Algebraic Framework for Design and Qualitative Evaluation of Interactive Musical Systems Temporality

M.Desainte-Catherine, I.Durand
Univ. Bordeaux, CNRS, Bordeaux INP,
LaBRI, UMR 5800, F-33400 Talence, France
myriam@labri.fr

J.Haury, B.P.Serpette
SCRIME, Inria, Bordeaux
jeanhaury@gmail.com
Bernard.Serpette@inria.fr

S.Schwer
USPN, CNRS, LIPN UMR 7030
F-93430 Villetaneuse, France
schwer@lipn.fr

ABSTRACT

In this paper, we propose an algebraic framework for organizing interactive musical systems in temporal space in a way that helps in their design. The objective is to structure this space in order to be able to represent the existing and to open the exploration of new configurations that could inspire new musical practices. To do so, we propose some properties of interactive music systems allowing to compare them in a temporal lattice. This framework should be extended to other musical dimensions and give rise to optimizations of musical systems according to the desired performative properties.

1. INTRODUCTION

Until today, many studies and experiments have been carried out to create and use interactive musical systems (IMS) and many issues were discussed concerning their design on the software and hardware level as well as concerning their use. This paper proposes a study quite different from the previous ones because it introduces a theoretical framework for studying the temporal aspects at the design level and their implication on the uses, within an extension of the formal languages framework, which is an original algebraic approach compared to generalized intervals systems like [1] or geometry/topology of [2]. This algebraic study focuses on the temporal aspects of interactive musical systems and aims at describing and comparing different systems according to properties related to expressiveness and playability. In order to do so, we formalize interactive musical systems according to their temporal dimension only, forgetting any spatial information. That means that the system knows neither the musical information (such as pitches, durations, timbre and volume) nor the sensors nor even their number. One of the perspectives of this work is of course to extend it to the spatial dimension. For this purpose, this model should be confronted and articulated with the framework proposed in [3].

Thus, an interactive musical system temporality is represented by a tuple of algebraic objects which are the static and interactive scores, the interpretation space, the device language and the mapping function associating the musician's playing with the unfolding of the score. These elements allow to describe representation spaces, and to measure some properties of the systems. In particular, we propose a lattice of interactive scores that allows to organize this space according to a partial order. Finally, we introduce the notion of playing modes which corresponds to certain configurations that are distinguished from the musical point of view. Many playings modes can be defined with this formalism, whether they are existing or not.

The content of this article is limited to the presentation of this formalism and its interest to describe playing modes and measure some properties. The longer term goal is to automatically compute from a score interactive musical systems according to the desired playing modes, which depend on given constraints, such as the performer's motricity or the available devices.

We first present the notation that are useful to formalize the considered objects, then we present the elements that constitute a temporal interactive musical system and finally we show how to express playing modes with this formalism and how to measure properties. We present the rhythmic mode by using one or more keys to show how this study can help in the design of a system. The case of Jean Haury's Metapiano [4] has strongly inspired this work and is the subject of a particular development in this formalism.

2. PRELIMINARIES

In this section we present the definitions and notation that are used to formalize interactive musical systems. S-words [5] are used to model the temporal relations between the beginnings (ON) and the ends (OFF) of the notes of the score.

Figure 1. The first two bars of a Scherzo of Beethoven without any articulation

Let us consider the first two bars of the score of Beethoven's scherzo in the figure 1. The score is composed of five notes that provides our basic alphabet $\mathcal{A} = \{A4, B4, F4, C5, D5\}$ Each note in the score is associated with a sequence of playing intervals. An interval has two limits, an ON limit and an OFF one. Each is represented by an occurrence of a letter denoting the note. In order to discriminate ON and OFF, the occurrences denoting OFF are overlined, and the rank of each interval in the sequence is also provided with the corresponding occurrences. Thus, any letter of $\mathcal{A}$ will occur twice as many times as the number of intervalles in the sequence. Half the occurrences - the odd ones are for the beginnings, the other half - the even one are for the ends and are overlined. In the Figure 1, the qualitative individual performance of A4 is modeled by the word : $[A4, 1][\overline{A4}, 1][A4, 2][\overline{A4}, 2]$, or, if their is no ambiguity, by $A4\overline{A4}A4\overline{A4}$; the word $B4\overline{B4}$ describes the one for B4. The traditional formal languages theory [6] is not able to take into account synchronization of events. S-languages does. A s-letter is a non-empty subset of $\mathcal{A}$. For instance, the s-letter $(\overline{A4}B4)$ [1] denotes that the end of the first play of A4 coincides with the beginning of B4.

This formalism allows to represent the following three cases of musical articulations :

- *Staccato* : $(A4)(\overline{A4})(B4)(\overline{B4})$
- *Legato* : $(A4)(B4)(\overline{A4})(\overline{B4})$
- *Esatto legato* : $(A4)(B4\overline{A4})(\overline{B4})$

The score in the figure 1 reduced to *esatto legato* articulations can be represented by the following s-word : $sb = (A4)(\overline{A4}B4)(\overline{B4}F4A4C5)(\overline{C5}D5)(\overline{D5}F4A4)$.

The s-language associated to the score describes all the possible articulations, it is denoted $\mathcal{S}(\mathcal{A})$.

3. MUSICAL PIECE

This study is based on traditional scores of notes as they contain a large quantity of events linked by numerous temporal relations that the interpreter must take into account during the performance which eventually requires virtuosity. These scores also have a large repertoire. But this study is also adapted to any style of written music or music on a support, the notes being able to be more general sound objects. Thus a piece of electroacoustic music segmented in sound objects could be represented by a s-word.

A musical piece consists of a score and a space of interpretation, that is, a set of modifications of the piece that are allowed by the composer. In this paper a score is represented by a s-word augmented with musical information. Each interpretation of a score is also represented by a s-word so that a space of interpretation is represented by a set of s-words. The letters of the score could be associated with spatial musical information, such as pitches, durations and velocities, or any other information allowing to produce the sound result of the score.

In this study, we separate the question of temporality from that of the agogic interpretation. The first one is defined in the score, while the second one is defined in the space of interpretation below.

3.1 Interpretation

An interpretation of a score is a s-word, derived from the one of the score following rules that the musician must respect. In this article, the musician is not allowed to add or remove any event, that is, the s-language of interpretations $\mathcal{I}(sb)$ is included in $\mathcal{S}(\mathcal{A})$. For example, $\mathcal{I}(sb)$ may be defined by the following algebraic formula where we omit the indices of the occurrences of the intervals : $\mathcal{I}(sb) = \{A4\}\{\overline{A4},\ B4\}\{\overline{B4},\ F4,\ A4,\ C5\}\{\overline{C5},\ D5\}\{\overline{D5},\ \overline{F4},\ \overline{A4}\}$

[1] We use parenthesis without comma instead of curly bracket because s-letters model leibniz'notion of instant, an equivalence classe for simultaneousness between instantaneous events. We use $\{[\overline{A4}, 1], [B4, 1]\}$ to mean no order between the two occurences, which covers the three cases of articulations.

where $\{a, b\}$ means that a and b are not ordered [2]. In this article, we consider that the musician cannot interpret the simultaneity of events by pressing two keys exactly at the same time. Thus, in this case, $\mathcal{I}(sb)$ is simply a language and the number of its words, that is, the number of interpretations is : $|\mathcal{I}(sb)| = 1!2!4!2!3! = 576$. However, if she or he could apply *esatto legato* and synchronism of chords, $\mathcal{I}(sb)$ would be a set of s-words, and the number of interpretations would be $|\mathcal{I}(sb)| = 1 \times 3 \times 75 \times 3 \times 13 = 8775$.

3.2 Distance Between Interpretations

The distance between two interpretations i_1 and i_2 is defined according to the cost of operations applied to i_1 and leading to i_2. In a musical context, it is convenient to consider that the costs of these operations increase from the first to the fifth ones in the enumeration below.

Operations on interpretations.

Let $opi =< out, in, per_e, per_{es}, per_s >$ where the operations apply to adjacent s-letters :

- out : output of a letter out of a s-letter into a single s-letter
- in : input of a single s-letter inside an adjacent s-letter
- per_e : permutation of two end events
- per_{es} : permutation of an end event with a start event
- per_s : permutation of two start events

Transformation of interpretations. Let i_1 and i_2 be two interpretations with the same alphabet, there exists at least one $r = o(r_1, r_2, ..., r_n)$ being a composition of operations in opi such that $i_2 = r(i_1)$. Let us notice that in general r is not unique.

Cost of transformations. Let us associate a cost to each operation in opi and let us denote this cost by the function $cost$. Let i_1 and i_2 be two interpretations with the same alphabet, and let $r = o(r_1, r_2, ..., r_n)$ be a composition of operations in opi such that $i_2 = r(i_1)$, the cost of r is defined to be the sum of the costs of all the operations that constitute r : $cost(r) = \sum_i cost(r_i)$

Distance between two interpretations. Let i_1 and i_2 be two interpretations with the same alphabet, let O be the set of all r such that $i_2 = r(i_1)$, then the distance between i_1 and i_2 is defined to be : $\Delta(i_1, i_2) = min_{r \in O}(cost(r))$.

3.3 Degree of Freedom

Let s be a score and let $\mathcal{I}(s)$ be the space of interpretation of s. The size of $\mathcal{I}(s)$ is related to the degree of freedom given to the musician by the composer. Indeed, the smaller set of interpretations is limited to the score s, and the largest admits any kind of transformation. Let us denote by $\mathcal{A}*(s)$ the latter. Between the two of them, there exists a lot of possibilities (for example the one given for sb in subsection 3.1). Let us define the following quantity to measure the degree of freedom : $df(s, \mathcal{I}(s)) = \frac{|\mathcal{I}(s)|}{|\mathcal{A}*(s)|}$

3.4 Lattice of Interpretations

With the operations presented in the preceding subsection, it is possible to build a lattice [7] starting from the score and reaching all interpretations by applying all possible operations. Therefore, an arrow from i_1 to i_2 would mean that i_2 can be obtained from i_1 by applying one of the operations on i_1.

[2] We can get ab or ba or a and b strictly synchronized

4. INTERACTIVE SCORE

Let s be a score, an interactive score of s is a coloring of the s-word s, i.e. a pair (s, i) where i is a s-word included in s [3]. Thus, for a given score, there are several interactive scores that can be defined, depending on the associated s-word that is included. The smallest is the empty word noted ϵ and the largest is the one which is equal to the score s.

4.1 Interactive Points

Given an interactive score (s, i), the s-word i is called the *interaction s-word*. The letters of i are called *interaction points*. These letters will later be associated to an interaction device that will allow to activate them. Thus, the musician will be able to decide the precise moments of triggering these events. This will allow her or him to control the rhythmic and agogic expressiveness of the musical piece. Moreover, it should be remembered that the interactive points that are positioned in the same s-letter correspond to synchronized events on the score. These interaction points will allow the musician to activate these events in any order of her or his choice, in particular to desynchronize attacks, or chose how to articulate successions of notes.

Example. Let (sb, i) an interactive score where sb is the s-word of the score in figure 1. We can represent the interactive score by coloring in red the interaction points within the s-word of the score : $sb = ([A4, 1])([\overline{A4}, 1][B4, 1])([\overline{B4}, 1][F4, 1][A4, 2][C5, 1])([\overline{C5}, 1][D5, 1])([\overline{D5}, 1][\overline{F4}, 1][\overline{A4}, 2])$. So, the interaction s-word is the following : $i = ([A4, 1])([B4, 1])([\overline{C5}, 1][D5, 1])$. This interactive score has 4 interaction points that are the 4 letters of i : $[A4, 1], [B4, 1][\overline{C5}, 1][D5, 1]$. Finally, we see that according to the s-word indices, we keep the correspondence between the letters of i and s : $i[1] = ([A4, 1]) \subset p[1], i[2] = ([B4, 1]) \subset p[2], i[3] = ([\overline{C5}, 1][D5, 1]) \subset p[4]$.

4.2 Degree of Interactivity

The degree of interactivity of an interactive score $is = (s, i)$ is equal to the number of letters of the s-word i, that is, the number of interaction points, divided by the total number of letters of the score : $di(is) = \frac{|i|}{|s|}$ Thus, a null degree of interactivity indicates that there is no interaction point and that it is therefore impossible to control the progress of the score, while a degree of interactivity equal to 1 indicates that all the events of the score, whether they are beginnings or ends of notes, are activated by the musician.

4.3 Partial Order of Interactive Scores

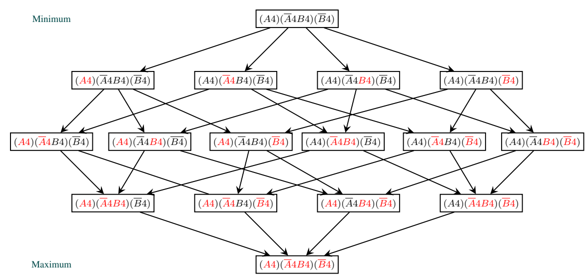
Let $is_1 = (s, i_1)$ and $is_2 = (s, i_2)$, we say that is_1 is less interactive than is_2 if and only if i_1 is included in i_2. $is_1 < is_2 \iff i_1 \subset i_2$. Let us notice that if $is_1 < is_2$ then $di(pi_1) \leq di(pi_2)$.

Thus, for a given s, there is only one interactive score of minimal order which is (s, ϵ), where ϵ denotes the empty s-word. There is no possible interaction with this interactive score. The next higher order interactive scores are those which admit one interaction point. There are as many of them as there are letters in s, since this interaction point can be located on any of these letters.

Finally, there is only one maximum order interactive score which is (s, s), because its s-word of interaction contains all the letters of s. So all possible s-words of interaction for the score s are included in it, by definition. This order is indeed partial and not total because some elements are not comparable.

4.4 Example of Lattice

Let us consider the following score : $s = (A4)(\overline{A}4B4)(\overline{B}4)$. It has 4 possible points of interaction. There are 15 interaction s-words that can be associated with this score. The interactive scores can be organised in a lattice according to the partial order introduced in the preceding paragraph, where $is1 \longrightarrow is2$ means that $is1 < is2$. In the following representation of the lattice, the minimum is located at the top and the maximum at the bottom. We have omitted the indexes of the letters in order to save space.



5. TEMPORAL INTERACTIVE MUSICAL SYSTEM

In this section, we define all the elements that constitute interactive musical systems and introduce several quantities to study their properties. This definition is limited to the temporal aspect of the interactive musical systems. This is why they are called temporal interactive musical systems (TIMS).

5.1 Definition of a TIMS

A temporal interactive musical system is a tuple $< is, \mathcal{I}(s), \mathcal{D}, m >$, where :

- s is a score
- $is = (s, i)$ is a colored s-word representing an interactive score
- $\mathcal{I}(s)$ is the space of interpretation of s
- $\mathcal{D}$ is the language of the device
- m is a command mapping that associates an interpretation of s to a word of $\mathcal{D}$

where $\mathcal{D}$ and m are defined just below.

5.2 Device Language

In order to interact with an interactive score, it is necessary to define the physical device that allows the activation of the interaction points. For discrete controls, we can use sensors such as buttons or keyboard keys. Depending on the type of interaction desired, several keys can be hold down simultaneously.

Two actions are possible on a key, the press and the release. Any sequence of actions on a key corresponds to a sequence of presses followed by releases.

[3] i is said to be included in s if it is obtained by deleting letters of s.

To describe the language of a key, we associate a letter x, an occurrence of x for pressing the key, and an occurrence of $\overline{x}$ for releasing the key. One cannot press twice the same key without releasing it before the second time. Thus, the key's language is described with the following rational expression :
$$T_x^1 = (x\overline{x})^*$$
If the sensor has n keys, let x_1, x_2, ... x_n, $\overline{x}_1$, $\overline{x}_2$, ... $\overline{x}_n$ the letters corresponding to the presses and releases of these keys, and let $T_{x_i}^1$ be the language of the key x_i, then the language of this sensor is defined by the shuffle product [4] (denoted by $\sqcup\!\sqcup$) of the languages $T_{x_i}^1$ of each of these keys :
$$T_{x_1,\ldots,x_n}^n = \bigsqcup_{i=1}^{n} T_{x_i}^1$$
In this article, we are interested in languages corresponding to the temporal dimension. To limit this study to this dimension, we consider that the system does not know the keys of the device and does not even know how many there are. Thus, all keys are combined into one generic key. In this case, the press or the release of a key is associated with an event in the score, via the s-word of interaction. This event does not depend on the key, but only on the position of the press or release in the interaction s-word.

Thus, let us denote $T_{x^n}^1$ the projection language of a n-key sensor in which all keys are represented by the same letter x. Thus the press of any key is denoted by x and any release by $\bar{x}$. In others words, we consider the trace of the presses and the releases. The only constraint is that the ith occurrence of $\bar{x}$ must occur after the ith occurrence of x.

This language is the set of Dyck words [5] of maximum height n, that is, the set of well-parenthesized words with a maximum of n embeddings of parenthesis. However, for these words, a closing parenthesis is always associated with the last free opening parenthesis (which has not yet been associated with a closing parenthesis). It may therefore be musically useful to study other kind of languages. Indeed, in the physical play of the keyboard, a release is rather associated with the press of the key that is released. However, it should be noticed that the number of words may differ according to the different associations of keypresses and releases that are considered. Let us notice in particular the following Dyck's word of $T_{x^2}^1$: $w = xx\overline{x}x\overline{x}\overline{x}$, there exist 4 words of T_{x_1,x_2}^2 that correspond to it :

$w_1 = x_1x_2\overline{x}_1x_1\overline{x}_1\overline{x}_2$; $\quad w_2 = x_1x_2\overline{x}_1x_1\overline{x}_2\overline{x}_1$;
$w_3 = x_1x_2\overline{x}_2x_2\overline{x}_1\overline{x}_2$; $\quad w_4 = x_1x_2\overline{x}_2x_2\overline{x}_2\overline{x}_1$;

Because we assume that the first non overlined letter that is read is x_1, the second one x_2 and so on. Let us notice here that the word w_2 corresponds to the choice of the command mapping used in the Metapiano that is presented in 6.4.2.

In the examples of the next section, the languages of the device $\mathcal{D}$ that will be considered are chosen among the languages of this subsection.

5.3 Command Mapping

Let s be a score, $is = (s, i)$ an interactive score and $< is, \mathcal{I}(s), \mathcal{D}, m >$ be a TIMS. Let us first define a *command* as a colored word of $\mathcal{D}$ and let $\mathcal{C}$ be the set of commands, that is the set of all colorings of the words of $\mathcal{D}$.
$$\mathcal{C} = \{(w, wi), w \in \mathcal{D} | wi \subset w\}$$
where w is a word in the language of the device $\mathcal{D}$ and wi is called the *command word*.

A *command mapping* is a function associating an interpretation in $\mathcal{I}(s)$ to a command in $\mathcal{C}$:
$$m : \mathcal{C} \longrightarrow \mathcal{I}(s)$$
Let us point out that the words w and wi are sequences of letters corresponding to presses or relesases of keys. Now, each letter of the command word wi is mapped to a letter of the interaction s-word i. Thus each letter in wi allows to activate the corresponding interaction point in i which itself allows to trigger the corresponding event in the score s. Let us notice that many choices have to be made in the design of the function m.

- In the definition of the mapping between wi and i, the kind of the letter of wi, that is press or release, is not necessarily related to the kind of the event in the score, that is start or end of a note. Indeed, a press may activate an end of note and vice-versa. However, for reasons of musical ergonomics, the types of the letters should match each other most of the time.
- To match the letters of wi with those of i, chronological order is preferred for reasons of real-time play. But in some cases, i can have s-letters including several interaction points. Then it is necessary to decide how to choose the interaction point of i depending on the letter of wi that the musician has just played.

Let us define the following sets :

- Let $\mathcal{C}(m)$ be the domain of m, $\mathcal{C}(m) \subset \mathcal{C}$
- Let $\mathcal{I}(m)$ be the codomain of m, $\mathcal{I}(m) \subset \mathcal{I}(s)$

Example. Let $< is, \mathcal{I}(s), T_x^1, m >$ be a TIMS. Let us consider the following interactive score :
$$is = (A4)(\overline{A4}B4)(\overline{B4})$$
The interactive s-word of is is $i = (A4)(B4)(\overline{B4})$
Let us define a mapping m reduced to one command c, so that the set $\mathcal{C}$ is defined as follows : $\mathcal{C}(m) = \{c\}; c = x\overline{x}x\overline{x}$
The command word of c is the word wi defined as follows :
$$wi = xx\overline{x}$$
The mapping m associates the red letters of is with the red letters of c in a chronological way. More precisely, m makes the following associations :

- the first x of wi is associated with the interactive point of the first s-letter of i, that is the event A4. Thus the first press of a key activates the beginning of the note $A4$.
- the second x of wi is associated with the interactive point of the second s-letter of i, that is the event B4. Thus the second press activates the beginning of the note $B4$, as well as the release of the note $A4$, because those two events are synchronised in the same s-letter of is.
- Finally the second release is associated with the interactive point of the third s-letter, that is the event $\overline{B4}$. Thus the second release activates the end of the note $B4$.
- The resulting interpretation is exactly the score s with the durations that the musician has played.

The first release $\bar{x}$ of the command c is inactive, but it is necessary so that the word c belongs to the language of the device $\mathcal{D}$ which guarantees that this word can actually be played by the musician on this device.

[4] The shuffle product of languages $L_1, L_2, ..., L_n$ is the set of words obtained by interleaving in all possible ways the words of $L_1, L_2, ...L_n$

[5] The number of Dyck word of size k is called the Catalan number and is denoted by $\mathcal{C}_{k/2}$.

5.4 Expressiveness and Difficulty of a TIMS

Let s be a score, and let $ms =< is, \mathcal{I}(s), \mathcal{D}, m >$ be a temporal interactive musical system, let n be the size of the command words in $\mathcal{C}(m)$ and let $\mathcal{D}^n$ be the set of the words of $\mathcal{D}$ of size n : $\mathcal{D}^n = \{w \in \mathcal{D}, |w| = n\}$. Let us define the following quantities associated to a TIMS :

- Degree of difficulty : $dd(ms) = \frac{|\mathcal{D}^n)| - |\mathcal{C}(m)|}{|\mathcal{D}^n|}$
- Degree of expressiveness : $de(ms) = \frac{|\mathcal{I}(m)|}{|\mathcal{I}(s)|}$

5.5 Improving a TIMS

According to a given objective, there are several ways to improve a TIMS by changing the definitions of its elements.

- To increase expressiveness one can increase $\mathcal{I}(m)$, eventually as well as $\mathcal{I}(s)$
- To decrease difficulty, increase $\mathcal{C}(m)$ or reduce $\mathcal{D}$
- The decrease of $\mathcal{D}$ may imply the decrease of $\mathcal{C}$, thus may imply the decrease of $\mathcal{I}(m)$. So that a balance must be found between difficulty and expressiveness.

6. PLAYING MODES OF TIMS

In this section, we show several playing modes that can be defined for TIMS in order to illustrate the formalism and show how to compare them according to the properties that were introduced before.

6.1 The Score

In all the examples below, we use the score which is displayed in figure 1. There are many ways to represent a score by a s-word. Let us recall the s-word sb :
$(A4)(\overline{A4}B4)(\overline{B4}F4A4C5)(\overline{C5}D5)(\overline{D5F4A4})$.
If we project this s-word on the onset letters, we obtain the successions of the beginnings of the notes and the synchronizations. We see that all the successions are represented and that the synchronizations are exact :
$(A4)(B4)(F4A4C5)(D5)$.
In particular, every succession is represented by *esatto legato*, whether the notes are actually linked or not, like in this example : $(A4)(\overline{A4}B4)(\overline{B4})$

6.2 Rhythmic Mode

This mode allows the musician to play with the rhythm using one or more keys. Let us first present the TIMS with one key and second the TIMS with two keys in order to compare them.
We want to define a TIMS allowing to play sb with a sequence of eight steps of key presses and releases. Let us define our TIMS as the following :

$$ms_1 =< is_1, \mathcal{I}_1(sb), \mathcal{D}_1, m_1 >$$

where the elements $is_1, \mathcal{I}_1(sb), \mathcal{D}_1$ and m_1 are defined below.

6.2.1 Space of Interpretation

Let us now define the space of interpretation $\mathcal{I}_1(sb)$ that we wish to allow the musician to play. This mode provides only one interpretation in term of s-word, because all the temporal relations of the score have to be strictly respected.

$$\mathcal{I}_1(sb) = \{sb\}$$

6.2.2 Interactive Score

In this case, the interactive score has exactly one interactive point per s-letter. As a consequence, the choice of the letter carrying the interactive point does not matter in practice. However, in this example, a maximum of start letters in the first voice have been selected for musical ergonomic reasons. The interactive score is defined as follows, the interactive points displayed in red, every line representing a bar and omitting the indices for the sake of clarity of the notation.

$$\begin{aligned} is_1 = (sb, i_1) = &(A4)(\overline{A4}B4) \\ &(\overline{B4}F4A4C5)(\overline{C5}D5) \\ &(\overline{D5}F4A4) \end{aligned}$$

6.2.3 One Key Device

Let us consider that the device is reduced to one key. Thus, the alphabet is $\{a, \bar{a}\}$ and $\mathcal{D}_1 = T_a^1 = (a\bar{a})^*$

6.2.4 Command Mapping

The command mapping is defined by the set of valid commands $\mathcal{C}(m_1)$, i.e. the set of commands leading to an interpretation in $\mathcal{I}(sb)$. It is defined as a set of colored word (w, wi) where w is in $\mathcal{D}_1$ and $wi \subseteq w$, the subword wi being represented by the subword colored in red below. In this example, we have only one command : $\mathcal{C}(m_1) = \{a\bar{a}a\bar{a}a\bar{a}a\bar{a}\}$
The command word wi is mapped to the interaction s-word i_1 of the interactive score is_1. Thus the three black letters are inactive, their purpose is to immerse the command into the language of the device $\mathcal{D}_1$.
Here is how the mapping works :
$a \longrightarrow A4$, $\bar{a} \longrightarrow \emptyset$, $a \longrightarrow \overline{A4}B4$, $\bar{a} \longrightarrow \emptyset$, $a \longrightarrow \overline{B4}F4A4C5$, $\bar{a} \longrightarrow \emptyset$, $a \longrightarrow \overline{C5}D5$, $\bar{a} \longrightarrow \overline{D5}F4A4$

The command mapping can be defined as follows :
$\forall c \in \mathcal{D}_1, m_1(c) = \perp \; if \; c \notin \mathcal{C}(m_1) \; else \; m_1(w) = sb$
Thus we have :

- $\mathcal{C}(m_1) = \{c\}, n = |c| = 8$
- $\mathcal{I}(m_1) = \{sb\}$

6.2.5 Properties

This TIMS allows the musician to play a piece of music with total respect for the temporal relationships with the *esatto legato* that are imposed for the first four notes. The musician cannot make mistakes so that the difficulty is minimum. Indeed, the reduction of the device to one key leads to only one command so that the degree of difficulty is minimum. The total respect of the relations implies that there is only one possible interpretation so that the expressiveness is at its maximum, while the degree of freedom is low.
Let us precise the quantities that are useful to measure the properties of our TIMS. The size of the commands is $n = 8$, the number of words of the device of size 8 is $|\mathcal{D}_1|_8 = 1$, the number of interpretations is $|\mathcal{I}_1(sb)| = 1$, and the number of all possible permutations of the score is $|\mathcal{A}*(sb)| = D(4, 2, 2, 2, 2)$ [6] with $D(4, 2, 2, 2, 2) >> D(2, 2, 2, 2, 2) = 2244361$. $|\mathcal{C}(m_1)| = 1, |\mathcal{I}(m_1)| = 1$.

- Degree of freedom : $df(s, \mathcal{I}_1(sb)) = \frac{|\mathcal{I}_1(sb)|}{|\mathcal{A}*(sb))|} < 4.45e^{-07}$

[6] This is a generalized Delannoy number, whose computation is given by the following recursive formula [see for instance (Durand-Schwer 2008) : Let $(p_1, p_2, \cdots, p_n)$ n natural integers, and $\mathcal{P}red(p_1, \ldots, p_n)$ = $\{(\widetilde{p_1}, \cdots, \widetilde{p_n}) \in \mathbb{N}^n | \; if \; p_i \neq 0$, then $\widetilde{p_i} \in \{p_i, p_i - 1\}$ else

- Degree of interactivity : $di(is_1) = |i_1|/|s| = 5/12 = 0.416$
- Degree of difficulty : $dd(ms_1) = \frac{|\mathcal{D}_1|_8 - |\mathcal{C}(m_1)|}{|\mathcal{D}_1|_8} = \frac{1-1}{1} = 0$
- Degree of expressiveness : $de(ms_1) = \frac{|\mathcal{I}(m_1)|}{|\mathcal{I}_1(sb)|} = 1/1 = 1$

6.2.6 *Device With More Than One Key*

Let us consider now that the device has two keys and let us consider that those two keys are generic. Thus, the language of our device is the projection of the language of two keys : $\mathcal{D}_2 = T^1_{x^2}$

This is the set of Dyck words of maximum height 2. Let us consider the following TIMS that differs from ms_1 only in the number of keys : $ms_2 = (s, is_1, \mathcal{I}_1(sb), \mathcal{D}_2, m_1\}$

This system has the same properties as ms_1 except that the degree of difficulty is much higher. Indeed, the difficulty increases with the number of keys because of the increasing number of possible words of the device. The reason for that is that a lot of words of the device are wrong. In our example, the number of words of length 8 of the device is 8 and only one of those words is correct. So the musician has a $7/8$ chance of being wrong. In particular, all the words beginning by two successive a, that is by two successive key presses, are wrong. For example, the following word is wrong because of the second letter that we have colored in red : $c' = aa\bar{a}\bar{a}a\bar{a}a\bar{a}$

In this example, only one word is correct and all other possible words the musician can play with the device are wrong. The situation gets worse if the number of keys increases. Moreover we omitted to count the series of presses and releases that would leave keys pressed at the end. This further increases the number of wrong words.

There are several ways to reduce the difficulty in order to improve this TIMS (see subsection 5.5). For example one may increase the number of right words by changing the mapping. The actual mapping associates only one word with the interaction word of the interactive piece, this word being a series of key presses and releases. We propose the two following ways to modify the mapping :

- One possibility is not to consider the nature of the action (press or release) and to associate any serie of actions to the interaction word. Thus, the word c' becomes correct with this mapping, and produces the same result as c.
- Another possibility is to reject every wrong letter. Indeed, since at any time, the system can predict the right letter, it may reject all letters that are different from the right one. The inconvenient of this choice is that the musician could think that the system is not working well because it does not always react to her or his actions. Thus, with this mapping, the word c' is the beginning of a correct word. Since the second letter a is rejected, the second letter $\bar{a}$ has to be rejected also, so that the word has to be extended by 2 correct letters, or even more if there are more upcoming wrong letters. For example, the following command is correct with this new mapping : $aa\bar{a}\bar{a}a\bar{a}a\bar{a}a\bar{a}$

$\widetilde{p_i} = p_i,\ 1 \leq i \leq n\} - \{(p_1, \cdots, p_n)\}$.

$$D(p_1, \ldots, p_n) = \sum_{\mathcal{P}red(p_1,\ldots,p_n)} D_n(\widetilde{p_1}, \cdots, \widetilde{p_n})$$

with $D_1(0) = 1$, $D_n(p_1, \ldots, p_{n-1}, 0) = D_{n-1}(p_1, \ldots, p_{n-1})$. For instance, if all $p_i = 2$, then, the sequence $(D(p_1, \cdots, p_n))_n$ is the sequence A055203 of the Sloane's on-line encyclopedia of integer sequences.

6.3 The Metapiano

Another way to reduce the difficulty of the previous TIMS is to increase its space of interpretation as well as its set of valid commands. This is the choice made in the *Metapiano* [7] of Jean Haury [4]. Like the temporal interactive musical systems we consider in this paper, the Metapiano has no timing information (rhythm and duration) and cannot play automatically. In fact, in the score of the Metapiano, are only associated to the s-letters note pitches and their belonging to a particular musical voice of polyphony; no other temporal (duration) or dynamic (velocity) information. However, many interpretations can be performed on a single or multiple sensors, or keyboard key.

Let us define the Metapiano by the following tuple :

$$mtp = \{is_2, \mathcal{I}_2(sb), \mathcal{D}_9, m_2\}$$

Only the score is the same as the previous examples, all the other elements change. Indeed, the space of interpretation is open to some permutations of beginning and ending notes in order to allow to play *legato* or *staccato* successive notes. Therefore, the interactive score admits more interactive points in order to allow those permutations, and the mapping has to be adapted. And $\mathcal{D}_9$ is defined below.

6.3.1 *Space of Interpretation*

The space of interpretation $\mathcal{I}_2(sb)$ of the Metapiano is a set of s-words made from the s-word of the score. In this space, the beginnings of notes are ordered between them as well as the ends of notes for all the s-words of $\mathcal{I}_2(sb)$. Moreover the end of a note always comes after its beginning. To do so, the beginnings and endings of notes that are in a same s-letter of the s-word s are desynchronised. More precisely, the s-letters of s containing ending and beginning letters are splitted into two s-letters, one containing the endings, and the other one containing the beginnings.

Let us project the s-word of the score sb on the two kinds of letters, as we did in the subsection 6.1, and consider the two sequences of s-letters, on the left are the s-letters corresponding to a projection on the onset letters, and on the right to the projection on the offset letters :

$a_1 = ([A4, 1])$
$a_2 = ([B4, 1])$
$a_3 = ([F4, 1][A4, 2][C5, 1])$
$a_4 = ([D5, 1])$

$\overline{a}_1 = ([\overline{A4}, 1])$
$\overline{a}_2 = ([\overline{B4}, 1])$
$\overline{a}_3 = ([\overline{C5}, 1])$
$\overline{a}_4 = ([\overline{D5}, 1][\overline{F4}, 1][\overline{A4}, 2])$

Now the a_i are considered as letters in an alphabet and we consider words without synchronization.

The temporal constraints between those letters in the space of interpretation are the following :

- For $1 \leq i \leq j \leq 4$, the s-letter a_i comes before the s-letter a_j
- For $1 \leq i \leq j \leq 4$, the s-letter $\overline{a}_i$ comes before the s-letter $\overline{a}_j$
- All the s-letters containing an *ON* letter x comes before the s-letter containing the *OFF* letter $\bar{x}$.

The temporal relations can be represented graphically by the lattice which is displayed on figure 2, where the arrows denote a temporal precedence relationship.

[7] see for example the small fugue of Förster: https://youtu.be/FdMWvQpQnPM or Beethoven Great Fugue Op. 133 https://youtu.be/0hDlWxA6SlY

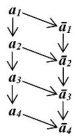


Figure 2. The lattice representing $\mathcal{I}_2(sb)$

In particluar, this lattice indicates that a_2 and $\bar{a}_1$ may be performed in any order but not synchronized, and both after a_1 and before $\bar{a}_2$.

The language corresponding the lattice is the set of the 14 possible articulations between the notes of the score :

$$\begin{aligned}\mathcal{I}_2(sb) = \{&a_1\bar{a}_1a_2\bar{a}_2a_3\bar{a}_3a_4\bar{a}_4, a_1\bar{a}_1a_2a_3\bar{a}_2\bar{a}_3a_4\bar{a}_4, a_1a_2\bar{a}_1\bar{a}_2a_3\bar{a}_3a_4\bar{a}_4,\\ &a_1a_2\bar{a}_1a_3\bar{a}_2\bar{a}_3a_4\bar{a}_4, a_1a_2a_3\bar{a}_1\bar{a}_2\bar{a}_3a_4\bar{a}_4, a_1\bar{a}_1a_2a_3\bar{a}_2a_4\bar{a}_3\bar{a}_4,\\ &a_1\bar{a}_1a_2\bar{a}_2a_3a_4\bar{a}_3\bar{a}_4, a_1a_2\bar{a}_1\bar{a}_2a_3a_4\bar{a}_3\bar{a}_4, a_1a_2\bar{a}_1a_3\bar{a}_2a_4\bar{a}_3\bar{a}_4,\\ &a_1a_2a_3\bar{a}_1\bar{a}_2a_4\bar{a}_3\bar{a}_4, a_1\bar{a}_1a_2a_3a_4\bar{a}_2\bar{a}_3\bar{a}_4, a_1a_2\bar{a}_1a_3a_4\bar{a}_2\bar{a}_3\bar{a}_4,\\ &a_1a_2a_3\bar{a}_1a_4\bar{a}_2\bar{a}_3\bar{a}_4, a_1a_2a_3a_4\bar{a}_1\bar{a}_2\bar{a}_3\bar{a}_4\}\end{aligned}$$

Figure 3 illustrates the temporal relations between three successive notes according to this space of interpretation. In this figure, for x and y being a letter a_i or $\overline{a_i}$ with $1 \leq i \leq 3$, an arrow going from a letter x to a letter y denotes that y is played just after x.

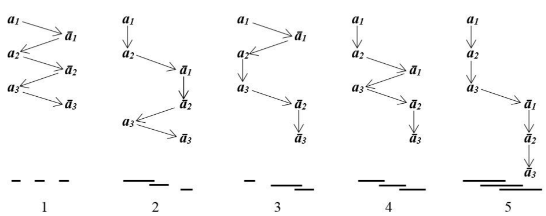


Figure 3. The five possible articulations of three notes and the qualitative temporal relations between the three associated intervals : the leftmost one for a_1, the middle one for a_2 and the last one for a_3

6.4 Interactive Score

There are many possibilities to define the interactive score. Let us consider the following one which is quite close to the previous examples.

$$\begin{aligned}is_2 = (s, i_2) = &(A4)(\overline{A4}B4)\\ &(\overline{B4}F4A4C5)(\overline{C5}D5)\\ &(\overline{D5F4A4})\end{aligned}$$

In this interactive score, three interaction points have been added compared to the previous examples so that the endings of the three first notes are interactive. The interaction s-word i_2 has 8 letters. This way, the parity between the beginnings and the ends of notes in i_2 is restored and the mapping will be simplified.

6.4.1 Device of the Metapiano

The Metapiano is usually played on a 9-key keyboard whose keys are generic. This number has been chosen for reasons of musical ergonomics, especially for fast playing. Let us base the study with such a device keys. Thus, the language of our device is the projection of the language of nine keys :

$$\mathcal{D}_9 = T^1_{x^9}$$

6.4.2 Command Mapping

In this mapping all the letters are active, so that the command word is always equal to the whole command. In other words, all the letters are red. The set of valid commands $\mathcal{C}(m_2)$ is the set of all the words of length 8 that can be obtained from the device with a succession of presses and releases of any key and when all the keys are relaxed at the end. However, the words that do not have as much a as $\overline{a}$ are not in this set. This set can be obtained from the word $(a\overline{a})^4$ by swapping the a letters with the $\overline{a}$ letters to bring them to the left or simply from $\mathcal{I}_2(sb)$ by erasing the indices.

$$\begin{aligned}\mathcal{C}(m_2) = \{\ &a\bar{a}a\bar{a}a\bar{a}a\bar{a}, a\bar{a}aa\bar{a}\bar{a}a\bar{a}, aa\bar{a}\bar{a}a\bar{a}a\bar{a}, aa\bar{a}a\bar{a}\bar{a}a\bar{a}, aaa\bar{a}\bar{a}\bar{a}a\bar{a},\\ &a\bar{a}a\bar{a}aa\bar{a}\bar{a}, a\bar{a}aa\bar{a}a\bar{a}\bar{a}, aa\bar{a}\bar{a}aa\bar{a}\bar{a}, aa\bar{a}a\bar{a}a\bar{a}\bar{a}, aaa\bar{a}\bar{a}a\bar{a}\bar{a},\\ &a\bar{a}aaa\bar{a}\bar{a}\bar{a}, aa\bar{a}aa\bar{a}\bar{a}\bar{a}, aaa\bar{a}a\bar{a}\bar{a}\bar{a}, aaaa\bar{a}\bar{a}\bar{a}\bar{a}\}\end{aligned}$$

The mapping between a command word and the interaction s-word proceeds as follows.

- the ith letter a of the command word corresponds to the ith beginning letter of the interaction s-word.
- the ith letter $\overline{a}$ of the command word corresponds to the ith ending letter of the interaction s-word.

Thus, the word c' that was considered wrong in the example 6.2.6 becomes correct with this mapping. This word is the third one of the set of commands $\mathcal{C}(m_2)$ above. Here is how the mapping works with this word :
$a \longrightarrow A4, a \longrightarrow B4, \overline{a} \longrightarrow \overline{A4}, \overline{a} \longrightarrow \overline{B4}, a \longrightarrow C5F4A4, \overline{a} \longrightarrow \overline{C5}, a \longrightarrow D5, \overline{a} \longrightarrow \overline{D5F4A4}$

We notice that the permutations of the letters in the command word imply permutations of the corresponding s-letters of the score via the interaction s-word. Since the device does not allow to release a key before it has been pressed, the temporal constraints between the a and $\overline{a}$ letters are always respected. Moreover the genericity of the keyboard keys implies that permutations between the letters a are not possible since they are not distinguishable. It is the same for the letters $\overline{a}$. Thus, all the words of the set $\mathcal{D}_9$ of size 8 are in $\mathcal{C}(m_2)$ and vice-versa.

The interpretation resulting from the execution of this command word is the following : $(A4)(B4)(\overline{A4})(\overline{B4})(C5F4A4)$ $(\overline{C5})(D5)(\overline{D5F4A4})$. Let us examine the distance of this interpretation to the score sb. We can see that the *out* operation must be applied three times to the score $sb2$ in order to obtain this interpretation. From left to right, the first application was made on the letter $B4$, the second one on the letter $\overline{B4}$ and the last one on the letter $C5$. Thus, the distance of this interpretation to the score is equal to three times the cost of the *out* operation.

6.4.3 Properties

This TIMS allows the musician to play a piece of music with respect for some temporal relationships between the events and with the possibility of swapping some events of the score. The mapping is such that its domain is equal to the set of words of length 8 of the device language, that is the 4th number of Catalan. The number of words that the musician can play on the device is much higher, because she or he can play prefixes of Dyck words of longer size. For example she or he can do more presses than the number of beginnings of notes of the score. So the musician can play some wrong words. At last, in this example, the codomain is equal to the interpretation space.

Let us precise the quantities that are useful to measure the properties of our TIMS. Let us denote by $|\mathcal{D}_9|_{8+}$ the number of words of size 8 that are prefixes of longer Dyck words. This number is much higher than $|\mathcal{D}_9|_8$. $n = 8$, $|\mathcal{D}_2|_{8+} \geq \mathcal{C}_4$, $|\mathcal{I}_2(sb)| = \mathcal{C}_4$, $|\mathcal{A}*(sb))| > 2244361$, $|\mathcal{C}(m_2)| = \mathcal{C}_4$, $|\mathcal{I}(m_2)| = \mathcal{C}_4$, where $\mathcal{C}_4$ is the 4th number of Catalan, that is 14.

- Degree of freedom : $df(s, \mathcal{I}_2(sb)) = \frac{|\mathcal{I}_2(sb)|}{|\mathcal{A}*(sb))|}$ < 6.23E-06
- Degree of interactivity : $da(is_1) = |i_2|/|s| = 8/12 = 0.66$
- Degree of difficulty : $dd(ims_1) = \frac{|\mathcal{D}_9|_{8+} - |\mathcal{C}(m_2)|}{|\mathcal{D}_9|_{8+}} > 0$
- Degree of expressiveness : $de(ims_1) = \frac{|\mathcal{I}(m_2)|}{|\mathcal{I}_2(sb)|} = 1$

It can be seen that in the two cases of TIMS considered, the rhythmic mode and the Metapiano, the degree of expressiveness is at its maximum. They both take advantage of the device to allow the musician to explore all the space of interpetation allowed by the composer. The metapiano offers a larger space of interpretation, which is expressed by the greater degree of freedom than the rhythmic mode. The Metapiano requires more gestures from the musician, which is expressed by a higher degree of interactivity. Finally it is a little more difficult to play than the rhythmic mode because the musician must anticipate the end of the score to stop all the notes in time.

7. CONCLUSION

In this paper we have presented an algebraic framework to describe the temporal aspects of interactive musical systems. For this purpose we have limited our study to systems using generic discrete control sensors. Indeed, we have eliminated all spatial knowledge concerning the musical information but also the sensor information. Thus, only the information concerning the temporal relations between the static events of the score and the real time events of the performance are considered. This study allows to dissociate the temporal constraints coming from the written score, from those coming from the interpretation domain opened by the composer to the musicians, and those coming from the sensors. The formal definition of these spaces opens the possibility to study them, to browse them, to measure them, and to compare various solutions for designing temporal aspects of interactive musical systems. Some examples are provided in this article, but many others could be studied and new system models could be identified.

This study opens the way to several new perspectives. One of them is to complete the study of playing modes that show musical interest. Distinguished playing modes could be formalized and integrated into interactive sequencers such as OSSIA score [8, 9]. Some questions arise, for example, how to simplify the playing of a score for a musician with a motor disability while ensuring a satisfactory degree of expressiveness? This requires detecting moments of interest in the score to place the points of interaction so that the interpretation is fluid and expressive. For this purpose, it could be useful to focus on models of expressivity like [10].

The extension to the space domain of this model can be done in several steps. First of all, a recognition of the device keys and the notes of the score will allow to know at any time which key is released and consequently which note is to be finish. This step may allow to model the second version of the Metapiano, which gave rise to the development of the Midifile Performer [11]. The study in [12] presents also various playing modes that are of interest.

Finally, a major objective is to be able to optimize TIMS according to certain criteria, in particular playability and difficulty, in order to find the best solutions in the algebraic spaces defined here according to the objectives of the design of interactive musical systems.

Acknowledgments

This research was hosted by the SCRIME of the University of Bordeaux (Studio of Creation and Research in Computer Science and Experimental Music) and funded by the french ministry of culture. The authors would like to thank wormly Gyorgy Kurtag Jr for bringing them together, and for helping them with his deep vision in artistic research.

8. REFERENCES

[1] T. M. Fiore, T. Noll, and R. Satyendra, "Morphisms of generalized interval systems and pr-groups," *Journal of Mathematics and Music*, pp. 3–27, 2013.

[2] G. Mazzola, S. Göler, and S. Müller, *The Topos of Music, Geometric logic of Concepts, Theory and Performance*. Birkhaus Verlag, 2005.

[3] J. Malloch, D. Birnbaum, E. Sinyor, and M. Wanderlay, "Towards a new conceptual framewok for digital musical instruments," in *DAFX'06*, 2006, pp. 49–52.

[4] J. Haury, "Un répertoire pour un clavier de deux touches. théorie, notation et application musicale," *Document numérique*, vol. 11, pp. 127–148, 2008. [Online]. Available: https://www.cairn.info/revue-document-numerique-2008-3-page-127.html

[5] I. Durand and S. Schwer, "A tool for reasoning about qualitative temporal information: the theory of s-languages with a lisp implementation," *J. of Universal Computer Science*, vol. 14 (20), pp. 3282–3306, 2008.

[6] S. Ginsburg, *The Mathematical Theory of Context Free Languages*. McGraw-Hill Book Company, 1966.

[7] G. Birkhoff, "Lattice theory. american mathematical society," *Providence, RI*, 1967.

[8] M. Desainte-Catherine, A. Allombert, and G. Assayag, "Towards a hybrid temporal paradigm for musical composition and performance: The case of musical interpretation," *Computer Music Journal*, vol. 37, no. 2, pp. 61–72, 2013.

[9] J.-M. Celerier and al, "Ossia: towards a unified interface for scoring time and interaction," in *First International Conference on Technologies for Music Notation and Representation, Paris, France*, 2015, pp. 81–90.

[10] C. Eduardo, Cancino-Chacón, T. Gadermaier, G. Widmer, and M. Grachten, "An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonicmusic," *Mach Learn, Springer-Verlag Nature*, pp. 887—909, 2017.

[11] J. Chabassier, M. Desainte-Catherine, J. Haury, M. Pobel, and B. P. Serpette, "Midifileperformer: a case study for chronologies," in *FARM*, 2021, pp. 13–22.

[12] T. Lucas, S. D. Laubier, and C. D'Alessandro, "Animer la musique sur support," in *JIM 2020. Université de Strasbourg/Faculté des Arts*, 2020.