The need for physical modeling.

Lorenzo Seno

http://www.acustica.org
lorenzo.seno@bigfoot.com

Centro Ricerche Musicali (http://www.crm-music.org)
EcOna (http://w3.uniroma1.it/econa/).

Starting from about 40 years ago – together with the beginning of Electronic Music - musicians and researchers gave raise to the long lasting effort to synthesize both the sound of acoustical instruments and to invent new ones in some way.
In order to investigate the rules following which sounds are built-up in actual instruments, synthesis techniques as well as the corresponding analysis techniques were developed. In this way it has became possible to analyze sounds of a particular instrument and to reproduce them by using the corresponding synthesis technique. Using the analysis-synthesis couple, amid these two steps of this process it is possible to introduce little modifications in some parameters in order to investigate their perceptual role.
These first-time studies have brought to important cognitive results about the mechanism of sound emission and of sound perception. Risset showed that what we call the “timbre” of a trumpet is due to the way in which partials raise in this instrument during the tone attack.
This kind of investigations crossed soon another mainstream in the research around sounds: those made by the more ancient studies on speech emission, a long lasting research which took place mainly in telephone company laboratories, in the quest of means to reduce the bandwidth for phone conversations.
Speech studies brought at very early times to the development of physical modeling, as opposed to “signal modeling” of sound. The widely used Cepstrum techniques belong to the signal modeling methods, as they deal with the intrinsic features of the speech sound. LPC techniques are based instead on the idea that speech is produced by some physical entities – an exciter, and a resonant cavity – and on the idea that the mathematical model has to represent these components and their behaviors.
In this way, the parameters of the model are related – or better: they directly represent – some physical quantity tied to the sound emitter.

Today synthesis techniques for musical instruments.

What it is actually used in today electronic musical instruments are the FM and wavetable synthesis techniques, although the first one in commercial instruments is now relegated to the low-end segment. This is not the case of cultivated electronic music, in which FM is today still extensively used because of the great flexibility in inventing new sounds offered by this technique.
The great success of FM is due to the circumstance that the use of frequency modulation makes relatively easy to obtain spectrally dense sound, whose spectrum is populated by a huge amount of evenly spaced lines. So, complex sounds can be obtained using very few oscillators – a circumstance greatly appreciated in the epoch in which oscillators were expensive in terms of hardware or software resources. Controlling these populated combs, or introducing in them some subtle behavior (as microfluctuations in amplitude and frequency, or small amounts of inharmonicity), is a totally different point, and can become a true challenge for the composer when using FM. This is the reason for the continuing quest for new synthesis techniques.
Wavetables are scarcely used by today cultivated composers because their use is confined to imitative sounds – a feature maybe important from an economic standpoint - not for artistic purposes. They survive in some way only in a sound composing technique, which is inspired by Gabor grains or by Wavelets, whose name is “granular synthesis”. A similar technique is used in today high-quality speech synthesizers (for what we can understand about undocumented purposes, this is the case of “Actor” by the Italian firm TelecomLab, the former CSELT).

Some years ago, mainly thanks to Karplus and Strong, then to Julius O. Smith III and Jaffe from Stanford's Center for Computer Research in Music and Acoustics (CCRMA), physical modeling synthesis has come to the scenes of the musical research, and it even appeared in some commercial equipment (Yamaha, Korg). This commercial premiere actually was a false start, and physical synthesis disappeared from the market after a while. This was – in my opinion- because of the roughness of these first attempts. Physical modeling is nevertheless continuing to be the mainstream in musical synthesis research.

Different approaches to physical modeling.

Waveguides are historically the first approach to the problem. The heart of any melodic musical instrument is made by a physical resonant device of some kind (for instance a string, or an air cylinder) excited in some way to become a multimode oscillator (for instance, plucking or bowing a string, or by means of a reed or a nozzle, as it is the case of wind instruments).
As a first approximation, string or air pipes can be considered as held by a two-dimensional wave equation of some kind, whose prototype is:

$$\textbf{(1)} \quad \frac{\partial^2 y}{\partial t^2} = c^2 \cdot \frac{\partial^2 y}{\partial z^2} + \ldots$$

Where y represents some significant magnitude (e.g. transversal displacement or speed in a string, pressure along a pipe, and so on …) and z is a longitudinal coordinate. Here c is the propagation speed, and is related to the string Tension T and linear mass Density μ :

$$c = \frac{T}{\mu}$$

It is a well-known result of Calculus that such a differential equation admits D'Alembert solutions in terms of superimpositions of forward and backward waves:

$$y(z,t) = y_b(z,t) + y_f(z,t)$$

in which the backward and forward functions are:

$$y_b(z,t) = y(z + c \cdot t, 0) \quad \text{and} \quad y_f(z,t) = y(z - c \cdot t, 0)$$

so that these two components are expressed in terms of the initial shape of the string.
In other words, the initial shape propagates toward left and right, being possibly reflected by the extremities of the string (or pipe).
These solutions work fine for the abovementioned raw equation. Without other terms, this is a conservative equation, namely describing a system in which the energy is conserved, where any perturbation from the rest will propagate for everlasting.
Actual resonators like pipes or strings are of course NOT conservative, so that, when the excitation is removed, the sound decays more or less quickly.
But if you are considering strings, and you take a period of oscillation as a measure of time, you can find that sounds in strings are close to be everlasting, so that there is a significant hope that the abovementioned raw conservative equation could be considered as a good first approximation of the actual string: more subtle behaviors can perhaps be introduced as small deviations from the conservative motion.
This matter is a complex issue, as we will find later, but what we are showing now is the basic idea behind the waveguides method.
The first implementation of the solution of the motion equation "à la D'Alembert" is the Karplus-Strong Algorithm.

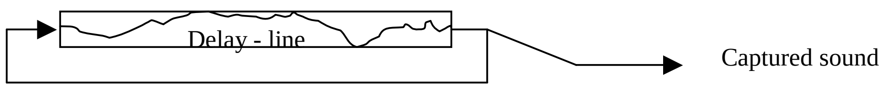

fig. 1.

If you load the delay-line with some waveform, and let the system evolve, the initial shape will be repeated for everlasting, with a frequency given by the total delay of the line. In this way you obtain a waveform composed by a fundamental partial (which is the one responsible of the pitch of the sound) and a set of harmonics determined by the shape of the initial curve.
This was historically the first algorithm by Karplus & Strong. Delay lines can be easily implemented, in numerical systems, as circular buffers in memory. Only a pointer is updated to read the sample stored in the delay line, so that the implementation of this algorithm is very inexpensive in terms of computational resources. This was a great issue, at the beginning of the computer music.
The original KS algorithm has only one set of propagating waves, so that this model is too raw. When using it, we are limited to play only initial conditions, but if you want to introduce more sophisticated excitations (like, hammering, or plucking, or bowing) you will need a more mathematically complete model, in order to cope with time-varying perturbations.

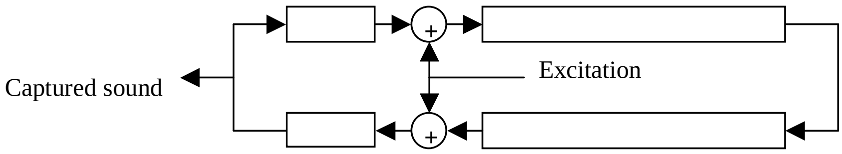

fig. 2.

This brings to the waveguide model, in which two delay lines are used to represent the two sets of traveling waves.
In this way, an excitation can be easily introduced by simply summing up its contribution to both waves. When dealing with strings, velocity waves are usually considered (because the sound emitted is tied to the speed of the motion of the string, due to the coupling mechanism with the atmosphere). As a matter of fact, the transforming of a physical excitation into an additive velocity perturbation can be quite tricky, but at present we will neglect this kind of technicalities. We now will deal with the drawbacks of this approach, which still retains the main pro of this ancestor: low computational needs.
Among the three main drawbacks of this approach, the first one that arises is the conservativity itself of the model. We are now dealing with excitations, which introduce energy into the system, so that not only every sound is in this way everlasting, but also amplitudes can raise-up everlasting, bringing to sounds of everlasting increasing loudness, when you repeatedly introduces excitations into the system.
The introduction of some dissipation into the waveguide model is in principle not difficult, and we will show in a while how this can be easily done. So it would at first glance appear that this should not be considered an actual drawback, only there is here some subtle problem cached behind the scenes, as we will show later.
For the moment, let’s show how dissipation can be introduced into the model.

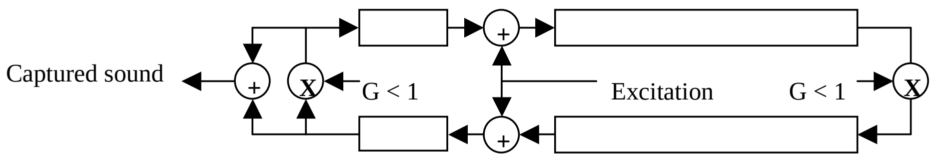

fig. 3.

We introduced a Gain (less that 1: indeed attenuation) into the loop, so that now the sounds will decay exponentially with time. Note that these gains are concentrated at the two extremities of the string – something like a way to take into account for the losses at the bridge. The exact location of the gain is actually not important at all – though placing it at the extremities can appear as having some special physical meaning. Our model is fully linear (it is made by delays, gains and sums), so that every processing block can commutate with any other. This is an important feature of the waveguide model: you can place ingredients in any order, the overall result will not change.

Commutativity has been extensively used to concentrate distributed behaviors into only one point, in order to reduce the computational complexity. Let's show this idea using exactly the non-conservative model we have just set-up.
Another interpretation of the model is given with the introduction into the wave equation of a non-conservative, of a viscous friction term. Let's suppose to deal with displacement waves:

$$\textbf{(2)}\quad \frac{\partial^2}{\partial t^2} y(z,t) = c^2 \cdot \frac{\partial^2}{\partial z^2} y(z,t) + S \cdot \frac{\partial}{\partial t} y(z,t)$$

The second term in the right-hand is something proportional to the speed, namely a typical viscous friction term. It can represent the friction of the string with the air. Please note that this is a distributed dissipation, but if you solve the equation, you will find the same exponential decay shown by the waveguide model. You can consider that there are infinitesimal losses distributed along the delay line (according to our physical representation of the phenomenon), and that you have commutated them, moved them and concentrated them in just one or two points.
With the introduction of the above dissipative correction, we are consequently setting-up a decay time which is identical for any partial, irrespective of the frequency. That is, a frequency-independent decay time.
This is not what happens in actual strings, in which the higher the partial frequency, the shorter the decay. This is a quite important point, as we will see later, because this feature is perceptually of great importance.
Let's for the moment go to the second announced drawback of the waveguide approach. In numerical system we can now use floating point calculations, so that the amplitude discretization error can be largely neglected. But also time is discretized, so that, given a certain sampling frequency, the overall delay of the waveguide is also discretized. As a consequence, the pitch of the emitted tone is in turn discretized. What is now arising is a "tuning problem": in order to cope with correct tuning (mainly for higher tones) or with glissandos, you have to deal with fractional delays. This is of course possible, because interpolation is always possible to any desired degree of accuracy, but fractional delays imply in any case some computational cost – about an order of magnitude more than the whole waveguide model itself - so that we are losing in this way - at least in part - the main appeal of the method - namely its computational cheapness. Approximate interpolation algorithms – suitable for musical purposes – have been developed mainly by J.O. Smith.

Perceptual features of sounds

Before introducing the third drawback of the waveguide method, let's have an overall glance to what we know about perceptual features of tones – namely their timbre –to state what features are important in order to be implemented into synthesis models.
The first theory of timbre is due to Helmoltz, one of the most interesting physicists and thinkers of the 19th century.
Following his first approach to a theory of sound perception, the timbre of a sound - i.e. the characteristic that allow us to segregate different sources irrespective from loudness, pitch and source position – is due to the distribution of energy between partials (harmonics) of sounds. Some laboratory experiment can easily apparently confirm this conjecture. A sawtooth waveform will sound much different from a square wave, and we can easily know that their spectral profiles are different.
But another (today) simple experiment will suddenly mess-up our certainty. The time reversal sound of a piano (just for instance) has the same harmonic profile that the original one has, but it can hardly resemble to a piano.
What is now clear (after more than a century) is that what we call "timbre", even of pitched sounds, is to a great deal more complex than one should expect from these simplified theorizations. Not only spectral, but also time-domain features are important, as well as mixed time-domain and spectral characteristics.
Let's turn back to the damping behavior of actual strings. Any string shows the features of a decay time, which decreases with frequency. This means, by the way, that during decaying, wave shapes are not preserved. Moreover, actual spectra are far to be perfectly harmonic, and some amount of inharmonicity gives to the sound its peculiarity.

Slightly superharmonic ratios in partials are perceived as “metallic”, as sounds produced by “stiff” strings are. This is the case of pianoforte strings, which are stiff, and consequently metallic.
Slightly subharmonic ratios give to the tone more and more the timbre of pot, or tin, sound. A high degree of inharmonicity (as happens in membrane sounds) can more or less destroy the pitch perception, bringing to percussive (toneless, a-melodic) timbres.
But there are more subtle features, which makes part of our daily perceptual experience of sounds. Till now, we have spoken about features, which are fully compatible with a linear model of the resonant object. Starting from now, we are dealing with non-linear components or behaviors.
For instance, the amplitude of partials can microfluctuate, as well as the frequency itself, both in synchronous and asynchronous fashion. The first feature can suggest some amount of energy transfer between partials, while the second one seems instead to suggest some interaction with parametric low-frequency oscillations. In all these phenomena, if you attentively listen to some acoustic instrument, you can hardly perceive any perfect periodicity. There is some amount of chaos (micro-chaos) superimposed to the main, linear, harmonic behavior of actual acoustic instruments. Sounds of actual acoustic instruments are much like the Heraclit’s river, in which no one can be washed by the same water.
These features are perceptually important to a great deal, because they are those, which make the difference between “natural” sounds, and “electronic” sound. One can even say that there is a particular “”timbre” of electronic sounds, which is due to the lack of these “micro-chaotic” components. Even with sampled sounds – as happens with wavetable synthesis – there is a well-known problem of this nature. As samples must be of limited duration, the mean to obtain long-lasting sounds from shorter samples is to put the end of the sound into a loop, so that in principle you can obtain a sound of any duration (think to a sampled Pipe Organ, for instance). But this loop (which is accurately chosen by the sound engineer) introduces a long-term periodicity (in timbre), which is suddenly recognized by any listener as an “artificial” feature.
No one, seriously involved in research about synthesis techniques, can ignore that the main problem of synthesis is to take into account of these non-linear, time-varying, phenomena. Many efforts have been made in order to cope with these quantitatively small, but perceptually important, non-linear effects. Results are partial and yet provisional.
The problem is that the “metrics” – so to say – of the physics or mathematics does not agree so much with the “metrics” of perception. Small physical phenomena can have a great perceptual counterpart, as well as quantitatively relevant ones can hardly be perceived. This is a quite well stated truth, in the audio compression epoch.
The problem with waveguides, in my opinion, is that there are problems also in coping with perceptually relevant, linear features, as decay times. Anyway, before going into a deeper discussion, let’s investigate more about the physical mechanisms, which are implied into these non-linear features.

Non linear and time-varying mechanism in resonant, sound-emitting systems.

Altough similar consideration can be applied to other resonators, let’s devote only to strings.
Our ground zero model (as any model actually does) is only a raw approximation of actual strings.
The classical two-dimensional wave equation (the “string equation”) is obtained under various approximations. We will discuss them in some detail.

Dimensionality.

First, comes the two-dimensionality. Actual strings motion is three-dimensional. If you take into account the centrifugal force, the equation becomes non-linear. You can consider of course that the coupling between the two coordinates (x and y) is weak, and that centrifugal force is negligible, so that you can start with the idea that a three dimensional string can be considered as the superimposition of a couple of two-dimensional strings, with some weak interaction mechanism between x and y directions, to be treated as slow and small perturbations. As a matter of fact, the sound of a three-dimensional, actual string greatly resembles to the sound of a two-dimensional one – with the exception of small peculiarities, though perceptually important.

Recalling force.

The recalling force in the string is considered as due to the tension projection for small displacements. This brings to identify a sinus with its arc, which is a forced linearization of the centripetal force. Moreover, the string is considered perfectly extensible, so that no extra recalling force is added due to the string stretch. These more or less slight non-linearities can of course give rise to (small) amounts of energy transfers between partials, and are the reasons of the well-known phenomenon of pitch increasing with the plucking strength. When a string is heavily plucked, at the beginning the pitch is slightly higher than in subsequent times. This gives the perception of an "sforzato" tone (for instance, in Harp).

Moreover, strings are considered not stiff at all, so the stiffness recalling force component is neglected. It is worth to note that stiffness is a linear term, which is responsible of the abovementioned inharmonicity. It isn't so hard to introduce a stiffness term, and it doesn't bring us outside linearity, so we will not go deeper in this direction. The stiff term is a fourth-degree derivative. Stiffness does not introduce dissipation (it is a conservative term), but it brings to dispersion, so that in terms of traveling waves, the phase propagation velocity is non-constant with frequency. This is the reason why this term can shift frequencies.

$$\textbf{(3)}\quad \frac{\partial^2 y}{\partial t^2} = c^2 \cdot \frac{\partial^2 y}{\partial z^2} + \Xi^2 \cdot \frac{\partial^4}{\partial z^4} y$$

Stiffness has been simulated in waveguide models (which inherently gives perfect harmonics), introducing frequency-dependent phase shifts. All-pass filters are appropriate for this purpose.

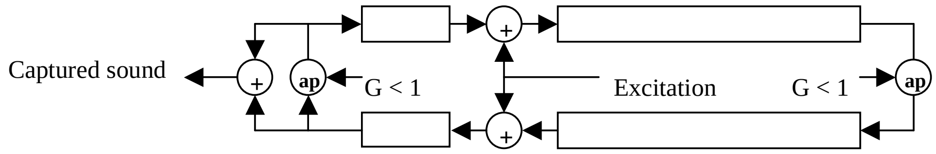

fig. 4.

Degrees of freedom.

Besides the dimensionality, the string has other "cached" degrees of freedom. Its motion is not confined to the transverse direction. If the string is capable of stretching, a quasi-independent set of longitudinal waves can propagate along the string. These waves are governed by a similar equation, but reasonably with higher propagation velocity (and consequently higher frequencies). They have both a direct (audible) effect, and a parametric effect.

The audible effect is due to the fact that longitudinal vibrations can propagate to the bridge (ponticello), via a cantilever action, being in this way radiated by the sound box. The parametric effect is due to local modifications in the centripetal force, which can slightly modulate the frequency of transversal motion.

As the string is not infinitely thin, it can also host torsional oscillations, to which the same considerations can apply. Moreover, torsional motions can become an important feature when the string is excited through a bow, due to the tangential friction mechanism of excitation.

Constraints.

Till now, we have considered the constraints at both the extremities as ideal, perfectly reflective bonds. They can also give rise to linear phenomena, as they transmit energy to the whole instrument (not necessarily behaving as being flat in frequency, during this operation), so that they can have a complex admittance. This aspect can be considered as a linear filtering behavior, so it is not so interesting in our discussion. But they are not ideal under further different aspects.

First, they are not symmetric. This means that the effective length of the string slightly changes during time, as a function of the direction of the motion, and also of the sign of the motion.

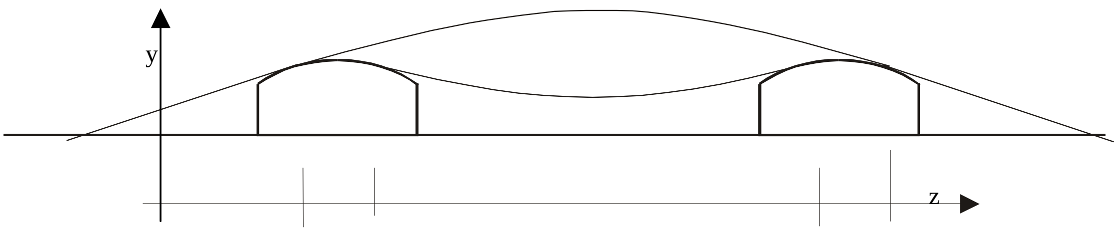

fig. 5. Vibrating string over its constraints – side view

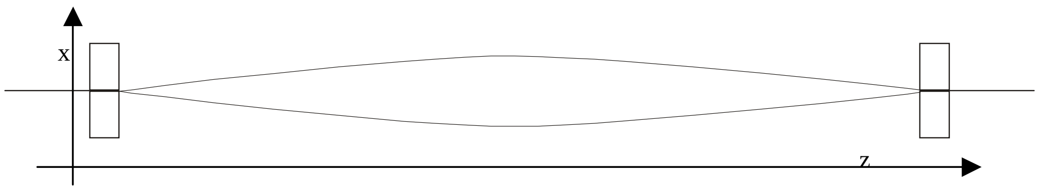

fig. 6. Vibrating string over its constraints – top view

The paradigmatic instrument in this respect is the Finnish Kantele, in which the fastening mechanism of the strings makes the length of vertical displacement quite different from the horizontal one, thus producing beats. The Kantele has been extensively investigated, and a waveguide model (using two slightly differently tuned waveguides) has been set-up.

Known perceptual and physical facts.

How much each of these factors is perceptually important? To my knowledge, not so much experimental work has been done in investigating the relative importance of these various factors. The circumstance that the decay time of partial, in strings, must decrease with frequency is a well-stated truth, but there is no knowledge of the empirical laws for various materials, length, and so on. This kind of measurements are not so difficult to implement, and can be made using a quite regular audio instrumentation. I hope I will be able in the future to experimentally investigate in this direction. From a theoretical point of view, this point will be investigated later in this paper.

It is a well-stated truth (from Risset) that the timbre of the trumpet is mainly due to the different (not monotonic in frequency) rise times of the partials. This can be considered as a general clue, that partials rise and fall times are important from a perceptual standpoint.

It is also a well-stated fact that partials micro-fluctuate both in amplitude and frequency, without any perceivable exact periodicity. The net effect of these fluctuations – in the frequency domain – is to make the spectrum denser (with respect to a "perfect", mathematically constant, set of partials).

It seems to me that our perceptual auditory system is particularly responsive to such "localized" (i.e. in the neighborhoods of spectral lines) densities. Sounds without this feature are perceived as "empty" sounds. It is reasonable to suppose that there must be some specific reason for such sensitivity, maybe tied to some fundamental (biological) teleonomy of our auditory system.

An exact knowledge of the operation of our auditory system is today not yet available, but maybe some conjecture can be made.

Many auditory mechanisms are based on the capability to identify "formant" spectral profiles - using comb-like sounds whose spectral profile is neither so well stated nor constant. This mechanism is deeply implied in speech, in which it is responsible of the identification of vowels, but also in sound spatial localization.

Many concurrent mechanisms contribute to our final capability to localize sound incoming direction, both in azimuth and elevation, also segregating front sources form rear sources.

Interaural differences in Time of Arrival (or intensities) are not capable to fully resolve ambiguities in sound direction, mainly front-rear and elevation ambiguities.

It is well known that a great contribution to the problem of avoiding ambiguities is done by the Head Related Transfer Functions, namely the filtering characteristics of the head, to which the main contribution is brought by the filtering capabilities of the Pinnae, with some contribution by the head shape. HRTF are direction dependent, so that the detection of their shapes allows the detection of the incoming direction.

In this way, also the process of detecting sources directions can be in part considered as a “formant extraction” process, a target hardly reachable when incoming signals have perfectly constant spectral lines but of uncertain profile (the knowledge of the spectral profile of the source is inherently approximate, due not only to the source variability, but also to the environment filtering behavior), and in presence of background noise.
Locally dense spectral lines can maybe help, allowing the detection of the local derivative of the formant profile, so that a sort of interpolation between consecutive lines can be envisaged. Only as a hypothesis, this can be the biological function of spectral density, and can be the reason for which too “perfect” sounds are hardly localizable, and the intelligibility of speech synthesis systems having too “perfect” spectral features is not robust.

Chaos.

Is chaos present in musical instruments sounds? There are few experimental studies about this matter; all conclude that chaos is present in acoustical instruments’ sounds. Here we can add only a few considerations.
First, we can observe the Lissajous figures that a plucked guitar string draws when observed near the ponticello on the plane perpendicular to the string itself. They have an unpredictable evolution, with sudden, irreproducible, changes in direction (we are speaking of the x-y plane), mainly near the end of the decay. It is very hard to avoid the temptation of calling this motion as “chaotic”. Of course, this motion might be recorded using adequate instrumentation, then submitted to an appropriate analysis (Liapunov coefficients, for instance, being the search of bifurcations in my opinion quite difficult or impossible).
Second, concerning the above-mentioned conjecture, it is worth noting that chaotic micro-fluctuations in amplitude and frequency of partials, when integrated over time, will bring to a spectrum more locally dense than any periodic parametric oscillations can do. So it is reasonable to think that the duties of our auditory system are made easier by micro-chaotic sounds.
As a final conclusion, it is worth noting that “reasonable” doesn’t necessarily mean “true” (whatever meaning you will assign to this word). This whole matter should be submitted to a series of investigations, both experimental and theoretical.

Damping of partials.

Some result in this matter was achieved in previous works I made a few years ago with Marco Palumbi.
The simplest mechanism you can envisage to physically explain this behavior is to suppose some inner mechanism in the string for which energy is dissipated due to internal friction, in such a way that the friction force is proportional to the speed of variation of the curvature of the string. It is a sort of viscous mechanism (because is tied to some speed), you can easily experiment when rapidly bending up and down a piece of iron wire. You can remarkably heat-up the bending point (in which the curvature variation during time is maximum), till melting and consequent break-up.

$$\textbf{(4)}\quad \frac{\partial^2 y}{\partial t^2} = c^2 \cdot \frac{\partial^2 y}{\partial z^2} + S_i \cdot \frac{\partial}{\partial t}\frac{\partial^2}{z^2} y$$

Such a mechanism will introduce into the string a term of mixed derivatives:
The last term can be easily read as the speed of variation (time derivative) of the second spatial derivative of the shape (curvature component normal to the z axis). It qualitatively works fine: no energy is dissipated if the shape doesn’t change in time, as expected by any internal dissipation mechanism. Moreover, S_i, the internal damping coefficient, is independent from the string length, and is a feature specific of the material of the string. We expect higher values for nylon and catgut than for metals.
If you go to the solutions of this equation, you will discover that motion is damped, but that the damping time-constant depends on the partial number (i.e., on the frequency).
This equations has the following general solutions:

$$y(z,t) = \alpha \cdot \sin(k \cdot z \pm \omega \cdot t) \cdot e^{\frac{-t}{\tau}}$$

The introduction of boundary constraints $y(0,t) = y(L,t) = 0$ will discretize k :

$$k_n = n \cdot \frac{\pi}{L} \text{ and } \omega_n = \pm n \cdot \pi \cdot \frac{c}{L} \cdot \sqrt{1-(n \cdot \pi \cdot \frac{S_i}{c \cdot L})^2}$$

For conservative strings (where $S_i = 0$), frequencies are instead:

$$\omega_n^c = \pm n \cdot \pi \cdot \frac{c}{L}$$

So that the modes of the dissipative string are:

$$y_{\pm n}(z,t) = 2 \cdot \alpha \cdot e^{\frac{-t}{\tau_n}} \cdot \sin(k_n \cdot z) \cdot \cos(\pm|\omega_n| \cdot t) \text{ if } \omega_n \text{ is real } (n < \frac{c \cdot L}{\pi \cdot S_i})$$

or

$$y_{\pm n}(z,t) = 2 \cdot \alpha \cdot e^{\frac{-t}{\tau_n}} \cdot \sin(k_n \cdot z) \cdot \cosh(\pm|\omega_n| \cdot t) \text{ if } \omega_n \text{ is imaginary } (n > \frac{c \cdot L}{\pi \cdot S_i})$$

The damping time constants are:

$$\tau_n = \frac{2}{S_i \cdot k_n{}^2} \text{ hence } \frac{\tau_n}{\tau_1} = \frac{1}{n^2}$$

Another useful way to represent these results is to express frequencies in term of the corresponding conservative "non-stiff" modes (i.e., frequencies having the same mode index n):

$$\omega_n = \pm n \cdot \omega_1^o \cdot \sqrt{1-\left(\frac{2 \cdot n}{\tau_1 \cdot \omega_1^o}\right)^2}$$

Stiffness.

Let's now take into account the stiffness of the string:

$$\frac{\partial^2}{\partial t^2} y(z,t) = -\Xi^2 \cdot \frac{\partial^4}{\partial z^4} \cdot y(z,t) + c^2 \cdot \frac{\partial^2}{\partial z^2} y(z,t)$$

The solutions of such an equation bring to the following modes (frequencies):

$$\omega_n = \pm\sqrt{\Xi^2 \cdot k_n{}^4 + c^2 \cdot k_n{}^2}$$

Or, in terms of the "non-stiff" modes ω_n^o:

$$\omega_n = \omega_n^o \cdot \sqrt{\frac{\Xi^2 \cdot \omega_n^{o2}}{c^4} + 1}$$

Cancelled modes.

Please note that modes having real frequencies are oscillatory, while modes having imaginary frequencies are "over-damped", non-oscillatory. In the following, we will call these as "cancelled modes", while the partial figure beyond that the modes are cancelled, is referred as the "canceling limit".
The condition for being oscillatory becomes now:

$n < \frac{\tau_1 \cdot \omega_1^c}{2}$ where ω_1^c is the conservative fundamental.

Another remark is to be done: this kind of damping introduces dispersion: partials' frequencies are not in harmonic ratios.

The sub-harmonicity is slight - at least out from a microtonal standpoint - as can be seen in the following figure, in which the inharmonicity of various partials is plotted against the fundamental frequency, for a damping time-constant of the fundamental $\tau = 0.5\,\text{sec}$

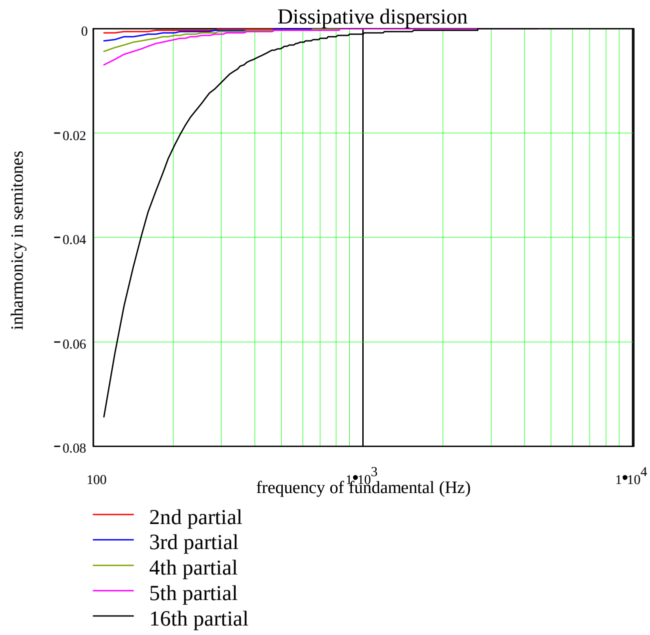

fig. 7.

The number of oscillating partials is high, whit 0.5 sec of time constant:

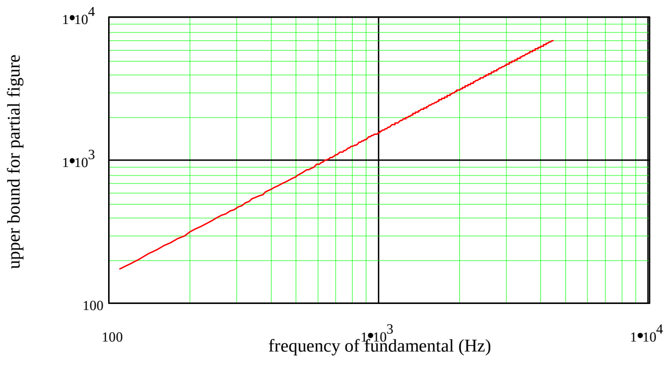

fig. 8.

Using the unrealistic value $\tau = 0.1\sec$, the inharmonicity starts becoming noticeable, mainly for higher partials, and the number of oscillating partials dramatically decreases:

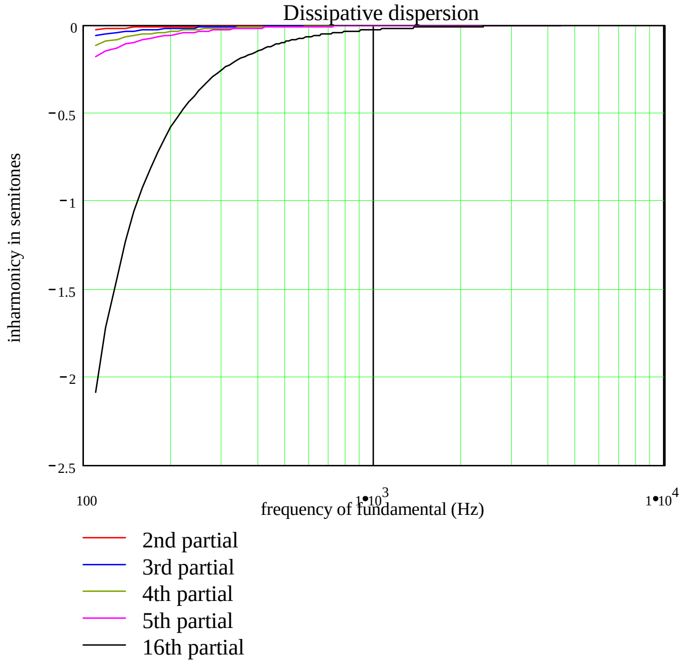

fig. 9.

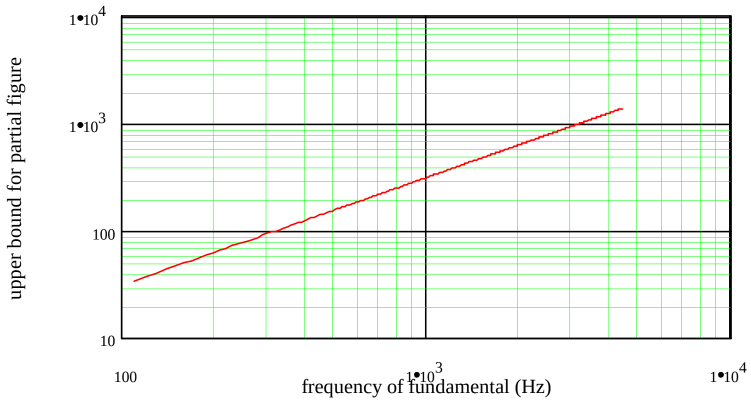

fig. 10.

We implemented a few years ago an algorithm to calculate in real time the solution of this equation, and a computer program was developed. The results of these simulations (in this case: synthesizations) were unexpectedly good. The decay of the sound was particularly natural (as you can hear from the sound examples), though yet perfectly periodic and reproducible. This can indicate that the law of decay with frequency is almost qualitatively correct.
Also waveguides were equipped with some trick to achieve a decay time decreasing with frequency. The problem was that any implementation we could hear gave results not so natural. It was not by chance, as you can see in a while.

To obtain a decay variable with frequency, the idea is to put a gain variable with frequency, namely a low pass filter instead of a simple gain. In this way, higher frequencies are damped faster than lower ones (because the loop gain is lower).

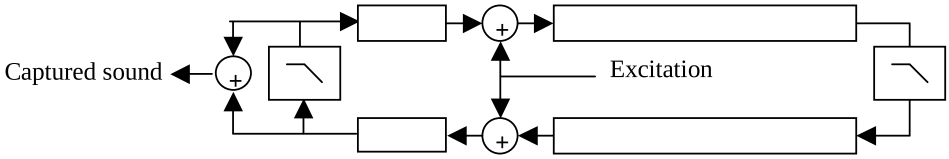

fig. 11.

Unfortunately, things are not so easy. One can ask: how has to be made this filter in order to match the behavior of the dissipative term we put into the equation? We can consider our distributed dissipative mechanism as an infinitesimal filter, then we can commute it moving it till the extremities, in the typical "concentrationist" fashion of waveguides.

The problem is that the resulting concentrated filter is transcendent, not-polynomial. It is, so speaking, not-concentrable in principle. We can get an idea of how things are, using a raw short-hand filter estimation. You can easily compute how much is the required gain as a function of the frequency in order to fit the above mentioned, inverse-square law, relationship:

$$\frac{\tau_n}{\tau_1} = \frac{1}{n^2}$$

Neglecting dispersion, this law is simply:

$G(\omega) \approx e^{-k\cdot\frac{\omega}{\omega_1}}$ where ω_1 is the fundamental pulsation. Considering it as the modulus of a transfer function, it is easily seen that it is not polynomial.

Here is a bilogarithmic plot of such a function.

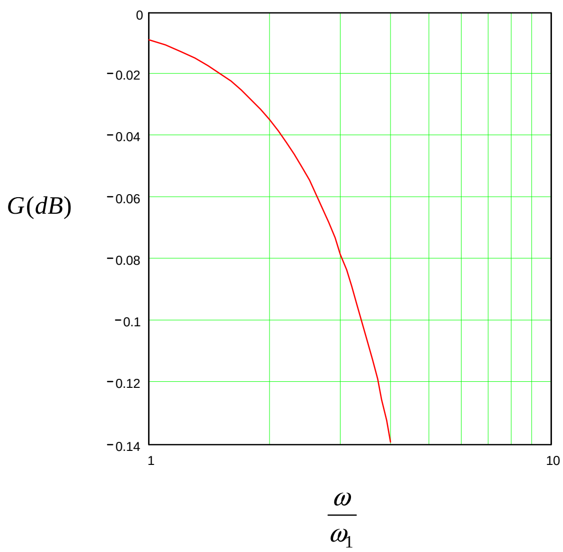

fig. 12.

One could think that such a curve could be piecewise fitted, using a limited (discrete and finite) number of poles. Of course it is in principle possible, even if not so easy, because any analytic function can be locally expressed as a Laurent's integer series. But there is another remark to be noted: the main problem is not, as you can expect at first glance, the gain approximation (which is not so critical, because of the great deal of

superimposing phenomena in the reality). The main problem is instead the phase approximation. The phase is responsible of the dispersion phenomenon, because a phase change acts in effect as a delay change, and things go exactly like the phase velocity of such partial was changed by the phase change. Any approximation in the phase characteristic will result in an error in the partial frequency, introducing some amount of inharmonicity. Now, to our perception, this is relevant to a great deal. To such an extent, that this was the main problem of our previous model, based in the direct numerical solution of the wave equation. Dissipation, as a matter of fact, does introduce dispersion (like stiffness does), but in a particular way. When dispersion is high, a fast decay makes inharmonicity less perceivable. When decay is very fast, the mode is not oscillating. ("cancelled modes"). Well, using only a finite number of poles you cannot cancel modes: any mode oscillates, even when it is quickly damped. Of course, if the time constant becomes significantly shorter than the period of the partial, the oscillations become quickly negligible, so that they cannot be perceived. But this moves towards higher frequencies the limit of cancelled modes, being this limit a function of the loudness, instead of being an intrinsic feature of the mode.
I never made an attempt to implement an approximation of this concentrated filter in order to evaluate perceptual results, mainly because in the meanwhile I choose, for all the abovementioned reasons, to develop another approach to the problem: modal synthesis. This approach will be introduced after a brief discussion of the method of direct solution.

A direct solution of the wave equation.

We will not show here the method we used to integrate the equation. A deep exposition can be found in the references. The method is based in a sinus transform decomposition of the spatial wave shape, plus an algorithm to integrate the motion in time.
The hardest problem to cope with – when numerically solving the wave equation – is due to the circumstance that damping is very low. To integrate the motion, high-order rules (as Runge-Kutta, for instance) are completely useless, because they introduce a "numerical damping" superimposed to the physical damping, of several orders of magnitude higher. Strings sounds are instead close to be everlasting. For this reason, we used a variation of a very ancient rule, the "Stoermer rule", which was used to compute the motion of conservative systems from the beginning of the XX century.
The main problem with this rule, when applied to a string, is that it shows small errors whose effects are not in the amplitude domain, but instead in the frequency domain: calculations slightly shift frequencies toward sub-harmonic ratios. This kind of inharmonicy is perceptually particularly boring, because gives to the sound a "pot" or " tin" timbre. To cope with these errors, small in magnitude, but dramatic in perceptual effects, the only solution is to over-sample in time, i.e. to compute smaller and smaller motions. We where forced to go to an "x 4" over-sampling with respect to the classic, CD-quality, 44.1 KHz sampling frequency in order to avoid these frequency shifts should be noticeable. Taking into account that our model was limited to only 16 partials, the sampling to the cutoff frequency ratio results even more dramatic. Moreover, the model intrinsically involves the computation of a vector-matrix product, whose dimensions are the number of partials you need.
So – even if the model gave perceptually good results – its computational complexity grows with the square of the desired number of partials, plus some extra cost due to the over-sampling, which has to follow the cut-off frequency (i.e. the frequency of the highest partial).
For this reason, after several unprofitable attempts to find some simplification in computations, I moved toward the adoption of a modal method.
The model, and the corresponding computer program, was used by the composer Michelangelo Lupone in two pieces: "Metal string" for string quartet and tape (Rome, Kronos Quartet, 1997), and "Canto di Madre", for string algorithm and female voice (Radio Vaticana, 1999).

Modal synthesis.

By means of the following approach to modal synthesis, you let each mode evolve as it happens in conservative systems by calculating the effect of excitation or dissipation as summing perturbations of this motion.

The main idea is as follows. Let's decompose the acceleration of the string in three components, so that the corresponding equation is also decomposed in three parts:

$$\textbf{(5)}\quad \frac{\partial^2}{\partial t^2} y(z,t) = \frac{\partial^2}{\partial t^2} y^c(z,t) + \frac{\partial^2}{\partial t^2} y^d(z,t) + \frac{\partial^2}{\partial t^2} y^e(z,t)$$

where the subscripts *c*, *d*, and *e* stand for "conservative", "dissipative", and "excitative".
We can write:

$$\frac{\partial^2}{\partial t^2} y^c(z,t) = c^2 \cdot \frac{\partial^2}{\partial z^2} y^c(z,t)$$

$$\frac{\partial^2}{\partial t^2} y^d(z,t) = S_i \cdot \frac{\partial}{\partial t}\frac{\partial^2}{\partial z^2} y^d(z,t)$$

$\frac{\partial^2}{\partial t^2} y^e(z,t) = F(z,t)$ where $F(z,t)$ is some excitation function, expressed in term of acceleration.

The system starts at t_0 with a known pair of conjugate variables, which are the initial conditions, common to the three equations. Let them be the displacement and the speed:
$(y(z,t_0), y'(z,t_0))$
From linearity, we have:

$$(y(z,t_0+\delta t), \frac{\partial}{\partial t} y(z,t_0+\delta t)) = (y_c(z,t_0+\delta t), \frac{\partial}{\partial t} y_c(z,t_0+\delta t)) + (y_d(z,t_0+\delta t), \frac{\partial}{\partial t} y_d(z,t_0+\delta t)) +$$
$$+ (y_e(z,t_0+\delta t), \frac{\partial}{\partial t} y_e(z,t_0+\delta t))$$

These solutions can be expressed as a linear combination of modes:

$$y(z,t) = \sum_n a_n \cdot y_n(z,t) = \sum_n Y_n(t) \cdot \sin(k_n \cdot z)$$

$$y'(z,t) = \sum_n \frac{\partial Y_n(t)}{\partial t} \cdot \sin(k_n \cdot z) \text{ where } k_n = n \cdot \frac{\pi}{L}$$

We can limit the summation over the number of partials we are interested in, or to the "canceling limit", or to the Nyquist frequency, or to any perceptually reasonable quantity, whatever comes first.
We are looking for new values, starting from the knowledge of the previous ones:
$(\, y(z,t_0)\,,\, y'(z,\delta t)\,) \rightarrow (\, y(z,t_0+\delta t)\,,\, y'(z,t+\delta t))$
We can put $t_0 = 0$, without loss of generality.
As to the first, conservative equation, the solution can be expressed in this way:

$$y_n^c(z,t) = A_n \cdot \cos(\omega_n \cdot t + \phi_n) \cdot \sin(k_n \cdot z)$$

$$y'^c_n(z,t) = -A_n \cdot \sin(\omega_n \cdot t + \phi_n) \cdot \omega_n \cdot \sin(k_n \cdot z)$$

The initial conditions being:

$$y_n^c(z,0) = A_n \cdot \cos(\phi_n) \cdot \sin(k_n \cdot z)$$

$$y'^c_n(z,0) = -A_n \cdot \sin(\phi_n) \cdot \omega_n \cdot \sin(k_n \cdot z)$$

So that, calling $Y_o^c = A_n \cdot \cos(\phi_n)$ and $V_0^c = -A_n \cdot \sin(\phi_n) \cdot \omega_n$, we can write:

$$y_n^c(z,t) = (\cos(\omega_n \cdot t) \cdot Y_o^c + \sin(\omega_n \cdot t) \cdot \frac{V_0^c}{\omega_n}) \cdot \sin(k_n \cdot z)$$

$$y'^c_n(z,t) = (-Y_0^c \cdot \sin(\omega_n \cdot t) \cdot \omega_n + V_0^c \cdot \cos(\omega_n \cdot t)) \cdot \sin(k_n \cdot z)$$

As a matter of fact, to compute the evolution of the conservative system, you have to know only the two (mutually related) quantities $\cos(\omega_n \cdot t)$ and $\sin(\omega_n \cdot t)$.

As to the second (damped) equation, the solution is quite straightforward:

$$\frac{\partial^2}{\partial t^2} y_d(z,t) = S_i \cdot \frac{\partial}{\partial t}\frac{\partial^2}{\partial z^2} y_d(z,t)$$

$$y_n^d(z,t) = ((1-e^{-S_i \cdot k_n^2 \cdot t}) \cdot \frac{V_n^o}{S_i \cdot k_n^2} + Y_n^o) \cdot \sin(k_n \cdot z)$$

$$y'^d_n(z,t) = V_n^o \cdot e^{-S_i \cdot k_n^2 \cdot t} \cdot \sin(k_n \cdot z)$$

As to the third equation, the task of calculating its contribution is easy, provided that you are able to express the excitation in terms of modes:

$$F(z,t) = \sum_n F_n(t) \cdot \sin(k_n \cdot z)$$

It is useful to express the general solution of our equation in terms of the sinus transform of its shape, truncated to some finite number of modes:

$$y(z,t) = \sum_n Y_n^o(t) \cdot \sin(k_n \cdot z)$$

The sound of the string can be taken as the speed of variation of the spatial derivative of the shape in the origin (supposed to be at the bridge):

$$snd(t) = \frac{\partial}{\partial t}\frac{\partial}{\partial z} y(0,t) = \sum_n k_n \cdot \frac{\partial}{\partial t} Y_n^o(t)$$

$$snd_n(t) = k_n \cdot \frac{\partial}{\partial t} Y_n^o(t)$$

Now the various contributions to the overall coefficient can be summed-up:

$$Y_n^o(t) = Y_n^c(t) + Y_n^d(t) + Y_n^e(t)$$

$$V_n^o(t) = V_n^o(t) + V_n^c(t) + V_n^e(t)$$

$$Y_n^o(t) = (\cos(\omega_n \cdot t) \cdot Y_n^o(0) + \sin(\omega_n \cdot t) \cdot \frac{V_n^o(0)}{\omega_n}) + (V_n^o(0) \cdot \frac{(1-e^{-S_i \cdot k_n^2 \cdot t})}{S_i \cdot k_n^2} + Y_n^o(0) + Y_n^e(t)$$

$$V_n^o(t) = (-Y_n^o(0) \cdot \sin(\omega_n \cdot t) \cdot \omega_n + V_n^o(0) \cdot \cos(\omega_n \cdot t)) + V_n^o(0) \cdot e^{-S_i \cdot k_n^2 \cdot t} + V_n^e(t)$$

So that the sound is:

$$snd_n(t) = k_n \cdot (-\sin(\omega_n \cdot t) \cdot \omega_n \cdot Y_n^o + \cos(\omega_n \cdot t) \cdot V_n^o + V_n^o \cdot e^{-S_i \cdot k_n \cdot t} + \frac{\partial}{\partial t} Y_n^e(t))$$

Algorithm for real time

We see that the computation of a mode contribution to the overall sound requires only few multiplications and sums, plus the knowledge of the quantities $\sin(\omega_n \cdot \delta t)$ and $\cos(\omega_n \cdot \delta t)$, where δt is the sampling time.

The evolution of the system requires the knowledge of this further quantity: $e^{-S_i \cdot k_n^2 \cdot \delta t}$. In order to avoid the use of divisions (which are expensive in some system), and of real-time trigonometry, we can make use of preloaded tables of these quantities:

$$\frac{\sin(\omega_n \cdot \delta t)}{\omega_n}, \ \cos(\omega_n \cdot \delta t)$$

Working on one cent basis (which is considered the minimum for microtonal purposes), this means 1200 values per octave per quantity. To cope with the entire audible spectrum 20Hz ÷ 20 KHz, we must deal with roughly 10 octaves, which brings to a couple of tables of 12,000 floating point values each. 24,000 values are today a reasonable quantity whatever system you will consider. In a PC, they can fit into a CPU cache even in the full precision of ten bytes per float.

Also the exponential term should be pre-calculated, in order to avoid real-time exponentials, but on a different basis. It depends on the damping friction coefficient S_i and on the wave number k_n. If you are coping with pitch variations due to the length variations, the latter depends on the string length, so that it can be modulated by any fundamental variation, together with the sinus-cosinusoidal terms. Remembering however that this term is responsible of the decay of partials, and of the dispersion, it comes quite natural to thing that a great deal of precision is not needed for this term. It can be thus put into a relatively loose table having the quantity $S_i \cdot k_n^2$ as entry, and then linearly interpolating to find intermediate values.

Conclusions.

Following the abovementioned guidelines, it is possible to implement a computer program capable of simulating strings in real time. Accepting reasonable simplifying hypothesis, some time-varying, non-linear feature can be introduced into the model, in order to investigate their perceptual importance. The results of the previous approach, based on a direct solution of the motion, encourage these efforts, because it is now a well-stated fact that such instruments can bring to results which are interesting from a musical point of view, representing thus a true improvement with respect to the previous synthesis techniques (namely: FM, wavetable and physical modeling based on a more or less sophisticated delay-line based model).
Moreover, such approach naturally brings to a better and deeper understanding of the operation of actual, acoustical instruments, suggesting thus at least new ways of using these instruments, or the implementation of new electro-acoustic instruments. The former was the case of "metal string", the musical piece for string quartet and magnetic tape by Michelangelo Lupone, in which the timbre framework of the piece was inspired by the algorithm used for the production of the tape part. The new performing techniques here introduced were previously experimented and conceived using the model, checking their perceptual results.
At the epoch (July 2002) in which this paper was written, the implementation work is not yet complete, but hopefully the first results will be shown at the workshop.

References.

[1] Raman, C.V. (Ramaseshan S. Editor), "Scientific Papers of C.V. Raman: Acoustics" 1989. *MIT Press.*
[2] Cremer L., Allen J.S., "The Physics of the Violin" 1985. *MIT Press.*
[3] Hiller H., Ruiz P., 1971. "Synthesizing Musical Sounds by Solving the Wave Equation for Vibrating Objects: Part 1" - "Part 2" *JAES* 12 July (6), Agust (7)
[4] Schelleng J.C. "The Bowed String and the Player" 1973. *JASA* 53 (1).
[5] McIntyre M.E. et al., "Fundamentals of Bowed String Dynamics", 1979. Acustica 43.
[6] Schumacher R.T., "Self-Sustained Oscillations of the Bowed String." 1979. Acustica 43.
[7] McIntyre M.E. et al., "Aperiodicity in Bowed String Motion", 1981. Acustica 49.
[8] McIntyre M.E., Schumacher R.T., Woodhouse J. "On the Oscillations of Musical Instruments", 1983. *JASA* 74 (5)
[9] Karplus, K., Strong, A. 1983. "Digital Synthesis of Plucked String and Drum Timbres." *CMJ* 7,2 pp.43-55
[10] Jaffe, D.A., Smith, J.O. III "Extension of the Karplus-Strong Plucked-String Algorithm." Summer 1983. *CMJ* 7,2
[11] Chafe, C. "Simulating Performance on a Bowed Instrument." 1989 *Current Directions in Computer Music Research* - Edited by Matew, M. and Pierce, J. MIT Press.
[12] Woodhouse, J., 1992. "Physical Modeling of Bowed Strings." *CMJ*, 16,4, pp. 43-56
[13] Smith, J.O. III 1996. "Physical Modeling Synthesis Update." *CMJ*, 20,2, Summer 1996 pp.44-56
[14] Smith, J.O. III, 1996. "Discrete Time Modeling of Acoustic Systems with Applications to Sound Synthesis of Musical Instruments." *Proceedings of the Nordic Acoustical Meeting*, Helsinki, June 12-14
[15] Välimäki V., Huopaniemi J., Karjalinen M., Jánosy Z. "Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis" 1996. *JAES* 44 (5)
[16] Mastropietro A., "Ritratto di musicista da sperimentatore: Michelangelo Lupone", Sonus n. 18, Sonus Edizioni Musicali, Potenza 1998.
[17] Lupone M., "Spazio, arco e metallo" in *Musica tecnologia interazioni*, Quaderni della Civica Scuola di Musica, anno XIV N. 26 – Comune di Milano, 1999.
[18] Marco Palumbi, Lorenzo Seno, "Physical modeling by directly solving wave PDE", ICMC 99 – November 1999 - Peking, China.