Edinburgh, Scotland
EURONOISE 2009
October 26-28

Discrimination of urban soundscapes through Kohonen map

Laurent BROCOLINI[a]
Catherine LAVANDIER[b]
Université de Cergy Pontoise, Laboratoire Mobilités Réseaux Territoires et Environnement, F-95000 Cergy Pontoise

Mathias QUOY[c]
Université de Cergy Pontoise, Laboratoire ETIS, F-95000 Cergy Pontoise

Carlos RIBEIRO[d]
Bruitparif, F-75018 Paris

ABSTRACT

Linking perceptive features and acoustic measurements in natural environments requires measurements obtained during field studies. These measurements can take into account global indicators and indices devoted to different sources, and generally, locations such as markets, parks or streets are selected with regards to their use by city dwellers. But what about the measurement differences along a street? Is the entrance of a park similar to the street beside it? How far from the street can a typical park sound environment be heard?
Therefore, in order to determine the optimized distance between urban locations, a series of recordings has been carried out. To establish a very fine space resolution along the streets, the parks and the transition areas between them, 15 minutes sequences were recorded every 7 meters. 70 locations in total were measured along a distance of 600 m. Sources on the recordings were identified by an expert listener. A set of 18 indicators was extracted from the recordings.
A classification of these locations, based on these indicators, was carried out to set a relevant space resolution. The clustering of the data was performed through a neural network using a Kohonen map. Four classes discriminate the urban soundscapes: park, thorough fare, pedestrian streets and “transition” (between park and street, or at crossroads). In particular, we show that the transition class extends for roughly 50 meters, thus far beyond its visual characterization. This means that park, thorough fare and pedestrian streets may only be acoustically well characterized beyond this limit.

1. INTRODUCTION

A soundscape can be defined as a sound environment which is perceived in a special context. Among features which characterize the context, the place and its morphology are particularly important[1]. The influence of the visual context on sound perception has already been studied[2]. For example, presence of bird twitters or vehicle sounds is not assessed in the same way in a park as it is in a street[3]. When two different types of places are nearby,

[a] Email address. laurent.brocolini@u-cergy.fr
[b] Email address. catherine.lavandier@u-cergy.fr
[c] Email address. mathias.quoy@u-cergy.fr
[d] Email address. carlos.ribeiro@bruitparif.fr

they are easily discriminated from a visual point of view, but is it so obvious from an acoustic point of view?

The aim of this work was to answer this question by determining homogeneous sound environments which correspond to typical urban areas (streets, parks, pedestrian precincts, etc.). Therefore, a measurement campaign in Paris was carried out. 18 acoustic variables were extracted from 70 recordings of 15 minutes long each. The recordings were taken at discrete locations separated by 7 to 8 meters. The data was analysed through a Kohonen map[4].

2. EXPERIMENTAL PROTOCOL

A. Locations

The neighbourhoods were chosen based on the following constraints: gathering on a limited area different types of places (thorough fare, pedestrian street, market, squares). A first site has been selected in the 12th district of Paris going from “Bercy” park, continuing through the thorough fare of “J. Kessel” (2x2 lanes with two lanes dedicated to the buses) and finishing in “Bercy” street (1 lane). The second site was chosen in the 5th district, beginning along a one lane street (“rue de l’Epée du bois”) and turning into the street “Mouffetard”, a pedestrian street where shops are concentrated and finishing in a square where a food market takes place every morning.

B. Measurement campaign

Successive 15 minutes sound tapes were recorded at each 7 to 8 meters interval. The recordings covered a total distance of around 600 meters (Figures 1 and 2) for a total of 70 locations. There were 46 measurements in site 1, and 24 measurements in site 2.

Figure 1: Description of site 1. Figure 1a presents the 46 locations; Figure 1b presents the “Joseph Kessel” thorough fare.

As we wanted to study the effect of the spatial resolution, we wanted to be sure that difference between two measurements was due to distance in space and not in time (day sound environment is different from the evening one). Actually, a previous study carried out in Aix-en-Provence showed that five periods can characterize the evolution of the soundscape during a day in a same location: the awakening of the town, the day period where measures are quite similar, the transition period toward the evening and finally, the evening and the night. In that study[5], it has been shown that during each period, 15 minutes recording is long enough to capture the soundscape of the location. So, in our presented work, all the recordings have been carried out from July to October (except August), on Tuesdays, Wednesdays and Thursdays between 9 AM and 11 AM.

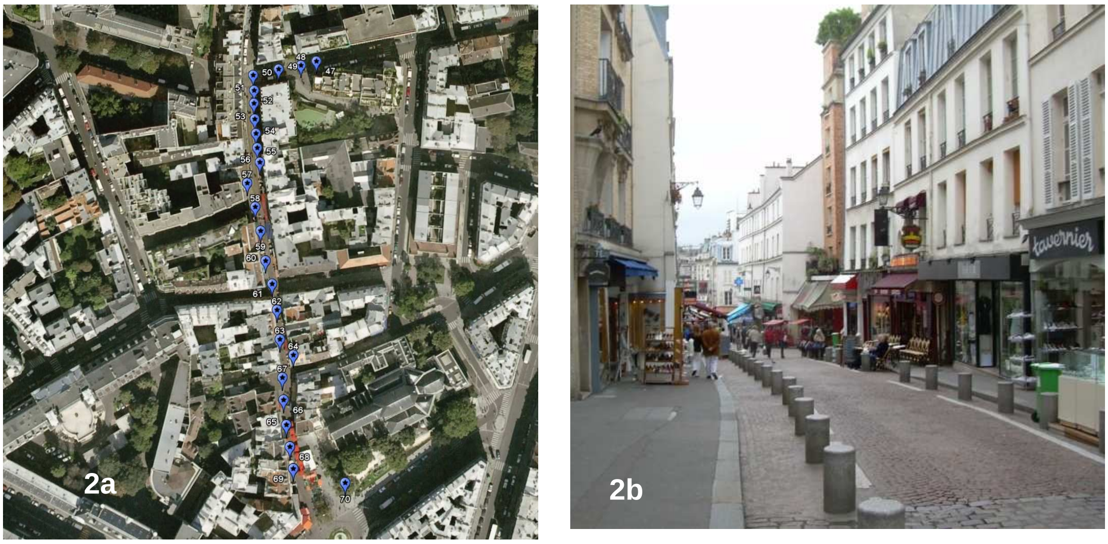

Figure 2: Description of site 2. Figure 2a presents the 24 locations; Figure 2b presents the pedestrian street “Mouffetard”.

C. Indicators

For continuous flow of vehicles, the equivalent level L_{Aeq} can discriminate different locations, but for fluctuating urban soundscapes, there is a need of source labelling[3]. In addition to global energetic indicators such as L_{Aeq}, σ (standard deviation of $L_{A,100ms}$ over the 15 minutes), L_{A90}, L_{A10}, we characterized all the following sources: light vehicles, buses and trucks, mopeds, horns, voices, birds and activity (food or goods delivery for shops and markets, etc.). Special indicators dedicated to each type of source are calculated from the time evolution $L_{A,100ms}$: time ratio of presence (%T, percentage of time where one type of source is present over the 15 minutes recording) and emergence (difference between particular L_{A10} for one type of source and global L_{Aeq}). The labelling of the sources on the tapes was carried out by an off-line audio scoring after an initial on-line label. If it is easy to label sources such as vehicles, it is quite difficult to label birds. Twitter is not loud enough to mask the other sources, but it is clearly noticed by listeners. So for birds, %T has been globally estimated by one “expert” listener, varying from 0 to 100%, and the emergence has been distributed over 6 clusters of perceptive units from 0 to 5. Moreover, when a source is not heard, %T is obviously fixed to 0. The emergence was determined by applying a -15 dB(A) offset, because we assumed that a source which is 15 dB(A) lower than the equivalent level is not heard. Thus we end up with a set of 18 indicators for each of the 70 recording locations, as may be seen in Table 1 showing a sample of the 18 indicators extracted from the recordings.

Table 1: *Extracted data from the data base displaying values for the 18 indicators.*

Indicators	Sources	Measurements	01	02	…	34	35	…
(1)	Global values	L_{Aeq} (dB(A))	54,0	56,8		69,5	68,9	
(2)		St. deviation (dB(A))	2,1	3,3		4,6	5,2	
(3)		L_{A90} (dB(A))	51,2	51,2		60,8	59,0	
(4)		L_{A10} (dB(A))	56,3	58,0		72,4	72,4	
(5)	Light Vehicles (LV)	%T	0,0	0,0		40,5	47,7	
(6)		Emergence (dB(A))	-15,0	-15,0		1,4	3,7	
(7)	Buses / Trucks (BT)	%T	0,0	0,0		11,1	17,8	
(8)		Emergence (dB(A))	-15,0	-15,0		6,4	5,2	
(9)	Moped	%T	0,0	0,0		4,1	1,9	
(10)		Emergence (dB(A))	-15,0	-15,0		4,8	11,4	
(11)	Horns	%T	11,6	1,2		0,3	0,1	
(12)		Emergence (dB(A))	4,6	-2,5		5,2	12,5	
(13)	Voices	%T	1,6	11,7		22,6	6,4	
(14)		Emergence (dB(A))	3,4	-2,3		4,7	4,4	
(15)	Birds	%T	100,0	100,0		50,0	10,0	
(16)		Em. (perceptive units)	5	5		1	1	
(17)	Working activities	%T	8,8	17,2		0,0	0,0	
(18)		Emergence (dB(A))	6,0	9,6		-15,0	-15,0	

D. Kohonen map classification

In order to discriminate the different types of soundscapes, the set of 70 (locations) x18 (indicators) data has been analysed through a Kohonen map[4] designed with 30 neurons (5x6 squared map). The map is organized with a reduced number of neurons[6] compared to the number of objects.

This analysis is a non supervised classification. It aggregates similar locations into a same neuron or into neurons which are near each other. More clearly, we consider 70 objects described by p variables (18 in our case). Each neuron is characterized by a weight vector of p dimensions too, resulting in 30 different vectors. This vector links each neuron with all of the indicators. For example, the first component of the weight of each neuron corresponds to the sound equivalent level L_{Aeq}. Similarly, the fourth is the fractile level L_{A10}. The algorithm may be described as follows. Randomly, an object is presented to the network. The neuron whose weights vector is the closest (in Euclidian distance) to the object, which is also a vector in dimension 18, is said to code for this object. The weights of this neuron are modified so that the distance between them and the presented object is reduced. The weights of the nearby neurons are also modified, the nearer the neuron, the higher the modification. Then a second object is presented, and the same computation is performed until the weights do not significantly change any more.

As opposed to k-mean clustering for instance, neighbour neurons code for close objects in variable space (thus the name “map”). In our case, this clustering method is interesting because the continuous transition from one recording location to another should be reflected by the activation of contiguous neurons.

4. RESULTS

A. Weights of neurons

The final weights of the neurons can explain why they respond to one location or another. On the panels of figure 3, each square represents the weight between one indicator and each neuron of the Kohonen map. The yellow colour characterizes high values of the indicators whereas dark colour characterizes low ones. The numbering of the neurons starts from bottom left side "1" to top right side "30" (see Figure 3a or Figure 6).

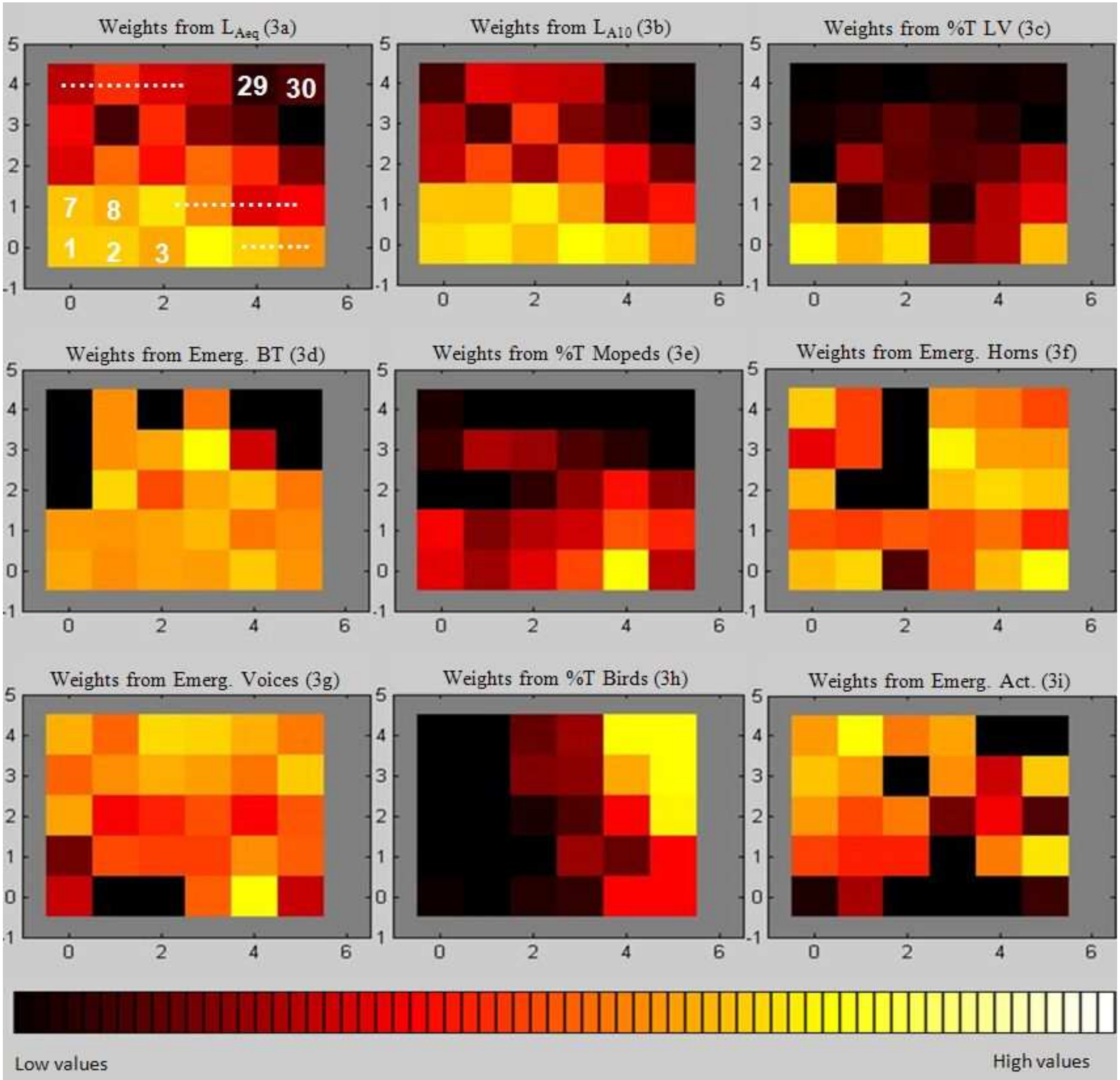

Figure 3: Weights of the neurons for different variables, with the colour scale.

On the top right of figure 3 (panel 3c), we can see that the percentage of light vehicles (%T LV) characterizes neurons 1, 2, 3, 6 and 7 with a high value and neurons 13, 19, 25, …, 30 with a low value. This last series is divided into two groups in regard to the percentage of birds (panel 3h). The left neurons 13, 19, 25 and 26 have a short time of bird presence whereas neurons 29, 30 present a long period of twitters. The three neurons situated on the bottom right side of each panel (neurons 5, 6 and 12) are characterized by presence or emergence of several types of sources (Light Vehicles panel 3c, Buses and Trucks panel 3d, Mopeds panel 3e, Horns panel 3f, Voices panel 3g, Activity panel 3i, and medium presence of birds panel 3h).

B. Kohonen and Ward classification

The results of the Kohonen map classification is reported on figure 6. In this figure, we indicated for each recording location the neuron that was coding that location. Thus for instance locations 1 and 2 are coded on neuron 24. We can see a pattern of successive activation going roughly clockwise and following the walking track of the two different sites.

In order to regroup the neurons into classes we use a hierarchical Ward classification (Figure 4). The advantage of carrying the classification on the neurons instead of initial objects (here the locations) is that, as shown above, we understand with the neuron weights which variables are responsible for gathering data into one cluster[7].

The classification proposed 3 or 4 clusters (Figure 5), but we decided to keep the fourth cluster (after the step number 56) in order to minimize the loss of information due to the fusion of groups. The final simplified representation of the network (Figure 6) based on the 18 acoustic indicators, presents the 4 typical areas.

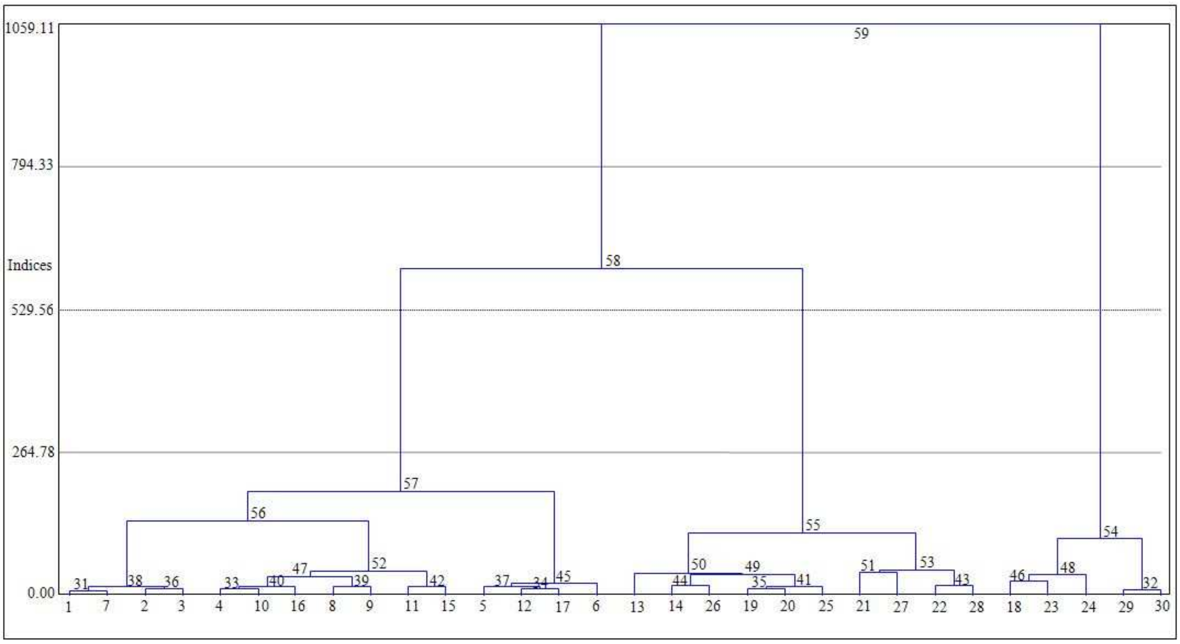

Figure 4: Dendrogram of the Ward's classification on neurons.

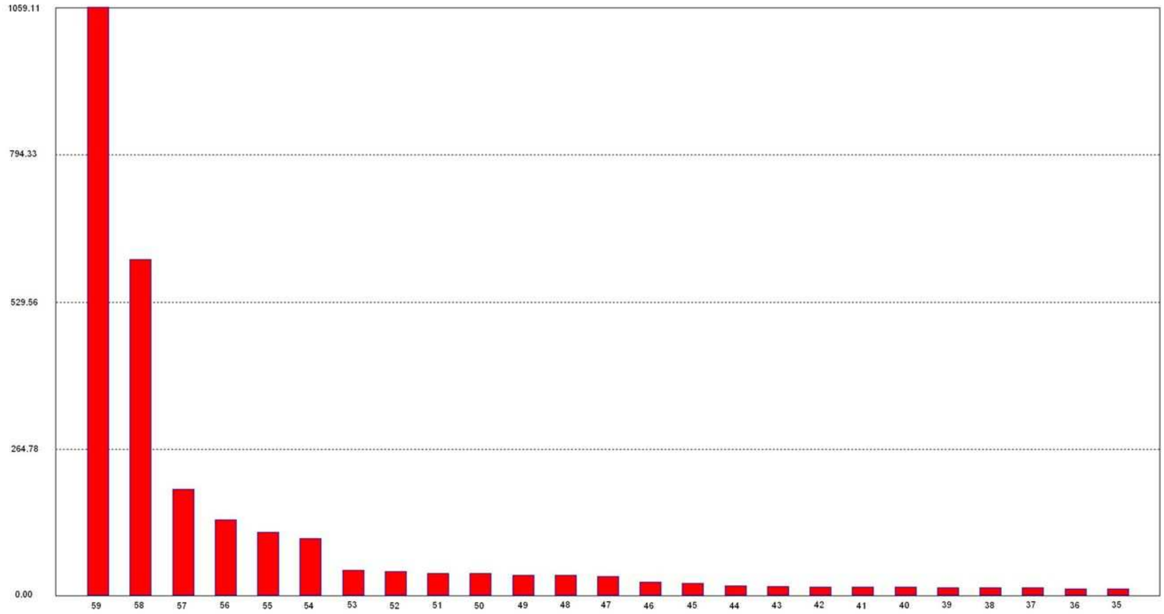

Figure 5: Inertia explained by the clusters at each step.

The first area (green) corresponds to the park. The park recordings (locations 1 to 12) are gathered into this area. The 47th location (neuron 18) is situated in this area too. This can be explained by a high value of bird presence in this recording (panel 3h of figure 3).

The second area (red) concerns the thorough fare “J. Kessel” from location 19 and the street of “Bercy”. It gathers recordings of high sound levels (panels 3a and 3b figure 3) where vehicles are identified for a long period (panel 3c). It can be noticed that the small street “Epée du bois” (locations 47 to 50) is spread over three different clusters. So it seems that a small street sound characterization is not so obvious. Another possible explanation is that there are too few recordings in this small street to allow a coding by a neuron of the Kohonen map. This remains to be further explored.

The third area (blue) gathers locations in pedestrian and animated street. It is characterized by high values of voices and activity presence (panels 3g and 3i figure 3).

Finally, the fourth area requires careful attention. Locations 13 to 18 (neurons 6 and 12) are gathered in this area and characterize the transition between the park and the thorough fare. Locations 30, 33, 34 and 36 (neurons 5 and 6) are located in this area too. They correspond to a crossroads with traffic signals where the vehicle flow is not continuous. Location 50 belongs to this area and corresponds also to junction between a small street and a pedestrian street. We have already noticed that these neurons are characterized by very different types of sources. It should be noted that the locations 23 and 24 which correspond to simple traffic signal on only one thorough fare, are associated within the road area and not within this last transition area, although the flow is not continuous. It seems that, even if the flow of vehicles stops at a traffic signal along a road, it is not enough to be acoustically classified in the transition area.

Neuron 25 60 66	Neuron 26 61 67	Neuron 27 64	Neuron 28 68	Neuron 29 6 10	Neuron 30 3 4 5
Neuron 19 55 58 59 62 63	Neuron 20 52 53 54 57	Neuron 21 48	Neuron 22 56 69	Neuron 23	Neuron 24 1 2
Neuron 13 65	Neuron 14 49	Neuron 15 51	Neuron 16 44 45	Neuron 17 46	Neuron 18 7 8 9 11 12 47
Neuron 7 21	Neuron 8 43	Neuron 9 40	Neuron 10 41 42	Neuron 11 70	Neuron 12 13 14 50
Neuron 1 19 23 24 25 26 27 29 32 35	Neuron 2 28	Neuron 3 20 22 31	Neuron 4 37 38 39	Neuron 5 36	Neuron 6 15 16 17 18 30 33 34

Figure 7: Neurons of the Kohonen map. Below the neuron number, we put the location number for which that neuron is responding the most. Colours correspond to 4 different classes obtained by Ward's Classification. Green is park, Red is thorough fare, Blue is pedestrian and animated street, and Tan is transition.

A space analysis shows that the acoustic transition from park to thorough fare is spread over 50m. For crossroads, the acoustic transition zone seems also to be spread over about 50m. Hence it is only at 50 meters away from crossroads or park boundaries that the soundscape corresponds to something other than a transition. This result is important since it does not correspond to the shorter visual transition. This also implies that one recording in a street between crossroads is enough to characterize a street soundscape.

5. CONCLUSION

As expected, three typical areas have been extracted from the sound recordings. The areas are characterized by homogeneous sound environments where typical sources are present: birds in a park, vehicles in the thorough fare and in the street, voices and activity in the pedestrian street. The forth area gathers transition sound environments, where different kinds of sources can be heard. Relying on these results, it would be now very interesting to study the human perception within transitional areas. Do people focus on sound or/and visual morphology? A perceptive study is currently in progress in Lyon, focusing on sound quality within the transition area between a park and a thorough fare. Traffic signal locations and crossroads will be also perceptively studied.

ACKNOWLEDGMENT

This research is part of a program funded by a PREDIT contract. It is supported by the French Energy Agency ADEME. We thank the urban office of Paris which allows us to measure all the sound environments in the different public areas.

REFERENCES

1. L. BROWN, J. KANG, T. GJESTLAND, Towards some standardization in assessing soundscape preference, submitted to *Euronoise 2009*, Ottawa, 2009.

2. VIOLLON S., LAVANDIER C., DRAKE C., "Influence of visual setting on sound ratings in an urban sound environment ". *Applied Acoustics*, Vol. 63 (5), pp. 493-511, mars 2002

3. LAVANDIER C., DEFREVILLE B., " The contribution of sound source characteristics in the assessment of urban soundscapes ", *Acta Acustica united with Acustica*, Vol. 92 (6), pp. 912-921, novembre-décembre 2006.

4. T. KOHONEN, Self-organization and associative memory, Springer.

5. C. LAVANDIER, B. BARBOT, Influence of the temporal scale on the relevance of acoustic parameters selected to characterize urban sound environments, *Proceedings of Euronoise 2003*, Naples, 2003.

6. J-P NAKACHE, J. CONFAIS, Approche pragmatique de la classification, Technip, pp. 165-170, 2005.

7. M. COTTRELL, S. IBBOU, P. LETREMY, SOM-based algorithms for qualitative variables, Neural Networks, New Developments in Self-Organizing Systems, Volume 17(8-9), pp. 1149-1167, October-November 2004.