Red Light Spotters

Images-Driven Sound and Rhythm Landscape

Philippe Codognet
Keio University, DMC
Research Institute for Digital Media and Content
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
philippe@dmc.keio.ac.jp

Gilbert Nouno
University of Paris 6, LIP6
IRCAM
1 place Igor Stravinsky, 75004 Paris, France
Gilbert.Nouno@ircam.fr

ABSTRACT

Red Light Spotters is a multi-contextual interactive project which aims to create an open and real-time musical environment sharing relations with and extending a natural ambient landscape. It encompass artistic creation process embedding image tracking, artificial intelligence with inductive original tempo tracking and beat prediction algorithms. We end up with emergent rhythmic processing for ambiant musical creation, interpreted and performed by sound artists.

Categories and Subject Descriptors

H.5.5 [**Sound and Music Computing**]: ModelingSignal analysis, synthesis, and processing; H.5.1 [**Multimedia Information Systems**]: Audio input/output

General Terms

Design, Algorithms, Experimentation

Keywords

Interactive sound installation, sensor data interpretation, digital art, computer-based music, creation techniques

1. INTRODUCTION

Red Light Spotters is a multi-contextual interactive project which aims to create an open and real-time musical environment sharing relations with and extending a natural ambient environment, namely an urban landscape. This interactive sound installation is intended to be exhibited in a "City View" observatory that are becoming more and more frequent in top floors of skyscrapers buildings, and to be an auditory complement to the visual stimulus from the elevated point of view over the city. In a first step, this project is specifically developed for Tokyo, as there the blinking red lights on top of buildings are creating a mesmerizing dynamic virtual landscape and can be used to create rhythms and music. But it can be further adapted for other cities. It encompass artistic creation process embedding image tracking, artificial intelligence with inductive original tempo tracking and beat prediction algorithms. We perform pertinent image and symbolic descriptors extractions, as pulsation and rhythm features, in order to synchronize both musical and control worlds with the natural visual environment. We end up with emergent rhythmic process for musical creation interpreted and performed to a certain extent by sound artists. This project is based on an idea by multimedia artist Maurice Benayoun and researcher Philippe Codognet and is developed by a collective of sound artists and researchers[1] as part of **Keio University DMC**[2]'s artistic and research projects.

In this paper we will present in Section 2 the natural visual framework of our project. We will then move to the constitutive elements and flowing data description of the installation in Section 3. Some details about the algorithmic intelligence linking the input data to the musical generation are discussed in Section 4. We will argue about the interpretation and interaction aspects in Section 7. Finally, the implementation of the computer environment is presented in Section 5 and the musical composition environment is considered in Section 6. A brief conclusion ends the paper.

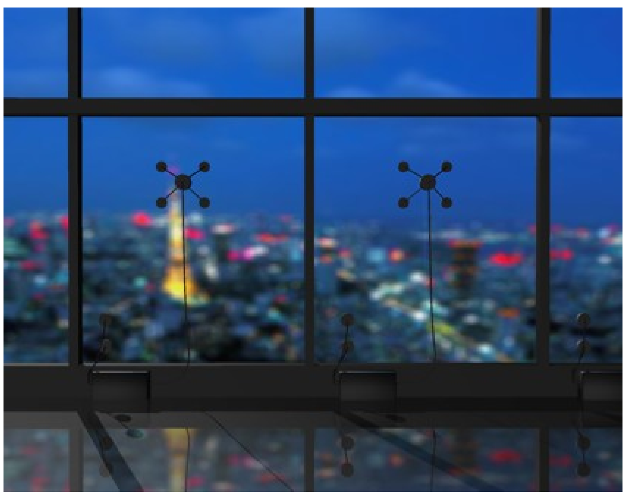
2. THE VISUAL POINT OF VIEW

The town of Tokyo at night Fig.**1(a)**, with its specific blinking red lights skyline ambiance[3], is the first experimental urban setup that we propose to use as an immersive visual and musical experience. Our concept is aimed to create intelligent real-time relationships between the context of the installation and the exhibition space itself; that is, between the environmental images of the town (i.e., somehow a landscape of lights) and the interactive musical surrounding occurrences and resonances. The spectators will thus be immersed in these related images and sounds and intuitively link them together, as the music is created by the city itself. However, we do not want to make this relationship too much explicit for the spectators, as this is not an "interactive technology demo" but a sound artwork, therefore our goal is to have the spectators first appreciating the musical ambience as such and then to have them slowly realizing the relationships and interactions between the city lights and the sound rhythms. We also wish to leave the music landscape and

[1]Alice Daquet, Gilbert Nouno, Philippe Codognet
[2]research institute for digital media and content
[3]those blinking red lights are used to signal the top and sides of various building and skyscrapers to helicopters flying over the city

composition tool open in order to renew itself in different contexts depending on the location, the time and the sound artists.

(a) The actual landscape as could be seen from the "Tokyo City View" observatory on the top of Mori Tower building, 52nd floor



(b) Tokyo view with detail of camera installation for image acquisition

Figure 1: Tokyo city view

3. FLOWING IMAGES TOWARDS SOUND

In this section we describe the real-time processing of the data flux and the transformations they undergo in the installation space. The emphasis is put on the perceptual experience that should result from this process, as the visual landscape would be the initial element offered to the eyes of the viewers who become listeners and potentially actors of their own multi-sensorial experience. By this, we aim to extract or reveal a certain amount of information handled in the image - the blinking red lights - and turn this information into musical output that we can control and modulate artistically.

The initial perceptive data is the dynamic view of the city with its intricate mesh of flashing red lights. It is turned by video camera grabbing into a series of image frames that we reduce after image-signal processing into a set of blinking dots (Fig.**2**) with characteristic parameters such as intensity, frequency and time-phase. These parameters extractions - features or descriptors - lead us to some rhythmic considerations and relations that we can achieve through different ways of grouping and correlating the blinking lights. We end up with an interdependent rhythmic structure in relation with the expectations that the viewer might have while contemplating the natural view offered to his eyes.

We relate the expectation of the spectator with some elements of anticipation: we use in an inner-predictive way parameters in the computation of the predictive algorithm that will be presented later in section 4. We implement the algorithmic process on the temporal flowing sequence of the flashing red lights so as to produce musical events synchronized with the quasi-random visual occurrences. Doing so, the musical content consists of real-time generated rhythms and sounds based on some predictability related to the quasi-random visual dots components. By adding very simple rules to the output of this on-the-fly-created compositional environment, we bring on emergent musical process [13]. We call the red lights quasi-random elements as the lights might have a quasi-constant time blinking interval - also called inter-onset if speaking about musical data. The way we decide to group the observed flashing dots can be dynamically modified and has a major influence on the tempo/synchronization decoding part.

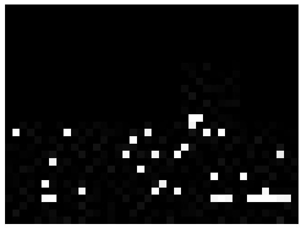
The tempo/synchronization decoding process is based on an inference stochastic framework which translates the visual impulses and turns them into musical rhythmic and textural elements. The way to look at the skyline landscape is manifold because of the too numerous lights, and thus we also keep a varied and open-ended interpretative musical environment that can be turned into either a compositional, improvisational or interpretative context depending on the choices made at a given time by the composer/interpreter of the musical session. Our musical environment is open to different sound artists and musical styles.

We are also considering in a further extension about the possibility of inviting the public to be part of the interpretative musical process. The quality of the interaction we are looking for is not only a simple and basic human-machine interaction, but rather a human to human interaction through machine technology. This idea is important as it considers the technology as a medium delivering and being the information as the same time. The point of view on the machines and their use evolves to a doubtless interactive context involving spectators at the foreground. In this installation, the technology is a factor of communication but does not receive the main focus, as opposed to so many digital art installations which focus on technology more than on the artistic process itself. We situate this project in a new media and technology-dependent way of conceiving and producing artistic processes and also relations between spectators, as exemplified in many work of contemporary art in the last decades [3]. This underlies that the sensorial emotion resulting form the attending experience is more important than the technology itself. Because computer technology and electronic-devices have now become basic elements of everyday life in the emerging ubiquitous society, we aim at designing with them and by them a contextual ambient media experience which might underline and emphasis the natural and interactive links between vision and audition.

(a) Tokyo's red light spotters

(b) after edge detection

(c) thresholding to reduce the number of flashing dots



(d) manageable number of flashing dots for tactus quest

Figure 2: Image processing on Tokyo video capture

4. FLASHING DOTS FOR TEMPO

We will discuss now about some more theoretical aspects that turn the red light spotters into a temporal flow of pulsations needed to drive a musical tempo. We recall some concept of musical time and come up with the foundations of an algorithm for tempo tracking and pulsation prediction regarding separate dots or groups of flashing dots.

4.1 Musical time

The notion of musical time, as simple as it seems to be, can be a surprisingly difficult concept to grasp. Music is composed of events which unfold within flowing time. The way they are organized creates a perceptual sensation of musical time. This organized time oscillates between two extremes, one completely loose and the other completely straight in its organization. Some composers refers to this quality of time as a continuous time and a striated time [2], and even if applied to harmony in a general sense of the music, it applies very well to the musical time. The reality of musical time lies between these two extreme boundaries. What we are interested in is the distribution of time which is closer to the striated time - namely a pulsed time - in which we can perceive a more or less strong cyclic sensation of organized time that we interpret as a pulsation and then as a tempo.

4.2 Input data and cognitive hypothesis

We present in the following the theoretical framework, with input and output data, and the method based on an optimal grid search we are proposing to solve the problem of pulsation estimation and tracking.

We consider the domain of symbolic events occurring in a series of onset time $\{t_n\}$ and focus our research on a temporal set of events which represent the instant polyphony or poly-rhythms of flashing dots. We can have the same approach if we consider the events as onset extracted from audio signal [4] [8].

The main problem to solve is to estimate in real time a pulsation which can be considered as the best matching pulsation of a temporal input sequence. This also imply there is not a single and unique master pulsation representative of the temporal data. Considering phrases of rhythm, this aspect is confirmed by the operation of transcription, when asking people to notate a rhythm they are listening to. When performing the imitation of musical tempo, it is often common to experience octave errors. Simha Arom [1] names one of these best underlying pulses the *tactus*, he also presents the incredible amount of confusing names trying to describe different and not always clear notions related to tempo. The *tactus* is presented as the greatest of the smallest pulses that *inducts* the *tempo*, and which enable the rhythmic structure to unfold on a sometimes unheard and hidden temporal grid. The *tactus* represents an underlying subdivided pulsation you can beat and which enables you to built and understand all the rhythms that are played. This consideration is one of the hypothesis of our research : if there is perception of a tempo, we presume there is a regularity which could be revealed by a filigree grid made by the regular *tactus* pulses. Of course this *regularity* is more or less accurate. When performing music, the tempo is moving, either because of unwanted inaccuracy, or because of a specific musical style. When extracting pulses from flashing dots considered as random variables, it is obvious that the resulting tempo, if found, would likely be non constant and depending on the (possibly changing) group of observed dots. The train of random incoming red-light-pulses is analyzed in real time. The observations are converted into some likelihood's of probable tempi. A Hidden Markov Model (HMM) framework is then use to decode the tempo estimation. The result is a tracking of the tempo and a prediction of the pulsation which can be used as a variable tempo-track to drive some musical hardwares or softwares modules.

4.3 Tactus and quantification

Considering the above hypothesis of perceptual regularity inducting a tempo, we are looking for a regular grid that can materialize the tempo. If we make this underneath grid appear, we also propose a quantification of the temporal events. This quantification is one of the solutions of the rhythmic transcription of the input phrase.

To consider onsets of the input sequence, we threshold the intensity of the flashing dots coming out of the image matrix analysis Fig.**2(d)**: when going above this threshold in a cyclic time, the trigger generate successive attacks, from which inter-onsets.

4.4 Paradigm for the tactus search

[14] proposes a tactus decoding approach based on the inter-onset set $\{o_i\}$ of a time events sequence. In an ideal and mechanical input data case, the tempo is absolutely still, and the *tactus* would result as a *greater common divisor* (*gcd*) of the inter-onset. For a real case, [14] considers the inter-onset to have random deviations, and defines an error function of the *tactus* period Δ and $\{o_i\}$:

$$e(\Delta) = \sum_{i=1}^{n} \left(\frac{o_i}{\Delta} - \left\lfloor \frac{o_i}{\Delta} + \frac{1}{2} \right\rfloor \right)^2 \qquad (1)$$

local minimum of Eq.(1) are potential candidates for the *tactus*. If an exact *gcd* exists, it is given by

$$gcd(o_1, o_2, ..., o_n) = \max\{\Delta | e(\Delta) = 0\}$$

The best candidates of Eq.(1) remains approximated as an empiric threshold is used to select the local minimum. Moreover, the graph of Eq.(1) can be unclear and does not always provide a best estimated candidate, namely Δ^*. The reason for this is that we can't differentiate Eq.(1) because of the integer part. We propose a slightly modified equation to replace Eq.(1) and keep the idea of *tactus* search which is widely used in music transcription [7].

4.5 Optimal grid search

This section regroups the mathematical concepts to solve the quest of the *tactus*. The original *Optimal Grid* approach discussed here is inspired by operational research mathematician Pr. Bruno Simeone [15]. We have choosen not to detail in this paper the proofs of some more technical mathematical results due to the limited space.

4.5.1 Mono-dimensional problem

Let's take a time onset series $\{t_1, ..., t_n\}$. Suppose a regular best approximation grid of step Δ and of length $p+1$ points which includes the origin : $0, \Delta, 2\Delta, ..., (p-1)\Delta, p\Delta$. We can show that we are looking for Δ^* such that it minimizes the following equation

$$f(\Delta) = \sum_{k=1}^{n} \min_{h=0,1,...,p} |t_k - h\Delta| \qquad (2)$$

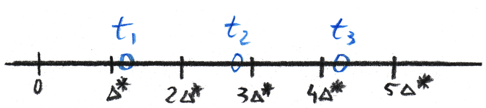


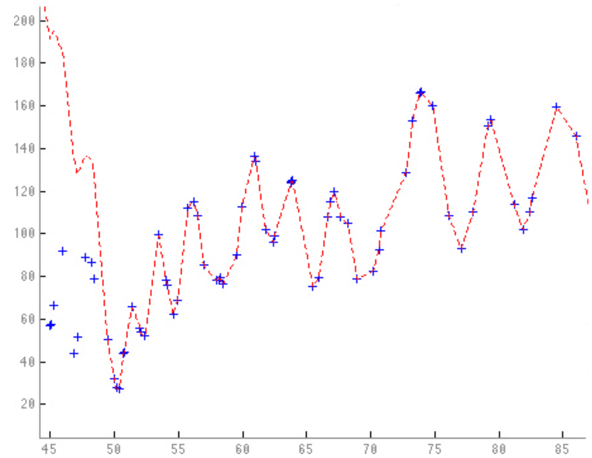
Figure 3: grid example with $k = 3$ et $p = 5$

We demonstrate that in a given close definition domain, Eq.(2) is the sum of n breakpoint functions. We demonstrate also that Eq.(2) admits $n(2p-1)$ singular breakpoints we note α_{kh}. As f is a breakpoint linear function, it find its minimum among the α_{kh}. Through this result, we built an $\mathcal{O}(n^2(2p-1))$ procedure which gives all the local minimum and maximum of f. We are now sure that Δ^* is among the finite number of candidates which consist of the local maxima and minima. We have solved Eq.(2) without thresholding. p, the length of the grid, is also a parameter that can be seen as a *quantification resolution* for the transcription. Figure 4(a) show a reconstruction of f with singular breakpoints in comparison with its numerically approximated computation.

4.5.2 Phase of grid and bi-dimensional problem

The Δ step grids we have found above include the origin point. But there is no requirement for the optimal grid to include the origin point. This leads us to extend Eq.(2) to a bi-dimensional equation of variables *step* Δ and *phase* or *centrer* s.

Let's consider a unidimensional grid of step Δ and center s : $\{... , s-2\Delta , s-\Delta , s , s+\Delta , s+2\Delta , ...\}$. For a given Δ, we want to find a center s^* which minimizes the grid phase function as the following :



(a) building f with $p = 26$. the best Δ^* candidate is 50.4 . The breakpoints are the $+$, dashed-line is the numerically approximated function

Figure 4: singular breakpoints of an onset $\{100, 203, 406, 702, 989, 1310\}$ sequence

$$\varphi(s) = \sum_{k=1}^{n} \min_{h\in\mathbb{Z}} |t_k - s - h\Delta| \qquad (3)$$

φ is Δ-periodic. We show that it reaches its minimum in a point of the form $s^* = t_q - l\Delta$ with $q \in \{1, ..., n\}$ and $l \in \{m_q, ..., M_q\}$ related boundaries we can also define.

We prove then the following theorem:

THEOREM 1. *Let's consider a Δ-step grid with center s^* which best coincides with a series of n points $\{t_1, ..., t_n\}$. Then there is at least one integer q^* , $1 \leqslant q^* \leqslant n$, such that t_{q^*} belongs to the grid.*

We will now use *optimal coincidence* or *best approximation* as equivalent terms.

From what precedes, we can reduce to a surface I the space area where to look for a minimum of the bi-dimensional function :

$$\varphi(s, \Delta) = \sum_{k=1}^{n} \min_{h\in\mathbb{Z}} |t_k - s - h\Delta| \quad , \quad \Delta \in I \qquad (4)$$

where I is a truncated triangle surface. Figure 5(a) and 5(b) show the periodicity observation and an overview of the local minimum and maximum of the *indices of coincidence* - indicators of how well a grid quantifies the input points - of Eq.(4) in its reduced definition domain.

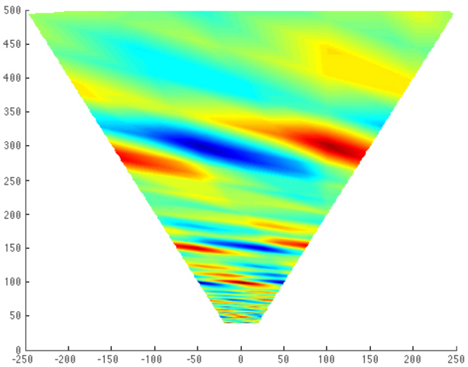
4.5.3 Bi-dimensional coincidence problem

With theorem Th.(**1**), we propose an $\mathcal{O}(n^3p^2)$ procedure to solve Eq.(4). We give in the Alg.(**1**) pseudo-code the resulting bi-dimensional optimization algorithm.

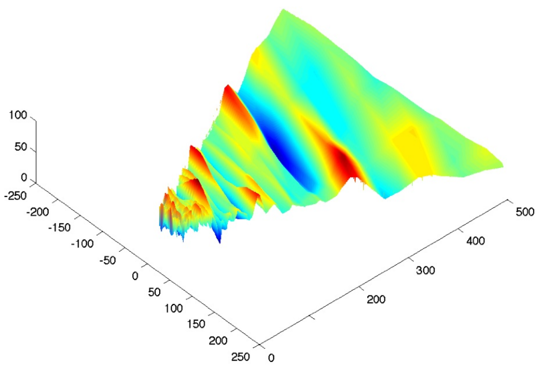
4.6 Probabilistic model

4.6.1 Observations and optimal grid search

The optimal grid search procedure provides us with some observations, namely some good candidates $\{\Delta_i\}$ and associated phase $\{s_i\}$. In dynamic programming environment,



(a) periodicity of φ on $\mathcal{I}(s,\Delta)$ and *indices of coincidence*



(b) detail for the optimal couple (s,Δ) with *indices of coincidence*

Figure 5: Approximation of φ on $\mathcal{I}(s,\Delta)$ for the sequence $\{100, 203, 406, 702, 989, 1315\}$

regarding a given observation window length, we obtain a series of observation vectors $\overrightarrow{\Delta}$ we consider as local density probability function.

4.6.2 Probabilistic HMM model

In a real-time context, we are interested in estimating and following the tempo to predict the next pulse. We use a Hidden Markov Model (*HMM*) to decode and infer the tempo-states transitions. The optimal grid paradigm provides us with some dynamic states sequences, updated on each new onset arrival. The unknown state is the tempo, given by Δ. We select some Δ candidates under the *index of coincidence* criterion. They provide the *likelihood probability* of the associated tempo for the observed sequence of time onset. This selection is done to avoid too many states among which too many would be improbable. It reduces also the amount of computations we need for the real-time implementation. With the selected Δ candidates we build on-the-fly a transition matrix between the states. We fill in the states transition probability matrix with gaussian probabilities centered on each candidate with respect to the others. This means the distance between two states affects directly their transition probability. We also use a log_2 scale for the tempo in order to keep equal probabilities for acceleration

Algorithm 1 Bi-dimensional best approximation grid

Require: $\{t_1, ..., t_n\}$sequence of n input points
p number of points of grid plus center point

for $q = 1$ to n **do**
  for $l = 0$ to p **do**
    for $k = 1$ to n **do**
      for $h = 0$ to $p, h \neq l$ **do**
        $\Delta_{kh} \leftarrow \frac{t_k - t_q}{h - l}$
        $\Phi_{kh} = \sum_{j=1}^{n} \min \left\{ frac(t_j - t_q, \Delta_{kh}), \Delta_{kh} - frac(t_j - t_q, \Delta_{kh}) \right\}$
      end for
    end for
    $\Phi_{ab} \leftarrow \min_{h,k} \Phi_{hk}$
    $\Delta \leftarrow \Delta_{ab}$
    $\Phi_{ql} \leftarrow \sum_{k=1}^{n} \min_{r=-l,\ldots,p-l} |t_k - (t_q + r\Delta)|$
  end for
end for
$\Phi_{q^*l^*} \leftarrow \min_{q,l} \Phi_{ql}$
$\Delta^* \leftarrow \Delta_{q^*l^*}$
$s^* \leftarrow t_{q^*} - l^*\Delta^*$
return s^*, Δ^*

and deceleration. The variance is chosen so as to allow more or less narrow tempo changes, a typical value being around 1% of the current tempo.

4.6.3 Tempo decoding, inference and prediction

Once we have defined the probabilistic model, we decode it with a forward Viterbi algorithm [12]. If we have no information on the initial tempo - as it would be mostly probable with the random flashing dots - we assume an initial flat prior. Running the HMM decoding in a dynamic programming environment gives us the tempo estimation and in the same time the phase of the *tactus* pulsation. As a direct consequence we can predict the anticipated tactus pulsation. It is interesting to note that the intrinsic phase computation of the optimal grid algorithm is used straight-forward as soon as we have made a decision for the tempo-state. To turn our paradigm into an anticipative behavior mode, we use the predicted pulse onset in the computation of the next incoming data : incorporating a future and forecast onset event into the present computation bring the algorithm in the anticipation domain [6].

4.6.4 Tempo tracking real-time implementation

We have chosen the real-time *MaxMSP*[4] and *PureData*[5] audio/control environment to implement the optimal grid algorithm as an object named *antepulse*, referring to pulse anticipation. The object is written with the *flext* `C++` facilities programming library which enables to share the code for *OS X* and *Windows* computers, both for *MaxMSP* (Fig.**6**) and *PureData* environments.

[4] `http://www.cycling74.com/`
[5] `http://puredata.info/`

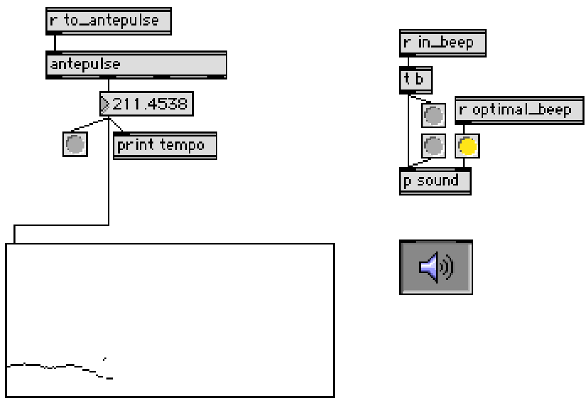


Figure 6: Antepulse in the *MaxMSP* environment

5. REAL-TIME FRAMEWORK

We present in this section some of the technical choices made in the *Red Light Spotters* installation.

The image analysis and its rhythmic decoding is done in the real-time *MaxMSP* environment. The image processing makes use of a specific edge detection algorithm [16] implemented in the *MaxMSP* graphic library *Jitter* in [10]. We end up with some rhythmic flux that can be used to drive other musical environments such as *Ableton Live*[6], *Reaktor*[7], *MaxMSP* or others, provided they support the *OSC*[8] communication protocol, which enable dataflow through Ethernet protocol. The temporal decoded pulsation is sent by *MaxMSP* as a tempo-track for tempo synchronization, or can be received by *OSC* Plug-Ins in some clients applications, so as to be used freely for tempo and music generation. As the light might vary through the night, and as the phases of the lights are always moving, the resulting tempo is also moving accordingly to the input pulses, creating new variations of tempo and rhythms.

6. MUSIC AND STYLES

The musical hybrid compositional and interpretative system we propose is open in the sense that it is independent of any given music genre or style. Our intent is to let the musician/artist free to use the sounds, samples, audio effects and musical textures and combine them as he likes. We are providing a framework where this musical process is easy to work with. Concerning the performative aspects, several control devices such as touch screens could be used to create an interactive interpretation of the mixing or sound control processes.

By the use of basic interaction rules applied to the parallel decoded tempo - like reaching some point together, being close to the same speed etc - we achieve some emergent behaviors which generate some musical forms.

It has to be noted that due the rhythmic sound-generation process, the most obvious music that can be created with *Red Light Spotters* is clearly in the spirit of Steve Reich' minimal music, e.g. early works *Piano phase* (1967) or later compositions such as *Six marimbas* (1973-1986) and *Music for 18 musicians* (1974-1976), although we are currently investigating a more contemporary style, also influenced by electro and dub style. Other musical styles could of course be investigated, but will require more development time in order to adapt and connect the rhythm generators to meaningful musical processes. Also it is worth noticing that as the rhythm generators used in *Red Light Spotters* can be considered as quasi-random, another influential style is the so-called random music, as pioneered in the 50's by John Cage with such musical pieces as *Music of Changes* (1951), in which the musical score is written using chance operations, and even more *HPSCHD* (1969), which uses computer-generated chance operations real-time for various sound sources. Indeed the technology available nowadays makes it possible for a more flexible and complex interactions.

7. INTERACTION

In the field of digital arts, the concept of interaction has repeatedly been identified as a fundamental characteristic and this notion is one of the fundamental advance brought by the use of computers in art installations [9, 5, 11]. In the paradigm of the interface, interaction has often been considered as a necessarily reduced and incomplete dialogue between the human and the machine, as a means of access that will always be frustrating since it is imperfectly codified according to the digital reality hidden at the core of the machine. In taking up this route, numerous artists have done their utmost to devise interfaces that are more or less natural to allow for an improved interaction with their digital works, as if the viewer's immersion should necessarily go through a complex technological apparatus. That approach, however, forgets that immersion is cognitive before being perceptive, the *reality* of a work clearly being invented and recreated by the viewer and not just perceived and undergone. A few artists have designed sound installations linking perception and imagination.

A pioneering digital artwork using contextual information and real-time generated music is *Ping* by Chris Chafe and Greg Niemeyer, first presented in the *Art and Technology* exhibition at San Francisco MOMA in 2001. The title refers to the `ping` command in Unix, which can measure the time needed to reach any specific TCP/IP address on the internet. Therefore in this work, several rhythms are created by several distinct computers, each one *pinging* a specific TCP/IP address. This environment could hardly be defined as musical but somehow aimed at representing in an auditory manner some (partial) topology of the internet, resulting in an intuitive notion of *close* and *distant* in this virtual network. Another more recent example is *Emotional Traffic (e-traffic)* by Maurice Benayoun, exhibited at V2, Rotterdam, in 2005. This work uses the measurement of the number of occurrences of some words representing emotions (like hope, fear, satisfaction, etc) in search engines on the web to produce a sound environment echoing the *emotional state* of the planet at a given instant. These words were also graphically represented within visual representations on projected screens.

In *Red Light Spotters*, we do not want to make explicit the mediation between the ambient context and the generated music. It has to be intuitive and immersive and therefore the fact that this is a *s*ituated installation, within an effective landscape that could be visually explored is utmost

[6] `http://www.ableton.com/`
[7] `http://www.native-instruments.com`
[8] `http://opensoundcontrol.org/`

important. The play between the natural visual stimuli and the artificial, computer-generated audio stimuli are one of the key features of this work, but they might remain fuzzy in the mind of the spectators. We do not aim at having, for instance, the spectators identifying each particular blinking light and its association to a precise sound. Nor do we need to explain to the spectators the details of the tempo-tracking machinery or of the light-grouping associations. This is more a *global* interaction: the nebula of the flashing lights creates a nebula of sounds and one can find an intuitive link between them, but there is no need to further explicit the precise links. If willing to do so, the spectator can use his eyes and ears to track down any particular relationship. In *Red Light Spotters*, the interaction that is created consists of a multi-sensory relationship between an external context, the urban landscape offered to the eyes of the spectators, and an internal exhibition space filled with rhythmic music, mediated through tempo-tracking, sound samples and audio-effects associations.

8. CONCLUSIONS

Red Light Spotters is an interactive sound installation that turns an ambient dynamic urban landscape into real-time generated musical soundscape. A key-point is the extraction of tempo in real-time for the groups of blinking lights that form rhythm generators, as this symbolic information can be used in many ways: not only to control the actual tempo of a musical instrument but to influence any parameter in the real-time sound generation (e.g. the waveform of an oscillator, the reverb delay, etc).

Another interesting point is that this installation is very flexible: it can either creates a sound environment in a purely *automatic* mode, or allow for a real-time performance by a musician or sound artist who could modify in real-time the grouping of blinking lights and the association between those rhythm generators and sounds. It can also make room for a mixed approach, where a musician blend his own music with the *Red Light Spotters*.

In its current stage, this installation is closely linked to the Tokyo urban landscape (Fig.**7**) and thus to be shown during night time in some Tokyo city view observatory, but could be further developed and adapted to other cities and urban landscape, as the only thing is to have some rhythmic flow that has to be captured by the camera sensors and analyzed by the tempo-tracking algorithm. It can be for instance the flow of people or cars in a particular urban location.

Figure 7: Tokyo's Red Light Spots

9. REFERENCES

[1] S. Arom. *Polyphonies et polyrythmies instrumentales d'afrique centrale, structure et méthodologies*. Selaf, 1985.

[2] P. Boulez. *Penser la musique aujourd'hui*. Gonthier médiations, 1963.

[3] N. Bourriaud. *Relational Aesthetics*. Les Presse Du Reel, Paris, France, 1998.

[4] N. Collins. *Beat Induction and Rhythm Analysis for Live Audio Processing*. 2004.

[5] E. Couchot. *La technologie dans l'art: de la photographie a la realite virtuelle*. Jacqueline Chambon, Paris, France, 1998.

[6] D. Huron. *Sweet Anticipation: Music and the Psychology of Expectation*. in press, 2006.

[7] A. Klapuri and M. Davy. *Signal Processing Methods for Music Transcription*. Springer US, 2006.

[8] T. H. A. Kristoffer Jensen. Real-time beat estimation using feature extraction. In *Computer Music Modeling and Retrieval*, volume 2771/2004, pages 13–22. Springer Berlin / Heidelberg, 2004.

[9] C. Paul. *Digital Art*. Thames and Hudson, London, U.K., 2003.

[10] J.-M. Pelletier. A shape-based approach to computer vision musical performance systems. In *NIME*, pages 197–198, 2004.

[11] F. Popper. *from Technological to Virtual Art*. MIT Press, Cambridge, USA, 2007.

[12] L. Rabiner. A tutorial on hidden markov models and selected applications in speech. *Proceeding of the IEEE*, 77:257–286, 1989.

[13] J. Seibt. Free process theory: Towards a typology of occurrings. *Axiomathes*, 14:23–55, 2004.

[14] J. Seppanen. Tatum grid analysis of musical signals. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.

[15] B. Simeone. Optimal grid search. private communications, 2006.

[16] S.-W. L. Tae-Yong Kim, Jihun Park. Object boundary edge selection for accurate contour tracking using multi-level canny edges. In *Image Analysis and Recognition*, volume 3212/2004 of *Lecture Notes in Computer Science*, pages 536–543. Springer Berlin / Heidelberg, 2004.