# Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance

Joseph Butch Rovan, Marcelo M. Wanderley, Shlomo Dubnov and Philippe Depalle
Analysis-Synthesis Team/Real-Time Systems Group - IRCAM - France
{rovan, wanderle, dubnov, phd}@ircam.fr

## Abstract

This paper presents ongoing work on gesture mapping strategies and applications to sound synthesis by signal models controlled via a standard MIDI wind controller. Our approach consists in considering different mapping strategies in order to achieve "fine" (therefore in the authors' opinion, potentially expressive) control of additive synthesis by coupling originally independent outputs from the wind controller. These control signals are applied to nine different clarinet data files, obtained from analysis of clarinet sounds, which are arranged in an expressive timbral subspace and interpolated in real-time, using FTS 1.4, IRCAM's digital signal processing environment. An analysis of the resulting interpolation is also provided and topics related to sound morphing techniques are discussed.

## 1 Introduction

A common complaint about electronic music is that it lacks expressivity. In response to this, much work has been done in developing new and varied synthesis algorithms. However, because traditional acoustic musical sound is a direct result of the interaction between an instrument and the performance gesture applied to it, if one wishes to model this espressivity, in addition to modeling the instrument itself - whatever the technique/algorithm - one must also model the physical gesture, in all its complexity. Indeed, in spite of the various methods available to synthesize sound, the ultimate musical expression of those sounds still falls upon the capture of gesture(s) used for control and performance.

In terms of expressivity, however, just as important as the capture of the gesture itself is the manner in which the mapping of gestural data onto synthesis parameters is done. Most of the work in this area has traditionally focused on one-to-one mapping of control values to synthesis parameters. In the case of physical modeling synthesis, this approach may make sense due to the fact that the relation between gesture input and sound production is often hard-coded inside the synthesis model. However, with signal models this one-to-one mapping may not be the most appropriate, since it does not take advantage of the opportunity signal models allow for higher level couplings between control gestures.

Additive synthesis, for instance, has the power to virtually synthesize any sound, but is limited by the difficulty encountered in simultaneously controlling hundreds of time-varying control parameters; it is not immediately obvious how the outputs of a gestural controller should be mapped to the frequencies, amplitudes, and phases of sinusoidal partials. [1] Nonetheless, signal models such as additive synthesis have many advantages, incuding powerful analysis tools [2] as well as efficient synthesis and real-time performance. [3].

Figure 1 shows the central role of mapping for a virtual musical instrument (where the gestural controller is independent from the sound source)[Mul94][VUK96] for signal and physical model synthesis. As shown in the case of signal models, the liaison between these two blocks is manifest as a separate mapping layer; for the physical modeling approach the model already encompasses the mapping scheme.

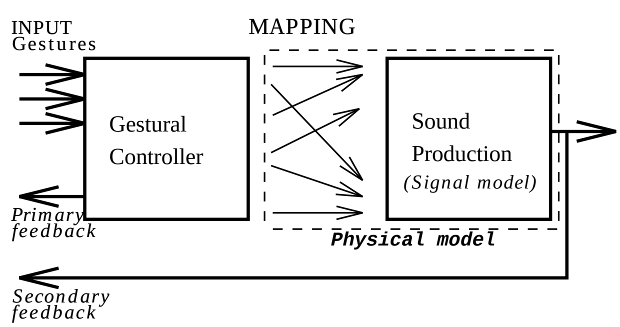


Figure 1: A Virtual Instrument representation

[1] For an example of a previous approach to this problem, see Wessel and Risset [WR82]

[2] The suite of analysis tools available at IRCAM include *additive* and *Audiosculpt*

[3] Our system uses an additive analysis/resynthesis method developed by X.Rodet and Ph. Depalle with synthesis based on the inverse FFT [RD92]

In the authors' opinion the mapping layer is a key to solving such control problems, and is an undeveloped link between gestural control and synthesis by signal models. Thus our focus in this paper on the importance and influence of the mapping strategy in the context of musical expression. We propose a three-layer distinction between mapping strategies: *One-to-One*, *Divergent* and *Convergent* mapping. Of these three possibilities we will consider the third - convergent mapping - as the most musically expressive from an "instrumental" point of view, although not always immediately obvious to implement.

We discuss these mapping strategies using a system consisting of a MIDI wind controller (Yamaha's WX7)[Yam] and IRCAM's real-time digital signal processing environment FTS[DDMS96], implementing control patches and an expressive timbral subspace onto which we map performance gestures. Departing from one of the author's experience as a clarinettist, we discuss the WX7 and its inherently non-coupled gesture capture mechanism. This is compared to the interaction between a performer and a real single-reed acoustic instrument, considering the expert gestures related to expressive clarinet/saxophone performance.

Finally, we present a discussion of the methods to do morphing between different additive models of clarinet sounds in various expressive playing conditions. We show that simple interpolation between partials that have different types of frequency fluctuation behaviour gives an incorrect result. Thus, in order to maintain the "naturalness" of the sound due to the frequency fluctuations, and to do the correct morphing, special care must be taken so as to properly understand and model this effect.

## 2 Mapping Strategies

We propose a classification of mapping strategies into three groups:

- *One-to-One Mapping* : Each independent gestural output is assigned to one musical parameter, usually via a MIDI control message. This is the simplest mapping scheme, but usually the least expressive. It takes direct advantage of the MIDI controller architecture.

- *Divergent Mapping* : One gestural output is used to control more than one simultaneous musical parameter. Although it may initially provide a macro-level expressivity control, this approach nevertheless may prove limited when applied alone, as it does not allow access to internal (micro) features of the sound object.

- *Convergent Mapping* : In this case many gestures are coupled to produce one musical parameter. This scheme requires previous experience with the system in order to achieve effective control. Although harder to master, it proves far more expressive than the simpler unity mapping.

Next we discuss the wind controller and compare its features to those of an actual instrument, offering coupling strategies that may aid in regaining some of the loss of fine control due to the wind controller's non-coupled design.

## 3 Comparative Analysis of Clarinet and MIDI wind controller

MIDI wind controllers have been designed to profit from the massive corpus of existing wind instrument playing technique, while at the same time providing the extra potential of MIDI control. Nevertheless, although MIDI wind controllers have the shape of and behave in a somewhat approximate manner to an acoustic instrument, they are drastically simplified models of real instruments (non-vibrating reeds, discrete [on/off] keys, etc.). In the WX7 controller, for instance, only three classes of woodwind instrumental gestures are sensed: breath pressure, lip pressure, and fingering configuration. These three classes of input are completely independent, sending three discrete streams of 8-bit MIDI data.

In contrast, acoustic instruments are obviously much more sophisticated. The reed of an actual wind instrument, for instance, has a complex behavior; many studies have shown the intricate and subtle non-linear relationships between the different instrumental gestures applied to the reed in woodwind instrument sound production. As one example, airflow through the reed of a single-reed instrument such as a clarinet or saxophone is a function of the pressure across the reed (i.e., the difference between the pressure inside the player's mouth and the pressure inside the mouthpiece) for a given embouchure.[Bac77][Ben90][FR91] (See Figure 2) In

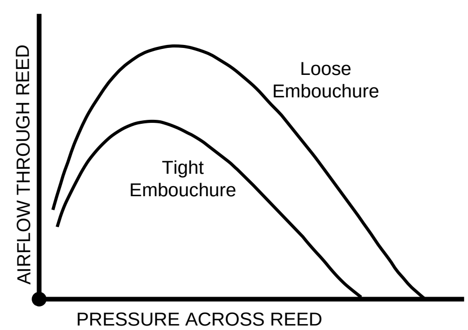


Figure 2: Flow through reed as a function of the pressure across the reed for a particular embouchure (Adapted from A. Benade [Ben90]).

an acoustic instrument, the reed actually behaves as a pressure-controlled valve, wherein increasing breath pressure tends to blow the valve closed. The closing point is thus a function of the embouchure, since the closing of the reed takes place earlier for a tighter embouchure than for a looser one, given the same pressure difference. Such couplings are not taken into account in available controller systems that mimic acoustic instrument interfaces, such as the WX7 or the Akai EWI, due to the fact that these systems do not include vibrating reeds. [4]

Furthermore, because of their role as controllers in two-stage systems that traditionally separate control from synthesis, the physical effects that account for sound production – and which are also very important as feedback for the performer – are intrinsically not modeled in wind controllers. These effects include feedback from the air pressure inside the instrument, symphathetic vibrations, etc. Although there is no means to simulate these physical feedback effects in a controller without the addition of actuators, one can simulate some of the behavior of the acoustical instrument through the use of specialized mappings.

## 4 Description of the System

The additive synthesis engine used for this project was implemented on a system consisting of an SGI workstation running IRCAM's FTS software. For the purpose of interpolation we constructed a 2 dimensional expressive timbral subspace covering a 2 octave clarinet range with three different dynamic levels. (See Figure 3). This additive parameter sub-

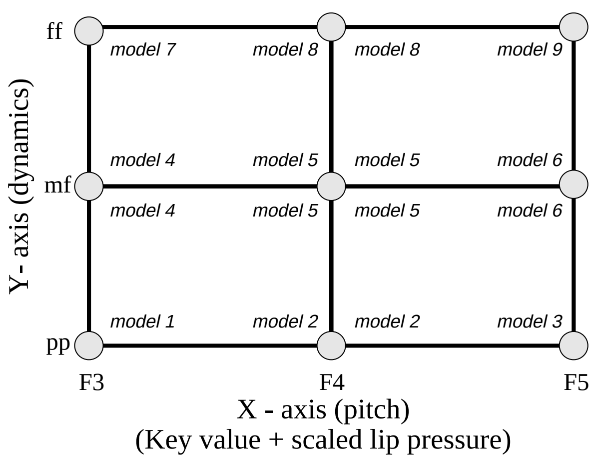


Figure 3: Expressive Timbral Subspace

space was built by analysing clarinet sounds from the Studio-on-Line project at IRCAM, recorded at high quality, 24 bits, 48 KHz and using six different microphone positions [Fin96]. Nine analysis files are obtained, three for each of three chosen pitches (pp, mf, and ff dynamics of F3, F4, and F5). Available synthesis parameters include global parameters such as loudness, brightness, and panning as well as the timbral space interpolation x- and y-axis values, and frequency shifting. An additional parameter - harmonic deviation - allows the scaling or removal of all frequency deviations from perfect harmonicity in the partials. The resulting output is an interpolation between the four additive model parameter files of each quadrant; first FTS performs two interpolations between the x-axis borders of each quadrant, and then a third interpolation between these two results is taken for the final output, according to the information received for pitch and dynamics from the controller/mapping.[5].

Although this approach seems very similar to the one taken in sample synthesizers – with the advantage of having a control (by interpolation) over the sustained portion of the sound – there is an important conceptual point to our approach which should be noted. By considering the additive method, we consider interpolation not between actual sounds but between models, and thus the issue of modeling is central to this work.

A simple noise source is also modeled in order to provide an approximation to the actual clarinet sound, since the models used for interpolation are issued from additive synthesis and therefore do not contain the noise components of the original sound. Within all mapping examples the noise level is controlled by a ratio of breath pressure to embouchure.

We should point out that our synthesis model considers "dynamics" to be strictly a timbral quality, based on the additive models for the normalized pp, mf, and ff clarinet sounds. Actual volume change is handled as an independent parameter.

## 5 Discussion of mapping implementations

In this paper we implement examples of One-to-One and Convergent mapping schemes. In order to develop these mappings, we recorded and analyzed various clarinet performance techniques, including non-standard examples such as overblowing and reed clamping. The couplings are then simulated by processing MIDI data from the controller.

The first example is a simple uncoupled One-to-One mapping, where airflow (breath pressure) data from the WX7 is mapped to overall volume and dynamics, lip pressure is mapped to vibrato, and fingering configuration is mapped to fundamental pitch.[6] In this case we consider the dynamic and volume change to be directly proportional to breath pressure.

With the second example we begin to consider different levels of dependency between parameters in an elementary implementation of Convergent mapping. Thus the input data for the synthesis engine may be dependent on the relationship of two or more gestural parameters. In this example embouchure information acts as a gating threshold for note production, apart from its normal application as a vibrato controller. If the embouchure is not inside a predefined range, no note is produced, as is the case

[4] For an up-to-date source of MIDI wind controllers, see the web sites *http://sunsite.unc.edu/emusic-l/info-docs-FAQs/wind-controllers-FAQ.html* or *http://www.ucs.mun.ca/ andrew/wind/*

[5] For the purposes of this paper we consider mf to be the middle point in the dynamic scale between pp and ff.

[6] The WX7 does provide some adjustments to change independently the response of its individual sensors, including the choice of different breath-response modes and lip-pressure curves.

with an acoustic instrument.

The third example investigates Convergent mapping further via the relationship between embouchure and breath pressure and their control of note production. Here we implement a "virtual flow" through the reed based on the acoustical behavior explained in section 3 (see Figure 2). (Note that with extremely high breath pressure levels, the loudness will actually decrease, due to the reed blowing closed.) We consider breath pressure data from the WX7 as directly proportional to the pressure inside the mouth, since the reed does not vibrate and the air pressure inside the controller's tube is not influenced by the activation of the keys. This information is sent through two tables, representing curves for loose and tight embouchure values. For all values between these two extremes an intermediate embouchure value is found by interpolation between the tables. For values outside this range, no note is produced. As a result of this coupling, loudness is a function of the "virtual flow." In this example we continue to consider the dynamic interpolation as a direct function of breath pressure.

From the analysis of the recorded clarinet performance technique examples we noticed that the dynamic interpolation is actually a function of the breath pressure *for a particular embouchure.* This fact leads to our fourth mapping implementation, where we improve upon example three by taking into account this interdependency. Example four (See Figure 4) adds another level of coupling, where variation on the timbral subspace y-axis is controlled by breath pressure, but scaled by the embouchure value. This effect is familiar to wind players when performing a crescendo; one must often progressively loosen the embouchure in order to increase the dynamic. One notices, for example, that for a tight embouchure the actual timbral and loudness variation is very limited. Loosening the embouchure accounts for an increase in both timbral and loudness ranges of our model; the maximum range for the y-axis is reached with a loose embouchure. (This maximum range is equivalent to the difference between pianissimo and fortissimo in our timbral subspace.) It must be noted, however, that although a tight embouchure restricts the timbral and loudness range, it does have advantages. Tightness of the embouchure also controls the timbral quality known to wind players as "focus." Focus appears to be related to the amount of noise component present in the sound; in our model we emulate its effect by varying the amount of noise added to the output.

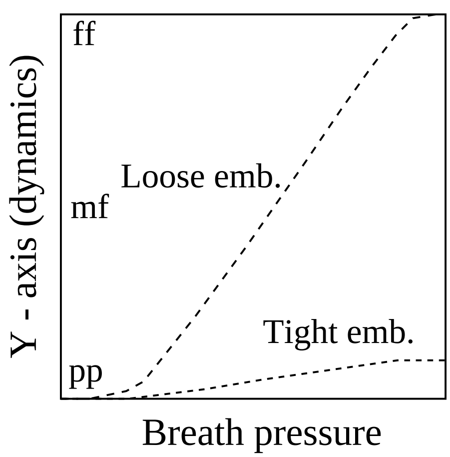


Figure 4: Mapping table for timbral subspace's Y-axis value

## 6 Analysis of the sound properties and problems with resynthesis

In the previous sections we have dealt with various ways to map gestural data in order to improve the espressivity of a controller, applied to a timbral subspace. After analyzing the synthesis results, however, it is evident that problems arise when interpolating between multiple additive models directly derived from sound analysis, such that it is difficult to capture the whole variety of the responsive behaviour of the sound. The purpose of this section is to consider these problems and discuss means to determine the correct synthesis model for interpolation.

Although the additive method allows a variety of transformations, two immediate problems arise in the context of expressivity control:

1. Change in register in the real instrument, resulting in a change of timbre, is not properly simulated by pitch shift.
2. Change in dynamics of the real sound, which is accompanied by a change in timbre and "texture"[7] of the sound, cannot be simulated by simple means such as changes in amplitude (loudness) of the sound.

When performing interpolation between additive models, it is exactly the textural properties that are the problematic ones. Let us explain the difficulty by simple example:

Let us assume that our system contains only pianissimo (pp) and fortissimo (ff) models. In order to reach an intermediate dynamic model, one morphs between the pp and ff models. In terms of amplitude relations, a close approximation to the mf spectral shape can be achieved by averaging the ff and pp sounds. In terms of the fine temporal behaviour, the situation is different: we observe in the morphed result a strong jitter of the high partials due to the interpolation of the frequency behaviour of the pp partials that are close to the noise floor, thus having a significant frequency jitter, with the originally stable frequency behaviour of the same ff partials. It is important to state that this effect is audibly significant and is heard as some unnatural, distortion-like behaviour of the high frequencies.

[7]By texture we mean the temporal behaviour of the sound components which are not captured by the powerspectra.

Investigating the frequency fluctuations of the three sounds reveals that the standard deviation of the mf sound is not only qualitatively closer to the shape of the ff model, but that the fluctuations in mf are smaller than in the ff sound and they cannot be apporximated by averaging between the pp and ff graphs[8] (see figure 5).

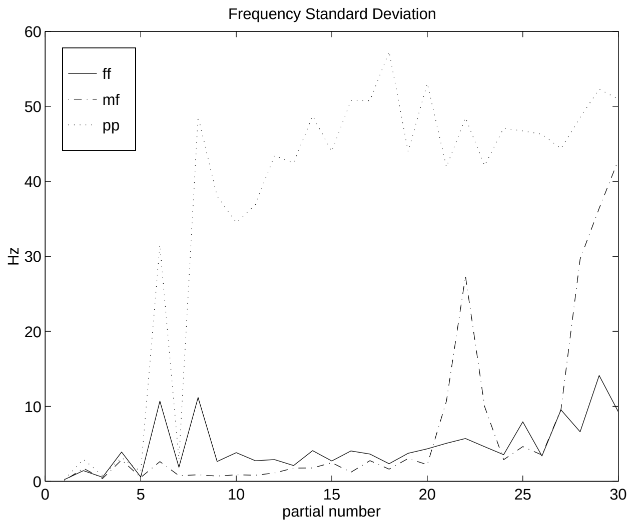


Figure 5: Standard Deviation of the Clarinet's F3 Frequency fluctuations of the first 30 partials in three different dynamics: ff, mf, pp.

Thus, superimposing wrongly the typical frequency jitter behaviour of the pp with the rather strong interpolated amplitudes creates an undesirable effect which is not present in the original mf sound. Let us now take a closer look at the frequency fluctuations of the partials in the three playing condintions.

## 6.1 Investigation of the Frequency Fluctuations

From the above experiment it appears that the problem lies in interpolation between partials that have very different regimes of fluctuations. Naturally, the first assumption about the origin of the big variance in frequency would be that partials close to the noise floor, i.e., the ones that are not sure to be actual partials but which are "forced" into the sinusoidal representation by the additive method, are the partials that have significant jitter. In such a case one might expect that:

1. There should be strong link between the amplitude of the partial and the amount of fluctuations.
2. The drop in fluctuations of the high partials should be proportional to the spectral brightness, i.e., increase in amplitude of the high frequencies.

[8] In terms of statistical analysis, linear combination of two independent random variables gives a new variable whose variance is the same linear combination of the original variables' variances. Thus morphing the frequency values is equivalent to averaging the variances.

It appears that the above assumptions **do not** hold for real signals and thus the whole mechanism of jitter stems from a different phenomena, which is apparently a non-linear one.

To see the dependence of frequency fluctuations on the playing condition, we have recorded a sound with gradually increasing dynamics [9].

For each one of the partials, the frequency standard deviation over 500 msec. segments was calculated as a function of time. As can be seen from figure 6, a drop in frequency fluctuations occurs **selectively** for some partials as a function of time and thus dynamics. For the other partials, the fluctuations never drop to be close enough to zero.

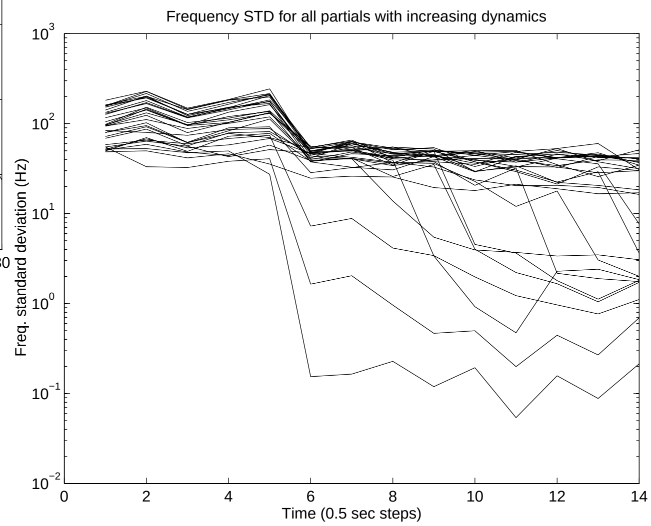


Figure 6: Standard Deviation of the Clarinet's Frequency fluctuations of the first 30 partials with increasing dynamics.

A closer look at the numbers of partials whose fluctuations drop as a function of time reveals the following interesting order (sorted according to fluctuation value, from low to high):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | | | | | | | | |
| 1 | 4 | 7 | | | | | | | |
| 1 | 4 | 5 | 7 | 3 | | | | | |
| 1 | 4 | 5 | 7 | 3 | 8 | 9 | | | |
| 1 | 4 | 5 | 7 | 8 | 9 | 3 | | | |
| 1 | 4 | 7 | 8 | 9 | 11 | 5 | 3 | | |
| 1 | 4 | 7 | 8 | 9 | 11 | 5 | 12 | 3 | |
| 1 | 4 | 7 | 8 | 11 | 9 | 5 | 12 | 3 | 15 |

Moreover, one can see that approximate harmonic relations exist between the different triplets of partials on the last line, according to the following combinations: (1 3, 4), (1 4, 5), (3 4, 7),(3 5, 8), (4 5, 9), (4 7, 11), (4 8, 12) and (3 9, 12), (7 8, 15) and (3 12, 15).

This phenomena is suggestive that the drop of variance is related to some sort of non-linear coupling phenomena that occurs between pairs of lower,

[9] In more precise terms, it was achieved by gradually increasing the air flow, keeping an almost constant loose embouchure

already exisiting and stable frequencies, and new partials that appear and their sum frequency. [10]

## 7 Conclusions

In this paper we presented a study of the influence of the mapping layer as a determinant factor in expressivity control possibilities. We introduced a three-layer classification of mapping schemes that proved useful in determining mapping parameter relationships for different performance situations; these mappings were applied to the control of additive synthesis. From this experience, the authors feel that the mapping layer is a key element in attaining expressive control of signal model synthesis.

Several mapping examples were presented and discussed. In an instrumental approach, the convergent mappings demonstrated in this paper have the potential to provide higher levels of expressivity to existing MIDI controllers. Without the need to develop new hardware, off-the-shelf controllers can be given new life via coupling schemes that attempt to simulate the behaviors of acoustic instruments.

Finally, regarding the interpolation between additive models, we showed that in order to achieve a "correct" morphing between models, the non-linear coupling phenomena must be considered. The interpolations between the partial frequencies thus must be allowed only among groups of partials having correponding "regimes" of fluctuations, i.e, coupled partials, non-coupled partials and "noise". In order to bypass this problem, we currently eliminate all inharmonicity from the models before performing the interpolations.

## 8 Future Directions

We plan to implement the fine control of texture in our additive models as suggested in Section 6.1, as well as to develop different mapping schemes. Also, we are considering using a custom data glove in conjunction with the WX7 in order to capture more detailed performance data.

Finally, this systematic investigation of gestural mapping uncovers interesting pedagogical uses of such an approach. One direction we are considering involves the application of such mapping strategies to methods that may improve the typical learning curve for an acoustic instrument through the use of MIDI controllers.

## Acknowledgments

We would like to thank Norbert Schnell of the Real-Time Systems Group/IRCAM for implementing custom FTS objects for this project. Also thanks to Xavier Rodet for his helpful comments.

[10] Although this method is not a direct proof of the non-linear coupling hypothesis, this effect can be shown more directly by application of Higher Order Statistical methods[DR97]. Due to limits of space in this paper, we will not present these results.

Parts of this work were supported by grants from the University of California at Berkeley Department of Music, CNPq (National Research Council) - Brazil, and AFIRST (Association Franco-Israelienne pour Recherche Scientifique et Technologique).

## References

[Bac77] J. Backus. *The Acoustical Foundations of Music.* W. W. Norton and Company, Inc, 2nd edition, 1977. Chapter 11.

[Ben90] A. Benade. *Foundamentals of Musical Acoustics.* Dover, 2nd edition, 1990. Chapter 21.

[DDMS96] F. Dechelle, M. DeCecco, E. Maggi, and N. Schnell. New dsp applications on fts. In *Proceedings of the International Computer Music Conference*, pages 188–189, 1996.

[DR97] S. Dubnov and X. Rodet. Statistical modeling of sound aperiodicities. In *Proceedings of the International Computer Music Conference*, 1997.

[Fin96] J. Fineberg. Ircam instrumental data base. Technical report, IRCAM, 1996.

[FR91] N.-H. Fletcher and T.-D. Rossing. *The Physics of Musical Instruments.* Springer-Verlag, 1991. Part IV.

[Mul94] A. Mulder. Virtual musical instruments: Accessing the sound synthesis universe as a performer. In *Proceddings of the First Brazilian Symposium on Computer Music*, 1994.

[RD92] X. Rodet and P. Depalle. A new additive synthesis method using inverse fourier transform and spectral envelopes. In *Proceedings of the International Computer Music Conference*, pages 410–411, 1992.

[VUK96] R. Vertegaal, T. Ungvary, and M. Kieslinger. Towards a musician's cockpit: Transducers, feedback and musical function. In *Proceedings of the International Computer Music Conference*, pages 308–311, 1996.

[WR82] D. Wessel and J.-C. Risset. *Exploration of Timbre by Analysis and Synthesis.* 1982. in: D. Deutsch, The Psychology of Music, Academic Press Inc, 1982, chapter 2, pp. 26-58.

[Yam] Yamaha. *WX7 Wind MIDI Controller.* Owner's Manual.