



HAL
open science

Inverse problem in sound synthesis and musical creation using mass-interaction networks

Jerome Villeneuve, Claude Cadoz

► **To cite this version:**

Jerome Villeneuve, Claude Cadoz. Inverse problem in sound synthesis and musical creation using mass-interaction networks. 9th Sound and Music Computing Conference, Jul 2012, Copenhagen, Denmark. pp.49-54. hal-00991405

HAL Id: hal-00991405

<https://hal.science/hal-00991405>

Submitted on 15 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INVERSE PROBLEM IN SOUND SYNTHESIS AND MUSICAL CREATION USING MASS-INTERACTION NETWORKS

Jérôme Villeneuve

ICA laboratory,
Grenoble Institute of Technology (France)
jerome.villeneuve@imag.fr

Claude Cadoz

ICA laboratory, ACROE,
Grenoble Institute of Technology (France),
Ministère de la culture et de la communication
claude.cadoz@imag.fr

ABSTRACT

Sound synthesis with mass-interaction physical modeling networks can be considered as a general paradigm capable of being the central part of complete software environments for musical creation. GENESIS 3, built around the CORDIS-ANIMA formalism and developed by ACROE/ICA Laboratory, is the first environment of this kind. Using it, the artist may be facing an inherent problematic of every creation process: how to use a given tool in order to obtain an expected result. In our context, the question would be: “Considering a sound, which physical model could produce it?” This paper especially aims at presenting the frame in which this inverse problem is set and at establishing its very own inputs and outputs. However, we will also present two different algorithmic resolutions applied on quite simple cases and then discuss their relevance.

1. INTRODUCTION

Sound synthesis is a matter of numbers. Max Mathews, its inventor, noted that a sound is nothing else than a pressure as a function of time, which, for the computer, means a stream of numbers, quite a large quantity of them each second. Since these first statements, a substantial amount of methods for generating those numbers and thus, sounds, has been developed. The physical modeling synthesis [1], [2], [13], is a category designating such methods. Our work takes place in the ongoing researches and development of one of them: GENESIS [4], in ACROE-ICA laboratory. This distinctive environment provides a mass-interaction network modeling approach and relies on the instrumental interaction paradigm. It is built around the CORDIS-ANIMA (CA) formalism [3] and is meant to be used with force feedback gestural transducers (TGR).

1.1 CORDIS-ANIMA

The CA formalism is a really effective solution for numbers production and suggests an original approach for controlling it. Its earliest algorithm (1), a second-order linear homogeneous recurrence relation, is a very simple “digital oscillator.” To control it in an intuitive way, it was used to describe a discrete version of the mechanic harmonic oscillator (Fig. 1). This analogy allows then a

sound generation using physical parameters (M for mass, K for stiffness, Z for damping), initial conditions (position or velocity) and more generally by making the Newton second and third laws ruling every elementary component of the formalism.

$$X_n = A \cdot X_{n-1} + B \cdot X_{n-2} \quad (1)$$

$$(A, B) = f(M, K, Z)$$

Where X_n represents the position of the moving part at each step.

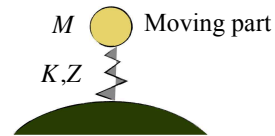


Figure 1. Model of a mechanic oscillator and its parameters

This concept is then extended with the notion of external forces.

$$X_n = A \cdot X_{n-1} + B \cdot X_{n-2} + C \cdot F_{n-1} \quad (2)$$

Finally, the algorithm (2) is split and distributed among each element, which we shall call modules (Fig. 2). Hence, they are distinguished in two categories regarding to the nature of their algorithms: The <MAT> kind modules have a force input, a position output and designate “matter.” As an example, MAS modules are comparable to punctual moving masses, SOL to an infinite inertia. The <LIA> kind modules have positions inputs, forces outputs and allow interconnections between <MAT> modules. As an example, REF modules are comparable to visco-elastic interactions and BUT modules to percussive interactions.

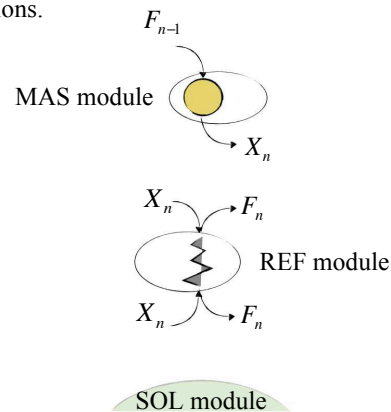


Figure 2. CORDIS-ANIMA formalism

Copyright: © 2011 First author et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

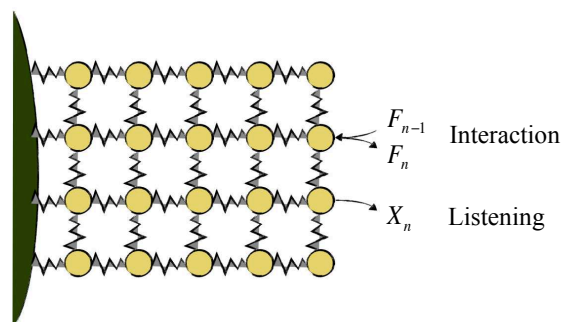


Figure 3. CORDIS-ANIMA model – Interact & Listen

CORDIS-ANIMA is thus characterized by its modularity, it allows building large models, by putting together basic physical elements. Those models are networks of interconnected moving matter. It also natively permits to interact with models and of course, “listening” to them by making the evolving position of chosen masses the normalized position of the loudspeakers membranes (Fig. 3).

1.2 Sound synthesis using GENESIS

Any sound creation approach by the mean of GENESIS thus supposes, at first, the elaboration of a mass-interaction model, according to the CA formalism. Each module receives a set of parametrical specifications. A simulation is then launched and a sound emerges. We shall call “direct approach” this initial process, from the model elaboration and setting to the sound phenomena production. We can imagine that this process leads to a permanent round trip between the emergent sounds, the model, and some adjustments made under both quantitative and qualitative constraints. A real difficulty is to establish relations between them. In the direct approach, we shall try to postulate the incidence of one or another model’s parameter on the perceptive attributes of the produced sound. Since its first actual implementation in 1996, numerous investigations and experiments were realized in the GENESIS environment in order to gather as much knowledge as possible on this matter and to suggest tools and pedagogical material to the users.

But this relation may be established differently and we can imagine an “inverse approach” as follow. If the users are able to describe precisely a sound phenomenon that they intent to synthesize, is there any means to compute and suggest them a model producing such a sound?

2. THE INVERSE PROBLEM

There is a large variety of inverse problems applied in numerous different disciplines. Most of the time, there is no direct access to a system that has to be defined. The only available data to rely on are what such a system produces. The inverse problem is then to theorize its nature considering those data. To achieve that, a lot of attention must be paid on a good evaluation and formulation of this very inverse problem.

2.1 Information, Determinism & Stability

Here follows a brief highlighting of three important notions.

Firstly, we might have to consider the amount of information available by gathering data that a system produces and whether or not it is sufficient for its description.

Secondly, in physics, a system is said “deterministic” when, for one of its configurations, it always has the same observable effect. But, since it can be hard to indentify all the components of a system and their relative influence on its global behavior, it is hardly possible to conceive an absolute determinism. Though, framed by physical modeling and digital simulation, and relying on the CORDIS-ANIMA formalism and algorithms, we can establish a more clear-cut definition of determinism. Indeed, all the parameters of a CA model are perfectly known and then both the model and the simulation are absolutely deterministic. The simulation of a given model always leads to the exact same signal. A given cause will always produce strictly the same effect. But, our purpose is to reverse this causality and in this case there’s no more determinism. Even with the finest analysis methods, two different causes can have the same measured effects. As a trivial example, the second order mechanical oscillator model in CORDIS-ANIMA will give two bitwise equivalent digital signals, if all its parameters are multiplied by the same coefficient. Consequently, it is necessary to consider that a given effect can have multiple and different causes. However, towards the context in which takes place our study, the research for a unique generative model of a given sound phenomenon is not required. What matters is to obtain at least one of them. If a choice has to be done, we might consider the most precise, the most physically coherent, maybe the most stable, efficient and usable model.

Thirdly, a model is said “stable” if, for very slight modifications of its parameters, the simulated effect is also slightly modified. In the case of CA models, it is not impossible to obtain really ill-conditioned (unstable) models when using non-linear interaction or negative “physical” parameters. Our intention is to provide models with which the users may play by making some changes on its settings and therefore take a look at sound phenomena close to their first idea. We can see here the particular interest of this approach, articulating the direct and inverse explorations in the general creation process. In conclusion, this forbids unstable models.

2.2 “Correctly-Set” Inverse Problem

Potentially unstable models, breach of determinism, and lack of relevant information, are unavoidable issues with which we have to deal. They will define the frame of our inverse problem and its formulation. Moreover, on that matter, J. Hadamard [6] says that we can consider a problem such as ours “correctly-set” on the condition that: 1) It has one and only one solution. 2) That this solution dependence on input data is continuous on a reasonable range.

Our first problem is not the “inverse problem” in itself, but its definition. We will afterward try to clearly establish what we shall expect from inverse resolution and of course, what data shall be given as an input.

3. SPECIFICATION OF OUR INVERSE PROBLEM

3.1 Outputs

In the real world, as within the GENESIS environment, producing a sound is associated with an energy exchange between two bodies: a “source” body that will be the “instrumentalist” and an “instrumental” body. The energy delivered by the instrumentalist onto the instrument will cause a modification of its state and its nature will characterize the resulting sound phenomenon. Of course, the applied gesture, the interaction between these two bodies is also significant in the specification of this sound. This trilateral configuration (Fig. 4) is necessarily involved in the sound phenomena production.

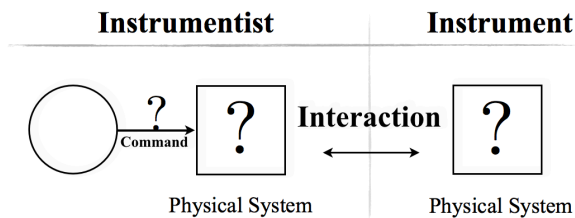


Figure 4. Generative configuration of sound phenomena production.

Therefore, any sound must be considered as resulting of such a configuration. As an output, the inverse analysis of a sound must lead to an instrumentalist model interacting with an instrumental model.

Furthermore, this generalization can be even wider if we decide to integrate what causes the instrumental gesture: the instrumentalist “acting decision.” Indeed, as we consider the instrumentalist component as an active mechanical system, we shall suppose it is fragmented in two components, one corresponding to purely physical system and another acting as a trigger. Here, the important aspect is that while the interaction between the physical system representing the body and the instrument must be considered as bilateral, the relation between the command and the instrumentalist physical system will be considered as unidirectional.

To sum up, two fundamental hypotheses define what's expected from a valid solution for the most general inverse problem within the framework of GENESIS: 1) The considered sound phenomenon results from the action of an instrumentalist (by gestures) on an instrument. 2) The instrumentalist is a physical system unilaterally controlled by "acting commands."

We have to keep these imperative considerations in mind as we seek for solutions. Nevertheless, in the first applications we present (developed in part 4. **First experiments and results**) this level of expectation is not yet satisfied.

3.2 Typology of the Inverse Problem Inputs

If the nature of the inverse problem outputs is relatively well defined and framed by the GENESIS environment, it isn't the case for its inputs. We thus sought the most exhaustive typology of their varieties and more generally of all the possible methods for describing sound phenomena (Fig. 5). First of all, we must distinguish two kinds of inputs:

- The low level inputs, directly in entry of the inverse resolution. Actually, they would be the parameters initiating the algorithmic resolution. We could think of numerical values identifying frequency, amplitude and damping time in the case of a simple signal.

- The high level inputs, such as the records of sound phenomena [12], the score or maybe the expression of what the artist could mentally imagine and want to produce. This type of entry will obviously correspond to a more common usage of the sound phenomena description. It is thus necessary to imagine adapted analysis tools allowing conversion from this kind of inputs to the first one.

This distinction reminds one of the one Nattiez made when he dealt with the notion of music semiology [11]. By extending his idea to the sound phenomenon analysis, without debate on musicality questions, we can identify esthetic and neutral levels. The neutral level would describe the inputs existing independently of perceptive notions (low level inputs, recording, sheet music). The esthetic level would correspond to abstract and subjective considerations. (The poietic level would appear essentially in the direct approach of musical creation with GENESIS).

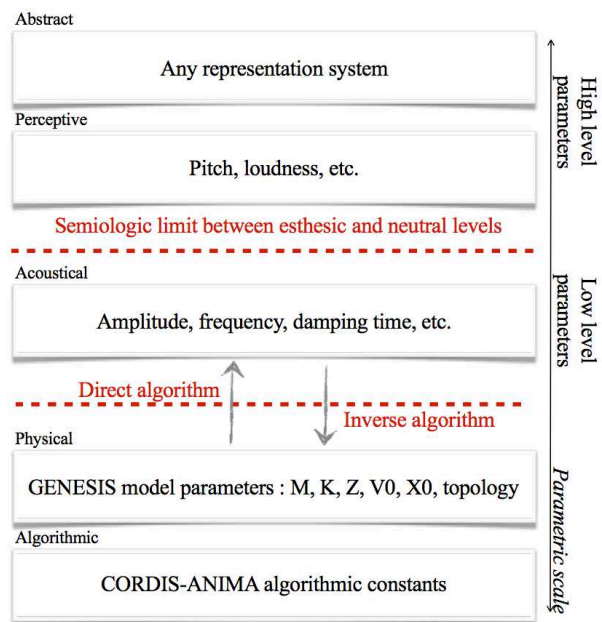


Figure 5. Parametric scale

4. FIRST EXPERIMENTS AND RESULTS

As a first practical approach of the inverse problem, let's consider the simplest CORDIS-ANIMA model as an

output: the harmonic oscillator as pictured in Figure 1. It is made of a mass (MAS) with an inertia parameter “M”, which is connected to the ground by a visco-elasticity defined with stiffness and viscosity parameters: “K” and “Z.” Furthermore, it oscillates within a single degree of freedom. This trivial object is well known and theorized in physics, the mathematical equations describing it are completely usable in our case on the condition that the shift from continuous to discrete expressions is quantified. Moreover, we focus only on the instrumental part of the problem.

Solving these equations knowing M, K and Z gives us its modal characteristics, i.e. its eigenfrequency and its damping time (amplitude and phase depending on initial conditions). We can easily reverse those linear equations and then find, with a frequency, a damping time, amplitude and phase, a description in inertia, stiffness, viscosity and initial conditions of a cell carrying these exact modal characteristics. We thus have an infinite number of solutions. To choose among them, K or M must be set.

In the simplest cases, a one frequency, over damped, critically damped or under damped signal, this mathematical reversal is a first resolution of the inverse problem considering low-level data. A direct extension of it would be to recompose complex sounds by reproducing what’s done in additive sound synthesis. Knowing the modal characterization of a sound phenomenon, we could recompose it by synthesizing as many sinusoidal signals (replicating as harmonic oscillators) as necessary. Nevertheless, as specified before, it is not what we expect of this resolution. We need a coherent model with interdependent components.

More complex models can be analyzed as well. Damped spring-mass modal systems analysis is a well-documented issue. Actually, a tool allows to compute the modal characteristics, a modal model, of a given CORDIS-ANIMA model, by solving the discrete version of equation (3).

$$M(\ddot{X}) = -K(X) - Z(\dot{X}) \quad (3)$$

With X , the position vector, M , K and Z matrices containing the masses, stiffness and viscosities information of the model. A model of n masses has an equivalent modal model of n independent oscillators and thus n eigenmodes. If we still consider a low-level input composed by three vectors (n amplitudes, n frequencies and n damping) then the instrumental part of a generative model must have at least n masses, and we can suppose that n masses are enough to define it. Thus, we suggest to choose its structure (also named “topology”). The simplest coherent model of n masses is then the linear “chain.” A “string” fixed at its extremities such as in Figure 6, at one extremity or none.

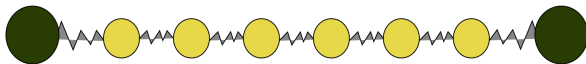


Figure 6. CORDIS-ANIMA linear chain model.

The instrumental part of the generative model will be considered later. The reduced inverse problem is then to define, given 3 vectors of n modes (or calculated fre-

quencies), n amplitudes and n damping time, a generative CA string of n masses connected by visco-elastic interactions. Such a topology imposes a certain nature on matrices M , K and Z defining the equation (3). M will be diagonal; K and Z will be tridiagonals as shown in (4). All the values contained in those matrices are the unknowns of our problem.

$$K = \begin{pmatrix} k_1+k_2 & -k_2 & 0 & \dots & 0 \\ -k_2 & k_2+k_3 & -k_3 & \ddots & \vdots \\ 0 & -k_3 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -k_n \\ 0 & \dots & 0 & -k_n & k_n+k_{n+1} \end{pmatrix} \quad (4)$$

Two different approaches have been followed. The first one, a numerical resolution, relies on an adapted optimization procedure. The second is algebraic.

4.1 Numerical Partial Resolution

This first attempt needs us to modulate our goal and just to focus on the frequencies input vector. For that, we have to assume that there is neither viscosity (because no damping time input) nor amplitude considerations. Thus, Z , is a null matrix and all masses are set to 1, M is then an identity matrix. We are now focused on the K matrix (4) and its characteristic polynomial (5).

$$P(\lambda) = f_1(K) + f_2(K)\lambda^1 + f_3(K)\lambda^2 + \dots \dots + f_n(K)\lambda^{n-1} + f_{n+1}(K)\lambda^n \quad (5)$$

Solving this equation by knowing all the coefficients contained in K would give us all the eigenvalues λ .

Here, frequencies are given allowing to calculate λ values according to (6).

$$\lambda_i = -2 \cdot (1 - \cos(\frac{2\pi \cdot f_i}{F_e})) \quad (6)$$

The unknown is K and we have to solve the equation (7) for all these λ values.

$$P_i(K) = f_1(K) + f_2(K)\lambda_i^1 + f_3(K)\lambda_i^2 + \dots \dots + f_n(K)\lambda_i^{n-1} + f_{n+1}(K)\lambda_i^n \quad (7)$$

For that, we aim to minimize $R(k)$ using Gauss-Newton algorithm combined with a gradient descent (8). This method is called Levenberg-Marquardt Algorithm [8], [9], [10], and is named after its inventors.

$$R(k) = \sum_{i=1}^n (P_i(K))^2 \quad (8)$$

This algorithm is quite sensitive to the initial conditions that have to be provided and might need some time to reach a solution. Its results will be discussed in the “4.3 Results” section.

4.2 Algebraic Resolution

This other method is separated in three parts: the first one handling the eigenmodes, the second one treating the

amplitudes of these and finally, the third one approaching their damping time.

4.2.1 Frequencies resolution

In this first part, viscosities are set to zeros. We thus suppose that M and K matrices carry the frequency information. The mathematical method used is similar to that detailed by G.M.L. Gladwell in [5]. It relies on the reconstruction of a Jacobian matrix (tridiagonal matrix including positive diagonal elements and negative co-diagonal elements, such as K in (4)) by application of an inverse version of the Lanczos algorithm [7].

$$0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_n \quad (9)$$

It needs a vector of n frequencies interpreted with (6) and verifying (9), and to be initialized by a vector of n non-zero values on which we shall return later. Matrices M , K and transformation matrix (which contains modal shape information) will then be directly interpretable.

4.2.2 Amplitudes resolution

Within the framework of the physical modeling, the amplitude, or we shall say the relative amplitude of each eigenmode, depends on how the energy has been distributed among them. But as mentioned before, the instrumental part of the generative model is inert. To make it move (to bring energy to it), it is necessary to introduce an instrumentalist part interacting with this instrumental part. Thus, we chose a simple interaction: the collision between a mass with an initial velocity and one of the linear chain's masses. Therefore, we have one structure of a given topology with fixed parameters (masses and stiffness), which has precise eigenfrequencies, and which we shall "excite" by percussion. Our goal is to control the amplitude of each one of these eigenfrequencies.

We can quickly make an analogy between this problem and that of piano makers when they chose to strike the strings of the instrument at a certain distance ($1/7$ and $1/9$ of the string) to enrich its harmonic spectrum. The amplitude of every mode directly depends on where the model is going to be "played." Furthermore, in GENESIS, users do not capture the whole sound produced by a model but proceed to a local listening of it. The listening point(s) of our structure will be an additional parameter.

We managed to solve this amplitude issue by going back to the frequencies resolution. Indeed, we previously spoke about a vector of n values necessary for the Lanczos algorithm initialization. This vector actually contains the first value of each eigenvector of the model to be computed (Fig. 7). Furthermore, the modal shape carries all the information necessary to estimate the relative amplitude of every mode and at any possible combination of excitation and listening points.

But, if we excite and listen to our model on a certain mass (here the first one), all we need to estimate the relative amplitude of each mode is the very vector of n values that we have to provide! Amplitudes and frequencies have to be considered in the same stages of the resolution.

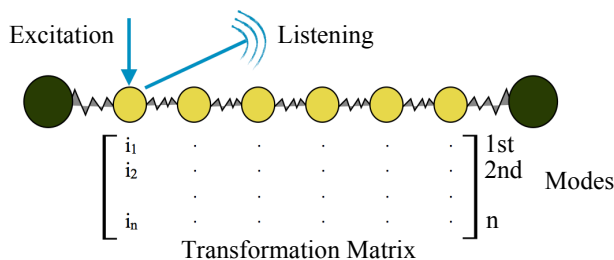


Figure 7. Representation of the transformation matrix and how to interpret it regarding to the model and its eigenmodes. (i_1, i_2, \dots, i_n) being the initial vector to provide

4.2.3 Damping times consideration

The trivial analysis of a simple model as for example a two masses chain, allows us to identify the constraints imposed by the physical coherence of the damping time of one or several modes. We particularly notice that the second eigenmode of such a structure cannot have a damping time longer than that of the fundamental mode unless we tolerate values of negative viscosities (Fig. 8).

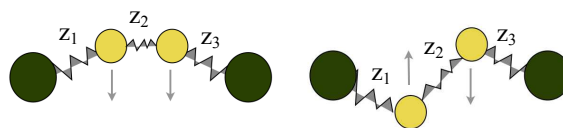


Figure 8. Representation of first (left) and second mode (right) of a two masses chain. If $z_2 = 0$: the two modes have the same damping time. The second mode can have a longer damping time than that of the first mode only if $z_2 < 0$.

Furthermore, it is impossible to preserve the precision of the frequency resolution by the Lanczos algorithm when we try to consider the viscosity parameters. To obtain the most coherent generative model, we thus ask the user to choose only one value of damping time, that of the fundamental mode. It also allows a better control of the viscosity influence on each mode frequency.

4.3 Results

The numerical resolution allows to efficiently obtain generative models defined with 10 or 15 frequencies. Its time of resolution increases dramatically with the number of frequencies to deal with. Its main advantage is that it doesn't depend on the kind of topology you may expect for the output model. It is also convenient to have all masses set to 1.

The algebraic approach is currently suitable for fixed, half-fixed or free linear chains. Any other topology for the output model would need to have its dedicated algorithm. However, with this approach we managed to obtain generative models defined by 40 frequencies for harmonic or even complex spectra. The global quality of this resolution, as well as the coherence of the resulting model, relies on the relative proximity of the wanted frequencies.

As an example, if 17 frequencies are harmonically sorted from 150Hz to 2550Hz, the resulting chain will have masses separated by a factor of two and stiffness values separated by a factor of one hundred, the frequency precision is close to $\sim 10^{-7}$ Hz.

If among these 17 frequencies, two of them get closer (to a 1Hz difference), then the resulting chain will have masses separated by a factor of one thousand and stiffness values separated by a factor of one million, with a frequency precision close to $\sim 10^{-2}$ Hz. Regarding the precision of frequency resolution, it can be adapted to the limits of our perception and of our frequency discrimination capacities.

Viscosities do have an influence on the modal model spectrum. But, they are quite insignificant, and again, this influence has to be evaluated regarding perceptual matters.

Furthermore, these two approaches are already implemented and used in the GENESIS environment.

5. CONCLUSIONS

This preliminary work on the inverse problem general definition, as well as the first concrete applications on very simple cases, allows us to set the bases of a vast and complex problem areas and to demonstrate its relevance thanks to some encouraging results.

The next stages are already framed. Still in simple cases, one of them will consist in separating the structural and event-based contributions from the generative model regarding to an input signal. Afterward, we could picture a complete automated inverse resolution from extended inputs (digital audiofiles) to CORDIS-ANIMA generative models.

6. REFERENCES

- [1] J.M. Adrien, "The Missing Link: Modal Synthesis", in G. De Poli, A. Picalli, and C. Roads, eds, *Representations of Musical Signals*. MIT Press, Cambridge, Massachusetts, 1991.
- [2] C. Cadoz, "Synthèse sonore par simulation de mécanismes vibratoires", Thèse de Doctorat, Spécialité Electronique - I.N.P.G. - Grenoble - France, 1979.
- [3] C. Cadoz, A. Luciani, J-L. Florens, "CORDIS-ANIMA : a modeling and simulation system for sound and image synthesis", *Computer Music Journal*, Vol. 17, No. 1 : 10-29, 1993.
- [4] N. Castagne, C. Cadoz, "GENESIS: A Friendly Musician-Oriented Environment for Mass-Interaction Physical Modelling". *Proceedings of ICMC'02*, Sweden, 2002.
- [5] G.M.L. Gladwell, "Inverse problems in vibration", 2nd édition, Kluwer Academic Publishers, 2004.
- [6] J. Hadamard, "Lectures on Cauchy's Problem in Linear Partial Differential Equations", Yale University Press, 1923
- [7] C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators", 1950.
- [8] K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares", *The Quarterly of Applied Mathematics* II (2) : 164-168, 1944.
- [9] K. Madsen, H.B. Nielsen, O. Tingleff, "Methods for non-linear least squares problems", *Informatics and Mathematical Modelling*, Technical University of Denmark, 2nd Edition, April 2004.
- [10] D.W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters", *Journal of the Society for Industrial and Applied Mathematics* 11(2) : 431-441, 1963.
- [11] J.J. Nattiez, "Fondements d'une sémiologie de la musique", Union Générale D'édition, Paris, 1975.
- [12] P. Schaeffer, "A la recherche d'une Musique Concrète", Seuil, Paris, 1952.
- [13] J.O. Smith, "Physical Modeling using Digital Waveguides", *Computer Music Journal*, Vol. 16, No. 4 : 74-91, 1992.