 1970; p. 244.
59. Kooi, F.L.; Toet, A. Visual comfort of binocular and 3D displays. *Displays* **2004**, *25*, 99–108. [[CrossRef](#)]
60. Arif, M.; Katafygiotou, M.; Mazroei, A.; Kaushik, A.; Elsarrag, E. Impact of indoor environmental quality on occupant well-being and comfort: A review of the literature. *Int. J. Sustain. Built Environ.* **2016**, *5*, 1–11.
61. Karimipannah, T.; Awbi, H.; Sandberg, M.; Blomqvist, C. Investigation of air quality, comfort parameters and effectiveness for two floor-level air supply systems in classrooms. *Build. Environ.* **2007**, *42*, 647–655. [[CrossRef](#)]
62. Weijters, B.; Geuens, M.; Schillewaert, N. The proximity effect: The role of inter-item distance on reverse-item bias. *Int. J. Res. Mark.* **2009**, *26*, 2–12. [[CrossRef](#)]
63. Weijters, B.; Baumgartner, H. Misresponse to reversed and negated items in surveys: A review. *J. Mark. Res.* **2012**, *49*, 737–747. [[CrossRef](#)]
64. Budescu, D.V.; Appelbaum, M.I. Variance stabilizing transformations and the power of the F test. *J. Educ. Stat.* **1981**, *6*, 55–74. [[CrossRef](#)]
65. Box, G.E.P. 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. *Ann. Math. Stat.* **1954**, *25*, 484–498. [[CrossRef](#)]
66. Nielsen, J.; Levy, J. Measuring usability: Preference vs. performance. *Commun. ACM* **1994**, *37*, 66–75. [[CrossRef](#)]
67. Shaki, S.; Fischer, M.H. Deconstructing spatial-numerical associations. *Cognition* **2018**, *175*, 109–113. [[CrossRef](#)]
68. Cipora, K.; Haman, M.; Domahs, F.; Nuerk, H.-C. On the development of space-number relations: Linguistic and cognitive determinants, influences, and associations. *Front. Psychol.* **2020**, *11*, 182. [[CrossRef](#)] [[PubMed](#)]
69. Felisatti, A.; Laubrock, J.; Shaki, S.; Fischer, M.H. A biological foundation for spatial–numerical associations: The brain's asymmetric frequency tuning. *Ann. N. Y. Acad. Sci.* **2020**, *1477*, 44–53. [[CrossRef](#)]
70. Rodway, P.; Schepman, A. A leftward bias for the arrangement of consumer items that differ in attractiveness. *Laterality* **2020**, *25*, 599–619. [[CrossRef](#)]