A FRAMEWORK FOR SOUNDSCAPE ANALYSIS AND RE-SYNTHESIS

Andrea Valle, Mattia Schirosa, Vincenzo Lombardo
CIRMA-Università di Torino
via Sant'Ottavio 20, 10124, Torino, Italy
andrea.valle@unito.it, mattiaschirosa@yahoo.it, vincenzo@di.unito.it

ABSTRACT

This paper presents a methodology for the synthesis and interactive exploration of real soundscapes. We propose a soundscape analysis method that relies upon a sound object behavior typology and a notion of "sound zone" that collocates objects typologies in spatial locations. Then, a graph-based model for organising sound objects in space and time is described. Finally, the resulting methodology is discussed in relation to a case study.

1 INTRODUCTION

The term "soundscape" was firstly introduced (or at least, theoretically discussed) by R. Murray Schafer in his well-known book *The tuning of the world* [8]. Murray Schafer and his associates of the World Forum For Acoustic Ecology studied for the first time the relation between sounds, environments and cultures. Then, the diffusion of the term has continuously increased, and currently the notion of soundscape plays a pivotal role at the crossing of many sound-related fields. It is worth noting that, despite the profusion of usages, there are neither models nor applications aiming at a simulation of a soundscape starting from the analysis of an existing one. To this goal, here we propose an analytical methodology for the description of soundscapes and a system for its re-synthesis, driven by the analytical results. The system allows for a real-time interaction.

2 A METHODOLOGY FOR ANALYZING EXISTING SOUNDSCAPES

The simulation of an existing soundscape requires to analyze the soundscape itself in order to provide data to be used in the re-synthesis process. The analysis aims at gathering data from the real environment. Our methodology is based on a multi-step process. We start by focusing on an "absent-minded" exploration of the soundscape: the analyst must be perceptually open and adhere to a passive listening strategy. In this way it becomes possible to identify the most relevant sound objects of the overall soundscape, i.e. the ones that are evident even to the least aware listeners. Traditionally, the soundscape studies have insisted on a tripartite typology of sounds in relation to their socio-cultural function: keynote sounds, signal sounds, soundmarks [8]. The identification of sound objects allows for a subsequent classification based on phenomenological and semiotic elements (more later). An active listening strategy is then performed, with the aim of locating the sound objects in the space. It is thus possible to create a sound map. Then, we focus on the analysis of the temporal organization of the soundscape, in order to retrieve specific sequences of sound objects. A database is produced containing the recordings of raw audio material related to the identified sound objects. On one side, large portions of soundscape are recorded with an omnidirectional microphone: so, a large quantity of raw material is available for editing and processing. On the other side, high directivity microphones are used to capture a wide variety of emissions while minimizing undesired background.
Two issues emerge from the analysis.
The first is related to sound classification. The one used by studies in acoustic ecology typically refers to socio-cultural and aesthetic aspects of sound, and it needs to be oriented toward the simulation of soundscape [11], [12]. Here we propose a supplementary classification that focuses on the perceptual and indexical properties of the soundscape and integrates elements from the theory of "sound object" [10] and from the research in "audiovision" [5]. Sound objects can be classified according to the following types:

- events: an event is a single sound object of well-defined boundaries appearing as an isolated figure. In this sense, it is similar to a signal as defined in soundscape studies.

- sound subjects: a sound subject represents the behavior of a complex source in terms of sequencing relations between events. In other words, a sound subject is a description of a sound source in terms of a set of events and of a set of sequencing rules.

- atmospheres: in relation to sound, Böhme has proposed an aesthetics of atmospheres [3]. Every soundscape has indeed one or more specific "scenic atmo-

SMC 2009, July 23-25, Porto, Portugal

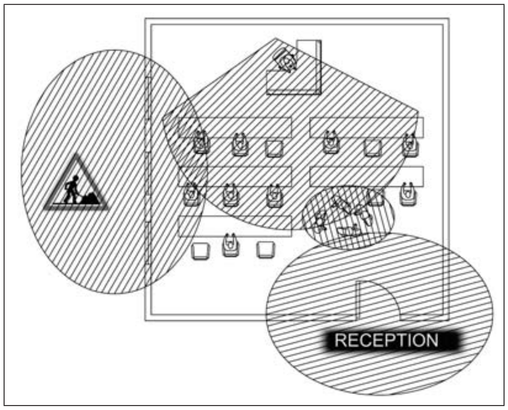

Figure 1. The classroom example: four sound zones.

spheres", which includes explicitly an emotional dimension. An atmosphere is an overall layer of sound, which cannot be analytically decomposed in single sound objects, as no particular sound object emerges. Atmosphere characterizes quiet states without relevant sound events.

A second issue concerns the distribution of sounds in space. Many scholars have noted that a soundscape can be decomposed as a group of several acoustic scenographies, which are then recomposed through the listener's exploring experience [1], [2], [14], [7]. As soundscapes are not uniform, the listener's experience is enhanced when s/he encounters sound aural transitions during her/his exploration of the environment [3]. When a listener is spatially exploring the soundscape, he can notice several perceptual differences in the sounds. In particular, the sound environment can be decomposed into internally homogeneous sub-parts. These sub-parts are here referred to as "sound zones". By the study of sound zones, sound aural transitions between them can be individuated, analysed and and re-synthesized in the simulation. Sound zones can differ in dimension and in number of elements, but they are characterized by typical sources, i.e. sound emissions are often present in a region and absent (or only rarely heard) in the others. The soundscape will then results from the summation of all the sound zones, that the listener will be able to explore. As an example, we can consider the following situation: in a university classroom with acoustic insulation walls, closed doors and windows, a professor is speaking in front of a very silent audience. The professor voice is loud and clear in all the classroom, without any relevant irregularity. By contrast, we can imagine the opposite situation: doors and windows are open, the thin walls are incapable of blocking any environmental sound, outside there are roadworks, a reception party is running in the hall just behind the door, a few students joke and laugh while the professor keeps explaining loudly. This second soundscape (represented in Figure 1) is completely different from the first one. Considering that the classroom is wide enough, it would be very simple to move around the space and run across several recognizable micro-soundscapes. Someone near the door can notice that reception party sounds are louder than any other sound source coming from the classroom. As s/he moves to the desk, s/he can hear the professor's voice, and so on. In the first case it is possible to identify a soundscape consisting of one only zone; in the second case four sound zones are clearly defined. Thus, even if their boundaries can be fuzzy, each zone can be considered as completely independent from another. This means that it is possible to describe the behavior of each zone. A soundscape results from the interweaving of each zone behavior.

3 A GRAPH-BASED MODEL FOR SOUNDSCAPE RE-SYNTHESIS

In this section we propose a model able to re-synthesize a soundscape starting from sound object introduced before. We discuss a system to organise sequences of sound objects in space and time, then we take into account the relation between the system and sound objects/sound zones.

The re-synthesis process meets two requirements. First, it must be generative, i.e. capable to create an infinite set of sequences of sound objects from a finite set of sampled sound objects. Second, the algorithm must be able to merge the information coming from the sequencing process with the user's navigation data. In this way, the simulated soundscape can be explored interactively. The generative model, named GeoGraphy, is based on graphs (for a more detailed description see [13]). Graphs have proven to be powerful structure to describe musical structures ([9]). Still, a common feature of all these graph representations devised for music is that they generally do not model temporal information: on the contrary, the model relies on time-stamped sequences of sound objects. The sequencing model is a direct graph (see Figure 2), where each vertex represents a sound object (sampled from the analysis phase) and each edge represents a possible sequencing relation on pairs of sound objects. This graph is actually a multigraph, as it is possible to have more than one edge between two vertices; it can also include loops. Each vertex is labeled with its relative sound object duration and each edge with the temporal distance between the onsets of the two sound objects connected by the edge itself. The graph defines all the possible sequencing relations between adjacent vertices. A sequence of sound objects (a *track*) is achieved through the insertion of dynamic elements, called "graph actants". A graph actant is initially associated with a vertex (that becomes the origin of a path); then the actant navigates the graph by following the directed edges according to some probability distribution. Each vertex emits a sound object at the passage of a

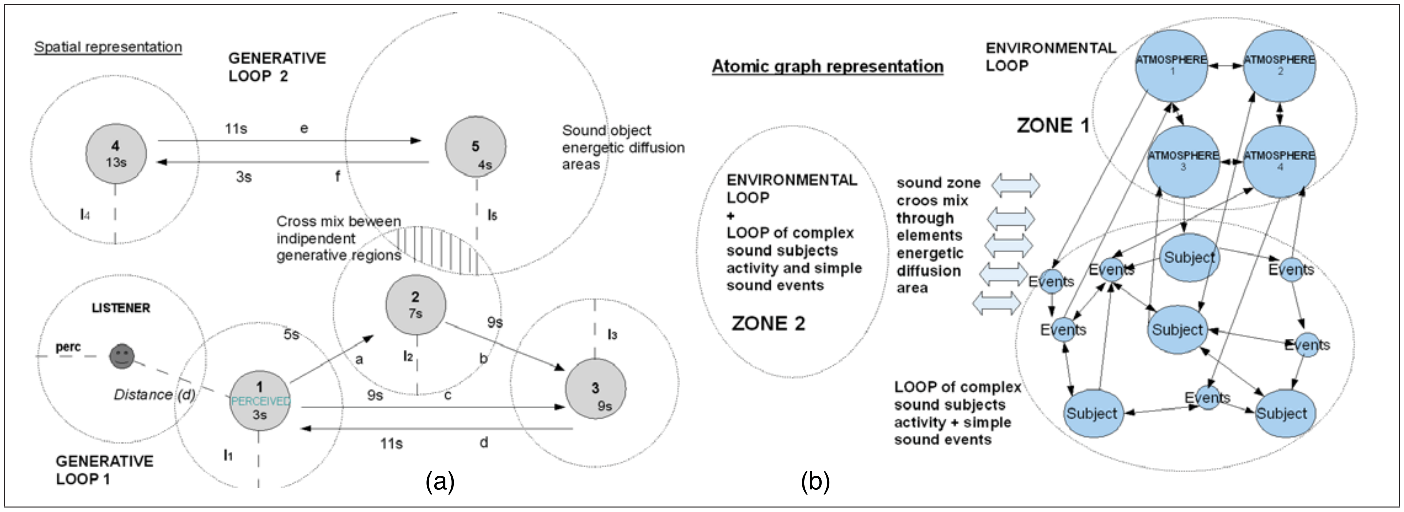

Figure 2. (a) Graphs and Listener. (b) A graph representing the relations among atmospheres, events and sound subjects in a zone.

graph actant. Multiple independent graph actants can navigate a graph structure at the same time, thus producing more than one track. In case a graph contains loops, tracks can also be infinite. As modeled by the graph, the sound object duration and the delay of attack time are independent: as a consequence, it is possible that sound objects are superposed. This happens when the duration of the starting vertex label is longer than the duration of the chosen edge (in Figure 2, the edge e between vertex 4 and 5). Thus, there will be as many superposed tracks as graph actants. In order to allow the inclusion of the exploration process graphs are placed in a two-dimensional space: in this way, the original location of a sound object can be represented. Each vertex is given a radiation area: the radius indicates the maximum distance at which the associated sound object can be heard. Inside the map of graphs, a "Listener" is defined. The Listener is identified by a position, an orientation and an audibility area (see Figure 2, a). The position is expressed as a point in the map; the orientation as the value in radiant depending on the user's interaction control; the audibility area defines the perceptual boundaries of the Listener. The Listener can be thought as a function that filters and parameterizes the sequences of sound objects generated by the graph actants. Every time a vertex is activated by a graph actant, the algorithm calculates the position of the Listener. If the intersection between the Listener's audibility area and the vertex's energetic area is not void, then the Listener's orientation and distance from the vertex are calculated, and all the data (active vertex, position, distance and orientation of the Listener) are passed to a DSP module retrieving from the database the recorded samples and processing them according to some spatialization model (e.g. reverberation, low-pass filtering, amplitude scaling, etc.).

To sum up, in our system a soundscape emerges as the relation between the set of tracks generated by the graph actant navigating the graphs and the filtering function defined by the Listener.

Through the GeoGraphy model and the sound zone modular description it is possible to generate a target complex soundscape. In Figure 2 (a), each zone is described by a dedicated graph. As discussed, zones can overlap their audibility with other zones, depending on the radiation of the sound sources that compose it. For that reason GeoGraphy provides the radiation area concept: it indicates the maximum distance at which the associated sound object can be heard. It is set according to dynamics annotations taken during the analysis phase. It is thus possible to regulate the radius of each element to interbreed the parent zone with the others. This modular description process allows to easily represent "sound pollution zones". In Figure 2 (a) a zone is represented by the generative loop 2 between vertices 4 and 5, respectively having radii I_4 and I_5. Similarly, "the generative loop 1" individuates one more zone, and the intersection between the elements 2 and 5 is the portion of space where the sounds of the two zones can be hearable. A soundscape is the summation of all the graphs apt to describe the sound zones that a listener will be able to explore. A zone in itself is the summation of all the sub-graphs representing atmospheres, events, and sound subjects. The soundscape is made by a continuous fusion process between sound figures and backgrounds. In this sense, each zone has a core granting continuity to the structure, and to the resulting auditory stream. The core is formed by the atmospheres, that consist several ambient sound materials (i.e. long field recordings), aiming at representing different natural "almost-quiet state" nuances. In Figure 2 (b) the atmospheres 1 to 4 allow the formation of a background against which semiotically and indexically relevant signals can emerge. Atmospheres can be connected to other atmospheres and to events and sound subjects. Their durations is typically set longer than the edges connecting them to events and sound subjects: in this way, atmospheres are still present when the events and sound subjects are activated by the actant. Atmospheres can then generate a background layer, while sound events and sound subjects reach the close-up perceptual level and then quickly disappear. Events can be thought as isolated signals. On the contrary, sound subjects feature a double nature. In Figure 2 (a) they are represented, for sake of simplicity, as single vertices, standing for a complex but unitary acoustic behavior. But a sound subject properly is a subgraph, organized recursively in a core surrounded by events, like a sub-zone. An example of sound subject is discussed in the next section.

4 A CASE STUDY: THE MARKET OF THE PORTA PALAZZO IN TURIN

The market is a typical case of a socio-culturally relevant soundscape. In particular, the market of the Porta Palazzo in Turin has a long tradition as it has been established more than 150 years ago. It is the greatest outdoor market in Europe, and it represents the commercial expression of the cultural heritage of the city of Turin. During the century, it has tenaciously retained its identity, characterized by the obstinate will of the workers of sharing its government's responsibility. It is probably the part of Turin where the largest number of social different realities and cultures coexist, both of Italian and of foreign origins [6]. As a consequence, its soundscape manifests an impressive acoustic richness. First, it includes languages and dialects from all the regions of Italy, South America, Eastern Europe, North Africa. More, there are many qualitatively different sound sources: every day the market serves 20,000 persons (80,000 on Saturday), and 5,000 persons are working in it every day. It can be said that the Porta Palazzo soundscape belongs to the Italian cultural heritage.

The analysis of the case-study initially focused on the socio-cultural dimension of the market and has started from the urban redevelopment survey by the Porta Palazzo public advisory committee [6]. More, short informal interviews with local workers, customers and workers' representatives have been realized. The interviews were done during the first "absentminded" explorations of the place, that made it possible to annotate the most pervasive sound objects. As an example, the sound of plastic shopping-bags is a unique keynote sound represented as a mass of sound events. The shouts of the merchants is another multi-particle keynote in which the listener of the marketplace soundscape is immersed: the most intense, vibrant, repetitive, significant advertising messages have been recorded to be simulated. In some sense, their sum is the pervasive call of the market: the Porta Palazzo voice. Finally, one can notice that there is a specific keynote sound in certain border regions that invades all the space: the noise of motor vehicles and carriages. As an example, in the customers opinion, the arrival of the streetcar number 4 is the unique sound source that can be heard throughout the whole soundscape, acquiring specific nuances in each zone (i.e. due to reverberation and to low frequency distance attenuation). Hearing this sound object makes one think immediately of the Porta Palazzo market. This is a very interesting strong semantic association picked up during interviews.

Then, sound signals have been taken into account, in particular those related to specific stands, giving origin to complex sound subjects: five typical stand sounds have been analyzed. It must be noted that a reduced set of samples has proven to be rich enough to describe different stands, as they could be differentiated by their grouping in different graph structures. A particular sequence of events has led to create the specific "shopping" sound subject: plastic rustle, paper rustle, clinking coins, cash opening, clinking coins, cash closing. The *stands of the anchovy sellers* have proven to be very different from all other stands: they include sounds of metal cans, of anchovies being beaten over wood plates, of olives thrown in oil, of noisy old scales.

Subsequently, we define the sound zones: the analysis of the soundscape led to five independent zones formed by distinctive elements. In this way, it has been possible to record specific atmospheres. In Figure 3 all the zones are assigned an identifying index.

The zones 1 and 2 are characterized by the sounds of motor vehicles. Zone 1 is mainly characterized by a sound atmosphere made up of little delivery trucks, hand-carts and gathering of packing boxes from stands. It is the only street accessible by any vehicles as bus, trams, cars and motorbike. Instead, in the zone 2 there are two important sound features: the load area of big delivery trucks and the street dedicated to public transport, with rail system allowing streetcar passage. Both the zones present sounds related to bread, mint and spice hawkers. Zone 3 is a diffused area showing a mixup of sounds related to market and to street/parking areas. This feature has required to aptly adjust the radius of sound sources to describe its fuzzy sonic boundaries. In addition, some emissions related to the daily process of assembling/disassembling stands are present. Zone 4 is formed by different and rare stands; it presents a less prominent sound density because the passage area is bigger, so the sound of walking costumers, hand-cart distribution, empty box collecting process, are louder than other sound objects. More, many atypical stands are positioned in this zone, making its atmosphere unique. The motor sound is almost imperceptible, with the exception of some very loud source (as

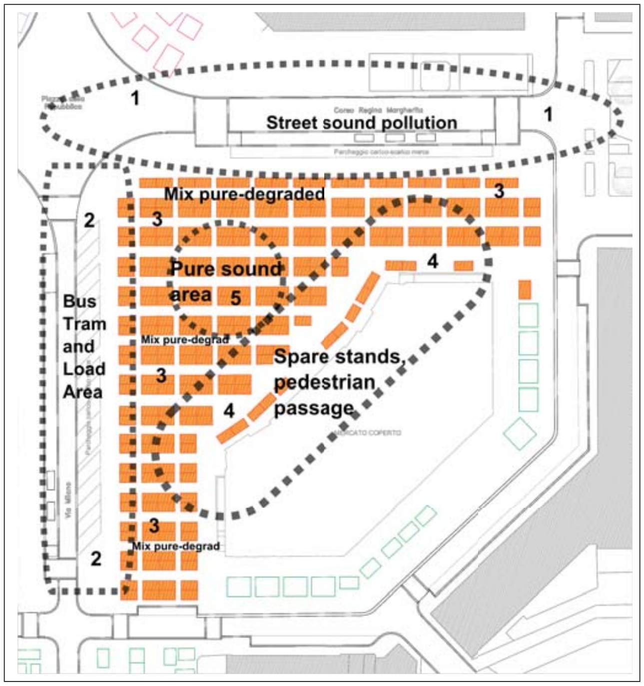

Figure 3. Plant of the Porta Palazzo market: organization in sound zones.

streetcar 4). Zone 5 presents only vegetable and fruit stands: transit ways are thin and rare, and only walking people can pass through. The shouts of the merchants reach the highest intensity and mask all the pollution sound coming from the other zones, while the many sound signals (activities and voices) make the soundscape particularly frenetic, a disarranged composition of sound objects making it a "pure" example of market soundscape.

The graph structure of a sub-part of zone 2 is showed in Figure 4 (a). The graph is a sub-part of zone 2. Edges define the sequencing rules between a possible atmosphere (labelled "1:XAtmo...", a four minute recording) of that specific subpart, and nine indexical sound events. The atmosphere describes the almost quiet state of that area, generated by the continuous walking of costumers and the activity of some mint hawkers. The sound events describe activities by different vehicles. The number of repetitions of a sound object (i.e bus, tram, delivery truck, motorbike) is proportional to its statistical relevance: there are four tram objects, then two for bus and trucks, and only one for motorbike. No car was noticed here. The graph is cyclic, thus generating potentially infinite tracks. In this case, each possible path is designed to have the same duration of the atmosphere. So the time duration of edge connection $Edur_{x_y}$ between vertices are set according to the following rule:

$$Edur_{Atm_2} + Edur_{2_3} + ... + Edur_{x_A}tm = Vdur_Atm$$

In this way a long, looping background is continuously varied by the superposition of different other sound objects. By only using nine objects it has been possible to represent a complex soundscape.

Figure 4 (b) depicts the graph of a sound subject. The graph represents the behavior of a delivery truck. The delivery trucks arrive at that zone, unload the products, and leave back. By connecting three "core" objects $1b$, $1c$, 5 and nine sound events, it allows the simulation of several instances of the truck. Here the sound subject reveals its sub-zone nature. The topological structure of the graph includes a start event (1) and an end event (10). The core objects are almost quiet recordings. As an example, $1b$ refers to a stationary truck with running motor while $1c$ refers to a truck making some accelerations. They are placed topologically in the center of the graph structure, providing a continuous background against which other smaller sound objects appear and disappear. As in the previous example, all the possible paths reactivate a core before its duration has finished. But there is an exception: after the end event the auditory stream stops, as the graph is acyclic. The graph can be made cyclic by the addition of edges connecting the end vertex 10 to the start vertex 1. These looping edges can have durations spanning over a large interval, from 15 to 530 seconds. After a path simulating the truck delivery has reached the end event, it is thus possible that the start event is emitted straight after that: the sound result will then be perceived as an activity of the same acousmatic subject. By contrast, when the path restarts a long time after the end vertex has finished, the result can be perceived as the arriving in the soundscape of a new truck.

5 CONCLUSIONS AND FUTURE WORK

The notion of soundscape is increasingly relevant not only in contemporary culture and acoustic ecology but also in many other fields, such as, e.g. audiovisual productions, in which soundscape is now a fundamental part of the whole soundtrack. The proposed system is able to generate soundscapes from original sound materials but without relying on loops. In this way, the typical "sound mood" of the original space is preserved: at the same time, the resulting soundscape is no more fixed, but undergoes a continuous variation thanks to the graph dynamics, due to probabilistic connections. The system was been implemented in the SuperCollider audio real-time synthesis programming environment, which allows the efficient real-time synthesis of the soundscape and its interactive exploration by the user [15]. A major issue in our system concerns the generation of multigraphs, actually to be carried out manually and potentially quite time-consuming. We are planning to extend the system so to include the automatic generation of graphs starting from information stored in the database or from sound-related semantic repertoires (see [4]). The

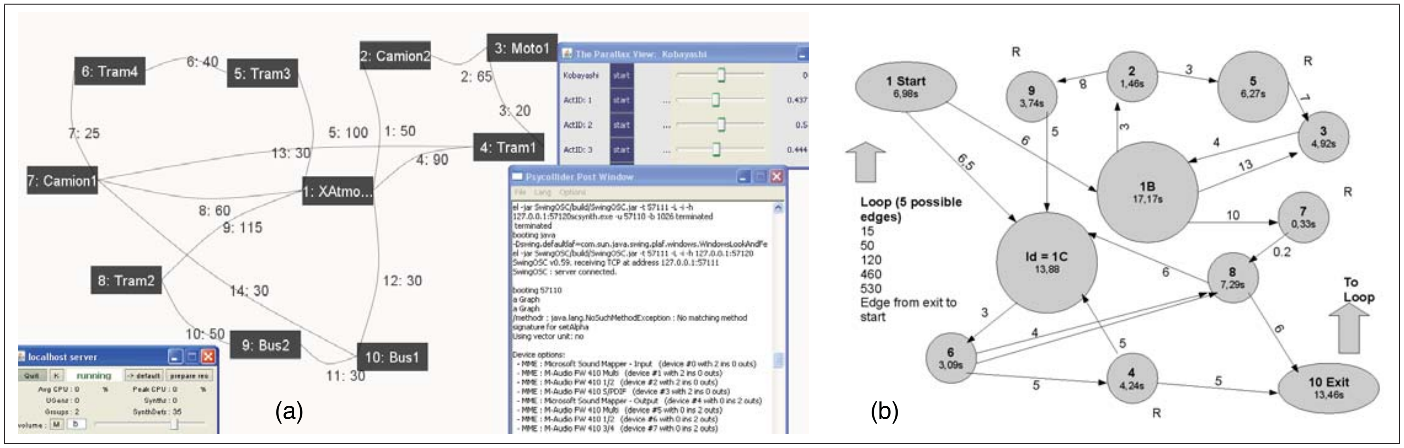

Figure 4. (a) Screenshot from the current SuperCollider implementation: the graph describes the sequencing rules of a part of zone 2. (b) Graph sequencing rules of the delivery truck sound subject.

database itself can eventually include not only sound samples created from direct recording but also from available sound libraries. Sound automatic recognition could led the automatic definition of some parameter (i.e. sample durations labelling vertices). An interesting perspective is to investigate user-generated, online databases such as Freesound [1] : in this case the graph generation process can be governed by social tagging.

6 REFERENCES

[1] Pascal Amphoux. *L'identité sonore des villes européennes, Guide méthodologique à l'usage des gestionnaires de la ville, des techniciens du son et des chercheurs en sciences sociales*. publication IREC, EPF–Cresson, Lausanne–Grenoble, 1993.

[2] Jean-François Augoyard and Henry Torgue. *Repertorio degli effetti sonori*. Lim, Lucca, 2003.

[3] Gernot Böhme. *Ecologia della musica: Saggi sul paesaggio sonoro*, chapter Atmosfere acustiche. Un contributo all'estetica ecologica. Donzelli, 2004.

[4] P. Cano, L. Fabig, F. Gouyon, M. Koppenberger, A. Loscos, and A. Barbosa. Semi-automatic ambiance generation. In *Proceedings of the International Conference of Digital Audio Effeccts (DAFx'04)*, pages 1–4, 2004.

[5] Michel Chion. *L'audiovision. Son et image au cinéma*. Nathan, Paris, 1990.

[6] Studio di Ingegneria ed Urbanistica Vittorio Cappato. 50 centesimi al kilo: La riqualificazione del mercato di Porta Palazzo dal progetto al cantiere. Technical report, Comune di Torino, Torino, 2006.

[7] Albert Mayr, editor. *Musica e suoni dell'ambiente*. CLUEB, Bologna, 2001.

[8] R. Murray Schafer. *The Tuning of the World*. McClelland & Steward and Knopf, Toronto and New York, 1977.

[9] Curtis Roads. *The computer music tutorial*. The MIT Press, Cambridge, Mass., 1996.

[10] Pierre Schaeffer. *Traité des objets musicaux*. Seuil, Paris, 1966.

[11] Barry Truax. *Acoustic Communication*. Greenwood, Westport, CT, 1984.

[12] Barry Truax. Model and strategies for acustic design. In H. Karlsson, editor, *Hör upp! Stockholm, Hey Listen! - Papers presented at the conference on acoustic ecology*, Stockholm, 1998.

[13] Andrea Valle, Vincenzo Lombardo, and Mattia Schirosa. A graph-based system for the dynamic generation of soundscapes. In *Proceedings of the 15th International Conference on Auditory Display*, pages 217–224, Copenhagen, 2009.

[14] VV.AA. Résumé de l'étude de conception et d'amenagement du paysage sonore du secteur de la Sucrerie - St. Cosme. Technical report, Acirene–atelier de traitement culturel et estetique de l'environnement sonore, 2007.

[15] Scott Wilson, David Cottle, and Nick Collins, editors. *The SuperCollider Book*. The MIT Press, Cambridge, Mass., 2009.

[1] http://www.freesound.org/