Categorization of urban sound sources: A taxonomical framework based on diegesis and intention

Kivanc Kitapci[1], Dogukan Ozdemir[2]

Department of Interior Architecture, Çankaya University, Ankara, Turkey
[1]kivanckitapci@cankaya.edu.tr
[2]ozdemirdogukan5@gmail.com

Abstract

The sound source composition of the urban sound environment varies depending on the geography and socio-cultural context. Current sound taxonomies in the literature categorize urban sound sources by their source types (e.g., human-made, natural, electromechanical) and respective semantic attributes. This study aims to add another taxonomical layer to the existing urban sound source categorization methods. The additional layer is a recently proposed sound source classification framework (CLIC). The CLIC framework identifies sound sources based on their Diegesis and Intention parametric attributes. The former parametric attribute, diegesis, was derived from film sound design. The geographical and socio-cultural context of the built environment can be considered as its narrative; hence, every event that happens within the functional context can be called diegetic, while the events that do not belong to that specific place can be called non-diegetic. The latter parametric attribute, intention, was derived from product design. One of the prominent sound source categorization methods in product sound design is to group the product emitted sounds as consequential and intentional. Combining these two parametric attributes with the existing taxonomies, the CLIC framework outputs a place-specific design guideline, clearly dictating the actions a sound designer should take. The two parametric attributes group the sound sources under four distinct areas, which dictates the degree of influence of designers on the specific sound source. The four zones are defined as the creation zone, limitation zone, isolation zone, and control zone. Each zone dictates step-by-step sound design instructions for the sound designer. This study consists of two main phases: field recordings and web-based listening tests. The sound sources present in the urban sound environment were identified in the field recordings phase. Later, in the web-based listening tests phase, the identified sound sources were evaluated based on the CLIC framework, and hence, their respective zones on the model were identified. The outcomes of the study propose clear step-by-step design guidelines and present action suggestions for environmental sound designers.

Keywords: soundscapes, sound design, diegesis, sound sources, sound taxonomy.

1 Introduction

Urban soundscapes host numerous sound sources; each of them varies in terms of physical and semantic properties. The geographical and socio-cultural context of a given environment dictates the sound sources found within the urban soundscape. Therefore, urban sounds are not a negligible design element and are also a natural part of urban life. In the literature, it is well-documented that the resident's well-being is directly influenced by the coherence of audial and visual design elements found in the urban environment. Therefore, the aim of the environmental designer (i.e., architects and urban planners) should be to achieve a holistic built environment designed specifically for a given socio-cultural and geographical context. The sound sources found within the urban soundscapes can be considered unique design elements to be carefully created, controlled, manipulated, or placed by the environmental designer. However, every design element should be empirically categorized to be used in the design process of urban environments. To control the sound sources as design elements, three properties should be analyzed. Each of the three sound source properties can be considered as a separate taxonomical layer. The first taxonomical layer is the sound source type, which has been extensively studied in soundscape research. The second taxonomical layer is the contextual, socio-cultural, and geographical coherence of the sound source with the built environment. The third and last taxonomical layer is the degree of control of a sound designer on the sound design element. The current study aims to present a sound design guideline for the environmental sound designers by creating two additional taxonomical layers in combination with the already existing (1) taxonomy on sound source types found in the soundscape literature. The additional layers are based on the two sound source categorization methods (2) *Diegesis*, which measures the contextual, socio-cultural, and geographical coherence of the sound source with built environment (derived from the sound design methods of film and video game disciplines) and (3) *Intention*, which measures the degree of control of a sound designer on the design element (derived from the product design discipline). It is hypothesized that the combination of the sound source type, diegesis, and intention will reveal the influence of sound designers on a sound source in a given environmental context. The complete taxonomical framework is presented in Figure 1.

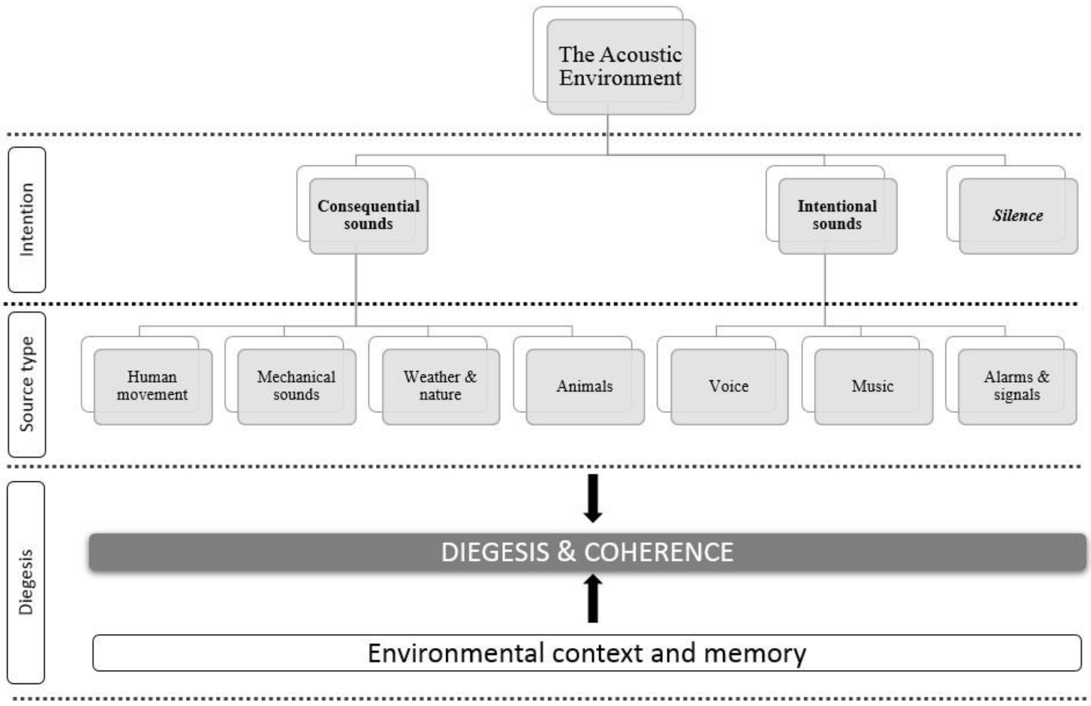

Figure 1 - The proposed taxonomy of sound sources in urban soundscapes.

In recent years, the first taxonomical layer, sound source types, has been substantial research interest in urban environments. There are many tools to be used for the categorization of sound sources; however, the most commonly used method in the soundscape literature is the semantic differential method in combination with factor analysis to reveal the principal dimensions or cluster analysis for identifying groups of similar sounds in terms of the attribute scales [1]. Most of these taxonomical frameworks were based on the studies of Schafer [2]. Environmental sounds were investigated under various categories in different studies found in the literature. Bones *et al.* explored the formation of categories for a set of everyday sounds that are found in the soundscape literature [1]. Moreover, the sound taxonomy of indoor sound environments was also investigated [3,4]. Additionally, Niessen *et al.* [5] linguistically investigated different types of sound source classifications from various disciplines and suggested explicitly stating the type of categorization applied. More recently, International Standards Organization published a taxonomy of the acoustic environment for soundscape studies [6]. In this study, the proposed taxonomical framework is based on the taxonomy of sound sources in soundscapes presented in ISO/TS 12913-2:2018 [6].

The second taxonomical layer, *diegesis*, is based on the diegetic/non-diegetic sound categorization interpreted from the film and video game theory. Diegesis is the narrative Spatio-temporal world described in a film. Anything within that filmic world is called diegetic, and anything outside of the filmic world is non-diegetic, including sound sources [7]. Diegesis is a well-balanced combination of preference, experience, and memory [8]. Similarly, the functional, geographical, and socio-cultural context of a built environment can be considered it's narrative. Every sound event within the functional, geographical, and socio-cultural context can be referred to as diegetic. In contrast, every sound event that happens outside of that context can be referred to as non-diegetic. Environmental sound designers should consider diverse sound sources and environmental factors to produce a proper diegesis for a specific environment. It should be noted that diegesis depends on both the users' and the designer's preferences; hence, the diegesis values of any given sound source should be collected as empirical data as a function of the environmental context.

The last taxonomical layer, *intention*, measures the degree of control of a sound designer on the design element. Fundamentally, sound design is a technical and artistic profession, and it is the art and practice of creating sounds for a variety of needs. It includes all non-compositional aspects of a movie, a play, a musical performance or recording, computer game software, or any other multimedia product [7]. It usually involves performing and editing pre-created or recorded sound, such as sound effects and dialogue, for the environment. Sound design exists in various disciplines, including filmmaking, television production, video game development, theatre, sound recording and reproduction, live performance, sound art, post-production, radio, and musical instrument development. More recently, sound design has gained increased research interest in industrial product design. It is a new multidisciplinary field with origins in acoustic measurements and product sound quality [9]. In one study, semiotics of product sounds were analyzed, and it was suggested that sound designers are engineers of communication [10]. Spence and Zampini also investigated the sounds of an everyday object, but from a multisensory perspective [11]. They found that the sounds of everyday objects influence users' perception of quality and efficiency. In product sound design, the product-generated sounds are categorized as consequential and intentional sounds [12–14]. The routine operation of a specific product produces consequential sounds. Intentional sounds, on the other hand, are created and controlled by sound designers. Therefore, determining the exact intention value of a given sound source would reveal the possible degree of impact an environmental designer has over a specific sound source.

This study aims to create a holistic design guideline for the environmental sound designer based on an extended sound source taxonomy. The proposed framework incorporates the coherence of a sound source with the environmental context (diegesis) and the degree of impact a designer has over a sound source (intention). Within the scope of the study, the framework was also applied to the sound environment of Kuğulu Park, Ankara, Turkey. The Environmental Sound Design Framework (CLIC), the experimental methods of the listening tests, and the results of the study are explained in the following sections.

2 Environmental sound design framework (CLIC)

The Environmental Sound Design Framework (Create, Limit, Isolate, and Control) was developed based on the two new sound source categorization methods derived from product sound design and film sound design theories: (1) Diegesis and (2) Intention. As previously mentioned, diegesis is the narrative world presented in a film or videogame. It can be suggested that every sound event is diegetic in the built environment. There is no audience to whom the non-diegetic sounds need to be designed and presented. However, it can also be argued that the context of an environment is strictly tied to its user. Therefore, if the subject is unfamiliar with the functional, geographical, or socio-cultural context of an urban sound environment, the unfamiliar sound events perceived by the residents should be categorized as non-diegetic sound events. Diegesis creates the first axis of the framework. The second sound source categorization method, intention, guide the designer in determining the level of control over sound sources. If a sound source is consequential, the designer will have limited control over it. If the sound source is classified as intentional, it may be possible for the designer to design or select the sound source in its entirety. Intention creates the second axis of the framework. The complete framework is shown in Figure 2.

A sound source can be located at any point on this model. To locate a specific sound source on the framework (Figure 2), intention and diegesis values should be identified. Table 1 shows the corresponding intention values for each action a sound designer can take on a 7-point scale. Furthermore, the diegesis values are determined by collecting empirical data from online listening tests. The methods used for listening tests are presented in the next section. It should be noted that when the same sound source is evaluated for different types of spaces of various functions, its position in the model will also change on the diegesis axis as the context changes. Therefore, the proposed framework differs from other categorization methods found in the literature. The two axes in the framework divide the chart into four distinct areas, which dictates the degree of influence of designers on the specific sound source. The four zones are called the creation zone, limitation zone, isolation zone, and control zone.

Table 1 - The corresponding intention values for each action a sound designer can take.

Category	Action of the designer	Intention value
Intentional	Design from scratch	+3
	Complete control by the selection of the product	+2
	Partial control by planning	+1
Neutral	Silence	0
Consequential	Partial control by the selection of the product or isolation	-1
	Full isolation	-2
	Uncontrollable	-3

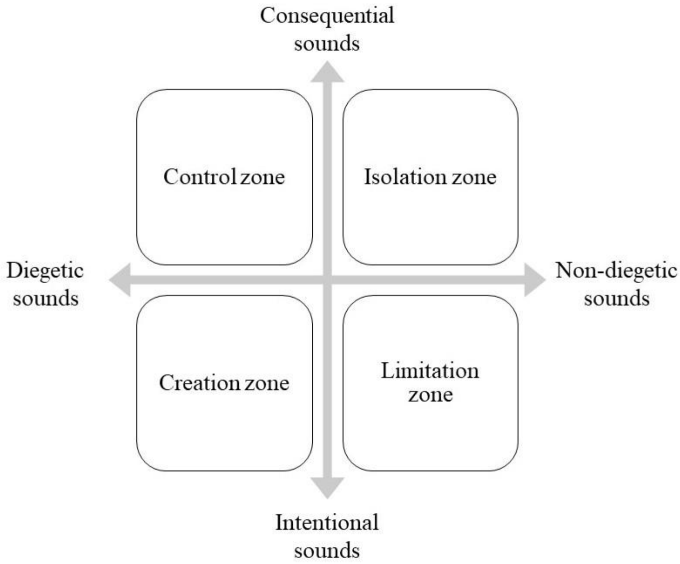

Figure 2 - The context-dependent CLIC (Create, Limit, Isolate, and Control) framework.

The first quarter in the chart is called the limitation zone. These sound sources are non-diegetic and intentional; therefore, the intensity of the sound sources should be limited by the environmental sound designer. Since these sound sources can carry high levels of important information, complete isolation is not suggested. The second quarter in the chart is the isolation zone. To locate a sound source in the isolation zone, it must be consequential and non-diegetic. Therefore, the sound sources located in this zone should be completely isolated by the sound designer. The third quarter in the graph is called the creation zone. For a sound source to be in this zone, it must be diegetic and intentional. In general, these sounds carry essential information. Digital or mechanical sound sources such as elevators, telephones, doorbells, and similar are usually located in this zone. These sound sources can be designed entirely (or selected) by the sound designer in accordance with the context of the space. The last quarter in the model is the control zone. A sound source is located within this area if it is diegetic and consequential. The sound designer has a limited influence on these sound sources. Since the sound sources in this zone are within the context of the space, they do not need to be isolated or limited. However, the sound source can be controlled within certain limits. For instance, footsteps in each space cannot be designed from scratch; however, a designer has limited and indirect control over the footsteps by changing the floor material.

The proposed sound design framework aims to guide environmental sound designers throughout the design process by dictating the degree and type of control they have over the sound sources, based on the sound source and context of the space. As a case study, the sound design framework was tested on the sound environment of Kuğulu Park, Ankara, Turkey. In the next section, the methods and results of the case study are presented.

3 Case study: Kuğulu Park, Tunalı Hilmi Avenue (Ankara, Turkey)

Within the scope of the study, sound sources present in Kuğulu Park were analyzed by using the proposed context dependent CLIC sound design framework. Kuğulu Park is located in Çankaya, Ankara (Turkey). Kuğulu Park is located in the Kavaklıdere district of Ankara. It is a First Degree Natural Protected Area and has an area of 7743 m². The Park was opened in 1958. Although Kuğulu Park is in the city center, where the population density is very high, it attracts more users compared to other parks. In the population census conducted in 2017, it was determined that 6,675 people lived in the Kavaklıdere district where the park is located.

3.1 Methods

In the first phase of the study, the environmental sounds were recorded during the working hours of the selected environments and when the user population is exceptionally high. The recordings were collected from three different recording positions to access a wide variety of sound sources. Each sound recording session was fifteen minutes long. The environmental sounds were recorded in 96kHz, 24-bit quality with Zoom SGH super-cardioid shotgun type microphone (individual sound source recordings) and Zoom XYH-6 unidirectional X/Y type microphone set 120 degrees recording angle (environmental sound recordings) connected to Zoom H6 hand-held audio recorder. The hand-held sound recorder is fixed on a height-adjustable tripod at the height of 140 cm. The collected sound recordings were monitored and analyzed with the Beyerdynamic DT770 Pro (80 ohms) closed cap isolated headphones connected to a Focusrite 2i2 third-generation audio interface. Each significant and identifiable sound source was time-coded and listed separately (Table 3).

The second phase of the study consisted of a web-based listening test. The aim of the listening tests was to determine the diegesis values for each of the previously identified sound sources. The listening tests were created by using the Web Audio Evaluation Tool [15]. A total of 32 subjects participated in the listening tests. Sound recordings were collected from 3 different zones (in the park, main street and near the pool). In the sound recordings collected from the park, human-induced sounds and nature sounds were more prominent. In the sound recordings collected from the main street, motorized transport sounds and electromechanical sounds were in the foreground. Finally, in the recording taken from the poolside, nature sounds, animal sounds, and human sounds were dominant. These three selected zones have been chosen because of the high human traffic, their frequent use in daily life, and the fact that there is high variability in the types of sound sources. As a result of the analysis of the collected sound recordings, 12 different sound sources were obtained.

Samples of sound sources to be used for listening tests were then selected from the International Affective Digitized Sounds-Extended (IADS-E) [16] database to be approximately 6 seconds long. Twelve auditory stimuli were selected from the International Affective Digitized Sounds - Extended database (IADS-E) [16] for the listening tests. The audio samples were selected according to the sound source lists created by the analysis of on-site audio recordings. The audio samples were sampled at 44.1 kHz and were approximately six seconds long. The participants were expected to listen to the sound events and evaluate each of them on a scale of 0-100 according to their preferences. The results were normalized by Modulus equalization to bring into standard the scale that participants used. The principle of this equalization is that all participants must have evaluated the same set of stimuli. It requires that all participants have evaluated the stimuli an equal number of times and the absence of 0 ratings in evaluating [17]. Their ratings are multiplied by a fixed, constant multiplier to achieve a geometric mean equal to the geometric mean of the group data. The formula for applying modulus equalization to the individual scores is given below (Eq.1).

$$S_{Eq} = S_I \left(\frac{G_{mean}}{S_{mean}}\right) \qquad \text{Eq.1}$$

where, S_{eq} is the equalized score, S_I is the raw (initial) score, G_{mean} is the geometric mean of the group data, and S_{mean} is the geometric mean of each participant's responses. The normalized diegesis scores were presented in Table 3.

The interaction between the sound sources were statistically analyzed by using the Principal Component Analysis (PCA) in IBM Statistical Package for Social Sciences (SPSS) software version 21. Varimax rotation was applied to the collected experimental data. The sampling adequacy of the data was tested by the

Keiser-Meyer-Olkin measure of sampling adequacy (KMO) and the suitability of data for dimension reduction was tested with Bartlett's test of sphericity [18]. The results of the PCA are presented in the results section.

3.2 Results

The complete list of identified sound sources in the sound environment and their corresponding intention and diegesis values are presented in Table 3. It should be noted that the diegesis values highly depend on the designer, customer profile, client profile, socio-cultural context, and geography. Therefore, these results solely apply to the environment under investigation. The analysis of audio recordings revealed that the majority of sound sources identified in the sound environment are motorized transport sounds. A total of five motorized transport were identified, two of them being intentional sound sources. Besides the motorized transport sounds, human movement sounds (i.e., footsteps), electromechanical sounds (i.e., construction and cell phone), social/communal sounds (i.e., car alarm), nature sounds (i.e., fountain), and animal sounds (i.e., birds and swans) were also present.

The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, with a KMO = 0.67, which is above the acceptable limit of 0.5. Additionally, Bartlett's test of sphericity revealed that the data is suitable for PCA or factor analysis (p = 0.00). The PCA was run to obtain eigenvalues for each component in the data. A total of four components had eigenvalues over Kaiser's criterion of 1 and the combination of these four components explained 72.7% of the variance. The largest component was mostly consisted of alarming sounds (car brake, car horn, car alarm, construction), which explained 34.5% of the total variance. The second component was consisted of natural and human sounds (birds, footsteps, and fountain), which explained 15.93% of the total variance. Car and motorcycle engine sounds were in the third component, explaining 12.99 % of the total variance. Lastly, cell-phone sounds were in the fourth component, explaining 9.35% of the total variance. After analyzing of the four components, Component 1 was labelled "alarming sounds". Component 2 was labelled "natural sounds", Component 3 was labelled "mechanical sounds", and lastly, Component 4 was labelled "digital sounds".

Table 3 - The list of sound sources identified in the recording of the case study and their corresponding intention and diegesis values.

Source type	**Sub-category**	**Sound Sources**	**Type**	**Intention**	**Diegesis**
Sounds generated by human activity/facility	Motorized transport	Motorcycle	Consequential	-2	103.11
		Car brake	Consequential	-1	25.91
		Car horn	Intentional	+1	1.99
		Car engine	Consequential	-1	512.20
		Ambulance	Intentional	+1	33.78
	Human movement	Footsteps	Consequential	-1	250.31
	Electromechanical	Construction	Consequential	-2	7.57
		Cell phone	Intentional	+2	148.54
	Social & communal	Car alarm	Intentional	+1	8.23
Sounds not generated by human activity	Nature	Fountain	Intentional	+3	2949.44
	Animals	Birds	Intentional	+2	4202.18
		Swan	Intentional	+2	229.91

In the last phase of the analysis, the sound sources identified in the previous phase were placed on the corresponding locations in the CLIC framework, based on their intention and diegesis values presented in Table 2. The finalized design guideline is also presented in Figure 3. The analysis of the sound sources located in the creation zone (diegetic and intentional) of design guideline revealed that the sound sources related to Component 2 (natural sounds) of the PCA, which are bird, swan, and water sounds (pool/fountain sounds), are highly expected and anticipated in the area. The environmental sound designer of the area should create a suitable environment that mimics a natural habitat, which should host animal life to potentially increase the restorative properties of the environment. Additionally, the sound sources located in the creation zone revealed an interesting finding, as well: the users preferred the sound sources related to Component 4 (digital sounds) of the PCA, (i.e., cell phone sounds) in the area. It is possible to argue that the users of an urban restorative environment do not prefer complete isolation from the urban community and modern lifestyle. Therefore, a careful balance between the modern lifestyle and restorative properties of a given environment might be an important issue in the design process.

Furthermore, the sound sources related to Component 1 (alarming sounds) of the PCA, which are car horns, car alarms, and ambulance, sounds should be limited (partially isolated) by careful regional planning. The environmental sound designer should also consider controlling the sound sources related to Component 3 (mechanical sounds) of the PCA, which are motorcycle and car engine sounds), as well as the footsteps. Lastly, it is highly advised to completely isolate car brakes and construction sounds, as these sounds were located in the isolation zone of the graph.

Table 4 presents the sound sources categorized depending on the extracted components from the Principal Component Analysis (PCA). It should be noted that all the natural sounds were located in the creation zone; however, the motorized vehicle sounds were scattered between the three remaining zones: limitation zone, isolation zone, and control zone. It was observed that the motorized vehicle sound sources located in the limitation zone carry essential information, especially in case of significant danger. Therefore, these sounds were not suitable for complete isolation. On the other hand, the motorized vehicle sounds located in the control zone were contextually appropriate to the area under investigation. The users expect to hear car and motorcycle engine sounds in the area. These sounds should be kept under control by legal legislation and careful city planning; however, complete isolation of such sounds might significantly decrease the sense of place.

Table 4 – The sound sources categorized depending on the extracted components from the Principal Component Analysis (PCA).

Creation zone	Component 2: Natural sounds	Pool/fountain and birds
	Component 4: Digital sounds	Cell-phone
Limitation zone	Component 1: Alarming sounds	Car horn, car alarm, and ambulance
Control zone	Component 3: Mechanical sounds	Motorcycle and car engine
	Component 2: Natural and human sounds	Footsteps
Isolation zone	Component 1: Alarming sounds	Car brake and construction

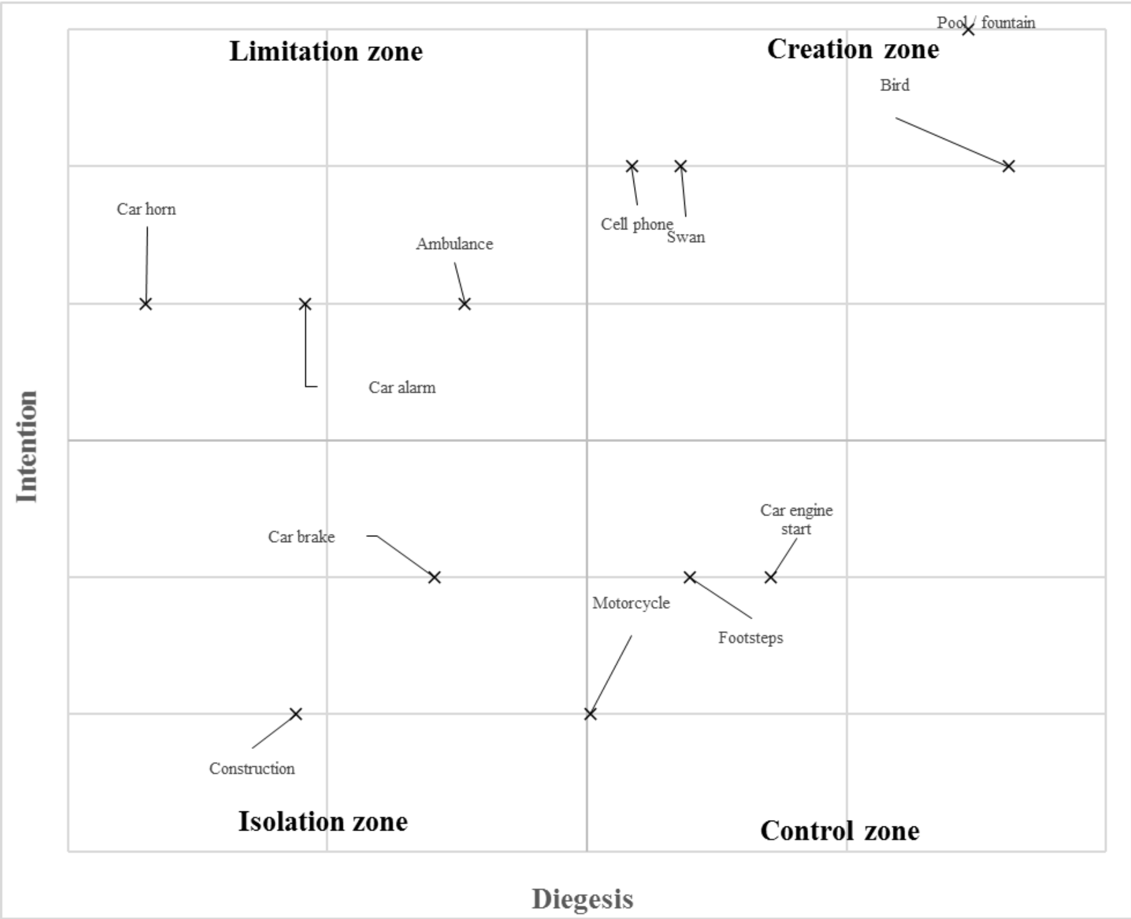

Figure 3 - The sound design guideline of the case study (Kuğulu Park, Ankara, Turkey).

4 Conclusions

In this study, we proposed an interdisciplinary sound source categorization framework for environmental sound design. The proposed framework was then applied in a case study: Kuğulu Park, Ankara, Turkey. The central aspect of the CLIC framework is its dependency on the environmental context. The sound taxonomies found in the previous literature did not take the context into account; hence they did not guide the sound designer throughout the design process. The initial results revealed that the preference of the users based on the context of a given space and the intention of the sound events could guide the environmental sound designers during the design process. In the scope of the ongoing research project, the suggested sound source categorization framework will be tested in different urban and interior environments to collect empirical data on contextual user preferences.

Acknowledgments

Research reported in this publication was supported by Çankaya University under award number MF.20.002. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the publication.

References

[1] Bones O, Cox TJ, Davies WJ. Sound Categories: Category Formation and Evidence-Based Taxonomies. Front Psychol 2018;9:1–17. https://doi.org/10.3389/fpsyg.2018.01277.
[2] Schafer M. The soundscape: Our sonic environment and the tuning of the world. New York: Alfred Knopf; 1977.
[3] Lindborg PM. A taxonomy of sound sources in restaurants. Appl Acoust 2016;110:297–310. https://doi.org/10.1016/j.apacoust.2016.03.032.
[4] Ercakmak B, Dokmeci Yorukoglu PN. A taxonomy of indoor acoustic environments and sound source preference in shopping centres. Inter-Noise 2020, 2020.
[5] Niessen M, Cance C, Dubois D. Categories for soundscape: Toward a hybrid classification. 39th Int Congr Noise Control Eng 2010, INTER-NOISE 2010 2010;3:2269–82.
[6] International Organization for Standardization. ISO 12913-2:2018 Acoustics — Soundscape — Part 2: Data collection and reporting requirements 2018.
[7] Dakic V. Sound design for film and television. Munich: GRIN Verlag; 2007.
[8] Lengen C, Kistemann T. Sense of place and place identity: Review of neuroscientific evidence. Heal Place 2012;18:1162–71. https://doi.org/10.1016/j.healthplace.2012.01.012.
[9] Kusitzky T. Cultivating Urban Sound as an Object of Design Cultivating Urban Sound as an Object of Design. Invis. Places 2017, Ponta Delgada, Sao Miguel Island, Azores, Portugal: 2017.
[10] Jekosch U. Assigning Meaning to Sounds — Semiotics in the Context of Product-Sound Design. Commun. Acoust., Berlin/Heidelberg: Springer-Verlag; 2005, p. 193–221. https://doi.org/10.1007/3-540-27437-5_8.
[11] Spence C, Zampini M. Auditory contributions to multisensory product perception. Acta Acust United with Acust 2006;92:1009–25.
[12] Özcan E, Van Egmond R. Product Sound Design : An Inter-Disciplinary Approach? Undisciplined! Des. Res. Soc. Conf., Sheffield: 2009, p. 16–9.
[13] Özcan E, Van Egmond R. Characterizing descriptions of product sounds. 11th Meet. Int. Conf. Audit. Disp., Limerick, Ireland: 2005, p. 55–60.
[14] Özcan E, van Egmond R. Audio-visual interactions in product sound design. Hum Vis Electron Imaging XV 2010;7527:75270P. https://doi.org/10.1117/12.856312.
[15] Jillings N, Moffat D, De Man B, Reiss JD. Web Audio Evaluation Tool: A Browser-Based Listening Test Environment. 12th Sound Music Comput. Conf., 2015.
[16] Yang W, Makita K, Nakao T, Kanayama N, Machizawa MG, Sasaoka T, et al. Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E). Behav Res Methods 2018;50:1415–29. https://doi.org/10.3758/s13428-018-1027-6.
[17] Howard R. Moskowitz. Magnitude Estimation : Notes on What , How , When , and Why To Use It. J Food Qual 1978;3:195–227.
[18] Field A. Discovering statistics using SPSS. 3rd ed. London: SAGE publications LTD; 2009.