Playing with Sounds as Playing Video Games

MYRIAM DESAINTE-CATHERINE, GYÖRGY KURTAG,
SYLVAIN MARCHAND, SCRIME, CATHERINE SEMAL,
University of Bordeaux 2, and PIERRE HANNA, SCRIME

In this article we propose a musical instrument that provides young children with a way to play in a virtual sound space by means of a joystick. We present the work of a multidiscipline team working on various aspects of this project. Sound models that are close to perception are used to define the sound spaces. Pedagogical experiments in public schools are done on the basis of cognitive pedagogy. Ethological and psychoacoustic experiments are carried out to improve the ergonomics of the instruments.

Categories and Subject Descriptors: H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing—*Modeling*; H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing—*Signal analysis, synthesis, and processing*; J.5 [**Computer Applications**]: Arts and Humanities—*Music*; *Performing arts*; K.3.1 [**Computers and Education**] Computer Uses in Education—*Collaborative learning*; *Computer-assisted instruction*; K.8.0 [**Personal Computing**]: General—*Games*

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Education, sound, music, interactivity, computer-assisted music education, sound models, electro-acoustic musical instruments

---

1. INTRODUCTION

The main aim of teaching musical games early is to develop the child's ability to listen to and to play music: in particular, to learn to listen to other musicians and to play together with others, either by using written scores or improvisation. Music is composed of sounds that are organized according to temporal and musical dimensions. Thus, musical pedagogy addresses two levels: those of sound and music.

Pedagogy at the level of sound consists in opening the ears of children to sound and encouraging them to create it by means of natural objects, synthesizers, and computers. Children learn how to listen to the world, select a sound from among all the other sounds, and analyze the sound from a perceptual rather than a production point of view. Reaching this goal requires the development of attentiveness, which is fundamental in many other areas besides music. First of all, due to the ear's alarm function, hearing a sound often automatically generates questions like:

Authors' addresses: M. Desainte-Catherine, SCRIME, LaBRI, ENSEIRB, University of Bordeaux 1, 351 cours de la Libération, F-33405 Talence Cedex, France; email: myriam@labri.fr; G. Kurtag, SCRIME; email: gkurtag@aol.com; S. Marchand, SCRIME, LaBRI, University of Bordeaux 1; email: sm@labri.fr; C. Semal, Laboratoire de Neurophysiologie, University of Bordeaux 2; email: catherine.semal@psyac.u-bordeaux2.fr; P. Hanna, SCRIME, LaBRI, University of Bordeaux 1; email: hanna@labri.fr

What is this? What is the source of this sound? For adults, listening is a matter of attention and education; but very young children have strong motricity so that their attention focuses spontaneously on their movements. They have to learn to listen carefully and analyze the sound they make in order to link their movements to the sound and control it – in other words, to *listen to their movements*.

In general, musical pedagogy consists in analyzing music by segmenting it and then organizing the segments. Listening is often associated with the practice of an instrument. This favors creation by production of music.

In this article we present a project based on a new pedagogy, which is at the border between sound and music: we call it *sound exploration*. We introduce a musical instrument that provides children with a way to play in a virtual sound space by means of a joystick. The pedagogical goals of the exploration of sound are to develop the children's listening strategies as well as their capacity to pay attention to their sensations. The main musical ability resulting from these developments is musical composition.

In the following, we first present this study's background, our pedagogical procedures, objectives, and methodology, and introduce sound models well suited for our purpose. We then present some details on the different parts of this multidisciplinary project, called Dolabip. The development of the software is told through several versions, and problems are presented. Some psychoacoustic and psychological experiments are then described.

2. BACKGROUND

François Delalande is a pioneer in musical pedagogy for young children. According to several of his articles (e.g., Delalande [1988]), children begin to spontaneously explore sound in the first months of life. Without any instruction children explore relations between sound and movement by shaking, scratching, rubbing, bouncing and listening. This attitude is very similar to one developed in musical research; it is close to the attitude of musicians who shake, scratch, rub, bounce, and listen, and then select the parts to be developed in a musical composition. Clearly, it is also close to composition by means of experimentation and research – a domain that has not yet been explored, in contrast to musical composition based on mental representation. The comparison between children and musical researchers is a historical one; musical pedagogy should take advantage of this fact.

The Dolabip project has its roots in the *Mélisson*, which was designed by Roland (GMEA, Albi, France) in the early 1980s. It is a modular and analog synthesizer for musical pedagogy, made up of three kinds of boxes: the oscillator, filter, and mixer. The boxes are connected to each other so that they describe a signal synthesis algorithm. A child can use the potentiometers of a box to control the parameters of the resulting sound.

For young children and nonprofessional musicians, Tod Machover[1] developed a new generation of interfaces that can be squeezed, pulled, stretched, and twisted, with various degrees of suppleness. The interfaces are used to build a coordinated collection of music toys useful for children and students, giving them direct tactile control over complex sound systems.

Other pedagogical experiments that we know of are further removed from our purposes because they mainly use visual representations or do not focus on sound

---

[1] URL: http://www.media.mit.edu/hyperins/

synthesis. For example, Baboni [2003] developed an initiation to musical composition via computer for children; *Frequency*[2] is a commercial software rhythm and music game played with audio triggers instead of visual ones.

To define the Dolabip project in this context, we can say that it is based on a musical pedagogy inherited from François Delalande. The concept of the Dolabip instruments is based on the *Mélisson*; but differs in that Dolabip is digital instead of analog, so that Dolabip instruments can be edited with software. Moreover, it is based on digital sound models permitting a wider range of synthesis techniques. The Dolabip project is also concerned with the kind of interfaces that children play with. Our experiments with children used a joystick because it is easy to equip schools with it. But interfaces developed by Tod Machover should also be very useful because they provide for various kinds of movement.

3. PEDAGOGICAL PROCEDURE

The creation of music comes from internal listening rather than representation. Music does not come from either visual representation or writing, but from sensation. Practicing an instrument involves only physical and sensual contact with the instrument. Thus, our pedagogical procedure is based on reflections about listening to music and on how music models thoughts and emotions. We are currently working on a project called *sound literacy*, which provides a methodological structure based on cognitive pedagogy. In this framework, active research and discovery are preferred to the passive transmission of knowledge. The creativity of the young musician is no longer subject to language and mental representations. Music is grasped in a sensory way, and action and sensation are the foundations of learning.

Teaching music has always been related to a particular aesthetic (classical music, jazz, rock, etc.). We address the question of the existence of a universal musical grammar. If it exists, we think that it is probably based at the level of sound. Thus, for teaching music, we suggest the term *acoustic art*; this new paradigm also provides a way to combine musical and technological knowledge.

3.1 The Action-Reaction Loop

Many early-learning programs are only oriented to listening and fail to develop creativity. We believe that creativity necessitates joining perception to movement in an action-reaction loop (see Figure 1). By doing so the learner is not just passively listening, but is also actively producing sound.

In this loop, the brain of the musician is always taking the nerve impulses coming from the auditory nerve into account in order to decide what to do. Memory and several other processes that anticipate the strategy and control of the instrument participate in this decision. The entire mechanism executes in real time, so that the simpler the processes are, the closer to perfection the result will be (based on the expectations of the musician). The complexity of the control process depends especially on the ergonomics of the device (see Figure 1) and the processes in the computer. Lots of parameters are necessary for music synthesis; but if the musician is given control of too many parameters, concurrent processes will certainly not be executed perfectly, and the musician will lose control of his instrument.

---

[2] URL: http://www.harmonixmusic.com

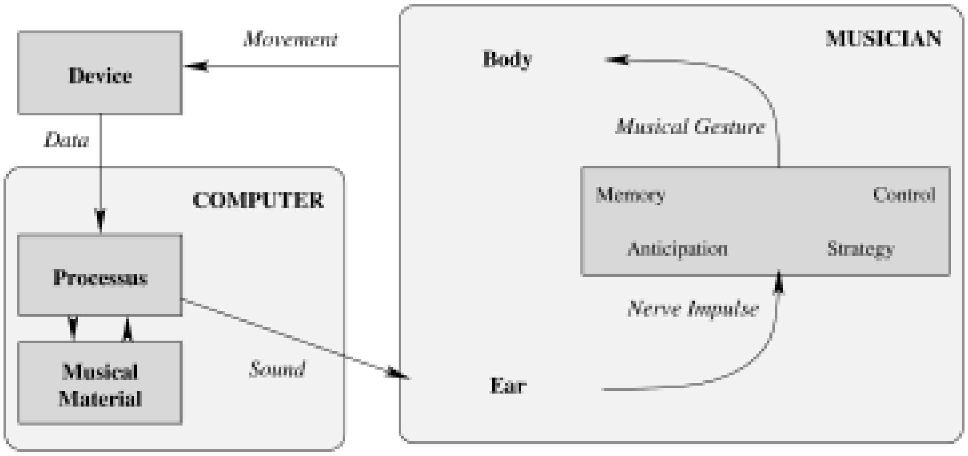

Fig. 1. Action-reaction loop.

Thus, in order to reduce the musician's cognitive overhead, it is important to enter as many pieces of information into the computer as possible. To provide real assistance to the musician, the question of the ergonomics of electro-acoustic instruments is key.

3.2 Practicing an Instrument

Traditional instruments often require special gestural skills (and it takes a long time to learn to play these instruments satisfactorily). As a matter of fact, at the beginning the correspondence between movement and listening is very uneven, so that a good interaction with the instrument is not immediate. As we stated in an introductory and general way in the previous section, the computer can help this process by assisting the musician's movements. In this way the musician can become a virtuoso very quickly, producing sounds of great quality with very simple movements. The computer may then add more and more complexity according to the progress made by the musician. Thus, the musician initially experiences pleasure from the quality of the sound he makes, and later from the virtuosity of his movements as they become more and more complex. With such a method everyone can experience the pleasure of playing a musical instrument, even when the musician is a very young child or a disabled person.

But even when the instrument is simple and movement is assisted by a computer, there is still a learning phase to master (which may be especially difficult for very young children). That is why we chose to make our first experiments with a joystick, assuming that a possible cognitive transfer from playing video games to playing music will occur and help reduce the cognitive overhead for young musicians.

3.3 Electro-Acoustic Instruments as Partial Musical Pieces

To assist the development of movement requires an analysis of the musical material in order to divide it into two parts: the musician's movements and their musical representation in the computer. The more the musician's movement is assisted by a computer, the more important the musical representation in the computer. The musical representation is used to compute the sound output from the input data, and constitutes an unfinished musical composition because input data is needed for completion**.** Thus, we can talk about partial musical pieces, that is, musical spaces within which the musician can evolve in various ways to produce a completely instantiated piece. This concept is very close to the one described by Brian Eno.[3]

3.4 Exploring Sound

In this project, we studied a special kind of partial musical piece that is well suited for the *exploration of sound*. When exploring sound, the whole system is considered a musical instrument rather than a musical piece. The musical representation in the computer is a sound space within which the musician can evolve by using the computer's discrete and continuous controls. In other words, the mapping between the input values (discrete or continuous) from the device and the musical internal representation is based on sound parameters (amplitude, frequency, etc.), rather than on musical parameters (tempo, harmony, melodic parameters, etc.). In this musical world, children are given a playing field to explore: when they discover a new variation in a sound, it is like reaching a new level in a game, along with the pleasure of discovery. Moreover, children are invited to give a purpose to their exploration by creating a musical discourse. Beyond the acquisition of musical skills, the children also experiment with orienting themselves in a multidimensional space.

4. GOAL AND PROJECT MANAGEMENT

The general goal of the project is to build electro-acoustic instruments, or unfinished musical pieces to satisfy our pedagogical requirements and to elaborate pedagogical programs around these instruments.

Within this framework, our first step is to define the instruments that are well suited for exploring sound. This step requires the definition of sound models and sound synthesis from inputs and sound parameters.

Instruments for exploring sound have to satisfy the following requirements:

- The instrument has to be easy to play, even for very young children. For this reason in our first experiments we chose to use the joystick, which has proven to be suitable for children in the context of video games. Moreover, this device is so attractive to children that teachers can easily persuade them to use it.
- The instrument must offer as various and rich musical possibilities as traditional instruments do. Hence we use models of sound that are close to the auditory system and based on analysis-synthesis. The models provide a procedure for analyzing natural sounds for translation, so that the sound palettes of instruments can be augmented by the instrument's maker.

---

[3] URL: http://www.wired.com/wired/archive/3.05/eno_pr.html
What people are going to be selling more of in the future is not pieces of music, but systems by which people can customize listening experiences for themselves. Change some of the parameters and see what you get. So, in that sense, musicians would be offering unfinished pieces of music – pieces of raw material, but highly evolved raw material, that has a strong flavor to it already. I can also feel something evolving on the cusp between "music", "game", and "demonstration" – I imagine a musical experience equivalent to watching John Conway's computer game of Life or playing SimEarth, for example, in which you are at once thrilled by the patterns and the knowledge of how they are made and the metaphorical resonances of such a system. Such an experience falls in a nice new place – between art and science and playing. This is where I expect artists to be working more and more in the future.

- Sounds and sound transformations provided by the instrument must be adapted to the perceptions of children, and the relation between the production of sound and movement must be easily controlled by children. In order to accomplish the goals above, we carried out this experimental research at school with several children.

To reach our first objective concerning sound exploration, we brought together specialists from several disciplines to address the multidisciplinary aspects of the project, as follows:

1. New results from research on modeling sound enrich the sound palettes of our instruments.
2. Software development is constantly evolving; it takes the experience of teachers who use software in the classroom into account. We develop a new software tool, Dolabip. Based on the models of sound resulting from point 1 above, the main problem of this development is the difficulty in designing the interface of this tool for making new instruments.
3. The programs are designed by a team of specialists in musical pedagogy for children; instruments adapted for this program are developed with the tool defined in point 2 above.
4. Early-learning sessions were carried out at school with children between the ages of 3 to 7, with instruments developed according to point 3 of this enumeration.
5. Psychoacoustic and psychological experimentation provides answers to questions about children's perceptions and their competence.

5. MODELS OF SOUND

In order to synthesize new digital sounds or manipulate existing ones using a computer, we need a formal representation for audio signals. A model of sound constitutes such a mathematical representation.

To produce sound, something must vibrate. More precisely, a mass located somewhere in space must oscillate around a position of balance. Such an acoustic source can be for example a pipe or a string, excited either once or periodically. If the sound is produced by a series of impulses occurring periodically, the associated spectrum is harmonic and the attack is quite soft. This is the case for the voice. The musician can then control the intensity, pitch, and timbre of such sounds. But if the sound is produced by a unique impulse, the spectrum is no longer harmonic and there are transients at the beginning of the sound, followed by a – often long – damping. This is the case for percussive sounds. The movements needed to produce such sounds are somewhat limited, since the musician controls only the very short attack and not the much longer damping period.

Regarding pedagogy, current research often considers only the percussive aspect of sound, whereas the originality of our approach is the focus on the sustained phases of sound. This lead us to design a specific sound model for our "virtual instruments" and to study relations between variations in movement and sound that are more suited for children.

Spectral models decompose sound in terms of frequency components, an operation which, to some extent, is similar to the decomposition process carried out by the human ear– more precisely by the resonance properties of the basilar membrane. Since spectral models are closer to perception, the design of audio effects should be more musically intuitive.

Additive synthesis is the original spectrum-modeling technique. It is rooted in Fourier's theorem, which states that any periodic function can be modeled as a sum of sinusoids at various amplitudes and harmonic frequencies. For stationary pseudo-periodic sounds, these amplitudes and frequencies continuously evolve slowly with time, controlling a set of pseudo-sinusoidal oscillators commonly called *partials*. This is the well-known McAulay-Quatieri representation [McAulay and Quatieri 1986]. The audio signal a can be calculated from the additive parameters using equations 1 and 2, where P is the number of partials and the functions f_p, a_p, and ϕ_p are, respectively, the instantaneous frequency, amplitude, and phase of the p-th partial. The P pairs (f_p,a_p) are the parameters of the additive model.

$$a(t) = \sum_{p=1}^{P} a_p(t)\cos(\phi_p(t)) \tag{1}$$

$$\phi_p(t) = \phi_p(0) + 2\pi \int_0^t f_p(u)\,du \tag{2}$$

The underlying real-time additive synthesis has been implemented in the *ReSpect* software tool [Marchand and Strandh 1999], whereas the parameters of the model can be extracted from sounds using, for example, the analysis method we proposed in Desainte-Catherine and Marchand [2000].

5.1 SAS: Structured Additive Synthesis

The first sound model we used is the SAS (Structured Additive Synthesis) model [Desainte-Catherine and Marchand 1999a].

This model can represent any pseudo-periodic sound provided that it is a monophonic source with no noise and no transients. One advantage of the SAS model is the independence of the musical parameters. It is possible to modify them and listen to them separately, and for example to stretch the sound in time without changing its pitch, to shift the pitch while preserving the timbre, to go forward, backward, and even to stop in time and "freeze" the sound, and so on. Another advantage of this model is its ability to produce hybrid sounds such as morphing or cross-syntheses.

This leads to a completely new way of listening to sound, where the listener is also the performer and needs to concentrate on auditory analysis.

The models based on additive synthesis are extremely difficult to use directly for creating and editing sound. The reason is the huge number of model parameters that are only remotely related to musical parameters perceived by a listener. The Structured Additive Synthesis (SAS) model keeps most of the flexibility of additive synthesis while addressing this problem. It imposes constraints on the additive parameters, giving birth to structured parameters that are closer to perception and musical terminology, thus reintroducing a perceptive and musical consistency into the model (see Marchand [2001] for a more complete presentation of the SAS model and its musical applications).

SAS consists of a complete abstraction of sound, according to only four physical parameters which are functions closely related to perception. These parameters – amplitude, frequency, color, and warping – are inspired by work on timbre by researchers like Risset [1986]; Wessel [1979]; and McAdams et al. [1999], and by the vocabulary of composers of electro-acoustic music. These parameters can be obtained from sounds in the classic time-domain representation using our InSpect [Marchand and Strandh 1999] software tool. We then describe a sound in the SAS model by means of the (A,F,C,W) notation. The first two parameters – amplitude A and (fundamental) frequency F – are one-dimensional, functions of time only, while the two others – color C and warping W – are two-dimensional, functions of both frequency and time.

Amplitude: A:*time→amplitude*. Human beings perceive amplitude on a logarithmic scale; amplitude is related to the intensity that corresponds to the volume in dB. In the additive representation, amplitude A corresponds to the sum of the amplitudes of all partials, and can be calculated from the additive parameters using Equation (3).

In order to consider the RMS (Root Mean Square) amplitude (closer to perception), Equation (4) must be used instead of Equation (3).

$$A(t) = \sum_{p=1}^{P} a_p(t) \tag{3}$$

$$A_{RMS}(t) = \frac{1}{\sqrt{2}} \sqrt{\sum_{p=1}^{P} \left(a_p(t)\right)^2} \tag{4}$$

Calculating the volume in dB from the amplitude is easy:

$$V(A) = 20 \log_{10}\left(\frac{A}{A_{0dB}}\right) \text{ dB} \tag{5}$$

where A_{0dB} is the amplitude of reference for 0 dB. In the remainder of this section we use the value $1/\sqrt{2}$ or 1 whether or not we consider the RMS amplitude A_{RMS}.

Frequency: F:*time→frequency*. Calculating the frequency from the additive parameters is trickier; the way to do so can be found in Desainte-Catherine and Marchand [1999b]. For harmonic sound, F coincides with the fundamental, possibly missing, or “virtual.” For these sounds, if the frequencies were integer multiples of this fundamental, we would have had something like this:

$$F(t) = \mathrm{gcd}(f_1(t), \Lambda\ , f_p(t))$$

where gcd is the greatest common divisor. In the general case this is the way to determine the F and W parameters: given a frequency f, let the harmonic frequency distortion for *f* be

$$W_f = ((rank_f(f_1) \cdot f, f_1), \Lambda\ , (rank_f(f_P) \cdot f, f_P)) \qquad (6)$$

with

$$rank_f(f_i) = \left[\frac{f_i}{f} \right]$$ ([*x*] being the nearest integer to *x*).

Denote by *F* the frequency for which the *y*=*x* line is closest to the ($rank_F(f_i) \cdot F, f_i$) points (in terms of the least-squared error). Then *F* is the frequency parameter of the model and the warping *W* is the continuous version (interpolation) of W_F. The frequency is also perceived on a logarithmic scale and corresponds to the pitch. For example, the MIDI pitch is a function of frequency

$$P(F) = 69 + 12 \log_2 \left(\frac{F}{440} \right) \qquad (7)$$

Color: *C*:*frequency*×*time*→*amplitude*. Color coincides with an interpolated version of the spectral envelope [Risset 1986]. Of course, for each partial *p* we have $a_p(t) = A(t)\, C(f_p(t), t)$. We call it color by analogy between audible and visible spectra. This analogy is already well known for noise (white, blue, etc.). Color and its manipulation are heavily used in contemporary popular music, although such manipulations are inherently present in the timbre of some ancient instruments like the didjeridoo, which shows very unusual color variations. For harmonic sounds the variations of color with time constitutes the timbre itself. For nonharmonic sounds, another parameter is necessary.

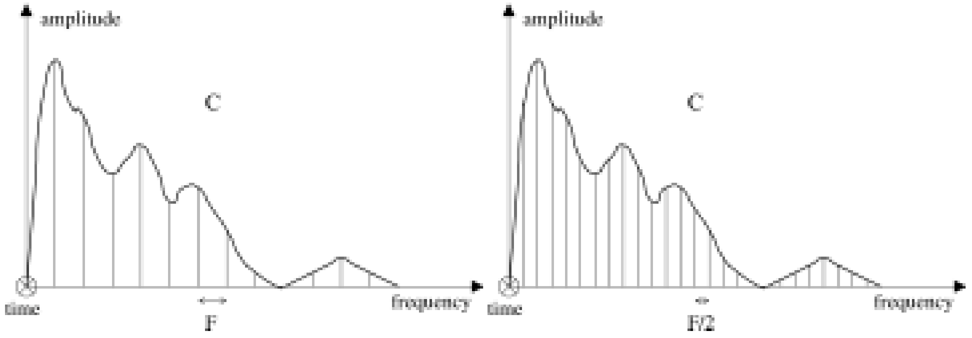

Fig. 2. Two harmonic sounds with the same color but different frequencies.

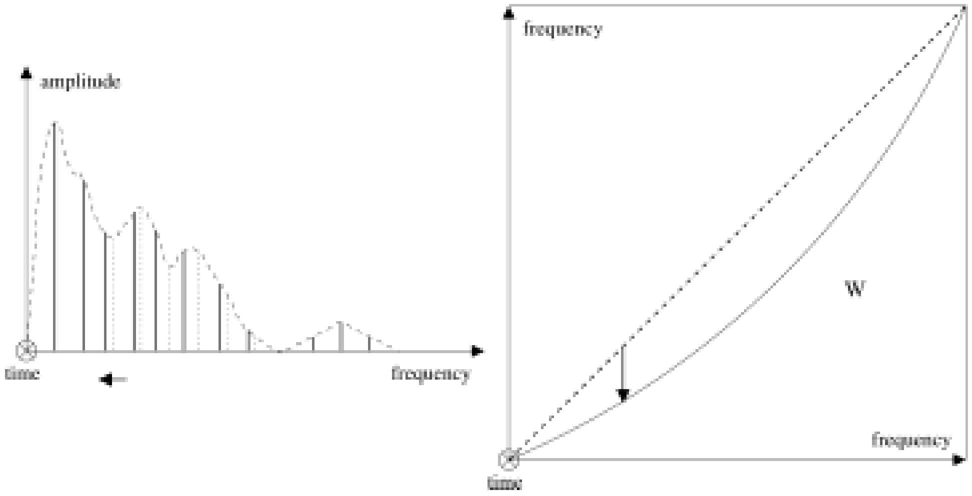

Fig. 3. A nonharmonic sound at time *t* (left) with its warping envelope *W* (right).

Warping: W:*frequency*×*time*→*frequency*. Harmonic sounds are totally defined by the *A*, *F*, and *C* parameters (see Figure 2). But when sounds are not perfectly harmonic [Mathews and Pierce 1980], the frequencies of the partials are not exact multiples of the fundamental frequency *F*. This is why the fourth parameter – called warping, after wavelet terminology [Evangelista and Cavaliere 1998] – gives the real frequency of a partial from the theoretical one it would have had if the sound had been harmonic, so that we have $f_p(t)=W(pF(t),t)$. Warping is related to nonharmonicity. Of course, for all harmonic sounds $W=\mathrm{Id}_1$, where $\forall f,t,\ \mathrm{Id}_1(f,t)=f$. Some sounds have a natural warping, such as those produced by pianos, gongs, or bells. Figure 3 shows such a nonharmonic sound.

All the SAS parameters vary slowly over time. More precisely, they are band-limited to a frequency around 20 Hz. Indeed, the variations of the parameters should remain inaudible, or else modulation phenomena will occur, which would put the perceptive consistency of the parameters into question.

The SAS parameters are closely related to the musical ones. Figure 4 shows the French *a* vowel sung on three notes. We can clearly see the dynamic and the melody of the song in, respectively, the *A* and *F* sound parameters. The arbitrary distinction between music and sound parameters tends to simply disappear.

Figure 5 shows a typical example of the application in Dolabip of the SAS model using the joystick to control the parameters of the sounds. The 10 joystick controls (axes, wheel, and buttons) at the bottom of the screen are used to control two SAS sound sources at the top of the screen in real time. On the first sound source, at the left, the amplitude *A* and frequency *F* can be changed – modulated – with a tremolo and vibrato, respectively. The color parameter *C* is filtered by a Gaussian-like band-pass filter, also controlled by the joystick. The warping or nonharmonicity factor *W* (*H* in the figure, which is the equivalent in French) can also be changed.

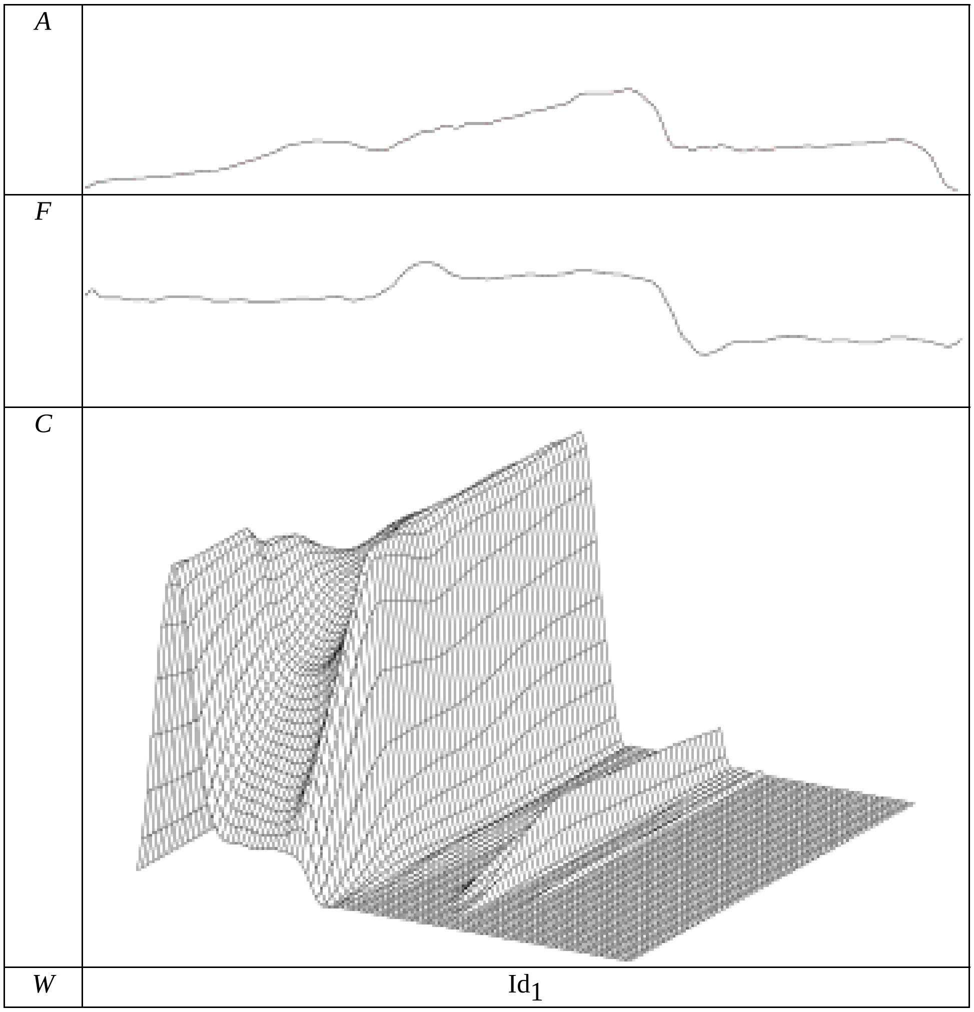

Fig. 4. Singing voice in the SAS model.

5.2 CNSS: Colored Noise by Sums of Sinusoids

In 2003 we began to investigate the synthesis of noise using the CNSS model (Colored Noise by Sums of Sinusoids). This analysis and synthesis model is adapted to monophonic sounds whose noisy part is high. A real-time implementation with existing free software (jMax) for real-time audio was already available, and is now incorporated in the general project. An example of a CNSS object for jMax is shown in Figure 6. The objective is to control all the synthesis parameters as fast as possible, while the sound is rendered.

The synthesis method is not presented in detail here. Readers can refer to previous articles [Hanna and Desainte-Catherine 2001; 2002; Hanna et al. 2002]. This synthesis

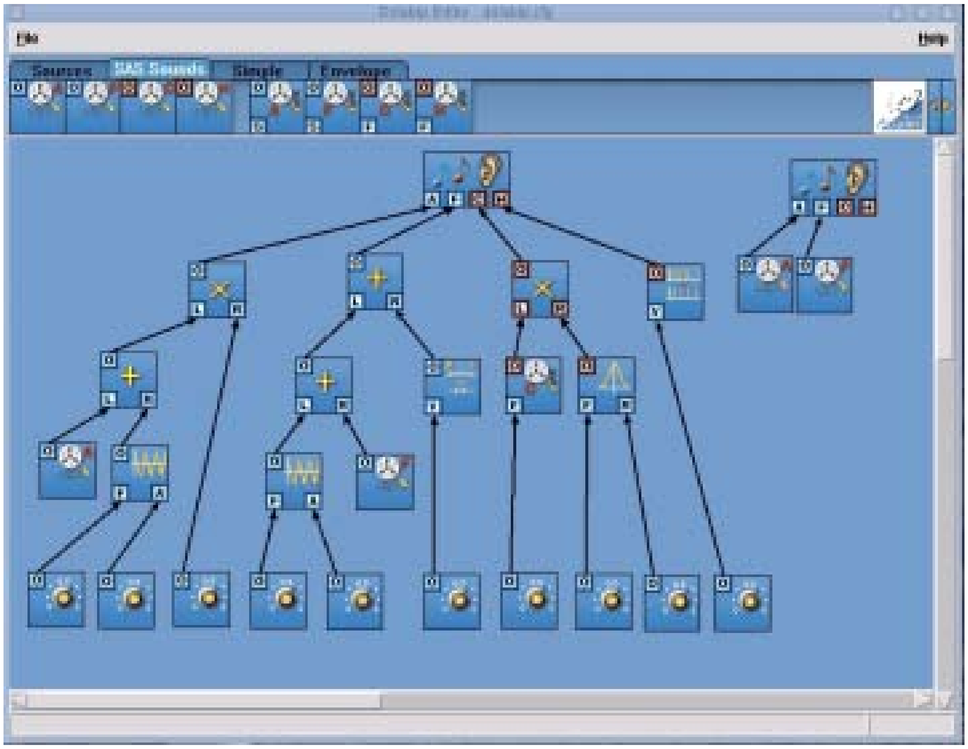

Fig. 5. Dolabip’s graphical user interface (first version).

method consists in generating successive short-term frames of noise in real-time. The CNSS model, both statistical and spectral, is based on the thermal noise model [Hartmann 1997] and considers sounds (with sample rate F_s) as random processes X. Each frequency component f_i is a random variable with fixed amplitude a_i and uniformly distributed phases ϕ_i:

$$X_k = \sum_{i=0}^{N} a_i \sin\left(2\pi f_i \frac{k}{F_s} + \phi_i\right) \tag{8}$$

where the frequencies f_i are distributed in a band whose width is ΔF (Hz). Once the oscillators are entirely defined, the real-time synthesis based on the SAS model synthesis is performed.

This model proposes original parameters to control the perceptual properties of noise.

Bandwidth. Users can set the bandwidth of the synthesized noise. So the band can be translated, broadened, or narrowed.

Spectral envelope. This parameter, also called *color*, is the one described in the SAS model, and is the successive smoothed spectral envelope. The only difference from the SAS model is the analysis process.

Spectral density. This original parameter is independent of the spectral envelope and is shown to be perceptually relevant through psychoacoustic experiments [Hartmann et al. 1986]. It is directly linked to the number and the distribution of the sinusoidal components. In order to control this distribution, we define two parameters: the number of bins M determines equal frequency intervals and the number of components N ($N \leq M$) sets the number of randomly selected bins inside which a frequency value is randomly chosen according to a uniform distribution.

Pulses. The model of thermal noise described in Hartmann [1997] considers each component phase as uniformly distributed. At the opposite end, noise synthesized with sinusoidal components with equal phases will result in noise with peaks of intensity. These peaks can be periodic, depending on the length of the synthesis window. Such noise is described as *impulsive*. By changing the width of the phase's probability density function, users can control the intensity of these peaks. By changing the length of the window size, users can modify the periodicity of impulsive noise.

Harmonicity. A parameter (denoted L) is related to the harmonicity of the synthesized sound. It implicitly defines the distribution of the difference between two successive sinusoidal components. Synthesized noise can thus be quasi-harmonic (*machine* noises) because sinusoidal components are equally spaced. Or, at the opposite end, increasing the randomness leads to *liquid*-like noises. This perceptual parameter cannot be modified using existing synthesis models based on filtered white noise.

Music or sound level. This parameter's attractiveness was unexpected, resulting from the work of composers who experimented with our synthesis tool. The length of the synthesis frame appears to vary with the scale in the structure of sound. If the length is long, the variations in the frame will be slow enough to allow control at the musical level. Users do not hear frequencies but pitches.

Figure 6 shows an example of an application of the CNSS model using the joystick to control the parameters. The joystick's buttons are used to start the synthesis and to modify the spectral density of the synthesized noise by defining the number of components N. The three axes of the joystick control the number of bins M, the higher frequency of the band $F_{\max}$ ($F_{\min}$ is fixed), and harmonicity. The joystick's wheel is used to modify the length of the synthesis frames.

This synthesis model defines original controls, in particular for the spectral density of noise. An analysis stage is being developed in order to extract parameters from natural sound. Once this analysis process is achieved, users will be able to transform natural sound. Although it is not completely related to this analysis process yet, it already provides a wide variety of soundscapes.

6. SOFTWARE DEVELOPMENT

The software acts as an instrument editor that provides a way to map device controls with sound parameters or variations. This software is used by musicians and teachers to build

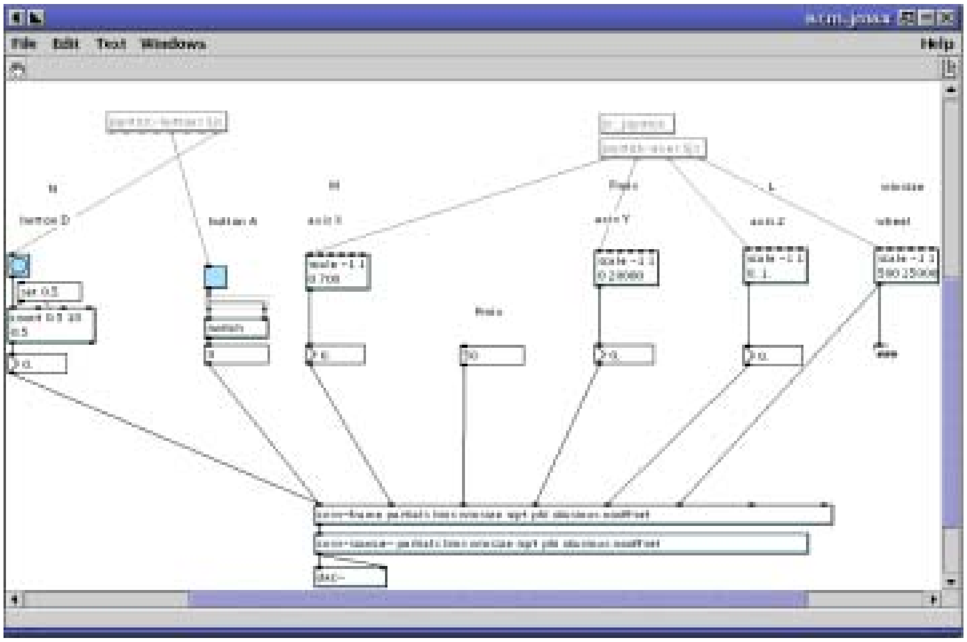

Fig. 6. Example of a CNSS object for jMax.

instruments. Children only play with the instruments, without knowing anything about their design.

Two versions of the Dolabip software have been developed and published under the GNU General Public License (GPL); the third version is under development.

6.1 First Version

The first version of the software was developed by the Algory company.[4] It runs under GNU Linux and is based on the SAS model and provides a way to design virtual instruments. A virtual instrument is software that associates movement that acts on a device to variations in sound parameters. For this version, the set of devices consists of a joystick, a MIDI table, and a keyboard.

The software's graphical user interface is presented in Figure 5. The software tool relies on a tree-like structure composed of interconnected processing nodes. Flows of SAS parameters travel between adjacent nodes; this structure computes the sound from the data either statically, stored in memory, or dynamically, sent in real-time by the external control devices (manipulated by the children). The SAS parameters produced by the root of the tree (top of the figure) constitute a sound in the SAS model, which is then synthesized in real-time. This real-time synthesis [Strandh and Marchand 1999] guarantees that there will be no clicks during live performance: in the worst case the sound will remain steady for a few milliseconds if the spectral data does not arrive on

---

[4] URL: http://www.algory.com

time. At last, teachers can dynamically edit (compose) the whole structure using a graphical editor.

6.2 Second Version

The second version also runs under GNU Linux, developed on the jMax platform[5] by A. Beurivé at SCRIME.[6] Its interface is harder to use for nonspecialists in jMax, but is more extensible and portable. Moreover, the sound model is no longer limited to that produced by SAS.

This version is a set of jMax objects that provides SAS and CNSS syntheses, as well as the Joystick input. The SAS synthesis is supplied by a set of objects for splitting an SAS frame into its four sound parameters or for building SAS frames from these four sound parameters. For now CNSS synthesis is a unique and very general object, which should, with experience, be specialized into several simpler ones.

6.3 Third Version

The third version of the software is still under development; it aims to simplify the interface while keeping its high expressive power. In this version we define a non-procedural interface, in contrast to the two preceding versions. We hope that with this version teachers will be able to edit existing instruments in order to adapt them to their pedagogical needs, and even make new instruments.

The expressive power of the first version of Dolabip is limited to SAS sounds, but it provides a very accessible interface. By contrast, the interface of the second version has the complexity of a programming language, but is no longer limited to SAS sounds. Our problem is to design software with a simple interface and reasonable expressive power, to provide a sufficient variety of possibilities in a pedagogical context. To solve this problem, we first analyzed the instruments developed by composers with Dolabip for pedagogical purposes. We observed that the initial idea is often a spatial one. The space explored by the two main functions of the joystick is divided into several regions [Wessel and Wright], each corresponding to a sound-synthesis technique which takes all real-time events as parameters. As a matter of fact, synthesis for exploring sound consists mainly in maintaining sustained phases of sound. It is necessary to maintain a kind of perceptual consistency around each point of the space, thus defining different spatial regions. For now, this requirement is satisfied in an empirical way, thanks to the skill of the composers. However, better knowledge of children’s perceptions will help us to define this property more precisely.

Two kinds of practical experiments were carried out in the public schools: practical experiments to test our pedagogical procedures in the real world and psychoacoustic and psychological experiments to try to answer questions about children’s perceptions and competencies.

7.PRACTICAL EXPERIMENTS

Two kinds of practical experiments were carried out in the public schools:: those to test our pedagogical procedures in the real world and psychoacoustic and psychological experiments to try and answer questions about children’s perceptions and competencies.

---

[5] developed at IRCAM (URL: http://www.ircam.fr)

[6] URL: http://www.scrime.u-bordeaux.fr

7.1 Pedagogical Experiments

Dolabip instruments can fit various pedagogical purposes. In the classroom, Dolabip is rarely used alone, but with other objects that make sounds. Many Dolabip instruments have already been designed by musicians and teachers with the instrument editor in the first version of Dolabip or with jMax objects.

We went on to experiment with early-learning games at school with children between the ages of 3 and 7. The groups of children were rather small (between 8 and 10 children). Educating children about sound consisted in tackling the notions of silence, resonance (attention), timbre, etc., presented from a perceptual point of view. The main objectives of the program were the development of perception, expression, and communication (composition and improvisation). Sounds were produced with voice, concrete objects, and Dolabip instruments.

A session began with the diffusion of a musical piece and was followed by a proposal for musical games. For example, several children were playing with the same Dolabip instrument and were faced with the following tasks: how to share a sound, how to make silence, how to make a sound (games made up of dialog, questions, and answers), how to communicate (watching other children, anticipating), how to progress from silence to making orchestral sound. Body movements and graphical expressions were often associated with musical work.

7.2 Ethological Observations and Psychoacoustic Experiments

The two main goals of the studies in this section are to adapt instruments to children's abilities and perceptions and to validate our method.

All these experiments were carried out in the context of a master's theses in cognitive science. The difficulty of experimenting with very young children is well known, necessitating the simplification of protocols, and doing and redoing experiments, so obtaining sound results takes a long time.

To reach the first goal, we made ethological observations on the ergonomics of the device and performed psychoacoustic experiments on sound timbre transformations. For the second goal, we began to study children's strategies for exploration.

Ergonomics. We observed young children and several ethological observations were done. The aim of the first experiment was to confirm whether the joystick is a good instrument for controlling the software (see Figure 7).

We developed three Dolabip instruments for this first experiment (see Figure 8).

- The first instrument is linked by the horizontal axis of the joystick to the temporal axis of an SAS sound. The SAS sound is a voice singing the three notes G A*b*, F. As the child moves the joystick from left to right and from right to left, the sound plays forwards and backwards. The speed of the child's movement controls the speed of the sound. The child can even stop his movement anywhere, to freeze the sound at that moment. The sound is always activated; silent areas are situated on the right and left extremities.
- The second instrument is divided into three areas. A silent area in the middle separates a pentatonic scale situated on the left from a harmonic scale situated on the right. The sound is triggered by the fire button, which is situated at the top and the back of the joystick.

Fig. 7. The joystick's stick, fire button, button 2, wheel, and base buttons (A, B, C, D).

- The third instrument links the X axis to a low-pass filter. The Y axis corresponds to the vowels (/*a*/, /∅/, /*i*/, /*o*/, /*y*/) extracted from SAS sound. Moreover, two sounds, S_1 and S_2, can be emitted simultaneously; S_1 situated in a low register and S_2 in a high register. Buttons A, B, C, and D located at the base of the joystick produce different frequencies for sound S_2 (494 Hz, 587 Hz, 659 Hz, and 784 Hz). Button 2 modifies the frequency of sound S_1 (73 Hz, 82 Hz, and 98 Hz). The wheel provides a global transposition of all the sounds.

Twenty-one young children were asked to play with the three instruments in a spontaneous way. Their hands were filmed during this exercise (between 1 and 7 minutes). The children were 5 to 6 years old and attended a 3-year school. The instructions were to play with the joystick; in particular, each child had to find out how to trigger a sound. We then analyzed the videos tapes by means of a behavioral grid and then went on to a variance analysis for each experiment. The composer was present during all experiments and also observed the videos tapes. He made some recommendations as a result of his own analysis. The behavioral grid consisted of the following four elements:

- The triggering and release of sound were measured according to four values: very slow, slow, fast, and very fast. The results provide strong evidence that, for the majority of children, the joystick allowed a fast rise in sound ($F(2,40)=6.55; P \leq 0.005$). The composer observed that even though the size of the joystick is better suited for adult hands, children were able to play with it, and found it attractive.
- Four different hand positions on the stick were observed: the right hand on the stick, the left hand on the stick, two hands on the stick, one hand on the stick and the other on the buttons. All the children were right-handed, and so prefer-

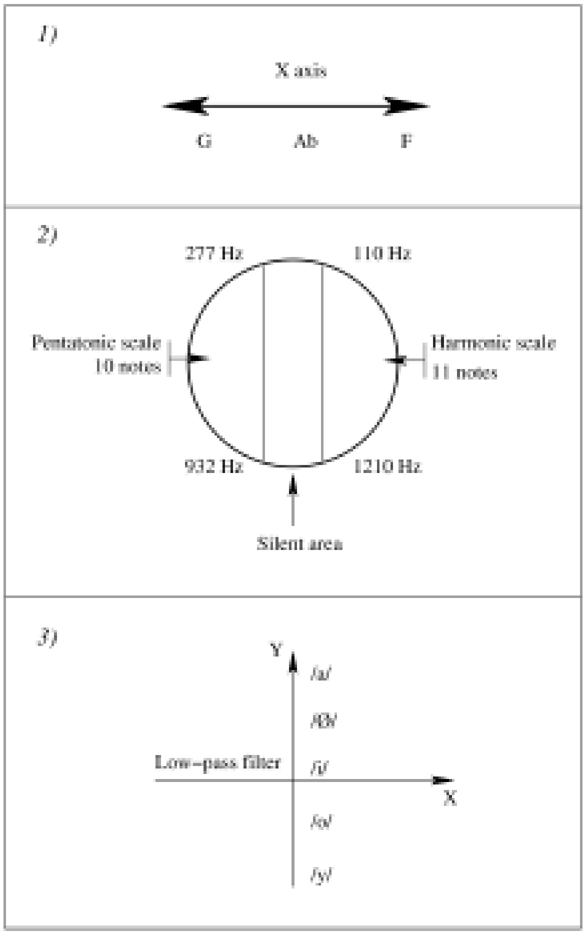

Fig. 8. Three instruments for the first experiment: definition of the 2D space explored with the joystick. (1) The horizontal axis of the joystick is linked to the temporal axis of an SAS sound. The SAS sound is a voice singing the three notes G A*b* F. (2) A silent area in the middle separates a pentatonic scale situated on the left, from a harmonic scale situated on the right. (3) The X axis is linked to a low-pass filter, and the Y axis to the French vowels (/a/, /∅/, /i/, /o/, /y/) extracted from an SAS sound.

red to use the right hand. For the third instrument, the majority of children used the right hand on the stick and the other on the buttons. Moreover, three hand positions were observed on the stick: the top, middle, and bottom. The children preferred to place their hands on the middle of the joystick ($F(2,40)=15.73; P \leq 0.005$). But the composer observed that toward the end of the experiment, the children often changed from one hand to the other, probably because they could not put their elbows on the table to rest their arms when they got tired.

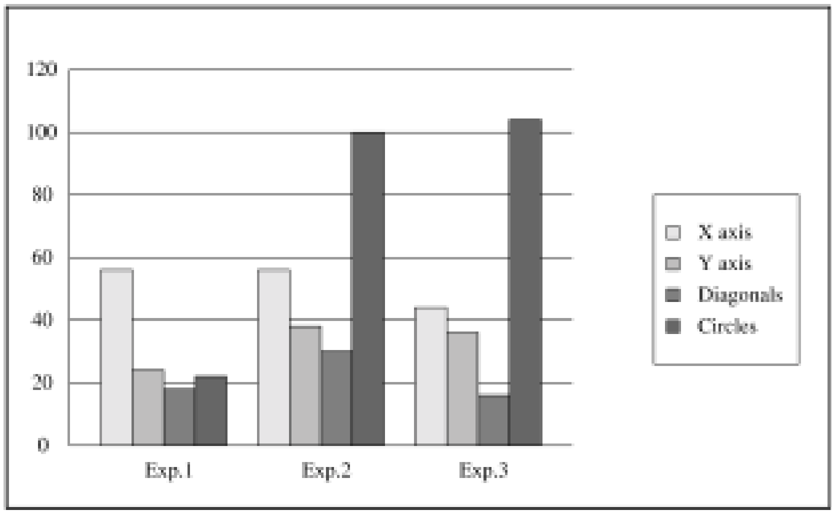

Fig. 9. A histogram of children’s motions for the three instruments along the X axis and the Y axis, that is, diagonals and circles. The first instrument causes lots of shifting movements along the X axis where the sound is situated. The second instrument induces lots of circular movements, providing interesting sound effects (filtering and vowel changes).

- Four kinds of movement were observed: along the X axis or the Y axis, and diagonal or circular. Most of the children spontaneously handled the joystick in a circular way; some used shifting movements along the X and Y axes; a few moved the joystick diagonally. We observed that movements were related to instruments (see Figure 9). As a matter of fact, the first instrument induced lots of shifting movements along the X axis, while the sound variations for this instrument are situated exclusively on this axis. For the second and the third instruments, circular motions were significantly more numerous ($F(3,60)=10.20; P \leq 0.005$). As a matter of fact, circular motions for the second instrument alternate the two kinds of scales, while for the third instrument they produce a rich sound by alternating two kinds of filters, low-pass and vowels.
- We observed the children’s exploration of all of the joystick’s buttons (A, B, C, D, wheel, inverse, 1, 2, 3, double-arrow, and fire); but this experiment did not provide any clear results. Nevertheless, if we refer to the (subjective) opinion of the composer, buttons that are situated on the joystick (1, 2, 3, double-arrow, and fire) were difficult for children to reach.

In conclusion, from the results of the first experiment, the joystick appears to be well-suited for this software and is attractive to children. In the second experiment, we observed correlations between motions with the joystick and the three instruments, showing that children related their motions to the resulting sounds. The third experiment gives the composer’s subjective opinion on the difficulty of using the buttons on the joystick.

Children and hearing. Several studies of babies' hearing have been carried out and lots of results have been obtained. Hearing in newborns is completely operational; children can hear and recognize very complex stimuli (see for example Thomas and Autgarden [1966]). Many studies show that the thresholds of hearing in children are rather similar, unlike those for adults. Children can distinguish very small variations in amplitude and frequency in the case of pure tones. By contrast, there are very few studies on children's ability to hear variations in timbre.

In 2003, we began a study on the perception of timbre. The goal is to understand the ability of children between the ages of 4 and 6 to classify timbre. For children, we decided to carry out a simplified version of Grey's experiments on perception of timbre distances in adults [Grey 1975]. We selected five sounds s_1, s_2, s_3, s_4 and s_5, each with the color of a French vowel, and built an experiment to study the perceptual proximities of these sounds. We presented each child with two different pairs of sounds, the first sound of each pair being always the same. The children were asked to tell which of the two following sounds was closer to the first one. For example, suppose the two pairs consist of the three sounds s_1, s_2 and s_3 in the following way: $P1 = \{s_1,s_2\}$ and $P_2 = \{s_1,s_3\}$. Thus, children were asked to choose between s_2 and s_3 for the sound that is closer to s_1.

Sounds were built by cross-synthesis between voice and violin SAS sounds. Let $S = \{s_a, s_e, s_i, s_o, s_u\}$ be the sounds of a female voice singing the five French vowels (/*a*/, /∅/, /*i*/, /*o*/, /*y*/). Let $v=(A_v,F_v,C_v,W_v)$ be the sound of a violin. Each sound s_k was defined by: $s_k = \{A_v,F_v,C_k,W_v\}$, where C_k is the color of the k-th vowel sound of S. With such a definition, sounds of S only differ from the color parameter that is the one we want to study.

Using this protocol we experimented with 15 children (6 boys and 9 girls) in two classes: class *A* for the second year of school and class *B* for the third year of school. With 5 sounds to compare, we obtained 30 pairs to be presented to the children. Very soon we observed that it was necessary to reduce the number of experiments from 30 to 15, so we selected pairs including the vowels /*a*/ or /*i*/ in the first position.

Conditions for a statistical analysis were not verified for this experiment so that we proceeded to a descriptive analysis, and thus results are rather limited. The children in class *A* all agreed that /*o*/ is closer to /*a*/ than /∅/, /*o*/ is closer to /*a*/ than /*y*/, and /*y*/ is closer to /*i*/ than /∅/. The children in class *B* all agreed on only one point: /*a*/ is closer to /∅/ than /*u*/. Moreover, results were rather heterogeneous for class *B*. Different results were observed for classrooms *A* and *B*, perhaps due to difference in the ages of the children. In conclusion, the protocol for the experiments has to be studied again, and more experiments with more children are necessary to validate and complete the results.

Validation of the pedagogical method. The validation of the method makes it necessary to evaluate the musical discourse produced by children. We began this study by analyzing their strategies in exploring sound. We carried out a first experiment whose aim was to analyze how many dimensions were explored at one time in a two-dimensional space. One sound target was given at the beginning of the experiment, and each child had to find the target within the sound space by moving the stick in the X-Y plane. Most of the children were able to find the target, but analysis of their movements did not permit any clear conclusions about their strategies. After this experiment, we reformulated the question as follows: Is there a correlation between creativity and the size of the movement? More generally, can we measure the creativity of children by analyzing their movements with Dolabip? Several analyses of motion are currently being developed to find a correlation with psychological tests of creativity [Torrance].

Summary. Practical experiments provide information to musicians and teachers for developing Dolabip instruments and map the joystick and the variations in sound to the ability of the children. Ergonomic studies of the joystick confirm that it is a good interface for the Dolabip software, and provide recommendations from a composer for eliminating the use of all the buttons on the joystick. Knowing the proximity of sound colors is useful in developing morphing operations between different sounds close to children's perception. Such morphing should give children a real sensation of continuity in variations. Finally, pedagogic experiments have permitted the definition of a new approach for developing software on the basis of spatial synthesis rather than procedural synthesis (see Section 6.3).

8. CONCLUSION

We presented a multidisciplinary project for developing early-learning games for electro-acoustic music in schools. We first developed a musical instrument as a toy for children to explore sound. The next step takes into account the musical organization of sound by defining temporal relations between sound objects, leading to the definition of partial musical pieces and giving rise to a system that provides a way to configure several interactions with the same piece, depending on the situation.

Pedagogical experiments will lead the development of this tool through concrete applications and psychoacoustic experiments, with a focus on children's perception of timbre and the musical organization of sound.

Teaching musical expression in schools necessitates more technology than graphical expression – the latter can require only a pencil and a sheet of paper. Making music with objects is very useful in developing children's attention, but organizing this material and recording it requires either the ability to write a score, which is too complicated for young children, or using special equipment for recording and editing, which is also too complicated and too expensive for school use. New technologies can bring progress by giving the child the ability to produce sounds, interacting with a musical piece as a conductor, musician, or composer. In classic Western music, conductors and composers occupy the top positions in the hierarchy, and the student has to study a long time and attend many classes before he can reach such a position. We address teaching young children to play any role in the musical hierarchy by means of games based on new technologies.

ACKNOWLEDGMENTS

This research was carried out in the context of the SCRIME project funded by the DMDTS of the French Culture Ministry, the Aquitaine Regional Council, the General Council of the Gironde Department, and IDDAC of the Gironde Department. The SCRIME project is the result of a cooperation convention between the Conservatoire National de Région of Bordeaux, ENSEIRB (school of electronics and computer scientist engineers) and the University of Bordeaux 1. It consists of composers of electro-acoustic

music and scientific researchers, and is managed by the LaBRI (laboratory of research in computer science of Bordeaux). Its main goals are research, creation, diffusion, and pedagogy.

REFERENCES

BABONI, J. 2003. Initiation musicale: vers la composition à l'aide d'outils informatiques. In *Proceedings of Journées d'Informatique Musicale* (Montbéliard, France).

DELALANDE, F. 1988. Le rôle des dispositifs dans une pédagogie de la création musicale enfantine. In *L'éducation musicale à l'école (actes du colloque départemental d'éducation musicale en Seine et Marne)*.

DESAINTE-CATHERINE, M. AND MARCHAND, S. 1999a. Structured additive synthesis: Towards a model of sound timbre and electroacoustic music forms. In *Proceedings of the International Computer Music Conference'99* (Beijing). International Computer Music Association (ICMA).

DESAINTE-CATHERINE, M. AND MARCHAND, S. 1999b. Vers un modèle pour unifier musique et son dans une composition multiéchelle. In *Proceedings of the Journées d'Informatique Musicale (JIM'99)*. CEMAMu, Paris, 59–68.

DESAINTE-CATHERINE, M. AND MARCHAND, S. 2000. High precision Fourier analysis of sounds using signal derivatives. *J. Audio Eng. Society 48,* 7/8 (July/Aug. 2000), 654–667.

EVANGELISTA, G. AND CAVALIERE, S. 1998. Dispersive and pitch-synchronous processing of sounds. In *Proceedings of the Digital Audio Effects Workshop* (DAFX'98, Barcelona, Spain). 232–236.

GREY, J. M. 1975. An exploration of musical tmbre. Ph.D. thesis, Dept. of Music, Stanford University.

HANNA, P., BEURIVÉ, A., AND DESAINTE-CATHERINE, M. 2002. Real-time noise synthesis with control of the spectral density. In *Proceedings of the Digital Audio Effects Conference* (DAFX'02, Hamburg, Germany), 151–156.

HANNA, P. AND DESAINTE-CATHERINE, M. 2001. Influence of frequency distribution on intensity fluctuations of noise. In *Proceedings of the Digital Audio Effects Conference* (DAFX'01, Limerick, Ireland), 125–129.

HANNA, P. AND DESAINTE-CATHERINE, M. 2002. Adapting the overlap-add method to the synthesis of noise. In *Proceedings of the Digital Audio Effects Conference* (DAFX'02, Hamburg, Germany*)*, 101–104.

HARTMANN, W. 1997. *Signals, Sound, and Sensation*. Modern Acoustics and Signal Processing AIP Press.

HARTMANN, W., MCADAMS, S., GERZSO, A., AND BOULEZ, P. 1986. Discrimination of spectral density. *J. Acoustical Society of America 79,* 6 (1986), 1915–1925.

MARCHAND, S. 2001. Musical audio effects in the SAS model. *J. New Music Res. 30,* 3 (Sept. 2001), 259–269.

MARCHAND, S. AND STRANDH, R. 1999. InSpect and ReSpect: Spectral modeling, analysis and real-time synthesis software tools for researchers and composers. In *Proceedings of International Computer Music Conference'99* (Beijing). International Computer Music Association (ICMA).

MATHEWS, M. AND PIERCE, J. 1980. Harmony and nonharmonic partials. Tech. Rep. 28, IRCAM.

MCADAMS, S., BEAUCHAMP, J. W., AND MENEGUZZI, S. 1999. Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *J. Acoustical Society of America 105,* 2 (Feb. 1999), 882–897.

MCAULAY, R. J. AND QUATIERI, T. F. 1986. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoustics, Speech, and Signal Processing 34,* 4 (1986), 744–754.

RISSET, J.-C. 1986. Timbre et synthèse de sons. *Analyse musicale,* 9–19.

STRANDH, R. AND MARCHAND, S. 1999. Real-time generation of sound from parameters of additive synthesis. In *Proceedings of the Journées d'Informatique Musicale* (JIM'99). ADERIM, Paris.

THOMAS, A. AND AUTGARDEN, S. 1966. Locomotion from pre- to post- natal life. *Clinics in Developmental Medicine* 24.

TORRANCE, E. P. URL: http://www.mhhe.com/mayfieldpub/psychtesting/profiles/torrance.htm.

WESSEL, D. AND WRIGHT, M. Problems and prospects for intimate musical control of computers. URL: http://www.cnmat.Berkeley.EDU.

WESSEL, D. L. 1979. Timbre space as a musical control structure. *Computer Music J. 3,* 2 (1979), 45–52.

Manuscript received October 2003.