Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces

D. ARFIB, J. M. COUTURIER, L. KESSOUS and V. VERFAILLE

Laboratoire de Mécanique et d'Acoustique (LMA), CNRS, 31 chemin Joseph Aiguier, 13402 Marseille, Cedex 20, France
E-mail: {arfib, couturier, kessous, verfaille}@lma.cnrs-mrs.fr

This paper is about mapping strategies between gesture data and synthesis model parameters by means of perceptual spaces. We define three layers in the mapping chain: from gesture data to gesture perceptual space, from sound perceptual space to synthesis model parameters, and between the two perceptual spaces. This approach makes the implementation highly modular. Both perceptual spaces are developed and depicted with their features. To get a simple mapping between the gesture perceptual subspace and the sound perceptual subspace, we need to focus our attention on the two other mappings. We explain the mapping types: explicit/implicit, static/dynamic. We also present the technical and aesthetical limits introduced by mapping. Some practical examples are given of the use of perceptual spaces in experiments done at LMA in a musical context. Finally, we discuss several implications of the mapping strategies: the influence of chosen mapping limits onto performers' virtuosity, and the incidence of mapping on the learning process with virtual instruments and on improvisation possibilities.

1. INTRODUCTION

This study started with an experiment where we linked several gestures to algorithms issued from a non-real-time musical work (Arfib and Kessous 2000). This has led us to define strategies that have been experimented in new virtual musical instruments and to examine the matter of mapping in a more general way. The work we present is a structured synthesis of this research project.

This paper is composed of three different parts: firstly an overview including a bibliography, secondly a description of our own experiments and the ideas associated with them, and thirdly a synthesis of the implications of our mappings on musical expressiveness and on musical play.

The first part is a general overview that deals with the relationship between gesture and sound; it includes a description of several mapping types. We more specifically describe a multi-layer mapping using perceptual spaces, a configuration that seems very important to us.

The second part first describes the general guidelines that govern our experiments, such as the modularity and the focus on perception; then it shows six experiments of sonic situations where this three-layers perceptual mapping is partly or fully used. The instruments we developed and experimented with use alternate controllers and classical synthesis models or effects. Moreover, our mapping chain uses perceptual spaces. The use of a perceptual layer is really specific to our approach: the counterpart of the complexity of mapping layers is the intuitive navigation in perceptual spaces, and an easiness to play music with those kinds of instruments. Finally, an evaluation of the mappings is done through the musical use of these instruments and effects.

The third and last part presents some consequences of the use of this mapping on the freedom and constraints that they bring up and on the learning process they induce.

2. GESTURE AND SOUND PERCEPTION REPRESENTATIONS

In humans, hearing involves perceptual representations thanks to its analysis system – the internal ear – and to its information abstraction system – the different integration stages in the brain. In the same way, a gesture can be, for example, visually analysed and recognised. We believe that it is important to have a perceptual layer appearing in the mapping. This is why we introduce gesture perception and sound perception in the mapping chain.

2.1. Mapping chain

First of all, let us present the mapping chain (figure 1). To describe the processing stages between a gesture and a sound created by the synthesis model, we use three main steps. The first mapping deals with the interpretation: it transforms gesture data into related-to-gesture-perception parameters, by means of a gesture extraction algorithm. This means that we transform measures, quantitative data, into more qualitative information, nearer to the perception. By *related-to-perception parameters*, we mean parameters that make sense to our perception. The subset consisting of the gesture measurement tool – a gesture transducer or a musical gesture analyser – and this mapping is a tool designed for instrument expressiveness. Let us specify that we consider two senses for gesture: a physical gesture, measured by a transducer, and a musical gesture done by the interpreter

Organised Sound 7(2): 127–144 © 2002 Cambridge University Press. Printed in the United Kingdom. DOI:10.1017/S1355771802002054

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

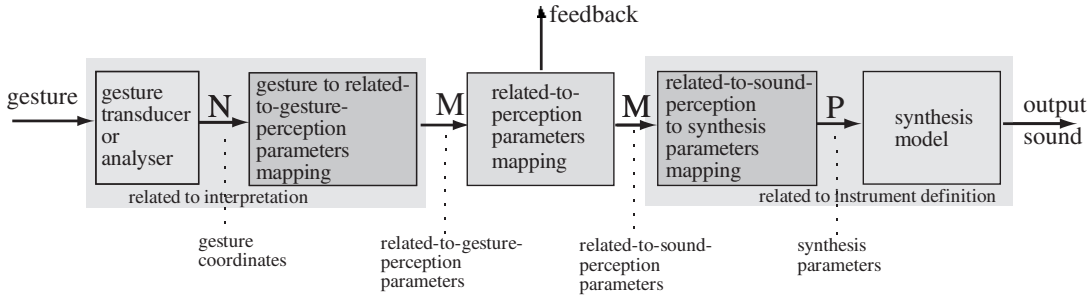


Figure 1. Information treatment from gesture to sound. Gesture is translated into physical data by the transducer; gesture data are then transformed into related-to-gesture-perception data; a second mapping transforms these data into related-to-sound-perception data. Finally, a mapping transforms related-to-sound-perception data into synthesis model data. N, M and P are the number of parameters.

or the composer, in our case measured only from a sound signal.

Sound can also be represented in a perceptual space. The second mapping transforms related-to-gesture-perception parameters into related-to-sound-perception parameters; this should involve a consistent hearing feedback.

On the right of figure 1, related-to-perception parameters, among which are psycho-acoustic parameters, can be computed from signal parameters and constitute the sound perceptual space. These parameters can also be used to compute synthesis parameters of the synthesis model. The third mapping transforms related-to-sound-perception parameters into synthesis-model parameters. The second subset consisting of this mapping and the synthesis model achieves the instrument definition.

This mapping chain designed in three parts is an improvement of the two-part mapping proposed in Wanderley, Schnell and Rovan (1998) and Wanderley and Depalle (1999). As proposed by Métois (1998), we use a perceptual layer; however, we explicitly develop the way to map from and to it.

The structure proposed is modular. This is very interesting in practice, because it allows the use of several gesture transducers considered as alternate controllers (Wanderley 2001) – graphical tablet, joystick, touch surface, data glove, driving wheel, etc. – or analysers, which can control several synthesis models (Wanderley *et al*. 1998, Hunt, Wanderley and Kirk 2000, Wanderley and Depalle 1999). It also offers the possibility of a visual feedback at different stages, such as the measure of the gesture, the related-to-gesture-perception parameters or the related-to-sound-perception parameters, or the synthesis parameters.

2.2. Sound descriptions

Let us now present descriptions of the sound through related-to-sound-perception parameters, sound perceptual space, and mappings between this space's parameters and synthesis parameters.

2.2.1. Related-to-sound-perception parameters

There are different kinds of related-to-sound-perception parameters. In physical modelling synthesis, some physical parameters have a direct relationship with perceptual parameters. Some signal parameters such as the spectrum centroid can also have this kind of relationship. Meta-parameters are used in synthesis to control a larger set of parameters; and psycho-acoustic parameters are descriptors of human perception of sound (figure 2).

First of all, some **physical parameters** of the sounding object can be related to perception: volume of the object, material, elasticity, etc. (MacAdams and Bigand 1994). In an article about timbre exploration by analysis and synthesis (Risset and Wessel 1999), the authors explain that an advantage of physical sound modelling is that its control parameters are simpler and more intuitive than certain control parameters for signal synthesis techniques – such as a modulation index for nonlinear distortion, for example – because they are correlated to a physical reality we are used to evaluating. These parameters are low-level features in the sense that they are really close to the model, but also high-level features when they have a strong perceptual significance. One can change their values – normalisation, translation, scale change, linear combination – to obtain higher level parameters, better related to perception.

There are **signal parameters** related to perception. For example, the energy measure using the root mean square value, even though different from loudness – the psycho-acoustic attribute for the perceived power – allows an evaluation of the evolution of the signal that can be close to the evolution of the loudness. The fundamental frequency of a harmonic sound is often assimilated to the pitch – its psycho-acoustical counterpart – because it is really close. The fundamental frequency can be substituted for the pitch when the mapping sensitivity is not too great. A last example is the spectrum centroid, correlated to the psycho-acoustic parameter known as brightness (MacAdams, Winsberg, Donnadieu, De Soete and Krimphoff 1995). These parameters

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

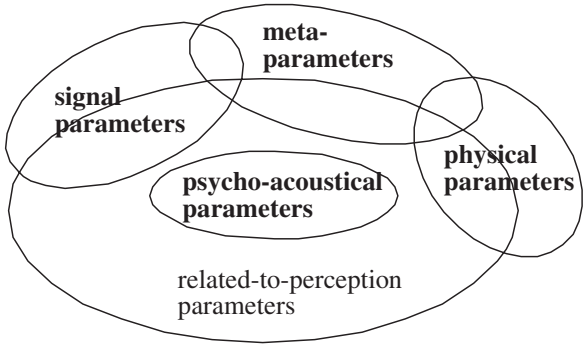


Figure 2. Parameters extracted from a sound. The set of related-to-sound-perception parameters includes all psycho-acoustic parameters, some signal parameters, some physical parameters (from the physical properties of the sounding object) as well as some meta-parameters.

should not replace their psycho-acoustical counterparts if one wants to control and modify a sound with precise real-time controls of psycho-acoustic features.

One may want to manipulate a set of parameters of the same level in a global way, thanks to a small subset of parameters, for example a curve providing its N parameter values. These high-level parameters relating sound and gesture are more strongly linked to the synthesis model than the physical parameters of the model: they are called **meta-parameters**. These meta-parameters can be described via a curve that can be translated or warped; they can also be curve description parameters – parts of the curve (Jensen 1999), order of the approximation polynomial for spline curves, for example. When manipulating meta-parameters with a profile, one can consider a particular mapping configuration. In that case, when the profile is changed, the mapping is changed. To qualify these meta-parameters as related-to-sound-perception parameters, their manipulation must have a perceptual effect on the sound.

Human auditory perception attributes are called **psycho-acoustic** parameters. They can be extracted from a time–frequency analysis of the signal. Some correspond to a characterisation of the instrument or sounding object – pitch, loudness, timbre, auditory space – at a micro-level, some to the musical performance – for example vibrato, legato, or loudness rhythm and timbre shades – at a more global, macroscopic level. Among those the psycho-acoustic literature proposes as descriptions of the sound, we keep a small subset of parameters, used in other studies. This subset will help give access to the expressiveness. It defines a four-dimensional space: pitch, loudness, timbre and spatialisation. Timbre is also considered to have many dimensions, among which we usually use the brightness and the attack-time logarithm. One may also use more timbre descriptors, such as roughness, signal-to-noise ratio; for nearly harmonic sounds, one may use vibrato, formant positions, synchronicity of partials during attack, harmonics' jitter and shimmer – both being at the limit between signal parameters and psycho-acoustical parameters, but having a sense in the source characterisation – partials' harmonicity, as well as the power ratio of odd and even harmonics. This set of parameters used to describe timbre has redundancies and does not consist of independent dimensions, but it corresponds to a useful signal and psycho-acoustical description, helping to analyse it as well as to synthesise it.

For these four kinds of parameters, we want to take into account the temporal evolution of the sound. In that case, we have to use the **derivatives** of some parameters. As an example, the loudness derivative helps to differentiate attacks and decays of sound portions having the same timbre, and it serves to keep or remove the temporal axis (Wessel, Drame and Wright 1998).

2.2.2. Sound perceptual space

The definition of a sound perceptual space – or related-to-sound-perception space – is ambiguous as long as we do not establish what is to be represented: sounds or patterns in the sound. The representations proposed in this paper concern mostly a low-level sound representation, describing the virtual instrument, and not higher-level musical patterns.

For the sound representation, the psycho-acoustic space – pitch, loudness, timbre, spatialisation – corresponds to the notation of written music, with several refining levels. Pitch changes correspond to the melody, loudness to shades, timbre to the instrument, and spatialisation to the localisation of the source in space. The timbre space corresponds to a sound-control space used a lot in contemporary music, but even more in electro-acoustic music. The first example of this space is in Grey (1975): for a given pitch, they use a two-dimensional space for brightness (related to timbre) and loudness. Wessel (1979) also resorts to timbre space to propose musical trajectories. Beauchamp (1982) does synthesis from loudness and brightness. Wessel, Drame and Wright (1998) control the sound through pitch, loudness and brightness. One can also create a sound colour space

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

(Slawson 1985) which can be explored, for example using a singing voice synthesizer.

2.2.3. Mappings between sound parameters and sound perceptual space

The mapping between related-to-sound-perception parameters and synthesis parameters can be considered in both directions. We know how to compute related-to-sound-perception parameters from a sound analysis, so we can go in the other direction by giving the good synthesis parameters of the synthesis model, in order to obtain a sound for which we control related-to-sound-perception parameters.

For extraction of psycho-acoustic parameters, different models exist for a given parameter; all use the input signal and knowledge about the functioning of the inner ear. They are implemented with algorithms given by authors such as Terhardt (1979) for pitch extraction from complex sounds and from harmonic ones, MacAdams *et al*. (1995) and Krimphoff (1994) for the use of spectrum centroid as the brightness, Zwicker and Scharf (1965) and Moore and Glasberg (1996) for loudness, Arfib and Delprat (1998), Rossignol, Depalle, Soumagne, Rodet and Collette (1998) and Desain and Honing (1996) for vibrato extraction, Daniel and Weber (1997), Pressnitzer (1999) and Leman (2000) for roughness computation, and Dubnov and Rodet (1997) and Jensen (2000) for harmonics jitter and shimmer. The mapping in the other direction, from related-to-sound-perception parameters to synthesis parameters, is more important for our study. It depends on the synthesis model and the kind of sound control we want: precise and accurate, or approximated, only taking the trends into account (this is the case when the pitch is given by the fundamental frequency and not the exact perceived pitch).

Synthesis methods began with non-real-time softwares such as the MUSIC V program or the Csound program. Now they can be implemented in real-time programs such as Max/MSP, but their modular principle is still the same so that given scores can still be used with these implementations. Traditionally, the synthesis methods are classified (Fischman 1999) as additive, with a special focus on grouped additive synthesis that reduces the data; subtractive, which includes many studies on vocal sound, but also musical applications of delay lines (comb filters, Karplus–Strong style instruments); global synthesis such as FM – frequency modulation – and waveshaping – also called non-linear distortion – and granular synthesis in which grains can be either synthetic or taken from acoustic recorder sounds. Some research has been done on imitation of instruments with synthesis techniques, either by trial and error, or by extraction of parameters from sounds (Wessel *et al*. 1998). Such extraction is used in phase-vocoder techniques, expanded to more complex methods extracting individual partials, noise and transients.

Each synthesis method makes assumptions on the signal to synthesise, and often uses many synthesis parameters to define the synthesised sounds precisely. The aim is not to control one by one each of these parameters; rather studies are done to extract a subset of meta-parameters, from which all synthesis parameters are computed automatically – Desainte-Catherine and Marchand (1999) for structure additive synthesis, Beauchamp and Honer (1992) for nonlinear distortion, Hélie, Vergez, Levine and Rodet (1999) for physical model inversion, and Jensen (2000) for a timbre model. We think that when this mapping layer exists, it can be considered part of the synthesis model, for modularity reasons. Whenever there is a synthesis model with such an access to a small set of meta-parameters, it can be used directly to explore the expressiveness possibilities of the instrument. Another way to create the mapping from related-to-sound-perception to synthesis parameters consists in first analysing a sound by a method for which we know the associated synthesis method, and by linking the analysis/synthesis parameters to the related-to-sound-perception parameters extracted from the analysis. This is the aim of imitative synthesis (cf. example section 3.2.1).

2.2.4. Mapping categorisation

We will now distinguish mapping types according to the explicit/implicit, simple/complex and dynamic/static criteria.

In Hunt *et al*. (2000), two types of mappings distinction are defined: explicit mappings and mappings using generative mechanisms. We make a distinction between explicit mappings with mathematical expressions clearly defined for each mapping connection, and explicit mappings with a general rule (e.g. with database).

Explicit mapping corresponds to the case in which one can exactly describe the links between the input and the output mapping parameters, thanks to mathematical formulae. Simpler relationships are also possible, such as first making a linear or nonlinear combination of parameters and then applying a linear or nonlinear function to it (Verfaille and Arfib 2001). Whenever one chooses to express the mapping relationships analytically, one gains insight into what happens in the mapping chain. The use of explicit mappings offers a better visibility on what is computed. It can also allow us to define in detail a specific expressivity of the instrument.

Implicit mapping corresponds to the case in which the mapping box is considered as a black box for which we define behaviour rules but not precise value rules. One describes the main aspects of the model and its way of functioning, and the mapping rules are set through parameters we cannot understand. The definition is more global and does not allow great visibility – since the inner parameters of the mapping are not understandable to the user. These mappings can be statistical tests for

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

decisions at different stages of classification, for example a Fisher discriminant test (Martin and Kim 1998), optimisations of probabilistic models used to predict the behaviour of an analysed system, or artificial neural network (Wessel *et al*. 1998); they can also be maxima estimation in classification models (Scheirer and Slaney 1997). In this case, mappings are clearly nonlinear.

In Hunt *et al*. (2000), **complex** mappings are defined as many-to-many; **simple** mappings as one-to-one. Most of the mappings presented in our three-layers mapping are complex mappings if related-to-perception parameters are to be linked to gesture data or sound synthesis parameters, as shown for instance by the use of artificial neural network or database. On the contrary, mappings between related-to-gesture-perception and related-to-sound-perception parameters are not necessarily complex. They could be simple, direct, linear, maybe more adapted to neophyte performer, or complex, probably more adapted to expert performer. The possibility of choosing this mapping layer, due to modularity, allows us to differentiate several levels, from beginner to expert.

The **dynamic** or **static** behaviour of the mapping appears at different interpretation levels. The first one corresponds to the ability of the mapping to evolve in time, to learn from the input data over time. The second interpretation corresponds to the use by the mapping of dynamic description parameters for gestures. Of course, it can also be a combination of these two ideas, thanks to an adaptive mapping using dynamic parameters. In the case where the mapping is adaptive, it can correspond to the evolution from a mapping to another, smoothly or abruptly – for example the morphing from one timbre to another, or the changing of timbre model – as well as the adaptation of one mapping. It can be the adaptation of the mapping to the performer's expressiveness. Focusing on a precise zone in a small part of the gesture perceptual space, one could use a kind of zoom to explore this zone more precisely, before proceeding to another portion of the space. This can be done through the computer mouse: slow movements of the mouse are associated with low-amplitude motions of the pointer, whereas rapid motions of the mouse are associated with greater motions of the pointer. On the other hand, optimisation methods can change the mapping in time.

In the case of implicit mappings such as artificial neural networks, such an adaptation of the exploration domain can be achieved by minimising the error between the perceptual parameter values to be reached and those predicted by the model. This is called the learning phase; it can be considered a preliminary task for the user – training before using the model – as well as a task going on while playing. One can also use genetic algorithms (Honer 1995) to try different mapping possibilities, keeping in the whole set of possibilities the mapping best fitting the constraints, according to the expressiveness wanted. One can consider several learning levels: the first one corresponds to the model adapting itself to an instrument, the second one extracts general rules, templates, from which the performer can change instrument or its expressiveness. Mappings taking into account dynamic parameters are very interesting because they take into account more information about the gesture and the intention, the way to evolve and to go from one part of the perceptual space to another. Expressiveness may be coded into instantaneous parameter values, long-term parameter values, or derivatives. To keep an access to expressiveness, we use temporal and spatial derivatives of parameters in mappings.

Mappings and their input parameters imply a **variation range** for input and output parameters. This variation range clearly defines the boundaries of the domain for the performer to explore. Adapting mapping to the performer creates evolutions of these boundaries, for example when gesture is confined in a small part of the gesture perceptual space whereas the sound explores a greater part of the sound perceptual space. We can, however, decide not to use such an adaptation. In both cases, limits are defined according to two criteria. The first one is technical: convergence and stability of algorithms, physical limits of the synthesis model and of the gesture transducer. The other criterion is subjective, and is about aesthetical limits. It needs the judgement of the composer or the performer using the instrument, and also the judgement of the instrument's designer. According to each one's knowledge, background and tastes, one can consider by listening to the synthesis model one plays which zones of the gesture space are of interest, and allow or disallow the exploration of such zones.

2.3. Gesture descriptions

We now come to the gestures and all that is related. As well as for the sound, we are going to deal with gesture data and descriptors, gesture perceptual space, and mapping linking gestures to the gesture perceptual space.

2.3.1. Related-to-gesture-perception parameters

We consider two senses for the word 'gesture' in this study: a **physical gesture**, measured by a gesture transducer – this gesture creates the sound on a usual instrument – it can also be **a musical gesture**, in our case extracted from the signal by an analysis.

The physical gesture is physically measured through a gesture transducer. Parameters given by the transducer can be static: position, distance from a point – using Cartesian, polar, cylindrical or spherical coordinates – orientation – azimuth and elevation. They can be dynamic parameters of the gesture as well: speed, acceleration, curvature. A detailed study of physical gesture

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

is presented in Cadoz and Wanderley (2000). The authors give a typology of gesture: excitation gesture, instantaneous or continuous, where gesture and sound co-exist; modification gesture, which can be parametrical or structural; selection gesture, allowing a choice among several similar elements of the instruments.

Using musical gesture relates to the musical meaning given by the sound performer, meaning which we can extract from the signal, with long integration time, for example transition type – portamento, legato, pizzicato – and modulations – vibrato, roughness. The musical analysis done by the human auditory perception system can, from the signal, identify physical gestures such as selection, for example timbre changing, excitation – plucked or rubbed string – or modification. Note that the effect of the physical gesture appears in the musical gesture, even though we cannot always notice it.

From the gesture data, whatever they are, we extract related-to-gesture-perception parameters, which are more qualitative. They can be, for example, the variation of the localisation, the more or less great variability around a mean curve, the relative or absolute amplitude, the combination (correlation) of evolution of several parameters, as far as a physical gesture is concerned. They can also be a melody, a playing style (rubato, legato), a pitch change or transition type, rhythmic information for musical gesture extracted from the signal, or identifiable patterns in time, such as slowing down/speeding up, or in space, such as regular or break point trajectory, for both gesture types. Several description levels exist: microscopic – the variability around a mean value or pattern – and macroscopic – the pattern itself.

2.3.2. Gesture perceptual space

We can create a gesture perceptual space by taking as axes all the related-to-gesture-perception parameters, or only a subset of those parameters, depending on the application. We introduce human perception in physical and musical gesture analysis to permit the design and the development of human–computer interfaces better adapted to interpretation and improvisation. As a matter of fact, the information we get from a gesture is mainly qualitative, whereas the physical measure we get is quantitative, and less adequate.

2.3.3. Mapping between gesture data and gesture perceptual space

Once the gesture has been measured in a geometric space – gesture coordinates, signal parameters – we change this representation into a more subjective and perceptual one, thanks to related-to-gesture-perception parameters such as those presented in the preceding parts, using mapping functions.

The link between gesture data and related-to-gesture-perception parameters is given by algorithms. Indeed, a gesture of any kind can be identified using methods such as the ones for the sound – multidimensional analysis, artificial neural networks – and then represented among several dimensions. It is an M to N mapping that can be explicit or implicit. In general, physical gestures are directly used with explicit mappings, whereas musical gestures need form recognition approaches or global information extractions, thus implying implicit mappings. However, the differences come more from working in several ways than from profound reasons: mapping techniques are more and more mixed for the two kinds of gestures. Indeed, we may want to recognise a form in a physical gesture – mimophony study in progress at LMA – as well as to use the related-to-gesture-perception explicitly to drive a model, as in example of the adaptive audio effects, in section 3.6.

2.4. Mapping between perceptual spaces

Perceptual spaces were investigated in several studies, as illustrated in section 3. As we already said, we believe that introducing human perception in the analysis and use of sounds and gestures allows the design of human–computer interfaces best fitted to interpretation and improvisation. Intermediary perceptual spaces, which are high-level information spaces, serve to describe a gesture and a sound in a more qualitative and easier way. This implies a direct mapping between the gesture perceptual space and the sound perceptual space (Arfib and Kessous 2001, Arfib 2002). That way, however complex the mappings from data to perceptual spaces are, a direct mapping between perceptual spaces makes the instrument easy to use, according to easy-to-learn schemes. Ease of learning implies that the gesture perceptual space and the sound perceptual space are of the same type, and that bijections between subspaces exist. This is equivalent to considering that axes of a subspace of gesture perceptual space are merged with axes of a subset of the sound perceptual space, for example when vertical movement is associated with a pitch change and horizontal movement with a timbre evolution. By introducing these two perceptual spaces, we separate what concerns the instrument's expressiveness – the gesture transducer or analyser and the gesture data to gesture perceptual space mapping – from what concerns the definition and modification of the instrument itself – from related-to-sound-perception parameters to synthesis parameter mapping and synthesis model.

Another important space is the user feedback. At several stages of the mapping chain, one can ask for a visual feedback to have a better mental representation of gestures during the instrument-learning phase. Indeed, learning to play a traditional instrument involves a multi-modal feedback that is of course auditory but also

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

visual and tactile. Auditory feedback concerns the learning of musical quality; visual and/or tactile feedback concerns the integration of relationships between gesture and produced sound, which help to understand the behaviour of the instrument as well as rules and limits induced by gestures. As far as virtual instruments are concerned, the auditory feedback is predominant. Then, visual feedback is sometimes used to show how the gesture analysis algorithm has decoded the information emitted by the transducer. Finally, tactile feedback is used, but not as widely as in the game industry. Research teams are developing multi-modal platforms, but industry is not ready to propose their use in virtual instruments (Cadoz and Wanderley 2000). We believe that such multi-modal virtual instruments will simplify the learning-to-play synthesis models.

3. EXPERIMENTS PERFORMED AT LMA

Let us now present some experiments we did in our team. After giving the guidelines of our work, we present six examples of the use of the three-layers mapping with perceptual spaces, and we then develop our musical use of these instruments.

3.1. Guidelines of our work: modularity and focus on perception

In the bibliographical work presented below, two main aspects stand out: the modularity of instruments, due to the use of mapping layers, and the use of knowledge about perception. Our approach uses a modularity concept: with the three-layers mapping, we clearly separate the control from the instrument, in order to use any alternate controller (Wanderley *et al*. 1998) to drive synthesis models, which are not physical models but signal models.

The control mapping, which is the first layer from the gesture transducer to the related-to-gesture perceptual parameters, is used with several gesture transducers, according to the expressiveness we want to impart to the transducer. The sound of the instrument (sounding palette) or the timbre space of the instrument is determined by the synthesis model and the third mapping layer, from the related-to-sound-perception parameters to the synthesis parameters. The ways to navigate in this timbre space are given by the second mapping layer, from the related-to-gesture-perception parameters to the related-to-sound-perception parameters. In our approach, the modularity is not only a useful feature of the whole system, but a desired possibility of high-level modular instrument making.

Moreover, our mapping chain uses perceptual spaces. The use of a perceptual layer is really specific to our approach: the counterpart of the complexity of mapping layers is the intuitive navigation in perceptual spaces, and an easiness to play music with those kinds of instruments. The gesture intention is highly consistent with the parameters.

3.2. Experiments: examples of the use of three-layer perceptual mapping

3.2.1. Imitative synthesis

Several works (Grey 1975, Wessel 1979, Beauchamp 1982, Wessel *et al*. 1998) dealt with perceptual sound representation to play music and re-interpret the sound from a harmonic instrument. The musical excerpt is first analysed and then represented according to perceptual and signal features, keeping a description of links between the two kinds of features. Then, we can move into the perceptual space representing the sound and use gestures to synthesise the sound from the perceptual features. The basis of imitative synthesis is analysis by synthesis: when we consider the extracted features are representative enough to resynthesise a plausible imitation of the original sound, we use it to create other sounds.

The perceptual features we use here are psycho-acoustic parameters: pitch, loudness, timbre, using the brightness. Other models make use of a timbre control space with the attack-time logarithm and the brightness.

A more complex model being developed at LMA uses a greater perceptual space, including features of the timbre dealing with harmonics (jitter and shimmer on harmonics, harmonicity), with the expressiveness (roughness, vibrato, portamento), and with time variation of these features, thanks to the temporal derivatives.

The implementation is the following: first, the imitation system is set up thanks to an additive analysis. We extract additive synthesis parameters of the sound: partials' modulus and frequency. We then calculate the psycho-acoustic features of this sound. The sound used is a whole sentence played by a harmonic instrument, but we also want to extend this synthesis method to non-harmonic sounds. The link between additive analysis parameters and psycho-acoustic features is created both explicitly (the two data sets in a database) and implicitly (thanks to artificial neural networks), in order to test both approaches. Then, we can perform more or less realistic imitative synthesis, by synthesising the sound through additive synthesis, using the parameters taken from the perceptual features, themselves extracted from a physical or musical gesture as input.

When synthesising with artificial neural networks, we use the training values of the network. For the database technique, we smooth the data by methods such as K-nearest neighbours, hierarchical K-nearest neighbours, or weighted clusters.

The use of the perceptual space enables the musical sentence analysed to be performed, but also new ones to be explored, keeping properties of the analysed instrument. Depending upon the context, however, one may

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

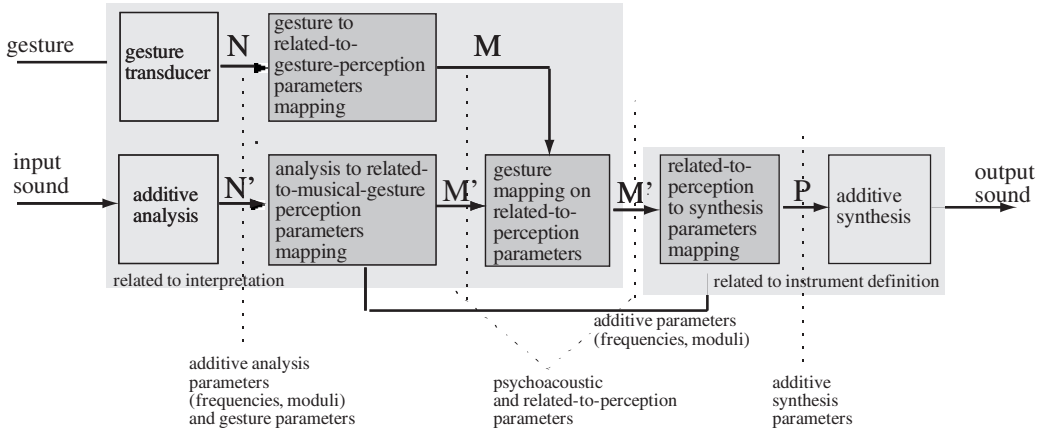


Figure 3. Imitative synthesis: additive analysis and synthesis, with re-interpretation of a sound using a gesture transducer. A mapping transforms gesture data into perceptive parameters; a second mapping computes psycho-acoustics and other perceptive features from the signal; a third modifies the perceptive features according to the gesture. The perceptive features are transformed into additive synthesis parameters using a multidimensional representation or artificial neural networks, first trained to link perceptive features to additive parameters. N, M, N', M' and P are the number of parameters.

want to play virtually with a real instrument; in that case, we want the best representation of the analysed performance. Mapping of the perceptual parameter modifications has to be very simple (linear, 1 to 1) when a gesture is linked to psycho-acoustic features. We may also want to warp, change, invent an instrument from another one: mapping between the gesture's data and psycho-acoustic features depends on the will and the creativity of the user. This last case is very interesting, with a host of the timbre explorations made possible.

It is logical to use a perceptual space for the synthesis of pre-analysed sound and for the representation and modification of timbre. Instead of giving the user the hard task of providing a mental representation of the process manipulated, it is the program itself that gives access to dimensions of human perception. The closer the gesture perceptual features to the way auditors analyse musical sounds, the easier the gesture manipulation of sounds. Mappings can be linear or not, using databases, multidimensional analysis, or artificial neural networks and evolving in time.

The first experiments we did, using database and neural network to synthesise flute sounds from the fundamental frequency, the centroid, the RMS and its derivative, sounds better than without the derivative. The attacks smoothing effect is less prominent and the sound better re-synthesised than in Wessel *et al.* (1998). We hope the whole model will help to create new expressivity tools.

3.2.2. The Voicer: a bi-manual control of moving formant synthesis with standard controllers

We have adapted a Wacom graphic tablet (Wacom 2002) equipped with a stylus transducer, and a game joystick (Saitek 2002) to use in a soloist expressive instrument. This is a bi-manual mapping experimentation with respect to leading hand/no leading hand motor behaviour relationship (Beaudouin-Lafon 1999). The choice of the two off-the-shelf controllers was influenced by low cost considerations, bi-manual possibilities, and focus on making an efficient mapping with off-the-shelf interfaces that could be used by everyone. This kind of interface has already been used in computer music (Wright, Wessel and Freed 1997, Serafin, Dudas, Wanderley and Rodet 1999). This instrument was implemented using the real-time musical programming environment Max/MSP (Zicarelli 1998). It has been technically described in another publication (Kessous 2002), so we shall describe it here from a mapping strategy point of view. The synthesis model simply consists of a rich harmonic source filtered by three second-order all-pole filters in cascade. This model can simulate a vowel singing voice. Here, we experiment with a mapping strategy that simultaneously allows melodic expressive control and spectral manipulation – the latter akin to navigation inside sound colour (Slawson 1985) or timbre spaces (Wessel 1979). For vowel singing voice, some people have used the vocal triangle representation with or without additional mapping (Rodet 1984, Wanderley, Viollet, Isart and Rodet 2000). According to Slawson (Slawson 1985), there are four perceptual attributes of sound colour: openness, acuteness, laxness and smallness. These dimensions can be seen as distinct directions along which we can perceive sound colour variations. Similarly, we have distributed the vowels according to the two principal dimensions in the plane of the parameter list interpolator, which is the *VTboule* Max object. The variations of sound colour are controlled via the joystick: x, y data assigned to the target position in the interpolator to which we could add several types of attraction functions.

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

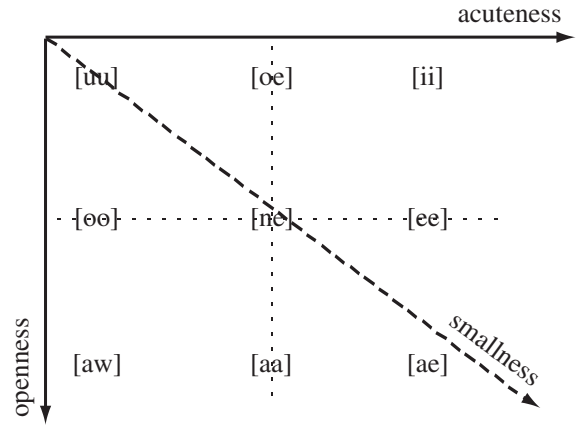


Figure 4. Vowel space according to the two dimensions, acuteness and openness (from Slawson 1985). Acuteness and openness are mapped to the *x* and *y* coordinates of the joystick. The dimension smallness is a linear combination of both axes' dimensions.

With conventional instruments (for example violin, saxophone), we often have several ways to move our hands (on the keys or strings) to cause a large change in pitch. The pitch can be changed from an octave to another by playing on another string or another key, or by modifying the pressure at the mouthpiece or by minor fingering change. To permit control within one octave and from one to the other, we divide the tablet's active space into twelve sectors, i.e. twelve equal angular parts where each part corresponds to a semitone of the chromatic scale. Turning clockwise changes pitch from low to high. We could go from a note to its lower or higher octave by pressing the stylus lateral button up or down. Fine tuning control, with gestures such as portamento or vibrato, is more powerful on limits of the control circle sectors. Categorising into sectors provides the users with stable control of the tone and forces them to make a conceptual effort to effect fine tuning variations. This principle of fundamental frequency control is conceptually inspired by the helical representation of chroma scale (Risset 1971, Shepard 1982).

Certain of the three mapping steps may not be always conspicuous or they could be considered simple one-to-one data scaling, but they are necessary if we decide to alter the instrument: changing a controller, modifying the expressive identity of the instrument, or the complexity of interaction or the synthesis model. Now we shall describe the mapping strategy step by step according to the three parts of the mapping chain defined in section 2.1.

To make up the instrument, we used *x* and *y* coordinates, pressure on the pen and pressure on the three-position lateral button of the stylus (released, up-pressed, down-pressed). The first mapping step for the tablet consist of defining in which of the twelve angular parts the pen is located, if it is more or less centred on it, and how many turns have been made around the centre of the tablet. It also defines the state of the buttons in a one-to-one mapping (released, down-pressed, up-pressed corresponding to states 0, –1, 1). If we have used another interface with three buttons that can be pressed at once and detected individually, we could define 2^3 different states in this first mapping. Pressure is passed onto a lookup table to enhance the strength and expressiveness of the gesture. Concerning the joystick we used *x* and *y* tilt. The first mapping step (data to related-to-gesture-perception parameters) simply consists of scaling the joystick position. It could have been different with another controller: for example with a Tactex controller we could have used a mapping 'based on interpreting the parameters of three fingers as a triangle' (Wright 2001). So the related-to-gesture parameters are now the following: region localisation L (integer 1 to 12), more/less centred on it C (continuous, float –1.0 to 1.0), turn number from start TN (integer –a to a), states S (integer –1, 0, 1), pressure P (continuous, float 0.0 to 1.0), X and Y scaled dimensions of tilt (continuous, float 0.0 to 1.0). The second mapping step will convert a combination of the related-to-gesture parameters from the tablet to pitch defined as values corresponding to MIDI note number extended to a continuum of pitch: pitch = pitch0 + L + (TN + S)*12 where pitch0 is an initial defined pitch (MIDI note number). Pressure is assigned to loudness. X and Y tilts of the joystick are assigned to the acuteness and openness perceptual attributes of sound colour (i.e. scaled to the 2D position in the interpolator plane). The last mapping step is to assign loudness to level, pitch to fundamental frequency of a sawtooth signal, and the two used vowel sound colour attributes (i.e. position in the interpolator) to all-pole filters coefficients and gain correction.

The control of information such as pitch (note, vibrato, glissando) and loudness (level and tremolo), is assigned to the dominant hand (left hand for a left-hander); the other hand is used to control spectral characteristics. Of course we know by observing conventional instruments and referring to Hunt *et al*. (2000) and Baudouin-Lafon (1999) that 'cross-coupling' (M to N complex mapping including two-handed co-operative action) plays an important role in instrument efficiency, and we will consider it to improve the instrument.

3.2.3. Meta-control of Scanned Synthesis

Scanned Synthesis has been recently developed by Verplank, Mathews and Shaw (2000). This is a new synthesis technique that can generate sounds from slow movements of mechanical systems. It corresponds to the creation of sound from a dynamic wave table. The synthesis model is described in figure 6.

Verplank, Mathews and Shaw have mainly used a circular string model with finite differences: a collection of masses, springs and dampers defines an object in motion.

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

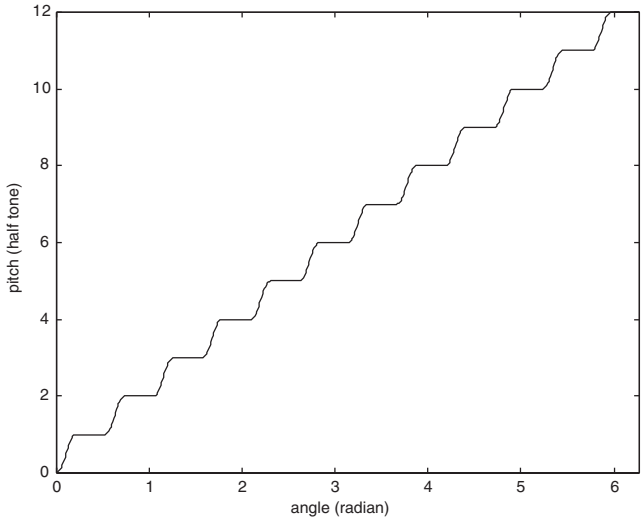


Figure 5. Chroma scale as a helical function: according to the angle (x: radian), the pitch (y) goes up or down by semi-tone, crossing a fixed frequency part and a gliding frequency part (in which vibrato and portamento are allowed). This representation in 2D corresponds to the helical 3D representation, using the angle cosine (x), the angle sine (y), and the pitch (z).

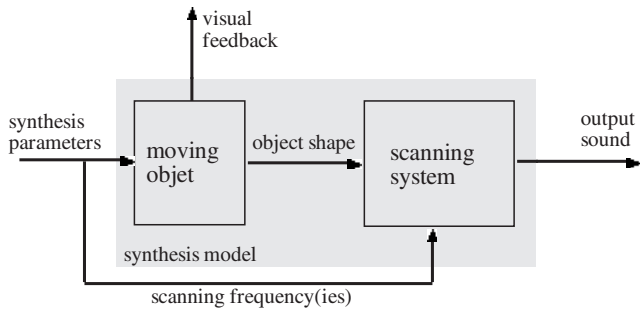


Figure 6. The Scanned Synthesis model consists of a slow dynamic system with vibration frequencies below 15 Hz (inaudible). To produce audible frequencies, the shape of the dynamic system is periodically scanned at a frequency corresponding to the sound fundamental. This model allows us to independently control the timbre and the sound fundamental frequency.

All the parameters of the system (springs, masses, damping) can vary along the string. The model is manipulated by shifting or striking the various masses or applying forces to them, and by controlling parameters. We have studied Scanned Synthesis using this string model: instead of trying to develop other Scanned Synthesis algorithms, our strategy was to focus on gesture control in the design of one complete musical instrument.

Scanned Synthesis gives the performer two points of view on what is manipulated: a mechanical one and an acoustical one. It provides two perceptual feedbacks: the sound produced by the algorithm and the visualisation of the string. An interesting feature of this synthesis model is that one controls a slow dynamic system. That way, there is a strong link between the shape of the string and the sound, because most of the synthesis parameters are close to psycho-acoustic parameters (damping is linked to the sound release time, spring stiffness to the roughness, etc.). As seen in section 2.2, physical parameters represent a physical reality, so they are more intuitive. Moreover, the string movements used in Scanned Synthesis have the same rate as human motions, so there is also a strong link between the physical object and the gesture.

Now the problem is to determine how to best controls the system. Our system has a large number of parameters due to the discretisation of the string (each element of the discrete string has at least seven control parameters, but all the elements give the same kind of parameters). We are in a situation where we have to control many parameters, but these can be put in different classes and manipulated together. As a consequence, we have focused our work on the control of the physical parameters via meta-parameters (see section 2.2). The solution we explored was to create profiles for the parameter distribution along the string. These profiles are manipulated by meta-parameters. As indicated in figure 7, we can consider that manipulating meta-parameters with one fixed profile corresponds to a definite mapping, but changing profile is like changing mapping.

To reinforce the relevance of meta-parameters, we

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

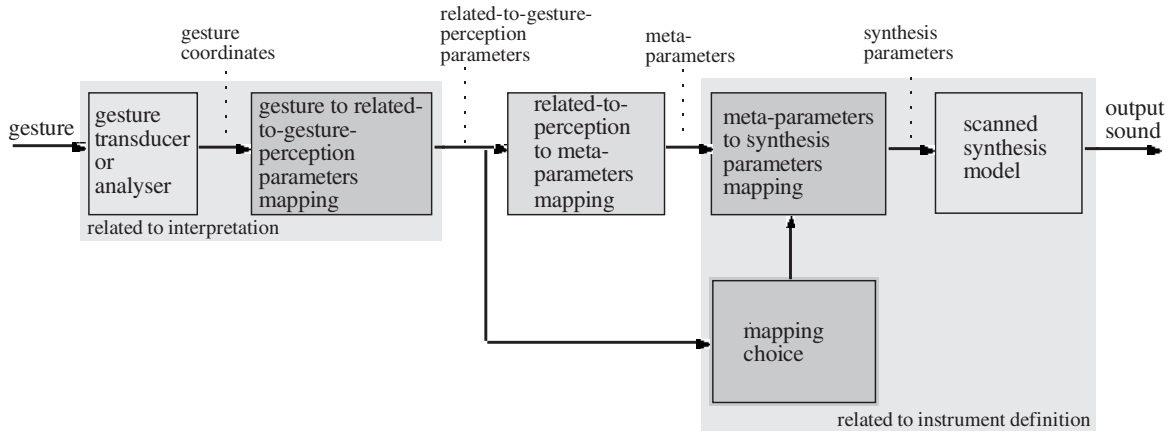


Figure 7. Meta-parameters are high-level control of synthesis model parameters. We choose them according to their relevance in gesture control and the effect they can produce on sound. With one gesture it is also possible to select one mapping among several pre-defined ones.

have applied some transformations on the model parameters. One is the definition of the variation range of synthesis model parameters. We have defined these limits taking into account constraints imposed by the model (stability constraints) and aesthetical criteria (to define the instrument playing space). In our implementation, some meta-parameters manipulate the logarithm of certain physical parameters instead of the parameters themselves: that way, their manipulation is more relevant.

Our mapping strategy is explicit: mathematical formulas link meta-parameters and lists of values containing profiles to model parameters. Relationships between synthesis model parameters and meta-parameters are not simple; the mapping is complex.

Our implementation of Scanned Synthesis proposes a virtual instrument with several pre-defined configurations offering different tones, and composers can establish their own configurations if they wish to. In analogy with an acoustic instrument, we have an instrument on which we can play normally but with extended possibilities, such as modifying intrinsic characteristics of the instrument (dimensions, thickness of strings) during the performance. These possibilities give a greater tone control than when the instrument characteristics are fixed. Therefore, performers can change the shape of their instruments (for example from the guitar shape to the violin shape), and create the shape they want. Other aspects are the possibilities to obtain sounds with instantaneous gestures (controlling the sound onset) or continuous gestures (exciting the sound in a continuous way as in a violin), and to use visual feedback of the string to help the performer.

To use Scanned Synthesis in real time with gesture control, we have created a C object, *scansynth*, for the Max/MSP software on the Macintosh. The inputs proposed by this object are a set of meta-parameters that can be directly connected to gesture devices, and a system that manages distribution profiles of the synthesis parameters on the string (stiffness of springs, damping, forces, initial position, etc.). This system proposes predefined profiles, and expert users can also define customised profiles and save them. Sound can be triggered by releasing the string from an initial position, or it can appear and be sustained by forces applied on the string. The object also proposes an inlet for the control of the sound fundamental frequency, and three outlets, the first one for the sound, the second one to display profiles and meta-parameters, the last one to display the string position.

For each gesture sensor, a Max/MSP patch was created, offering outlets that can be directly connected to the Scanned Synthesis object. Those outlets contain normalised data from the sensor, and data calculated from the previous ones. It allows the use of data from specific gestures (for example, polar coordinates to take into account circular motion onto the string). Thanks to this modular organisation into separate objects, the user can easily choose gesture devices and connect them to the synthesis model. We have created a complete Scanned Synthesis instrument for a demonstration and for a musical performance (Couturier 2002). The meta-parameters were linked to gestures in a musical context, which was the best way to evaluate the instrument.

We will now present an example of gesture control with the A4 Wacom Tablet. This device, composed of one tablet and one pencil, provides a lot of information: the position of the pencil lead on the tablet, the pressure on the lead, the pencil tilt angles, and various buttons. We assign the control of force amplitude to the lead pressure, which is the most relevant mapping since we have to apply force on the lead. We use a rotational motion to control the offset of the force profile (the angle between the pencil and a line on the tablet plan). This gesture is the same as the one we can produce with a joystick, but here the pencil is not fixed (so we can use

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

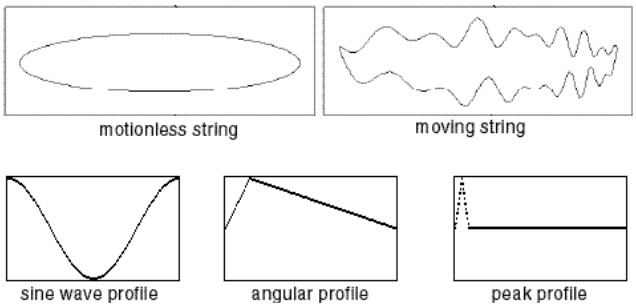


Figure 8. Some examples of 2D and 3D visual feedback of the Scanned Synthesis circular string model and of profiles of parameter distribution.

the leading coordinates to manipulate other parameters). This kind of mapping makes it possible to control four continuous parameters with one single hand.

In the Scanned Synthesis implementation we developed, we used a type of related-to-perception parameters called meta-parameters, which controls physical parameters through the manipulation of profiles. Future work will explore further sound possibilities of the model, and attempt to find psycho-acoustical parameters to control the model, instead of meta-parameters.

3.2.4. Real-time control of Paradoxical Sounds

Paradoxical Sounds were demonstrated in computer music by Shepard and Risset (Risset 1989) in the 1960s. They are based on a particularity in the human perception of pitch, referred to as pitch chroma in psychoacoustics (Shepard 1982). One of these sounds consists of ten sinusoidal components separated by one octave with amplitudes controlled by a constant spectral envelope, a bell-shaped curve. When we simultaneously and continuously move the frequencies of the ten sinusoids to one octave above, we return to the same spectrum. In fact, the highest frequency component has disappeared and a new low component has been added. We have the same effect when going downward (figure 9). This has yielded the endless glissandi used by Risset in his pieces.

We have implemented this example in the Max/MSP software in order to control it by gesture. A simple mapping was done to control the sinusoid's frequency. The glissandi can be obtained with a circular gesture, in a mapping where the frequencies are linked to the polar angle of the pencil on a graphical tablet. When the angle goes continuously from 0 to 2π, each component frequency goes continuously to the lower or higher octave, and the final sound is identical to the initial one.

The glissando velocity corresponds to the gesture velocity. When the gesture is slow and regular, we hear a paradox, but when the gesture is faster or irregular, the variation of pitch becomes the prominent perceptual cue. A slow movement is required to produce the sonic paradox.

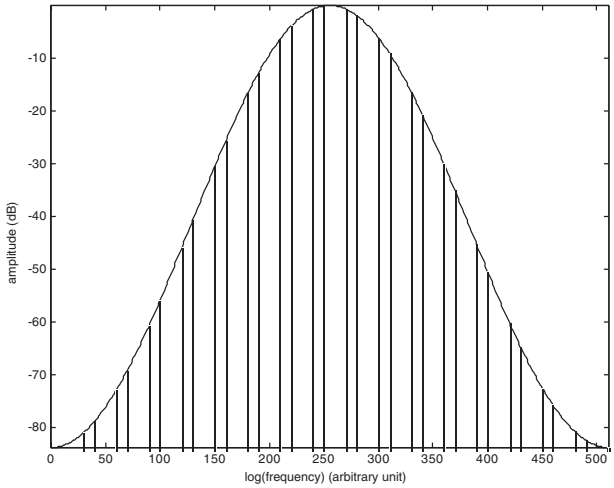


Figure 9. An example of Risset's paradoxical sounds. Ten sinusoidal components separated by one octave can slide under a bell-shaped envelope with a log(frequency) unit for the *x*-axis and magnitude (dB) for the *y*-axis. We can connect the rate of the frequency glide to the speed of a gesture.

3.2.5. Nonlinear distortion control

Here is an implementation instance of a source-filter synthesis system, two parameters of which are mapped to gesture data. The source consists of an amplitude-controlled oscillator connected to a nonlinear distortion module (Arfib 1979). The control parameter of the filter is its resonance frequency.

When the source amplitude increases, the harmonic amplitudes increase, and this can represent the sound power. The filter parameter plays on the harmonic tuning of the sound by imposing a formant structure on it. We can connect any gesture sensor to those parameters, but the choice of the device must also relate to the musical idea that is behind the sound. For example, on a specific driving game controller, we can connect the source amplitude to the accelerator pedal and the resonance frequency to the driving wheel. Any other device is usable, and so we can explore the impact of different gestures on the expressiveness of the sound and the virtuosity of the performer.

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

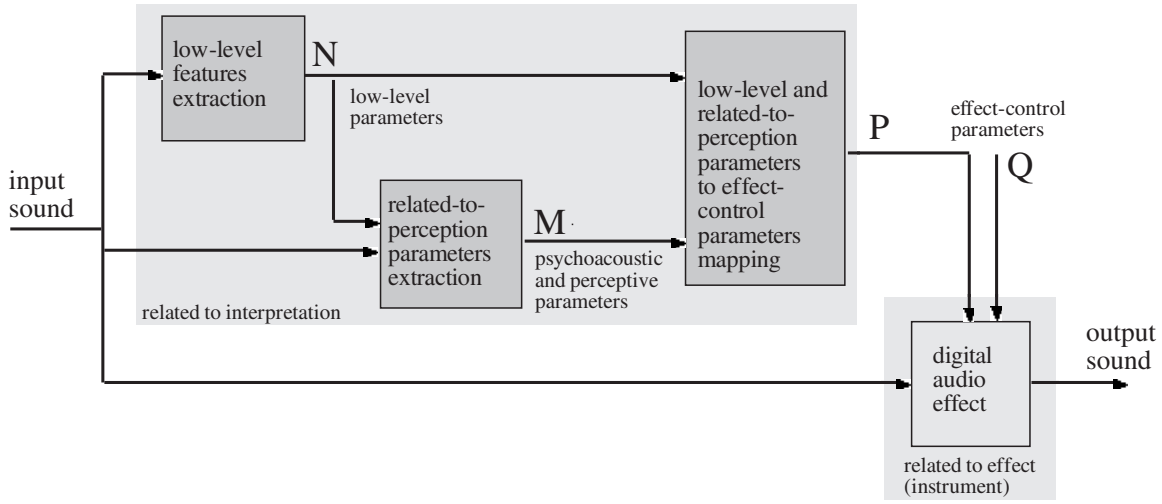


Figure 10. Adaptive digital audio effect. Low-level and high-level features are extracted from the input sound; a mapping is done between these features and the effect control parameters. The effect is then applied on the input sound according to its control values. N, M, P and Q are the number of parameters.

3.2.6. Adaptive digital audio effects

Adaptive digital audio effects are effects for which control parameters are driven by features extracted from the sound (Verfaille and Arfib 2001). The user decides what features control what control parameters, and according to what mapping laws. Gestures can be added to control the mapping itself, giving a second, more general, control level to the effect.

Figure 10 shows mappings used in the A-DAFx. The first one is about the low-level feature extraction. The second one is about high-level and perceptual feature extraction from the input sound and from the low-level features. Both are explicit, algorithmically defined mappings; one cannot really change them in any significant way. The third mapping transforms the low-level and the perceptual parameters into effect-control parameters. This is typically the place where the performer should experiment. An effect driven via a perceptual feature of the sound is more consistent than an effect applied with any signal parameter. We used linear combination to link perceptual features, and we applied a nonlinear function to the linear combination to obtain the effect-control parameters.

Low-level features such as the RMS energy, the spectrum centroid, the fundamental frequency, and the voiced/unvoiced status were extracted. Psycho-acoustic features such as pitch and loudness were also extracted. Nonlinear rules such as truncation, sinusoidal, exponential or logarithmic transformation of the effect-control curve were applied. Effects obtained, such as adaptive robotisation or the selective time stretching, implemented with a phase vocoder, are very consistent with the input sound, and they allow new musical gestures in the effect itself. This is not surprising since a musical gesture extracted from the signal is used as an input of the effect-controls.

Gesture data drive effect-control parameters or can be used to change the mapping between extracted features – low-level and perceptual features – and the effect-control parameters. This is presented in figure 11.

A mapping layer is added to figure 10, from the gesture transducer parameters to the effect-control parameters and also to the mapping between low-level and perceptual features to effect-control parameters. This gesture mapping is used only for real-time effects, such as adaptive granular delay or adaptive robotisation, but not for non real-time effects such as selective time stretching. Any kind of mapping, explicit or implicit, can be used; the user decides by listening to the effect which mapping is the best adapted (we only experimented with explicit mappings). This procedure allows a higher-level control on the effect and its performance counterpart.

Adaptive digital audio effects serve to re-interpret musical gestures in sounds at different levels: at the sonic micro-level (with a timbral effect, for example), or at a macro-level, with the progressive modification of inference control rules of the effect. This macro-level can be accessed through gesture, allowing a global musical gesture and re-interpretation of the sound. The two transformation levels of a sound provide for finer use of digital audio effects, and help to produce effects that do not overload the sound.

3.3. The evaluation of these instruments

Defining sound algorithms, mappings and peripherals does not insure that music can be played on these virtual instruments. One must emphasise the fact that the true

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

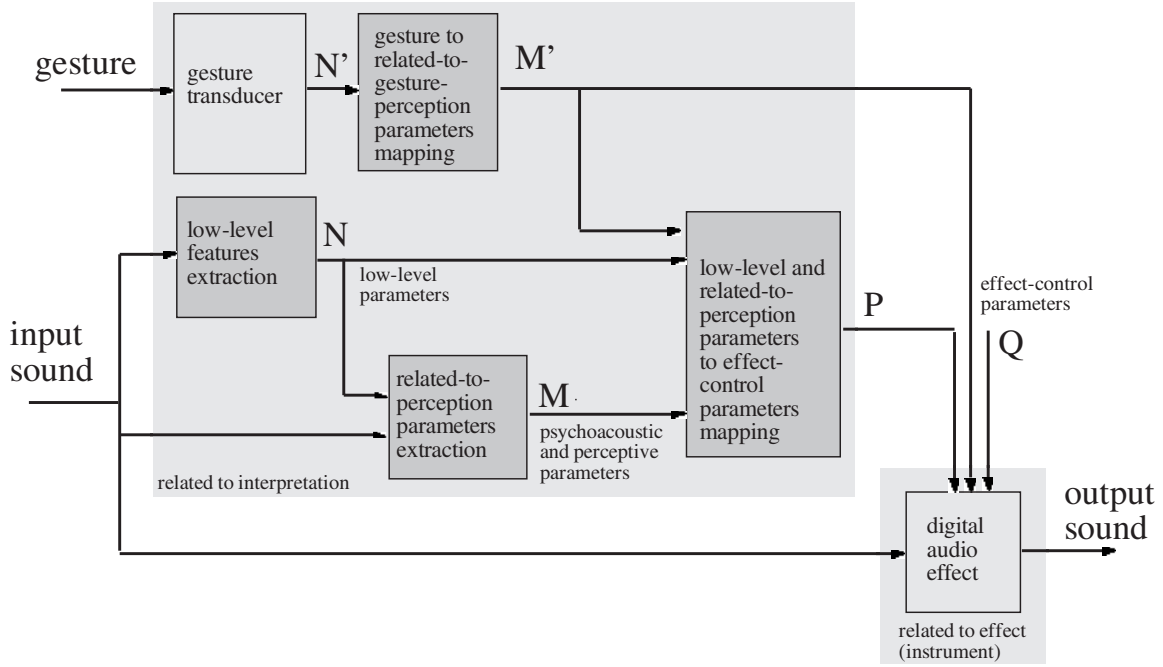


Figure 11. Adaptive digital audio effect with gesture interpretation, by a gesture transducer. A mapping stage is added to figure 9, from the gesture transducer data to the effect-control parameters and also to the mapping between low-level and perceptive features to effect-control parameters. This addition allows a higher-level control on the effect and its performance counterpart. N, M, P and Q are the number of parameters.

test is music, and this fact is clearly shown in the difficulty of relating papers and musical works in conferences dedicated to interfaces. One can of course evaluate instruments through physical measurements, or psychological test appreciation. We have bypassed this, not because we deny it, but because it is in a way outside our present goal. Our goal is clearly to design instruments that can be played in a small ensemble with acoustic instruments, be they electrified or not. Also the musical use of these instruments immediately reflects the pluses or minuses of the ergonomy, and of course the mapping emerges as the key point where one can act to improve things or focus the attention.

The 'voicer' (section 3.2.2) has been played either solo, which can give a direct demonstration of its possibilities – in which case it is good to have a multiple oscillator as a source – or in conjunction with saxophone and percussion or flute, guitar and percussion. The vocal characteristic of this instrument makes it easy to lead, and its timbre emerges as a recognised pattern for everyone. The easiness of vibrato makes it particularly useful for extra-European style, and the mapping of the angle of virtual circle on a graphic tablet has proven very efficient to reproduce voice inflexions, providing that the pitch-to-angle function is nonlinear. Depending upon this function, the gesture changes and of course the playing of the instrument. The mapping of vowels has also been subject to evaluation: like in a geographic map, you may want to have the north up, but where is the north in a vowel map? Such questions arise continuously, and the best answer is the music that is played.

From the implementation of Scanned Synthesis, a bimanual musical instrument has been designed (section 3.2.3). This instrument provides a 'traditional' control on pitch and envelope and a control on timbre: in addition to the mapping of pitch to the angle, the mapping of the movement and forces on the string with movements on a tactile graphic pad has to be learned and experimented with. As regards music, this instrument has been played in a situation of a harmonic dialogue with another acoustical instrument, and we have also explored Scanned Synthesis timbre in an electroacoustic context. In the first case, the specific timbre of Scanned Synthesis patch we used enhances the traditional instrument timbre, and a duo with a banjo reveals possibilities of imitations, dialogues and harmonisation which are both traditional and very intriguing. In the second case of a piece in a dialogue with electroacoustic samples, Scanned Synthesis provides a great facility for providing textures, giving a powerful counterpart to these sounds. While in the case of playing with a traditional instrument it is quite easy to separate the two sources, here we go into situations where sometimes fusion can exist and sometimes a clear distinction arises, which of course is a real challenge that helps making good and fresh music.

The term 'musical use' is an understatement for the research issued from the attempt to control by gestures *Le Souffle du Doux*, of which one example is given in section 3.2.5: the origin was musical and the realisation came after. *Le Souffle du Doux* is a musical work that was realised in 1979 with the Music V synthesis program. The idea to make a 'live version' of it using the

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

Max/MSP real-time implementation gave rise to two main questions: which peripherals should we use, and what kind of mapping should be done. As a matter of fact, each sound sequence had to pass through this questioning, and the answers were very different for each. As an example, for the 'nonlinear instrument' we chose to use a Mathews' radio-drum (Boie, Mathews and Schloss 1989), now called radio-baton, as the device, and to look for natural movements that could symbolise spectral navigation not as a consequence but as a primal fact. So we decided to map the distortion index value on the vertical position of one stick and the beating frequency on the horizontal value of the second one. We consider we have a good mapping when we 'feel' the sound with easy movements. The performance of this entire piece gave us the opportunity to see some implications of the mapping upon performance possibilities.

4. THE IMPLICATIONS OF MAPPINGS

We now discuss the musical implications of using mapping strategies from gesture to sound.

4.1. The role of mapping limits

One of the roles of a mapping is to define limits, which are also boundaries within which instrumental playing or composition can exert itself. This is quite clear with the voicer instrument (see section 3.2), for which a bi-dimensional timbre space is defined, associated with the joystick device coordinates. Each point of the (x, y) space corresponds to one and only one configuration of three filters, so the domain is totally defined by the performance space.

Nevertheless the static limits are not the only ones the mapping immediately brings, but also the limits on the motions this mapping allows: some gestures are made physically feasible or not by a specific mapping. So we can say that a static mapping also brings a dynamic use of it. Dynamism comes then from the gesture and not from the mapping.

An interesting case is the way to control a vibrato on an electronic instrument. Usually one has the choice between two mappings: either we define it by its rate and its width, and we connect two sliders to these values, or we map only one slider on the fundamental frequency of the sound and it is the gesture – as with the Theremin – that makes up the vibrato. It is possible to propose innovative mappings such as the one in voicer, where the tablet's mapping towards frequency is a mixture between fixed parts of a circle and gliding ones. In each of these three mappings, we get a different gesture, hence a different virtuosity and also different limits. In the case where the vibrato rate comes from a gesture, the bounds are given by the gesture, so there is a muscular feedback; conversely, a slider can give any rate, but the haptic feedback is nearly non-existent. The expressiveness of a musical electronic instrument really depends upon the mappings that are used.

4.2. Freedom, constraints and boundaries: what pedagogy?

The musical use of electronic systems that include mapping strategies relies on the definition of a 'usual field', which is the central part of the performance domain of this instrument, and the 'boundaries of the field', which define the limits of its musical use. This distinction is quite explicit when it comes from a naturally built instrument: there is a common ground to a given instrument as well as deviations and explorations.

The common ground can be defined as what we discover at first glance or first gesture. It is very important that it exists, it is the first contact with the instrument. As an example, if the mapping is totally randomised, instrumental creativity can become lower because any gesture will give rise to the same kind of sounds. Extreme mappings that try to mimic and possibly overcome the difficulty one has with a regular instrument (as an example, clarinet sounding is not easy) may prevent free space from which the future virtuoso can start.

So the central domain imposed by a mapping can be called 'normal playing'. Newcomers are very cautious when exploring an instrument: the first gestures allow them to 'get an idea', to make a mental map – in the same way that an explorer would in a Chinese city without any knowledge of Chinese. This learning phase is very important because it affects the future ones. The central space has a good probability of becoming the 'rest place', the centroid where the performer will return from time to time. In the learning process, it is essential to find cues that one can remember. Should they be sonic, visual or haptic, they serve to come back to a known state. The consistency of the mental map we associate with a specific algorithmic mapping is crucial; it does not necessarily have the same logic, but the representation must have a direct cognitive link with the gesture representation. This can be applied for continuous fields as well as for discontinuous ones: the act of selection is also a part of the mental map, similarly to the new mental map that comes from another man–computer interface, so it is a multi-layer one.

The domain defined by all the possibilities allowed by using a specific mapping has one central base, where one comes in a 'natural way'. But it also has boundaries, which can be explored by choosing extreme values of the map parameters. If the mapping does not include perceptual 'valleys and hills' – in the sense of contrasted values within the field – exploring the border leads to the musical possibilities of the static mapping. But once again this does not mean we thus explore the limits of the musical playing. Many gestures are possible inside the medium range of parameters. We come back here to

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

a separation that has reigned for centuries: the performer can make unusual sounds come out of a usual instrument. So a distinction must be made between boundaries provided by the static mapping and boundaries implied by the gesture using this mapping. The previous example of the vibrato helps make clear these two kinds of boundaries.

Haptic feedback is essential to sense the boundaries of a given instrument playing. This feedback can come from the instrument itself, for instance the force feedback of an electronic device, or the percussive plane for a percussion instrument. It can also be linked solely to the body of the performer. The ensemble of the muscles and nerves are part of the instrument as far as the boundary problem is concerned. There is no such thing as a free gesture: the possible variations of a free hand are directly linked to the rest of the body and the earth. In a way the art of dance and the science of motion can bring new insights in the pedagogy of new gesture systems, not only for the functionality of the body but also for the 'art of moving'. Recall that Augustinus defined music as 'ars bene movandi', the art of good motion.

4.3. The pleasure to learn

The pedagogy that stems from the use of mappings in electronic instruments may be different from that of traditional instruments. Mappings include or at least suggest a range of materials, a range that calls for learning how to arrange these materials, but also defines colours of sound, in which case we are closer to language and discourse.

This distinction is not new – we can find it in the typology of gestures as well as in the sorting between composers and performers – but here we have the prospect of a truly experimental pedagogic research on the way the choice of a mapping conditions a learning process. Up to now the connection methodology was empirical: with our modular strategy, we can 'plug' several devices, that is several gesture situations, to the same synthesis patch, and carry out musical experiments on what fits best. A study is planned on how well mappings fit some musical applications.

4.4. Outside the limits – breaking the boundaries

It is paradoxical that some kinds of music may rely more on the breaking of the rules than on the rules themselves. Some composers are literally looking for the possible deviations in a computer music program. When planning a mapping, one must take this into account: an instrument is interesting only if it allows the discovery of unknown features. This relates to 'encounter entropy': there are some codes, but the prolongation of an encounter needs a hidden garden, otherwise the first encounter has been the whole story.

So what is a new field of experience, in contrast to a limited field? Should we set up 'exotic mappings' to get exotic sounds? This is not a condition: in a limited field one can also explore gestures that make new strategies emerge through the mapping itself. In fact, one of the keys would be to define the exploration made possible by the mapping. Creativity comes from the fact that we are not bored with exploring. What we find needs at some point to be cartographic, and nothing unknown will remain as such forever. But if the instrument does not allow, or even does not suggest any way to explore new paths, it is a limited one. It may be a good one, classical, sensitive, but not very innovative.

4.5. Virtuosity

When it comes to the professional use of an instrument that includes mappings, the issue is no longer learning, but virtuosity. Some gesture devices and some associated mappings constrain the gesture in such a way that virtuosity is impossible because no expert gesture can occur. However, one must not identify the virtuosity of a gesture with its appearance. As Jean Haury demonstrated on stage, it is possible to have a mapping with only two keys pressed by fingers in order to control a virtuoso interpretation of a concerto. This brings forth an entire field of research: now that we know this possibility, to what extent do we want to introduce a 'facility' or 'laziness' in the mapping strategy in order to keep some kind of virtuosity with this mapping? This may be more than a technical matter: educators keep reminding us that motivation depends also upon the obstacle.

4.6. Expressiveness and improvisation

So far we have spoken about virtuosity as the capacity of executing an action that seems nearly impossible for human beings. Another issue is improvisation and expressiveness, which often relate to virtuosity but do not totally depend upon it. Indeed, a good virtuoso is often at ease and so has some availability for expressiveness and improvisation, but one can compensate a lack of virtuosity by a sensitivity that can be enhanced by a nice instrument.

The difference between expressiveness and improvisation is very narrow: when do we consider that we still have an expressive performance and when does it become improvisation? The choice of a mapping can constrain the possibility of improvisation. For example, a mapping with only selection gestures limits improvisation to the possible choices. This is the case for keyboards, marimbas and xylophones, but these instruments can also be expressive. A mapping that includes modulation gestures allows the exploration of paths in 2D or 3D physical spaces. Following a path suggested by a composer or by the system, we are in the frame of interpretation and expressiveness. If we go into a creation of a new path, we improvise. There can be improvisations

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

based on themes and free improvisations. Mappings involving the use of a sequencer do not allow improvisations on notes but only rhythmic variations. Whether a mapping includes the capacity to work on the sound itself or not conditions the possibility of improvisation either on sound events or on timbre.

To conclude, with these thoughts on the musical implications of mapping in gesture controlled systems, one can widen the focus and see what is happening with other sensory modalities such as vision or haptics. Little research has dealt with multi-modality, and the use of mappings may lead to this. The strength of metaphors is especially true for sound: some strong cues can at once give a reference for a meaning. This may be also true for gesture devices: a gesture intention is also the expression of a metaphor, and this metaphor is perceived in the sound we hear. Mappings can then be seen as ways to link senses in order to converge toward the same metaphor.

5. CONCLUSION

In this article we have tried to describe how the emergence of theoretical ideas, such as the use of perceptual spaces and the definition of a modular structure for the mapping from gesture to sound, can help define virtual instruments that are sensitive and efficient. Like every emergent technology, mapping strategies go back and forth from theory to experimentation: experiments bring material that can be theorised, such as the need for strong perceptual cues, the interchangeability of devices, or the importance of developing a pedagogy of gesture. Conversely, theory brings ideas that can be applied in many situations: the static/dynamic or implicit/explicit mapping functions are the basis of every mapping strategy, and experiments reflect a few fundamental ideas in a variety of implementations.

The perspectives we see, along with a deeper theorisation and its application to new instrumental situations, lead to the confrontation with the real world of music: it is important to see how other musicians act and react to new instruments they have not devised, but which they use as composers and performers. Instruments, old or new, have to be evaluated on a musical and aesthetical basis, and not only on the basis of technological features.

ACKNOWLEDGEMENTS

We would like to thank the 'Conseil Général des Bouches du Rhône' for the financial support of our project 'Creative Gesture in Computer Music', as well as the CNRS (Centre National de la Recherche Scientifique), where it was developed.

REFERENCES

Arfib, D. 1979. Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves. *Journal of the Audio Engineering Society* **27**: 10.

Arfib, D., and Delprat, N. 1998. Selective transformations of sound using time-frequency representations: an application to the vibrato modification. In *104th Convention of the Audio Engineering Society*, Amsterdam.

Arfib, D., and Kessous, L. 2000. From Music V to creative gesture in computer music. In *Proc. of the VIIth SBC Conf.*, Curitiba, available in CD format or at: http://www.niee.ufrgs.br/SBC2000/eventos/sbc&m/sbc&m3.pdf

Arfib, D. 2001. L'Espace perceptif dans la relation du son au geste. *Journées d'Informatique Musicale*, Bourges.

Arfib, D., and Kessous, L. 2002. Gestural control of sound synthesis and processing algorithms. *Gesture Control Workshop*. Springler-Verlag.

Baudouin-Lafon M. 1999. Contrôle gestuel de la synthèse sonore. In H. Vinet and F. Delalande (eds.) *Interfaces homme – machine et création musicale*, pp. 145–63. Paris: Hermès Science Publishing.

Beauchamp, J. W. 1982. Synthesis by spectral amplitude and brightness matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society* **30**(6): 396–406.

Beauchamp, J., and Horner, A. 1992. Extended nonlinear waveshaping analysis/synthesis technique. In *Proc. of the Int. Computer Music Conf. (ICMC'92)*. San Francisco.

Boie, B., Mathews, M., and Schloss, A. 1989. The Radio Drum as a synthesizer controller. In T. Wells and D. Butler (eds.) *Proc. of the Int. Computer Music Conf. (ICMC 1989)*, pp. 42–5. San Francisco: International Computer Music Association.

Cadoz, C., and Wanderley, M. 2000. Gesture – Music. In M. Wanderley and M. Battier (eds.) *CD-rom Trends in Gestural Control of Music*. Publication Ircam.

Couturier, J. M. 2002. A Scanned Synthesis virtual instrument. In *Proc. of the 2002 Conf. on New Instruments for Musical Expression (NIME-02)*, Dublin, Ireland.

Daniel, P., and Weber, R. 1997. Psychoacoustical roughness: implementation of an optimized model. Acustica – *Acta Acustica* **83**: 113–23.

Desain, P., and Honing, H. 1996. Modeling continuous aspects of music performance: Vibrato and portamento. In *Proc. of the 4th Int. Music Perception and Cognition Conf.*, Montreal.

Desainte-Catherine, M., and Marchand, S. 1999. Structured additive synthesis: towards a model of sound timbre and electroacoustic music forms. In *Proc. of the Int. Computer Music Conf. (ICMC'99)*. Beijing.

Dubnov, S., and Rodet, X. 1997. Statistical modelling of sound aperiodicities. *Proc. of the Int. Computer Music Conf. (ICMC'97)*, pp. 43–50, Tessaloniki.

Fischman, R. 1999. A survey of classical synthesis techniques in Csound. In R. Boulanger (ed.) *The Csound Book*. MIT Press.

Grey, J. M. 1975. *An Exploration of Musical Timbre*. Ph.D. dissertation, Dept. of Psychology, Stanford University. CCRMA Report STAN-M-2.

Hélie, T., Vergez, T., Levine, J., and Rodet, X. 1999. Inversion of a physical model of a trumpet. In *Proc. of the Int. Computer Music Conf. (ICMC'99)*. Beijing.

Horner, A. 1995. Envelope matching with genetic algorithms. *Journal of New Music Research* **24**(4): 318–41.

Hunt, A., Wanderley, M., and Kirk, R. 2000. Towards a model for instrumental mapping in expert musical interaction.

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press

Proc. of the Int. Computer Music Conf. (ICMC'2000), pp. 209–12. ICMA: Berlin, Germany.

Jehan, T., and Schoner, B. 2001. An audio-driven, spectral analysis-based, perceptual synthesis engine. *Audio Engineering Society Convention.*

Jensen, K. 1999. Envelope model of isolated musical sounds. In *Proc. of the DAFx99 Conf.*, Trondheim.

Jensen, K. 2000. *The Timbre Model of Musical Sounds.* Ph.D. thesis, Datalogisk Insitut, Kobenhavns Universitet, Denmark.

Kessous, L. 2002. Bi-manual mapping experimentation, with angular frequency control and sound color navigation. In *Proc. of the 2002 Conf. on New Instruments for Musical Expression (NIME-02)*, Dublin, Ireland.

Krimphoff, J. 1994. *Analyse acoustique et perception du timbre*. D.E.A thesis, Université du Mans, France.

Leman, M. 2000. Visualization and calculation of the acoustical musical signals using the Synchronization Index Model (SIM). In *Proc. of the DAFx00 Conference*, Verona.

MacAdams, S., and Bigand, E. 1994. Penser les sons. *Psychologie cognitive de l'audition*. Paris: P.U.F.

MacAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. 1995. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research* **58**: 177–92.

Martin, K. D., and Kim, Y. E. 1998. 2pMU9: Musical instrument identification: A pattern-recognition approach. *136th meeting of the ASA.*

Métois, E. 1998. Musical gestures and audio effects processing. In *Proc. of the DAFx98 Conf.*

Moore, B. C. J., and Glasberg, B. R. 1996. A revision of Zwicker's loudness model. *Acustica – Acta Acustica* **82**: 335–45.

Moorer, J. 1979. The use of linear prediction of speech in computer music applications. *Journal of the AES* **27**(3): 134–40.

Pressnitzer, D. 1999. *Perception de la rugosité psychoacoustique: d'un attribut élémentaire de l'audition à l'écoute musicale*. Ph.D. thesis, Université Paris VI.

Risset, J.-C. 1971. Paradoxe de hauteur sonore: le complexe de hauteur sonore n'est pas le même pour tout le monde. *Proc. of the 7th Int. Congress of Acoustics, Budapest*, pp. 613–16.

Risset, J.-C. 1989. Paradoxical sounds. In M. Mathews and J. Pierce (eds.) *Current Directions in Computer Music Research*, pp. 149–58. Cambridge, MA: MIT Press.

Risset, J.-C., and Wessel, D. L. 1999. Exploration of timbre by analysis and synthesis. *The Psychology of Music*, second edition, pp. 113–69. Academic Press.

Rossignol, S., Depalle, P., Soumagne, J., Rodet, X., and Collette, J.-L. 1998. Vibrato: detection, estimation, extraction, modification. In *Proc. of the DAFx98 Conf.*

Scheirer, E., and Slaney, M. 1997. Construction and evaluation of a robust multifeature speech/music discriminator. In *ICASSP Proc.*

Serafin, S., Dudas, R., Wanderley, M., and Rodet, X. 1999. Gestural control of a physical model of a bowed string instrument. In *Proc. of the Int. Computer Music Conf. (ICMC99).* Beijing.

Shepard, R. N. 1982. Structural representation of musical pitch. *The Psychology of Music*, pp. 343–90. New York: Academic Press.

Slawson, W. 1985. *Sound Color*. Berkeley: University of California Press.

Tactex. 2002. *http://www.tactex.com/*

Terhardt, E. 1979. Calculating virtual pitch. *Hearing Research* **1**: 155–82.

Thrustmaster. 2002. http://www.thrustmaster.com/

Verfaille, V., Arfib, D. 2001. A-DAFx: adaptive digital audio effects. In *Proc. of the DAFx01 Conf.*, Limerick.

Verplank, B., Mathews, M., and Shaw, R. 2000. Scanned Synthesis. *Proc. of the 2000 Int. Computer Music Conf. (ICMC'2000)*, pp. 368–71. Berlin.

Wacom. 2002. *http://www.wacom.com*

Wanderley, M., Schnell, N., and Rovan, J. B. 1998. Escher – modeling and performing composed instruments in real-time. *Proc. of the 1998 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC'98)*, pp. 1,080–4. San Diego, CA.

Wanderley, M., and Depalle, P. 1999. Contrôle gestuel de la synthèse sonore. In H. Vinet and F. Delalande (eds.) *Interfaces homme – machine et création musicale*, pp. 145–63. Paris: Hermès Science Publishing.

Wanderley, M., Viollet, J.-P., Isart, F., and Rodet, X. 2000. On the choice of transducer technologies for specific musical functions. *Proc. of the Int. Computer Music Conf. (ICMC'2000).* Berlin.

Wanderley, M. 2001. *Intéraction Musicien-Instrument: application au contrôle gestuel de la synthèse sonore*. Thèse de doctorat, Université Paris VI, IRCAM.

Wessel, D. L. 1979. Timbre space as a musical control structure. *Computer Music Journal* **3**: 2.

Wessel, D., Drame, C., and Wright, M. 1998. Removing the time axis from spectral analysis-based additive synthesis: neural networks versus memory-based machine learning. *Proc. of the Int. Computer Music Conf. (ICMC'98).* Ann Arbor.

Wright, M., Wessel, D., and Freed, A. 1997. New musical control structures from standard gestural controllers. *Proc. of the Int. Computer Music Conf. (ICMC'1997),* pp. 387–90. Thessaloniki.

Wright, M., Freed, A., Lee, A., Madden, T., and Momeni, A. 2001. Managing complexity with explicit mapping of gestures to sound control with OSC. In *Proc. of the Int. Computer Music Conf. (ICMC'2001).*

Zwicker, E., and Scharf, B. 1965. A model of loudness summation. *Psychological Review* **72**: 3–26.

Zicarelli, D. 1998. An extensible real-time signal processing environment for Max. In *Proc. of the Int. Computer Music Conf. (ICMC'1998).* Ann Arbor.

https://doi.org/10.1017/S1355771802002054 Published online by Cambridge University Press