Translation of soundscape perceptual attributes from English to Turkish

Papatya Nur Dökmeci Yörükoğlu[1*], Özlem Türker Bayrak[2], Nilgün Akbulut Çoban[3], Uğur Beyza Erçakmak Osma[1], Francesco Aletta[4], Tin Oberman[4], Andrew Mitchell[4], Jian Kang[4]

[1] Department of Interior Architecture, Faculty of Architecture, Çankaya University, Ankara, Turkey
[2] Statistics Unit, Inter-Curricular Courses Department, Çankaya University, Ankara, Turkey
[3] Department of Environment Management and Audit, Antalya Provincial Directorate of Environment, Urbanization and Climate Change, Antalya, Turkey
[4] Institute for Environmental Design and Engineering, The Bartlett, University College London, Central House, WC1H0NN London, United Kingdom

*Corresponding author: papatya@cankaya.edu.tr

Abstract

The International Standard Organization (ISO) published the standard series on soundscape for the identification, data collection and data analysis. However, since all these standards are in English language, the reliable standardized usage in other languages and its applicability is questionable. Thus, this two-staged study aims: i) to determine the Turkish equivalences of the 8 soundscape perceptual attributes that are published in ISO/TS 12913-2:2018 and ISO/TS 12913-3:2019, ii) to analyze if the determined Turkish attributes have concept equivalence to the original ones, and iii) to determine if the translated Turkish scale is reproducible. The first stage involved the translation of the attributes by focus group discussions and finalization by Turkish linguistic experts. As a result, the attributes 'eventful', 'vibrant', 'pleasant', 'calm', 'uneventful', 'monotonous', 'annoying', and 'chaotic' are translated to Turkish as 'hareketli', 'coşkulu', 'keyifli', 'sakin', 'durağan', 'tekdüze', 'rahatsız edici', and 'karmaşık', respectively. The second stage involves the analysis for the reproducibility in terms of inter-rater reliability and conceptual validity. It is found that the Turkish scale is reproducible due to high inter-rater reliability in all attributes. Context validity at conceptual level is analyzed both in terms of the difference between the average scores given to the English attributes and their corresponding Turkish equivalences and the correlation between the English and Turkish scores given to each attribute. The highest difference between the average scores (around 10 points on a slider scale of 0 to 100) is found to be in the translation of 'vibrant' while the lowest correlated one (slightly lower than 0.5) is found in 'chaotic' attribute as in line with literature. Despite of this result, when the scores are reduced to 2 dimensions as pleasantness and eventfulness, it is seen that there is a high correlation between the English and Turkish scales. It is considered that the results obtained from this research could act as a base in the future for the establishment of Turkish Standards on soundscape and standardization of the translated and validated Turkish soundscape perceptual attributes and the 'perceived affective quality' scale defined under ISO/TS 12913-2:2018 in English.

Keywords: ISO 12913; acoustics; perceived affective quality; listening tests; reliability and context validity; conceptual equivalence

1. Introduction

The soundscape approach, which is a rapidly developing research field in recent years, focuses on how individuals perceive, experience, and understand sound environments [1]. In this manner, the soundscape field aims to evaluate and/or enhance an acoustic environment with a holistic attitude, including both physical characteristics of sound and how individuals contextually interpret these characteristics. ISO established a working group for perceptual evaluation of soundscapes in 2008 (ISO/TC 43/SC 1/WG 54 – Perceptual Assessment of Soundscape Quality). This working group published the first standard ISO 12913-1:2014 in 2014 [1], which gives the definition of soundscape and related concepts. In 2018 and 2019, the second and the third part are published respectively as technical specifications. ISO/TS 12913-2:2018 document is related to the data gathering and reporting protocols and methods [2], whereas ISO/TS 12913-3:2019 [3] is related to data analysis methods.

The tendency for the use of standardized data gathering methods for soundscape research has increased in different regions after the series of ISO standards are published. However, it is known that sounds and their interaction with the environment are defined differently in various languages [4, 5] yielding the need for translation into other languages, conducting studies at national level and gathering data which can be universally compatible. Though the studies regarding the translation of data-gathering methods are limited, preliminary ones [6-10] identified some critical aspects for adapting the English version to other languages. However, consistency and comparison of questionnaires designed and used in other languages remain problematic despite using the English protocols recommended in ISO/TS 12913-2:2018 [5].

The basis of the ISO/TS 12913-2:2018 document [2] was structured mainly on initial works, one of which is the study of Axelsson et al.[11] recommending the measurement system for soundscape quality including a two-dimensional space as Pleasantness-Eventfulness defined by the eight attributes in a circumplex model of soundscape perception. The other baseline work is the study of Cain et.al [12] which revealed that the two independent emotional dimensions of a soundscape relate to its "Calmness"

and “Vibrancy”. There have also been studies focusing on certain attributes, such as the concept of ‘vibrancy’, to examine the ‘perceived affective quality’ attributes of the soundscape approach[13]. In the light of these studies, 8 attributes are finalized as ‘eventful, vibrant, pleasant, calm, uneventful, monotonous, annoying, chaotic’ and published in the ISO/TS 12913-2:2018 document in 2018 and indicated in the ‘perceived affective quality (PAQ)’ scale. In order to provide globally usable and comparable PAQ attributes, ‘Soundscape Attributes Translation Project’ (SATP) initiative has been conducting an international multi-partnered study to translate and validate the 8 soundscape perceptual attributes in languages other than English [5].

Some of the studies focusing on translation of the PAQ attributes from English to other languages are summarized in **Table 1** together with their methodology and the number of participants included [8-10, 14-17]. **Table 1** shows that different methodologies are used according to the varying aim, objectives and expected outcomes. Studies concentrating on the other languages across the world are still continuing in terms of translation and validation of the PAQ scale, which is included in ISO/TS 12913-2:2018 in English. This study is designed to perform the Turkish translation and validation of the scale as part of the SATP team and conducted as a ‘Scientific Research Project (SRP)’ funded by Çankaya University, Turkey.

Table 1
Methodologies Used for translation of PAQ Attributes in the literature.

Research	Target Language	Translation of PAQ attributes		Evaluation of translated PAQ attributes	
		Methodology	Number of participants	Methodology	Number of participants
Jeon et al. [8]	Korean	Based on the authors’ experience and knowledge of soundscape research	6	Using semantic descriptors in their own native language at the laboratory of each country	30
	French				30
	Swedish				35
Nagahata et al. [9]	Japanese	Using six popular English – Japanese dictionaries	1	Assessing soundscapes in a class via assessment sheets	156
				Watching videos of six parks in laboratory	26
				Soundwalk on six places in urban site	11
Tarlao et al. [10]	French	Based on the authors’ previous experience and knowledge	4	Questionnaire applied in a small park	197
				Questionnaire applied in a large park	247
				Questionnaire applied in a restaurant	41
Antunes et al. [14]	Portuguese	Based on the authors’ previous experience and knowledge	4	Online surveys applied to people	245
Papadakis et al. [15]	Greek	Combined technique of forward translation, synthesis, back translation, pre-test and committee approach is implemented by four independent translations groups	16	Listening tests in laboratory	62
Watcharasupat et al. [17]	Thai	Focus group discussion with 5 linguistic experts and additionally 1 soundscape researcher	6	Conducting online survey. to people in Thailand, Singapore, and UK	31
Sudarsono et.al [16]	Indonesian	Focus group discussion with researchers and students working on soundscape	5	Soundwalk in situ	5

Modern Turkish is the national primary language in Turkey and the Turkish Republic of Northern Cyprus, and is also spoken in various Turkish communities in Europe and Middle East [18]. Turkish is one of the most widely spoken language members of the Turkic language family [19] which is spoken in Turkey and Azerbaijan. Turkmen, Uzbek, Kazakh, Kyrgyz, Uyghur, Karakalpak, Yakut, Tuva, Altai, Caucasian, Tatar, Bashkir, Karachay, Malkar, Gagauz, and Chuvash speak Turkic originated languages [20]. Mustafa Kemal Atatürk, the founder of the Turkish Republic (1923) started the language modernization movement to purify the Turkish vocabulary from the foreign elements and words that led to the replacement of Arabic script by the Latin alphabet in 1928. In today’s world, approximately 250-300 million people speak Turkish [21, 22]. Therefore, the translation of soundscape attributes into Turkish is important to develop and compare the studies on soundscape research field.

In Turkey, the soundscape studies are started mostly after 2011. As an initiator, Özçevik [23, 24] carried out research conducting in four study areas consisting of two enclosed shopping centers and two urban streets with commercially active property in Istanbul, which aims to investigate the flexibility of using soundscape approach in both open urban spaces and enclosed ones. The research of [23] includes field study for subjective assessment by a survey including a questionnaire and a semantic differential test. As a second part of [23], the study was conducted to get proper data decided for using to assess if the subjective evaluation of soundscapes in laboratory environment is consistent with the data obtained from the field one [24]. Additionally, in the scope of urban spaces, the project “Evaluation of Soundscape Perception and User Preferences in Urban Parks in Frame of Life Quality: Case Study Ankara

– Turkey" was accepted by The Scientific and Technological Research Council of Turkey (TÜBİTAK) within the European Cooperation in Science and Technology (COST) Action "Soundscape of European Cities and Landscapes" (TD0804) in 2010 [25]. In that research, face-to-face questionnaire surveys and sound level measurements are conducted in five urban parks of Ankara city [26-28]. In order to determine the perceived soundscape characteristics of four key locations as Gezi Park, Taksim Square, Galatasaray Square, and Tunel Square which are all located in İstanbul, soundwalks were carried out based on the methodologies used in the 'Positive Soundscape Project' and Schafer's 'World Soundscape Project' [29]. Besides these, in the study conducted in Diyarbakır Walled City (Suriçi) Region, the perception of sound environment of users in urban spaces with sound quality metrics as loudness, sharpness, roughness, and fluctuation strength are obtained through the binaural sound recordings performed simultaneously with the questionnaires [30]. In those studies, carried out in 3 different cities of Turkey, there are more than 20 adjective pairs, generally with bipolar scales, are used in questionnaires to obtain the perception of the sound environment. However, those adjective pairs do not completely agree with the 8 adjective pairs defined under the ISO/TS 12913-2:2018 since the studies are carried out before 2018 and different questionnaire scale is used.

There are two doctoral theses [31, 32] that use the Turkish equivalences of the 8 soundscape perceptual attributes in their questionnaires. They both focus on the identification of the quiet areas in the scope of environmental noise management and integration of this term with soundscape approach. Akbulut Çoban [31] use the objective parameters as acoustic and air quality and the subjective parameters as soundscape perception and environmental quality pleasantness from four urban parks in Antalya. The soundscape attributes used in that study was based on the project conducted on quiet areas [33]. Akın Güler [32] propose a comprehensive quiet/calm area determination approach and model based on Analytical Hierarchy Process, the Weighted Sum Analysis with Geographic Information Systems in line with the legislation. The model includes the main factors that are effective in the perception of quietness/calmness that revealed with soundscape approach. The Turkish translation of the attributes used in these studies are given in **Table 2** in line with the ones given in [3].

Table 2
The Turkish translation of the soundscape perceptual attributes in previous studies.

Soundscape attributes	Aletta et. al. [5]	Akbulut Çoban [31]	Akın Güler [32]
Eventful	Hareketli	Hareketli	Hareketli
Vibrant	Heyecan verici	Keyifli/Canlı	Heyecan verici
Pleasant	Keyifli	Hoşa giden	Keyifli
Calm	Dingin	Sakin	Sakin
Uneventful	Durağan	Durağan	Durağan
Monotonous	Sıradan	Sıkıcı	Sıradan
Annoying	Keyifsiz	Hoşagitmeyen	Keyifsiz
Chaotic	Kargaşalı	Karmaşık	Kargaşalı

As regards to the standardized soundscape perceptual attributes [2, 3], in several studies [26, 28-31, 34, 35], conducted in Turkey, scales containing some or all of the attributes were translated into Turkish and presented to the participants in the field studies. However, as these working groups used different translations and scales, there are limitations occurred such as, inability to compare and discuss Turkish research outputs. Similar to the drawbacks of translation validation of the soundscape perceptual attributes, noise annoyance surveys did not have a standard scale to be used in noise studies across the world in different languages. Therefore, necessity had arisen to develop a standardized protocol for noise annoyance studies to lead reliable research and comparable results [36]. Consequently, the standardized protocol ISO/TS 15666:2003 [37] had been published first in 2003, then it has been revised by ISO/TS 15666:2021 [38] in light of these necessities. The new version of the standard had revisions on protocol's scales and measures [39], also aimed to translate in different languages to provide international access and comparable datasets [36]. As a result, ISO/TS 15666:2021 [38] has been replaced with a content of seventeen different translations of protocols including Turkish language.

To this extent, there is also a need for Turkish translation and validation of the soundscape perceptual attributes defined under ISO/TS 12913-2:2018 and it is considered that the results obtained from this research would make great contribution to establish the scientific knowledge for the future research and publication of Turkish Standards on soundscape field.

Thus, this study aims to answer the following research questions:

1. Which Turkish equivalences can be used to translate the 8 soundscape perceptual attributes that are published in ISO/TS 12913-2:2018 and ISO/TS 12913-3:2019?

2. Do the respondents interpret and respond in the same manner as they do to English attributes? That is, are the translated Turkish attributes have concept equivalence to the original ones?

3. Is the translated Turkish scale reproducible across native Turkish speaking participants?

2. Methodology

Translating the soundscape perceptual attributes in the standardized 'perceived affective quality' scale into other languages have challenges, which requires the caution in linguistic level on both conceptual and terminological aspect [9]. The study by Papadakis et al. [15] presents the literature review on the translation methodology such as expert panel or committee [40], working in a team of at least three or four persons [41], using focus groups [42], and pilot testing [40]. For example, Antunes et al. [14] used an approach based on qualitative methodology including the researchers' previous experience, survey via the experts and lay people.

Nagahata et al. [9] used six different dictionaries in order to make proper translations of the soundscape attributes. Papadakis et al. [15] used a combined technique of the forward translation, synthesis, back translation, pre-test and a committee approach. Watcharasupat et al. [17] carried out the translation in two phases, one of which is the parallel translations by linguistic experts and a group discussion to obtain a shortlist of candidate translations. Then at the second phase, they proposed a quantitative framework to assess the translation candidates and select the final set based on the evaluation scores. The study of Cha et al. [43] reveals that a combined technique is an appropriate method to maintain the content equivalences between the original and translated instruments in international research. While there are many different methodologies, none of them has been referred to as the gold standard.

An instrument must be both valid and reliable. Validity looks at the accuracy and relates to the question 'Does it measure what it is supposed to measure?' whereas reliability looks at consistency and relates to the question 'How reproducible is the measurement?' A translated instrument must ensure both linguistic validity and content validity. Linguistic validation is necessary to ensure that the translated words state what the original words intended. These steps constitute the first phase of the used methodology given in **Fig.1**. The translation of attributes from English to Turkish in [5] is based on the approach of focus group consisting of four experts as translation methodology. In this study, we take this work one step further through the focus group with 8 experts for translation. After this phase of linguistic validation, it is important to test the linguistically validated instrument to ensure that the respondents interpret and respond in the same manner as the instrument in the original language, i.e., content validity in terms of the conceptual equivalence. Also, it must be reliable, i.e., the results must be reproducible. For these purposes, bilingual listening tests are conducted and analyzed at the second phase of the methodology given in **Fig.1**.

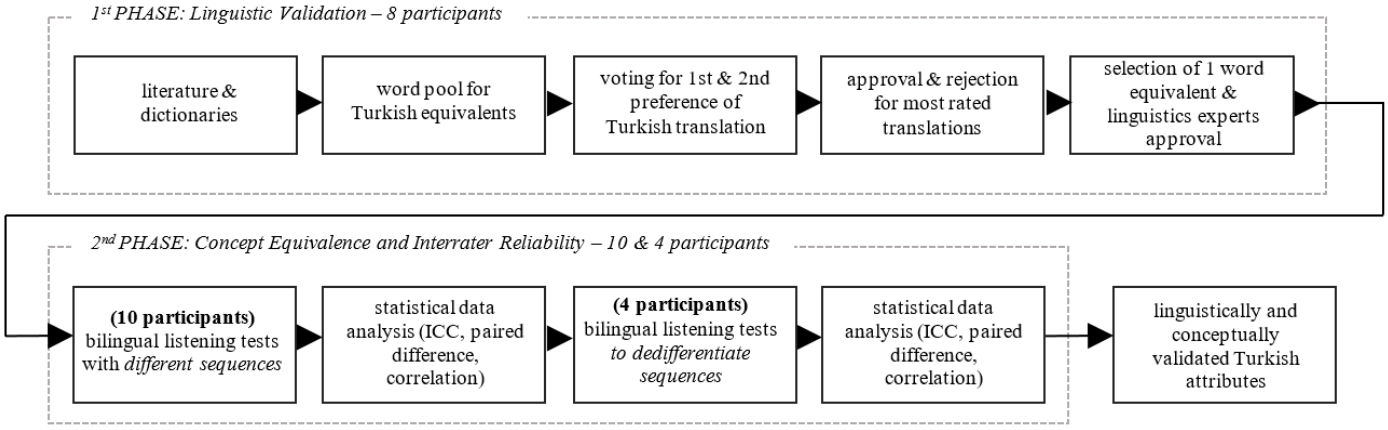

Fig. 1. Methodological framework of the study.

2.1. First Phase: Linguistic Translation

Within the context of focus group studies five steps are organized. Firstly, a detailed literature survey is conducted related to the SATP project and Turkish uses of the soundscape attributes with the additional alternatives from dictionaries. As a result of the survey, extensive word pool for Turkish equivalents of the attributes is formed to be used in focus group studies. In focus group studies, it is aimed to reduce the Turkish equivalents into two alternatives by voting the words from the word pool, then to receive approval for consensus on one of the two Turkish alternatives for each attribute. As a last step of focus group study, linguistic experts' approval is taken to determine one Turkish equivalent for each attribute to be used in the second phase of the study.

Focus group translation studies are conducted with a total of eight participants; four of them are the Turkish authors of this study who are specialized in soundscape, architectural acoustics, noise management and statistics. The other 4 participants are Turkish-speaking experts in the various fields of acoustics most dominantly in the fields of architectural acoustics, soundscape, speech and noise. All participants of the focus group have advanced/mastery level of language proficiency in English, which are approved by the Higher Education Institute of Turkey with an equivalence with International English Proficiency exams such as TOEFL or IELTS.

Before the focus group meeting, literature about the national soundscape studies that uses questionnaires including Turkish attributes are examined [26, 27, 29, 31, 34, 35] and a list for each attribute is prepared to be discussed in the focus group meeting. At the beginning of the meeting, authors informed the participants about the purpose and the scope of the project, and the prepared list of Turkish attributes by a presentation.

2.2. Second Phase: Concept Equivalence and Inter-Rater Reliability

In the second phase of the study, bilingual listening tests are conducted to analyze the content validity at a conceptual level as well as the reproducibility of the Turkish translations of each soundscape attribute obtained at the end of the first stage.

2.2.1. Audio stimuli

The 27 different 30 second binaural soundscape recordings obtained by SATP dataset [44] each including equal number of scenarios dominated by the sound of nature, people and traffic/construction; 1 signal sound file to be used for calibration; information about the installation of the necessary equipment; the profile of the participants for the study; and the survey software using slider scale are obtained from the SATP project data repository [44] for the listening tests.

The Turkish interface of the software (REDCAP) is prepared to be used for the survey applied during the listening tests. Before starting the listening tests, SATP approved headphone (Sennheiser HD650) calibration is done using the signal recording with 1

kHz pure sine tone, as recommended SATP project protocols [45], through measuring the appropriate voltage of headphone sensitivity with the use of a computer, soundcard (Focusrite Scarlett 2i2), and voltmeter (**Fig. 2a**), so that the participants listen to the recordings with the original sound pressure levels (SPL-dBA) (**Fig. 2b**). More in detail, to provide the headphone calibration, a signal to headphone sensitivity ratio is calculated from equation (1) as 355 mV. Here, the signal value SPL calibration is taken as 94 dB obtained from SATP project data repository [44] and Sennheiser HD650 has 103 dB sensitivity at 1kHz. Thereby, 355 mV should be measured from the voltmeter, which is connected to the soundcard to headphone input with ¼ TRS-to-TRS cable. When the measured value reaches this value calibration is completed and headphone is plugged to the sound card.

$$V_{RMS} = \frac{10^{\frac{SPL_{calibration}}{20}}}{10^{\frac{SPL_{sensitivity}}{20}}} = \frac{10^{\frac{94}{20}}}{10^{\frac{103}{20}}} = 0{,}355\ V = 355\ mV \quad (1)$$

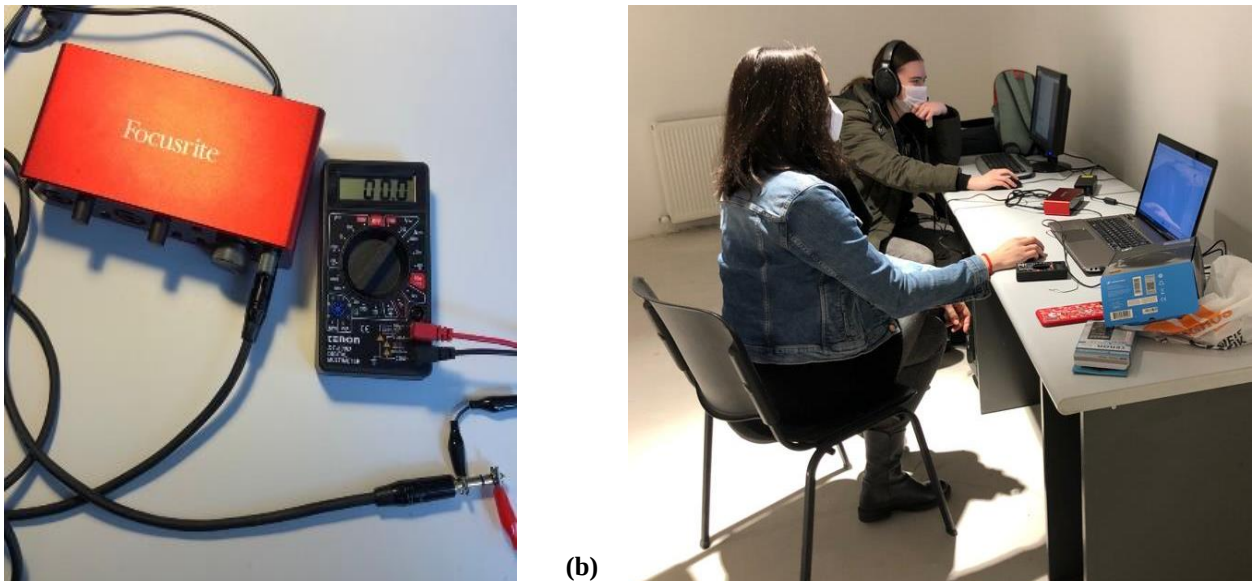

Fig. 2. (a) Soundcard (Focusrite Scarlett 2i2) and voltmeter setup (b) Bilingual listening tests in the laboratory.

2.2.2. Participants

The number of participants for the listening tests is determined to be 10 as described in Section 2.3.1. As recommended by the SATP aiming to ensure that each international project partners of SATP follow the similar procedure in their national study; 5 females and 5 male participants are selected randomly among undergraduate and graduate students between the ages of 18-30 satisfying the study requirements of age, gender, and English language proficiency.

Participants are randomly selected among the undergraduate programs of Faculty of Architecture at Çankaya University giving education in English. To represent the graduate students who are older, participants have been selected among the research assistants of the same Faculty. Some of the selected assistants are students of other universities teaching in English. The participants speak modern Turkish and have either passed the English Proficiency exam of Çankaya University that have a B2 equivalence in Common European Framework of Reference (CEFR), or have advanced/mastery level of language proficiency in English that was determined by standardized tests applied by the Turkish Higher Education Institute, which have equivalence with International English Proficiency exams such as TOEFL or IELTS.

2.2.3. Listening tests

5 randomly determined sequences of the 27 audio scenarios are prepared and assigned randomly to the participants in such a way that each sequence is assigned to 1 female and 1 male participant to eliminate the gender effect. Participants are first given a listening test using the Turkish survey, and to mitigate memory effects, after 1 week, listening tests are conducted again using the English survey. A slider scale of 0-100 is used in the survey and the participants are asked to give a value between 0 (not at all) and 100 (extremely) to evaluate all 8 attributes and perceived loudness for each scenario. Thus, 27*10=270 pairs of evaluation are obtained for the statistical analysis. Since the sequence of the audio scenarios listened for the Turkish scale is different from the one for English scale for each participant at this stage and feedback about the difficulty of evaluation of some sequences is obtained from the participants, a second study is conducted to eliminate the effect of sequencing. It is important since if the scores to English and Turkish attributes differ, the reason cannot be assigned totally to inappropriate translation of the attribute since it might be due to this sequence effect. Thus, if there is such an effect, the analysis will be more reliable when the evaluators give score to each scenario at the same sequence for both Turkish and English scales. Due to the time limitation, 4 of the previous participants whose English proficiency level is higher based on the standardized tests mentioned in Section 2.2.2, are asked to evaluate the scenarios again. First, they evaluated the scenarios based on the Turkish survey in the same sequence they listened for the English survey before. Then, they are asked to evaluate the scenarios based on the English survey in the sequence they listened for the Turkish survey before. Thus, each participant evaluated scenarios with two different sequences, which are the same for both English and Turkish surveys.

2.3. Statistical Methods

There are different ways to measure the reliability of an instrument:

- *Test-retest reliability*: measures the consistency through time,
- *Inter-rater reliability:* measures the consistency across the evaluators,
- *Parallel forms reliability:* measures the consistency across theoretically-equivalent measurements,
- *Internal consistency reliability:* measures the consistency of the answers to the different parts of the measurement.

An instrument does not need to possess all of these reliability forms. The choice depends on the design and purpose of the instrument. In the translation of the soundscape perceptual attributes, the consistency among the participants is an important issue since the comparison of Turkish and English attributes would be meaningless if the participants do not agree with the scores for each attribute in each language. Thus, we focus on the inter-rater reliability in this study. There are different statistics used for this purpose distinguished according to the data type. Since we use a slider scale, ICC that is calculated by the mean squares obtained through analysis of variance is the suitable statistic.

There are different ICC forms according to the underlying model, the type and the definition of the relationship considered to be important. A guideline for the selection of the correct ICC statistic can be found in [46]. In this study, the participants are randomly selected as described in section 2.2.2 so that the results can be generalized to a larger population of native Turkish speakers. Also, to involve the variation due to participants in the analysis, the study is conducted under the scheme that each participant evaluates all audio scenarios. Consequently 2-way random effects model is chosen in ICC analysis. Further, each scenario should theoretically have the same value for each attribute and since, regardless of the reason, the differences in the absolute value of the scores should be viewed as disagreement, the concern is about absolute agreement rather than consistency. Finally, since the assessment is based on the mean score of the participants in soundscape studies, we use the ICC statistic based on a mean-rating, absolute agreement, 2-way random effects model calculated as;

$$ICC = \frac{MS_{audio\ scenarios} - MSE}{MS_{audio\ scenarios} + \frac{MS_{participants} - MSE}{n}} \tag{2}$$

where $MS_{audio\ scenarios}$, $MS_{participants}$ and MSE are the mean squares for audio scenarios evaluated, participants and the error; respectively obtained from the ANOVA table and n is the number of audio scenarios (27 in this study). It takes values between 0 and 1 where values close to 1 indicates stronger reliability. If the judges frequently disagree, the inter-rater reliability will be low meaning that the measurement system is invalid.

The corresponding 100(1-α)% confidence interval is

$$\left(\frac{n(MS_{audio\ scenarios} - F_* MSE)}{F_*(MS_{participants} - MSE) + nMS_{audio\ scenarios}}, \frac{n(F^* MS_{audio\ scenarios} - MSE)}{MS_{participants} - MSE + nF^* MS_{audio\ scenarios}} \right) \tag{3}$$

where F_* is the $(1-\alpha/2)100^{th}$ percentile of F distribution with n-1 numerator degrees of freedom and ν denominator degrees of freedom and F^* is the $(1-\alpha/2)100^{th}$ percentile of F distribution with ν numerator degrees of freedom and n-1 denominator degrees of freedom. Here

$$\nu = \frac{(aMS_{participants} + bMSE)^2}{\frac{(aMS_{participants})^2}{k-1} + \frac{(bMSE)^2}{(n-1)(k-1)}}, \quad a = \frac{kICC}{n(1-ICC)}, \quad b = 1 + \frac{kICC(n-1)}{n(1-ICC)}, \tag{4}$$

and k is the number of participants.

2.3.1. Choice of the Number of Participants

Determination of the number of audio scenarios and the number of participants is an important issue for reliability analysis. There are two main approaches to determine the required sample size in the literature. First one is based on the hypothesis testing perspective [47, 48], while the second one is based on the estimation perspective [49, 50]. Since our aim in this study is to estimate the ICC value not to test it, second approach is the appropriate one.

The problem in this approach is that the width of the confidence interval cannot be expressed as a simple function of sample size thus cannot be solved to yield the desired width. To overcome this issue, Bonett [49] approximates the interval width and suggests the approximate sample size that yields an exact 100(1-α)% confidence interval having the desired width for the 1-way and 2-way single measurement type ICC measures as:

$$n = 8z_{\alpha/2}^2 \frac{(1-\widehat{ICC})^2(1+(k-1)\widehat{ICC})^2}{k(k-1)w^2} + 1 \tag{5}$$

which is rounded up to the nearest integer. Here, $z_{\alpha/2}$ is the point on a standard normal distribution exceeded with probability $\alpha/2$, n is the number of audio scenarios, k is the number of participants, w is the interval width, and $\widehat{ICC}$ is the planning ICC value.

Besides, simulation studies are performed [51-53] to have an insight for the required sample size. However, all these studies are done for the single measurement type of ICC.

It is known that the confidence interval for the mean of k measurement type ICC is narrower than that of single measurement type of ICC and the bounds for the confidence interval can be obtained from the bounds of single measurement type of ICC as [54]:

$$\frac{k\rho^*}{1+(k-1)\rho^*} \tag{6}$$

where ρ^* represents the lower (or upper) bound of 2-way single measurement confidence interval. According to equation [6], even if the width of the single measurement type ICC is 0.3, the width of the interval for the mean of k measurement type of ICC will be less than 0.1 even for the 3 participants.

For 95% confidence interval with the planning ICC value of 0.8 due to the high reliability of the English scale, n=27 audio scenarios yield a width of w=0.2 with k=6 participants according to equation (5). If the number of participants is taken as 10, the required number of audio scenarios become 93 and 24 to obtain a width of 0.1 and 0.2, respectively, indicating that the width with 27 audio scenarios will be between 0.1 and 0.2 which will yield a width less than 0.1 according to equation (6) for the mean measure type of ICC used in this study. Thus, to be conservative, number of participants is determined as 10, which is also found to be appropriate according to the simulation results of [51-53]. JASP 0.16.3 (free R based statistical package) is used for all statistical analysis.

3. Results and Discussions

3.1. Results of the First Phase Analysis

Using the semantic equivalents of attributes in the literature and their English-English and English-Turkish equivalents obtained through different reliable dictionaries, a large Turkish attribute pool given in **Table 3** is gathered at the first focus group meeting.

Table 3
Word pool formed by the focus group.

Eventful	Vibrant	Pleasant	Calm	Uneventful	Monotonous	Annoying	Chaotic
Hareketli	Canlı	Keyifli	Dingin	Durağan	Sıradan	Rahatsız edici	Kargaşalı
Olaylı	Heyecan verici	Hoş	Sakin	Durgun	Tekdüze	Can sıkıcı	Karmaşık
Hadiseli	Coşkulu	Çekici			Sıkıcı	Sinirlendirici	Düzensiz
Dinamik	Hayat dolu	Memnuniyet verici			Monoton	Sinir bozucu	Karmakarışık
	Heyecan dolu	Hoşa giden					Kaotik
	Coşturucu	Hoşnut edici					Karışık
	Enerjik	Memnun eden					
	Enerji verici						
	İlgi çekici						

Next, the focus group members are asked to choose the two most appropriate semantic equivalences of each attribute among the ones given in **Table 3** through online tables prepared as google sheets. Then based on the most frequently written two equivalents in this round (**Table 4**), each participant is asked to rate, which one to be used as the primary translation and which one is the alternative semantic equivalence in a second online round. The most approved equivalence for the primary translation is chosen as the Turkish attribute and the most approved equivalence for the alternative translation is determined to be investigated by further analysis if needed. However, there were ties (some of these words got the same score) in the equivalence for the primary translation (monoton/tekdüze). Thus, the list was also shared independently with the two Turkish linguistic experts, who have advanced/mastery level of language proficiency in English, and asked to choose the most suitable equivalence of each English attribute as a primary translation. The feedbacks from the linguistic experts were collected separately and compared. Both of them selected the same words as the focus group's primary equivalence except both preferred 'tekdüze' to 'monoton', and 'keyifli' to 'hoşa giden/hoş'. Therefore, 'tekdüze' has been chosen for the equivalence of the primary translation for 'monotonous', and respectively 'keyifli' for 'pleasant'. The results of this stage is presented in **Table 4**. Finally, Turkish translation model is prepared according to the final decision for each attribute's translation as a comparison to 2D model presented in ISO/TS 12913-3:2019 (**Fig.3**). Although in the literature, many languages have provided more than one word for the translation of the soundscape perceptual attributes, the priority is given to conclude translations with one translated attribute choice according to ISO and SATP suggestions. Therefore, the initial focus has been to decide on one word, which would best fit to describe a sound environment.

Table 4
Results of the 1st phase; linguistic translation.

Attributes	Primary equivalents	Approval percentage	Alternative equivalents	Approval percentage	Final decision
Eventful	Hareketli	100	Dinamik	62.5	Hareketli
Vibrant	Coşkulu	75.0	Heyecan verici	62.5	Coşkulu
Pleasant	Keyifli	75.0	Hoşa giden/Hoş	100.0	Keyifli
Calm	Sakin	100.0	Dingin	62.5	Sakin
Uneventful	Durağan	75.0	Durgun	75.0	Durağan
Monotonous	Monoton/Tekdüze	75.0	Sıkıcı	100.0	Tekdüze
Annoying	Rahatsız edici	87.5	Hoşa gitmeyen	62.5	Rahatsız edici
Chaotic	Karmaşık	75.0	Kargaşalı/Kaotik	75.0	Karmaşık

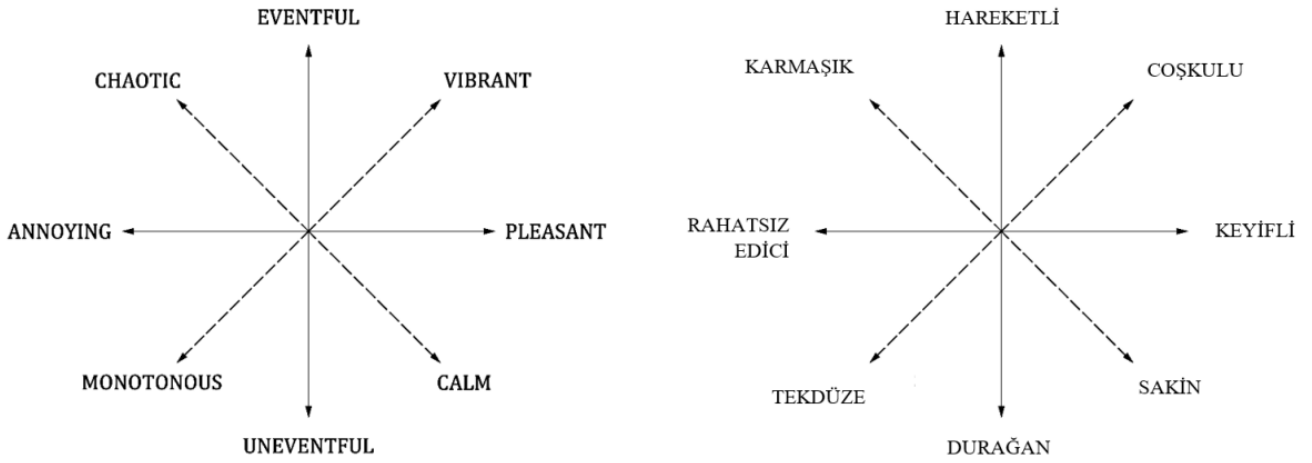

Fig. 3. Two-dimensional model for English and Turkish attributes adapted from [11], ISO/TS 12913-2:2018 [2], and ISO/TS 12913-3:2019 [3].

3.2. Results of the Second Phase Analysis

To investigate the reliability of both Turkish and English survey data, ICC estimates and their 95% confidence intervals are calculated based on a mean-rating agreement (k=10 participants) and reported in **Table 5** for each attribute. It is seen that the agreement between the raters is very high, so the translated scale is reliable. Besides, note that while the ICC estimate for the "vibrant" is relatively lower than the rest, its Turkish equivalence "coşkulu" has higher estimate as the other attributes. This can be due to culture difference or the native Turkish participant might be less familiar with this word's use in soundscape. Whatever the reason is, this might yield a difference between the Turkish and English scores of "vibrant".

Table 5
ICC estimates and their 95% confidence intervals for 10 raters (k=10) measuring the inter-rater agreement levels.

English attributes	ICC	95% confidence interval	Turkish attributes	ICC	95% confidence interval
Eventful	0.923	(0.904, 0.939)	Hareketli	0.908	(0.887, 0.926)
Vibrant	0.740	(0.682, 0.788)	Coşkulu	0.925	(0.907, 0.940)
Pleasant	0.950	(0.938, 0.960)	Keyifli	0.950	(0.933, 0.962)
Calm	0.959	(0.945, 0.969)	Sakin	0.958	(0.944, 0.969)
Uneventful	0.911	(0.890, 0929)	Durağan	0.920	(0.898, 0.937)
Monotonous	0.928	(0.911, 0.942)	Tekdüze	0.906	(0.877, 0.927)
Annoying	0.941	(0.924, 0.954)	Rahatsız Edici	0.958	(0.949, 0.966)
Chaotic	0.934	(0.912, 0.949)	Karmaşık	0.898	(0.873, 0.919)

To evaluate the concept equivalence of the translated attributes, the differences between the scores given for the Turkish and English attributes (as Turkish grade – English grade) are examined. If the raters understand the same concept as the original one, they should give similar scores to the corresponding attributes in each language. The summary statistics for the differences in scores are given in **Table 6**. It is seen that the highest difference is in the translation of 'vibrant', and the score given to it is higher than its corresponding Turkish attribute 'coşkulu' on the average. Considering the relatively low absolute-agreement between the raters for the attribute 'vibrant', this is an expected result. Although the variability of the scores for all attributes is close to each other, when the minimum and maximum values are examined, it is remarkable that there are extreme values such as -100 ('chaotic', 'vibrant', 'uneventful'), 99 ('annoying'). Even these values could be evaluated as outliers; it is not appropriate to exclude them from the analysis since they contain valuable information for the validation. Considering whole summary statistics, the most inconsistency is in the 'vibrant/coşkulu' pairing.

Table 6
Summary statistics on the differences between the scores given to Turkish and English attribute pairings.

	N	Mean	Median	Standard Deviation	Min	Max	Q1	Q3
Eventful/Hareketli	270	-3.06	-2.50	25.22	-68.00	75.00	-15.75	9.00
Vibrant/Coşkulu	270	-16.62	-11.50	25.53	-100.00	45.00	-30.00	0.00
Pleasant/Keyifli	270	-1.38	-0.50	19.48	-60.00	66.00	-13.00	10.00
Calm/Sakin	270	0.52	0.00	19.47	-55.00	78.00	-10.00	10.00
Uneventful/Durağan	270	2.23	3.00	27.14	-100.00	73.00	-8.00	20.00
Monotonous/Tekdüze	270	0.83	0.00	24.50	-87.00	89.00	-10.00	13.75
Annoying/Rahatsız Edici	270	0.39	0.00	19.57	-88.00	99.00	-10.00	10.00
Chaotic/Karmaşık	270	-3.40	-0.50	25.04	-100.00	70.00	-15.00	8.750

When the Spearman correlation coefficients between the Turkish and English scores calculated without removing the outliers are examined (**Table 7**), the value for the pairing 'vibrant/coşkulu' is seen to be below 0.50 as in line with the previous results given in **Table 6**. Correlation is slightly higher than 0.50 for the pairings 'eventful/hareketli' and 'uneventful/durağan' and around 0.60 for the pairings 'chaotic/karmaşık' and 'monotonous/tekdüze'. In other attributes, correlation is found to be above 0.70. This finding is in line with the other studies. For example, the difference in translation occurs for the attributes 'eventful' and 'uneventful' in the study [17] where it is associated with the fact that there is no corresponding word in the target language, Greek. The difficulty in translation of these two attributes is also faced in the translation process of Thai [17] and French [10]. In addition to these attributes 'chaotic' is also found to be differently rated for Indonesian [16].

Table 7
Spearman correlation coefficients between the scores given to Turkish and English pairings.

Attribute	Correlation Coefficient
Eventful/Hareketli	0.515
Vibrant/Coşkulu	0.440
Pleasant/Keyifli	0.781
Calm/Sakin	0.814
Uneventful/Durağan	0.559
Monotonous/Tekdüze	0.636
Annoying/Rahatsız Edici	0.794
Chaotic/Karmaşık	0.620

To investigate the reason for low correlations, participants are asked for their experience in evaluating the scenarios. Important feedbacks regarding some of the sequences such as 'more tiring' and 'difficult to evaluate than the others' are obtained. This is interpreted as the values given to a calm scenario listened after an annoying environment can be higher than the one listened after a pleasant environment. Similar sequence effect can be thought for the other scenarios' grades as well. In [55], it is examined whether the meaning that individuals attribute to the concept of tranquility has an effect on the perception of the sound environment, and found that those who associate calmness with nature sounds become more aware of mechanical sounds. Therefore, it is concluded that in the process of perceiving the sound in an environment, people notice the sounds they do not expect to hear more, and these sounds attract their attention more. In this context, it is decided that additional analyzes are needed to determine how the sequence in the audio recordings (scenarios) containing different types of sound sources affected the perception of the participants in the study and how this is reflected in the scoring given to the attributes.

When the Kruskal-Wallis test is applied on all scores, the sequence is found to be statistically significant (p-value = 0.021). Thus, a second study is conducted as described at Section 2.2. The ICC estimates based on a mean absolute agreement (k=8; 4 participants with 2 sequence evaluation, once in Turkish then once in English) with 2-way random-effects model and the corresponding 95% confidence intervals of this second study are given in **Table 8**, and it is seen that the reliability is still high. Since it can be expected to have relation between the ratings of the same participant for the two sequences, which yields violation of independence assumption in the calculation of ICC estimates, the analysis is repeated by taking the average grade for the two sequences of each participant. It is observed that the new ICC estimates (k=4) are still high (between 0.849 and 0.954) except for the 'vibrant' attribute, although the estimates decrease due to the nature of ICC calculation as expected. The ICC estimate for 'vibrant' decreases to 0.682, while it is 0.885 for its Turkish equivalence, 'coşkulu'.

Table 8
ICC estimates and their 95% confidence intervals with k=8 measuring the inter-rater agreement levels.

English attributes	ICC	95% confidence interval	Turkish attributes	ICC	95% confidence interval
Eventful	0.930	(0.877, 0.964)	Hareketli	0.897	(0.821, 0.947)
Vibrant	0.820	(0.678, 0.909)	Coşkulu	0.928	(0.877, 0.963)
Pleasant	0.953	(0.915, 0.976)	Keyifli	0.954	(0.912, 0.977)
Calm	0.941	(0.890, 0.971)	Sakin	0.941	(0.890, 0.971)
Uneventful	0.932	(0.881, 0.965)	Durağan	0.898	(0.826, 0.947)
Monotonous	0.947	(0.910, 0.973)	Tekdüze	0.925	(0.869, 0.962)
Annoying	0.970	(0.944, 0.985)	Rahatsız Edici	0.970	(0.947, 0.985)
Chaotic	0.942	(0.897, 0.970)	Karmaşık	0.901	(0.821, 0.950)

The Spearman correlation coefficients between the Turkish and English scores as well as the summary statistics for the differences between them are given in **Table 9**. When compared to the values given in **Table 6**, the differences as well as the outliers decrease as expected and when the correlation coefficients are compared with the ones given in **Table 7**, an increase can be observed, which justifies the effect of the sequencing on the scores. The only exception is in the pairing 'chaotic/karmaşık' whose correlation coefficient falls under 0.50.

Table 9
Spearman correlation coefficients and summary statistics for the differences between Turkish and English scores given by the 4 raters evaluating each scenario in 2 different sequences.

	N	Mean	Median	Standard Deviation	Min	Max	Correlation Coefficients
Eventful/Hareketli	216	-0.53	-1.00	24.51	-55	81	0.578
Vibrant/Coşkulu	216	-12.96	-10.00	19.81	-71	40	0.719
Pleasant/Keyifli	216	-2.03	0.00	16.55	-70	55	0.869
Calm/Sakin	216	2.63	0.00	17.76	-45	90	0.850
Uneventful/Durağan	216	-0.47	0.00	27.94	-85	70	0.570
Monotonous/Tekdüze	216	-2.07	0.00	19.21	-64	50	0.789
Annoying/Rahatsız Edici	216	0.62	0.00	16.67	-80	65	0.862
Chaotic/Karmaşık	216	-4.64	0.00	31.70	-100	70	0.428

The analyses are repeated by taking the average of each participant's grades for the two sequences as before, and the results are given in **Table 10**. Although the correlation values increased as expected in general, equivalent results are obtained with the previous analysis.

Table 10
Spearman correlation coefficients and summary statistics for the differences between the averaged Turkish and English scores given by the 4 raters.

	N	Mean	Median	Standard Deviation	Min	Max	Correlation Coefficients
Eventful/Hareketli	108	-0.53	-1.25	19.45	-47.5	68.0	0.680
Vibrant/Coşkulu	108	-12.96	-10.00	15.17	-67.5	17.5	0.810
Pleasant/Keyifli	108	-2.03	-2.50	11.20	-40.0	42.5	0.934
Calm/Sakin	108	2.63	0.00	13.34	-30.0	55.0	0.910
Uneventful/Durağan	108	-0.47	1.50	23.99	-82.0	47.5	0.631
Monotonous/Tekdüze	108	-2.07	-2.50	13.06	-36.5	32.5	0.891
Annoying/Rahatsız Edici	108	0.62	0.00	12.10	-45.0	37.5	0.908
Chaotic/Karmaşık	108	-4.64	0.00	27.96	-100.0	55.5	0.505

When the studies related to the translation of the English attributes to other languages are examined, it is seen that the results are parallel to our findings. In the validation study for French, Korean and Swedish languages [8], it is revealed that while there is no big difference in the adjectives describing the 'pleasantness' dimension, there is a difference in the adjectives describing the 'eventfulness' dimension. In addition, the adjective 'chaotic' could not be translated into Korean and this word is replaced with the word 'noisy'. In the study by Sudarsono et. al. [16], within the scope of the translation of attributes used in soundscape studies from English to Indonesian, it is found that the consistency is lower especially in the attributes 'chaotic', 'eventful' and 'uneventful'. In addition, it is reported that there were difficulties during focus group study process in determining a single word that can represent attributes that are a combination of two emotions (e.g. 'vibrant', which is a combination of 'eventful' and 'pleasant', or 'chaotic', which is a combination of 'annoying' and 'eventful'.), which are placed in the diagonal directions in two dimensional model [3]. Although the data gathering method and the statistical analysis applied in [16] differs from our study, the findings are in line with ours, namely, there is inconsistency in the translations of the attribute 'chaotic', thereby its exact equivalent is not easy to determine.

Although the aim of this round was to investigate the effect of the sequence on the differences between Turkish and English scores, the results of the first and second round for these four participants can be used to evaluate the test-retest reliability if sequence effect is ignored. For this purpose, ICC statistic based on a mean-rating, absolute agreement, 2-way fixed effects model which is again calculated as in equation (2) [46] as well as the summary statistics for the differences between the scores of these two rounds for the translated Turkish scale are given for each participant in **Table 11**.

Table 11
Summary statistics for the differences between the Turkish scores of the two rounds and the corresponding ICC values for test-retest reliability.

Attribute	Participant	Mean	Median	Standard deviation	Min	Max	ICC
Eventful/ Hareketli	1	12.04	5.00	22.76	-20	65	0.819
	2	-0.89	0.00	14.78	-40	21	0.809
	3	7.03	10.00	14.23	-30	25	0.900
	4	16.30	20.00	20.02	-20	50	0.863
	Overall	8.62	5.00	19.15	-40	65	
Vibrant/ Coşkulu	1	6.11	0.00	13.75	-10	45	0.945
	2	0.74	-1.00	19.11	-34	41	0.761
	3	-2.15	0.00	11.63	-35	19	0.900
	4	-0.15	0.00	19.02	-40	40	0.860
	Overall	1.14	0.00	16.28	-40	45	
Pleasant/ Keyifli	1	-2.59	0.00	12.04	-35	20	0.972
	2	-3.15	-2.00	17.36	-44	23	0.886
	3	-0.82	0.00	11.67	-30	15	0.935
	4	-1.67	0.00	21.79	-50	40	0.857
	Overall	2.06	0.00	16.05	-50	40	
Calm/ Sakin	1	-0.70	0.00	11.86	-30	25	0.973
	2	-8.70	-12.00	14.04	-33	31	0.917
	3	4.26	0.00	14.85	-25	40	0.814
	4	-6.96	0.00	21.66	-45	50	0.898
	Overall	3.03	0.00	16.63	-45	50	
Uneventful/ Durağan	1	-0.19	0.00	16.02	-35	35	0.944
	2	-6.07	-6.00	16.58	-37	30	0.777
	3	2.22	0.00	21.94	-45	69	0.703
	4	-3.56	0.00	21.12	-50	50	0.899
	Overall	1.90	0.00	19.09	-50	69	
Monotonous/ Tekdüze	1	-3.48	0.00	16.82	-50	30	0.925
	2	-5.96	-5.00	16.21	-41	31	0.829
	3	-1.48	-5.00	17.91	-30	30	0.869
	4	-1.04	0.00	22.57	-48	50	0.891
	Overall	2.99	-2.50	18.39	-50	50	
Annoying/ Rahatsız Edici	1	5.00	5.00	11.69	-20	40	0.970
	2	-0.19	-3.00	19.62	-35	47	0.871
	3	-4.26	0.00	14.02	-35	28	0.955
	4	11.00	5.00	16.74	-27	50	0.951
	Overall	2.89	0.00	16.60	-35	50	
Chaotic/ Karmaşık	1	6.63	0.00	14.94	-10	50	0.921
	2	-12.19	-13.00	14.03	-38	25	0.849
	3	-3.89	0.00	19.38	-50	50	0.797
	4	17.96	18.00	23.75	-45	70	0.824
	Overall	2.13	0.00	21.44	-50	70	

Although it is seen from **Table 11** that the agreement between the rounds is high meaning that the translated scale is consistent over time, there are differences exceeding 10 points for some attributes of some participants which can be attributes to the sequencing effect. Thus, test-retest reliability has to be investigated by a more proper design eliminating sequencing effect in the future.

Finally, the data is reduced to 2 dimensions as Pleasantness (P) and Eventfulness (E) as [3]:

$$P = (p - a) + cos45°(ca - ch) + cos45°(v - m),$$
$$E = (e - u) + cos45°(ch - ca) + cos45°(v - m), \quad (7)$$

where a is annoying; ca is calm; ch is chaotic; e is eventful; m is monotonous; p is pleasant; u is uneventful; v is vibrant so that the 45-degree angles of the "vibrant"/"chaotic"/"monotonous"/ "calm" axes with respect to the "Pleasantness"/"Eventfulness" axes as shown in the theoretical models in Figure 3 is preserved.

The Spearman correlation between the Turkish and English scores of these two dimensions with the 10 participants is 0.849 and 0.761 for P and E respectively while they become 0.887 and 0.790, respectively with 4 participants where the sequence effect is eliminated. Thus, it can be stated that the relatively low correlation coefficient of the attribute 'chaotic' do not seem to pose a problem and could be used in Turkish studies.

4. Conclusion

Turkish is one of the most commonly and widely used Turkic Language spoken by more than 250-300 million people in the world. Thus, it is important to translate the soundscape perceptual attributes to Turkish for providing a common validated translation for the studies carried out in Turkey and other Turkish speaking regions or communities. The main objective of this is to figure out

the Turkish equivalences of the English attributes that are reproducible and the respondents interpret and respond in the same manner as they do to corresponding English ones.

In this context, the study is conducted in two phases. The focus group studies are conducted as the first phase for Turkish translation of attributes to answer the first research question. Then, with the Turkish attributes obtained from the first phase, the second phase of the study is conducted for the context validation at conceptual level and reliability analysis through bilingual listening tests in relation with the second and third research questions, respectively.

The attributes are translated to Turkish as Hareketli (eventful), Coşkulu (vibrant), Keyifli (pleasant), Sakin (calm), Durağan (uneventful), Tekdüze (monotonous), Rahatsız edici (annoying), and Karmaşık (Chaotic). As a result of the bilingual tests, it is found that the respondents interpret and respond in the same manner as they do to English attributes. However, the most problematic translation is in the attribute 'vibrant' in terms of the difference between average scores given to English and Turkish scales and in the attribute 'chaotic' in terms of the correlation between the scores given to English and Turkish scales as in line with the literature. Despite this fact, all Turkish attributes are found to be reproducible according to the high inter-rater reliability. Besides, a high correlation is seen between the Turkish and English scores reduced to 2 dimensions as Pleasantness (P) and Eventfulness (E) preserving 45-degree angles as in ISO/TS 12913-3:2019 document [3]. Thus, it can be stated that the relatively low correlation coefficient of the attribute "chaotic" and the relatively high difference between the average scores obtained from English and Turkish scale for the attribute "vibrant" would not pose a problem in 2-dimensional analysis and could be used in Turkish studies.

In addition to the translated and validated Turkish attributes, the methods applied in the different phases of the study and the statistical tests have led to interesting findings that points to further analysis. One such finding is on the statistical significance that has been identified on the scenario sequence in the binaural listening tests. According to the feedbacks of the participants taken the bilingual tests, the sequence effect on the attributes in the 'perceived affective quality' scale is investigated and found that it is effective. More specifically, as an example, a score for calmness will be different for the same scenario listened after an annoying scenario than the one listened after a pleasant one.

Although the results obtained are promising, there are limitations of the study. Due to the pandemic conditions, the participants are selected among university students who were in Ankara. Thus, there is a need to repeat the analysis by covering participants from different regions of Turkey as well as Turkish Republic of Northern Cyprus. The design of the study must consider the sequence effect in a proper way. Second, during listening tests, first Turkish, then English scales are used but not the reverse. Thus, it could be interesting to figure out whether this influenced the results. Next, test-retest reliability must be investigated by controlling the sequence effect. Finally, in this study although we have prepared alternative translated equivalences for each attribute, we focused on the performance of the selected primary equivalences due to the mentioned limitations and in line with the ISO's one-word preference. Therefore, it would be beneficial to test the alternatives especially for "vibrant" and "chaotic". This work will continue with the application of Turkish scale to more participants and analyze the results to see if there is a need for the revision of the translation of the attributes.

5. Acknowledgement

This study is conducted as the Turkish partner of the 'Soundscape Attribute Translation Project' (SATP) coordinated at University College London (UCL) and funded by the European Research Council (ERC) Advanced Grant (no. 740696) on 'Soundscape Indices' (SSID) and supported by the UCL Cities Partnership Program 2021-2022. Besides, it is funded by Çankaya University as a 'Scientific Research Project' entitled; "Turkish translation, validation and testing of soundscape evaluation scale" (Project no. MF.20.016 [D9]). The authors would like to acknowledge all soundscape experts and language experts who have participated in the focus group studies, Çankaya University students and assistants who have participated in the listening tests, and reviewers for their valuable comments.

References

[1] ISO, Acoustics-Soundscape-Part 1 : Definition and conceptual framework BS ISO 12913-1:2014 BSI, Genève, Switzerland, 2014.
[2] ISO, Acoustics-Soundscape-Part 2: Data collection and reporting requirements ISO / TS 12913 - 2 : 2018 Genève, Switzerland, 2018.
[3] ISO, Acoustics-Soundscape-Part 3: Data analysis PD ISO/TS 12913-3:2019, Geneva, Switzerland, 2019.
[4] J. Kang, Comparison of speech intelligibility between English and Chinese, The Journal of the Acoustical Society of America, 103 (1998) 1213-1216.
[5] F. Aletta, T. Oberman, Ö. Axelsson, H. Xie, Y. Zhang, S.-K. Lau, S.-K. Tang, K. Jambrošić, B.D. Coensel, K.v.d. Bosch, P. Aumond, C. Guastavino, C. Lavandier, A. Fiebig, B. Schulte-Fortkamp, J. Sarwono, A. Sudarsono, A. Astolfi, K. Nagahata, J.-Y. Jeon, H.-I. Jo, J. Chieng, W.-S. Gan, J.-Y. Hong, B. Lam, Z.-T. Ong, P. Kogan, E.S. Silva, J.V. Manzano, P.N.D. Yörükoğlu, T.L. Nguyen, J. Kang, Soundscape assessment: towards a validated translation of perceptual attributes in different languages, Inter-Noise 2020Seul, 2020.
[6] F. Aletta, C. Guattari, L. Evangelisti, F. Asdrubali, T. Oberman, J. Kang, Exploring the compatibility of "Method A" and "Method B" data collection protocols reported in the ISO/TS 12913-2:2018 for urban soundscape via a soundwalk, Applied Acoustics, 155 (2019) 190-203.
[7] P. Delaitre, C. Lavandier, C. Cance, J. Pruvost, What is the Definition for the French Word calme in the European Directive Related to "Quiet Areas"? A Lexicographic Study from the 16th Century Until Today, Acta Acustica united with Acustica, 98 (2012) 734-740.
[8] J.Y. Jeon, J.Y. Hong, C. Lavandier, J. Lafon, Ö. Axelsson, M. Hurtig, A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments, Applied Acoustics, 133 (2018) 107-117.

[9] K. Nagahata, Examination of Soundscape-Quality Protocols in Japanese, Inter-Noise 2019Madrid, 2019.
[10] C. Tarlao, D. Steele, P. Fernandez, C. Guastavino, Comparing soundscape evaluations in French and English across three studies in Montreal, Inter-Noise 2016Hamburg, Germany, 2016.
[11] Axelsson, Nilsson, Berglund, A principal components model of soundscape perception, J Acoust Soc Am, 128 (2010) 2836-2846.
[12] R. Cain, P. Jennings, J. Poxon, The development and application of the emotional dimensions of a soundscape, Applied Acoustics, 74 (2013) 232-239.
[13] F. Aletta, J. Kang, Towards an Urban Vibrancy Model: A Soundscape Approach, Int J Environ Res Public Health, 15 (2018).
[14] S. Antunes, R.L.X.N. Michalski, M.L.d.U. Carvalho, S. Alves, Validated translation into Portuguese of perceptual attributes for soundscape assessment, Euronoise 2021Madeira, Portugal, 2021.
[15] N.M. Papadakis, F. Aletta, J. Kang, T. Oberman, A. Mitchell, G.E. Stavroulakis, Translation and cross-cultural adaptation methodology for soundscape attributes – A study with independent translation groups from English to Greek, Applied Acoustics, 200 (2022).
[16] A.S. Sudarsono, W. Setiasari, S.J. Sarwono, N.P.A. Nitidara, The development of standard perceptual attributes in indonesian for soundscape evaluation: Result from initial study, Journal of Applied Science and Engineering, 25 (2021) 215-222.
[17] K.N. Watcharasupat, S. Jaratjarungkiat, B. Lam, S. Jitwiriyanont, K. Ooi, Z.-T. Ong, N. Pichetpan, K. Akaratham, T. Suthiwan, M. Rojtinnakorn, W.-S. Gan, Quantitative evaluation approach for translation of perceptual soundscape attributes: Initial application to the Thai Language, Applied Acoustics, 200 (2022).
[18] Britannica, T. Editors of Encyclopaedia, Turkish language. Encyclopedia Britannica, Retrieved from 2022, September 15, https://www.britannica.com/topic/Turkish-language, 2022.
[19] Johanson, L.,Turkic languages. Encyclopedia Britannica. Retrieved from 2022, September 30, https://www.britannica.com/topic/Turkic-languages, 2022.
[20] E. Yaman, Bir Ulu Çınarın Dalları Türk Dili ve Lehçeleri, Türk Kültürü, (1994) 378.
[21] İ. Karakuş, Türkçenin Gücü ve Atatürk, Türk Dili Araştırmaları Yıllığı-Belleten, 54 (2009) 155-164.
[22] E. Yaman, Türkçenin güncel söz varlığı (Current Vocabulary of Turkish), Milli Eğitim Dergisi, 45 (2016) 85-91.
[23] A. Özçevik, Z.Y. Can, Işitsel peyzaj kavramı ve kapalı mekanların akustik konfor değerlendirilmesinde kullanılabilirliği, Megaron, 6 (2011) 52-59.
[24] A. Özcevik, Z. Yüksel Can, A field study on the subjective evaluation of soundscape, Acoustics 2012Nantes, France, 2012.
[25] J. Kang, K. Chourmouziadou, K. Sakantamis, B. Wang, Y. Hao, Soundscape of European Cities and Landscapes, EU COST, Oxford, 2013.
[26] I. Kaymaz, User perception of soundscapes of urban parks in Ankara, AESOP2012, 26th Annual congressAnkara, Turkey, 2012.
[27] I. Kaymaz, N. Belkayalı, Peyzaj mimarlığı kapsamında işitsel peyzaj kavramı, Peyzaj Mimarlığı 5. KongresiAdana, 2013, pp. 182-194.
[28] I. Kaymaz, N. Belkayalı, Current Status of Soundscape Research and Practice in Turkey, International Congress on Landscape EcologyAntalya, Turkey, 2014.
[29] S. Bahalı, N. Tamer-Bayazıt, Soundscape research on the Gezi Park – Tunel Square route, Applied Acoustics, 116 (2017) 260-270.
[30] D.Ç. Aydin, S. Yilmaz, Evaluation of Diyarbakir Gazi Street Sound Environment Perception by Soundscape Approach, Middle East Journal of Science, 3 (2017) 15-25.
[31] N. Akbulut Çoban, Çevresel Gürültü Yönetimi Kapsamında Sessiz alanların Belirlenmesine Yönelik Model Önerisi: İşitsel Peyzaj Yaklaşımının Kullanımı (Model Proposal for Identification of Quiet Areas Under Environmental Noise Management:Soundscape Perspective), PhD Dissertation, Akdeniz University, 2021.
[32] G. Akın Güler, Kentsel Alanlarda Sessiz/Sakin İşitsel Peyzajların Belirlenmesine Yönelik Bir Yaklaşım ve Türkiye İçin Bir Model Önerisi (An Approach to Determining Quiet Soundscape In Urban Areas and A Model Proposal for Turkey), PhD Dissertation, Eskişehir Teknik Üniversitesi, 2022.
[33] I. Aspuru, I. Garcia, C. Bartalucci, F. Borchi, M. Carfagni, L. Governi, R. Bellomini, S. Luzzi, H. Wolfert, P. Gaudibert, LIFE+2010 QUADMAP Project: a new methodology to select, analyze and manage Quiet Urban Areas defined by the European Directive 2002/49/EC, Noise Mapping, 3 (2016).
[34] A. Özcevik, Z. Yuksel Can, A laboratory study on the evaluation of soundscape, Soci´et´e Fran¸caise d'Acoustique. Acoustics 2012Nantes, France, 2012.
[35] A. Özçevik Bilen, Z. Yüksel Can, An applied soundscape approach for acoustic evaluation – compatibility with ISO 12913, Applied Acoustics, 180 (2021).
[36] J.M. Fields, R.G. De Jong, T. Gjestland, I.H. Flindell, R.F.S. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felscher-Suhr, R. Schumer, Standardized General-Purpose Noise Reaction Questions for Community Noise Surveys: Research and a Recommendation, Journal of Sound and Vibration, 242 (2001) 641-679.
[37] ISO, ISO/TS 15666:2003 Acoustics - Assessment of noise annoyance by means of social and socio-acoustic surveys, 2003.
[38] ISO, ISO/TS 15666:2021 Acoustics - Assessment of noise annoyance by means of social and socio-acoustic surveys, 2021.
[39] C. Clark, T. Gjestland, L. Lavia, H. Notley, D. Michaud, M. Morinaga, Revising ISO/TS 15666 – the noise annoyance standard, The 13th ICBEN Congress on Noise as a Public Health ProblemStockholm, Sweden, 2021.
[40] R.K. Hambleton, L. Patsula, Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices, Applied Testing Technology Journal, (1999).
[41] B.M. van Widenfelt, P.D. Treffers, E. de Beurs, B.M. Siebelink, E. Koudijs, Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families, Clin Child Fam Psychol Rev, 8 (2005) 135-147.
[42] J. Epstein, R.M. Santo, F. Guillemin, A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus, J Clin Epidemiol, 68 (2015) 435-441.

[43] E.S. Cha, K.H. Kim, J.A. Erlen, Translation of scales in cross-cultural research: issues and techniques, J Adv Nurs, 58 (2007) 386-395.
[44] T. Oberman, A. Mitchell, F. Aletta, Almagro Pastor, J. Antonio, K. Jambrošić, J. Kang, Soundscape Attributes Translation Project (SATP) Dataset (1.2) [Data set], in: Z. 10.5281/zenodo.6914433 (Ed.), 2022.
[45] B. Lam, K. Ooi, Z.-T. Ong, K.N. Watcharasupat, T. Wong, W.-S. Gan, Assessment of a cost-effective headphone calibration procedure for soundscape evaluations, 24th International Congress on AcousticsGyeongju, Korea, 2022.
[46] T.K. Koo, M.Y. Li, A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research, J Chiropr Med, 15 (2016) 155-163.
[47] A. Donner, M. Eliasziw, Sample size requirements for reliability studies, Stat Med, 6 (1987) 441-448.
[48] S.D. Walter, M. Eliasziw, A. Donner, Sample size and optimal designs for reliability studies, Statistics in Medicine, 17 (1998) 101-110.
[49] D.G. Bonett, Sample size requirements for estimating intraclass correlations with desired precision, Stat Med, 21 (2002) 1331-1335.
[50] B. Giraudeau, J.Y. Mary, Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient, Stat Med, 20 (2001) 3205-3214.
[51] G. Doros, R. Lew, Design Based on Intra-Class Correlation Coefficients, American Journal of Biostatistics, 1 (2010) 1-8.
[52] L.B. Mokkink, H. de Vet, S. Diemeer, I. Eekhout, Sample size recommendations for studies on reliability and measurement error: an online application based on simulation studies, Health Services and Outcomes Research Methodology, (2022).
[53] Y. Saito, T. Sozu, C. Hamada, I. Yoshimura, Effective number of subjects and number of raters for inter-rater reliability studies, Stat Med, 25 (2006) 1547-1560.
[54] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychol Bull, 86 (1979) 420-428.
[55] K. Filipan, M. Boes, B. De Coensel, C. Lavandier, P. Delaitre, H. Domitrović, D. Botteldooren, The Personal Viewpoint on the Meaning of Tranquility Affects the Appraisal of the Urban Park Soundscape, Applied Sciences, 7 (2017).