SIGNAL PROCESSING IV: Theories and Applications
J.L. Lacoume, A. Chehikian, N. Martin, and J. Malbos (Editors)
Elsevier Science Publishers B.V. (North-Holland)

DYNAMIC CONTROL OF FIR FILTERS FOR SOUND SYNTHESIS

Giovanni DE POLI

C.S.C. Dept. Electronics and Informatics
Via Gradenigo 6A
35131 PADOVA Italy

Aldo PICCIALLI

Dept. of Physics
Mostra d'Oltremare Pad. 19-20
80100 NAPOLI Italy

The global spectral envelope is particularly important in the perception of sound. The use of a parallel bank of linear phase FIR filters for a versatle and high quality acoustic synthesis is suggested. The dynamic parameter control techniques (center frequency, bandwidth, amplitude and shape) starting from lowpass prototype filters are studied. Finally, two implementation models with different control modes are presented.

1. INTRODUCTION

The importance of global spectral envelope of sounds in timbre perception is well known. Many musical instruments, including the human voice, can be modelled as resonating cavities excited by an acoustic waveform.The sound is therefore marked by the energy concentration in certain areas of the spectrum called formants. Thus digital filters having a suitable frequency response, can be successfully used for sound synthesis. The features of the filters are not stationary but time varying. To make use simpler it is advisable to formulate the model in such a way that the parameters that influence the sound can be separately controlled. In our case the perceptually significant parameters are those of the formants.

Two main types of models are used, cascade and parallel [1]. In the cascade model, a change of a formant parameter (eg. amplitude) often influences in an unexpected way the other formants. An easier control is obtained in the parallel models where interferences are much less. The use of a 2nd order IIR filter for each parallel branch presents the inconvenient that formant spectra often do not sum up with the desired phase relationships when the parameters vary. Thus, disturbing interferences and cancelations occur in the spectrum which are not easy to avoid.

To overcome this problem we suggest the use of linear phase FIR filters. The model here considered consists of a parallel structure of M bandpass linear phase FIR filters, where M = number of formants, centered on the resonance frequency f_c of each formant. The convolution of the total impulse response with a train of impulses supplies the output signal. The convolution is computationally efficient due to the null samples between one impulse and the next. It is interesting to note that this model is related to the idea of using sums of elementary waveforms, traslated in time and frequency domains, to produce a varying sound. This idea is implicit in various approaches to suond synthesis [2-12]. This work illustarates the dynamic control techniques of FIR filters for a versatile and high quality acoustic synthesis.

2. FORMANT SYNTHESIS BY MEANS OF FIR FILTERS

The model of sound production source and linear filter implies that the signal be defined by $s(n) = e(n) \times g(n)$, that is, the convolution of an input signal $e(n)$ and the impulse response $g(n)$ of the filter. In our model we approximate $s(n)$ with

$$(1) \quad r(n) = u(n) \times h(n) = \sum_{1}^{M} {}_j\, u(n) \times h_j(n)$$

where $u(n)$ is a quasi periodic, unit impulse sequence; $h(n)$ is the overal response of the filter bank, and $h_j(n)$ the reponse of the j-th parallel branch.

The hypothesis of a quasi periodic excitation results to be reasonable for the nature of many types of sounds. The unit impulse hypothesis is justified by the possibility of blending into the impulse response different shapes and amplitudes of the input. Furthermore, these hypothesis make the implementation of the model particularly efficient. The spectral envelope of $r(n)$ must be perceptively a good approximation of the one of $s(n)$. Therefore, the spectral envelope of the input and the transformation is globaly approximated. The phase linearity and parallel structure hypothesis helps to supply an indipendant and perceptively significant control of the sound parameters. Note that natural systems are generally not linear phase. But imposing a linear phase can resolve many problems in dynamically controlling the parameters. On the other hand, it's not very damaging being the spectral envelope much more important than the phase.

3. CONTROLLING THE PARAMETERS OF A FORMANT

We have to design a linear phase FIR filter for each formant, in which central frequency, bandwidth, amplitude and shape should be dynamically controlled. We suggest to start off from some low pass FIR filter prototypes and changing their features in a dynamic manner, for major flexibility, by means of appropriate transformations. Here, to simplify we consider non-causal filters, symetric to zero, with an odd length. The corresponding causal version is the one suitably time delayed.

Let $h_{LP}(n)$ be the impulse response of the low pass filter prototype; then equivalent band-pass version $h_{BP}(n)$ will have the following impulse response

$$(2) \qquad h_{BP}(n) = h_{LP}(n) \cdot \cos(2\pi f_c n) \qquad |n| \le \frac{N-1}{2}$$

where f_c is the normalized center frequency of the formant region. Being BW the bandwidth of h_{LP}, the bandwidth of h_{BP} will be 2·BW if BW>f_c; otherwise it will be f_c + BW.

The bandwith control can be obtained in the following way. From the theorem of scale change of the Fourier tranform we have:

$$(3) \qquad F(\, h(b \cdot t)\,) = 1/b \cdot H(f/b)$$

where F(·) is the Fourier transform. Since a reduction of the temporal scale determines an inverse widening of the formant's bandwidth. Let b be the bandwidth increment (or decrement) which is to be obtained, the ratio between the desired bandwidth and 2·BW. The impulse response of the low pass filter which will supply the desired bandwith becomes:

$$(4) \qquad h'_{LP}(n) = h_{LP}(b \cdot n)$$

If the $h_{LP}(n)$ samples are stored in a table, the $h'_{LP}(n)$ impulse response can be obtained by reading the table scanned with a step value b. The frequency amplitude of the formant will deminish by factor B. It is advisable to allow for this effect in the final sum of the various formants. For non-integer b values the $h'_{LP}(n)$ coefficients are obtained from the table with a suitable interpolation.

It is often useful to have more detailed tables containing various intermediate values between the initial values. This implies the designing of FIR filters for higher sampling rates and operating with multi-rate systems. Being these low pass filters, the added values can be normally obtained by linear interpolation.

The shape control of each formant is obtained by designing the prototype filter in a suitable manner or by modification techniques that will be stated further on.

The amplitude control of the various formants is performed when we sum the various parallel branches. Consider the prototype amplitude normalized so that $H_{LP}(0) = 1$. Let a_k be the amplitude of the k-th formant and

$$(5) \qquad h_k(n) = h_{LP}(b_k \cdot n) \cdot \cos(2\pi f_{ck} n)$$

the impulse response of the k-th parallel branch, obtained by means of the above mentioned techniques. The global impulse response of the stated system, composed by M formants, will be

$$(6) \qquad h(n) = \sum_{k=1}^{M} c_k \, h_k(n)$$

where $c_k = a_k \cdot b_k$ to consider the band variation.

The dynamic control of the system is ensured by the updating of the parameters a_k, b_k, f_{ck}. It is advisable that these parameters be constant in the construction $h_k(n)$ thus not hampering the filter's operations. Therefore, the updating must be performed in a synchronised manner to the pitch period or by one of its multiples and only for the $h_k(n)$ that is about to start. It can be observed that normally the filter's length is longer than the pitch period. Therefore, close to a variation, one impulse response according to the former parameters overlaps one impulse response according to the latter parameters. Thus we obtain a smooth interpolation between the parameters and a good dynamic behaviour.

4. IMPLEMENTATION OF THE SUGGESTED MODEL

To implement convolution two different models are suggested to which two different control modes of formant amplitude apply. Often it is useful to destinguish the overal amplitude envelope of the sound from the relative formant amplitude. The former may vary very rapidly in many cases while the latter varies more slowly. This difference has its operative importance as it corresponds to a common thought among the users of sound synthesis techniques. We will call A(n) the overal amplitude envelope and a'_k the relative amplitude of the k-th formant. The overal impulse response will be once again given by (6) but in which $c_k = a'_k \cdot b_k$. The overal amplitude will be considered after the convolution with the excitation. The synthesised signal will be

$$(7) \qquad r(n) = A(n) \cdot [\, h(n) \times u(n)\,]$$

The first implementation model which we will call global realizes this expression. It firstly calculates the overal response h(n) of the system and then makes the convolution. The second model, called additive, performs the convolution for each formant independantly then its sums the various results. The amplitude control is performed on each parellel branch. The following expression is calculated:

$$(8) \quad r(n) = \sum_{k=1}^{M} C_k(n) \cdot [h_k(n) \cdot u(n)]$$

Both models first calculate the $h_k(n)$ and thus do not differ for their control modes of the formant center frequencies and bandwidths. They differ in the formant amplitude control. The first model has a low updating parameter rate and performs one convolution. On the other hand intermodulation may occur between audio signal and the parameter control. The second model allows major versatility and control flexibility, due to the fact that we can define continuous amplitude envelope for each formant. It requires a larger amount of data for this control and furthermore M convolutions. Normally in this case the overal envelope is absorbed in those of the single formants.

It can be observed from (5) that the computation of $h_k(n)$ can be accomplished as a product of two tables scanned by a non-unit step. The first contains the chosen prototype, the second the cosine and it is cyclicly scanned with a step proportional to f_{ck}. The $h_k(n)$ of each formant are synchronised with respect to the central point thus maintaining the overal phase linearity. The symmetry can be exploited by calculating only half $h_k(n)$ starting from the central point, symmetrically copying the remaining. This operation requires of a suitable buffer to perform these calculations before convolution, that is before summing with other prior impulse responses yet to be terminated. Thus, a buffer is necessary for each different impulse response that occurs at the same time. Furthermore, h(n) must be ready L instants before the impulse, where 2L + 1 is the maximum length of h(n).

The excitation impulse frequency determines the pitch of the resulting sound. Therefore, apart from matching the melody it must be carefully controlled in the transitions. In the steady state it will vary around the medium value periodically and in an aleatory manner (often 1/f type). These variations, as stated, allow the perception of the spectral envelope and the fusion of the sound's components. It can be observed that the excitation impulses are only a temporal reference of the impulse responses and don't need to exist physically. Therefore, the impulses can be fixed in intermediate instants between two effective signal samples. Being h_{LP} and the cosine value defined by the tables, just displace the central point reference, which will no longer coincide with a sampling instant. This allows to avoid the pitch period quantisation, which can be troubling particularly at low sampling rate.

5. FORMANT SHAPE CONTROL

The formant shape is characterised by the passband and by the adjacent minor energy area, which we will call skirt. The amplitudes of the two parts often have a different ratio and transition occurs in a differentiated mode. Therefore, one prototype is not sufficient. A first formant shape control is obtained by designing various prototypes suitable for the various circumstances. Standard methods such as window techniques, and minimax approximation can be used. Analytic formulae can also be employed, some of which are used for window design [13]. Figure 1 shows a spectrum with four formants produced using a simple gaussian window as prototype.

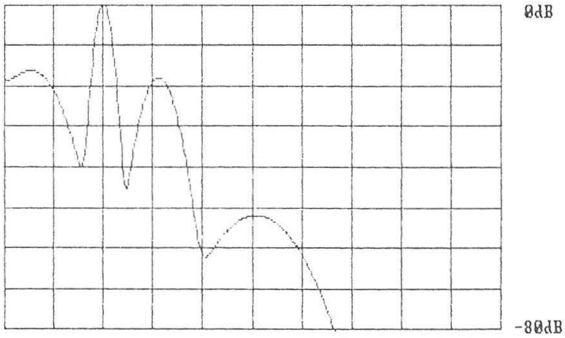

Figure 1

Often more continuous variations and independant control of various parts of the shape are required, without having an excessive number of prototypes. Due to this aim some dynamic transformation techniques of the prototype were used, partly derived from currently employed musical synthesis techniques [14]. Among the most promising we point out: summing of different prototypes; look-up, having a variable step, of the table containing the prototype; amplitude modulation; non-linear distortion. An example of non linear distortion of h_{LP} is given in figure 2.

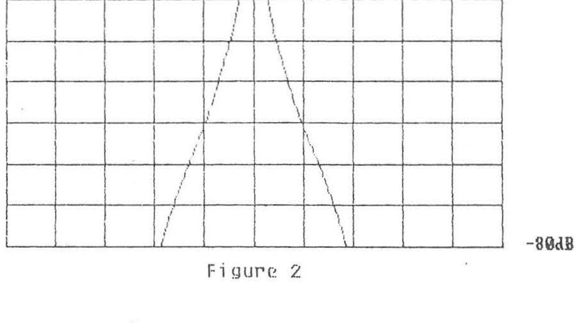

Figure 2

Figure 3 shows an example of amplitude modulation applied to a gaussian prototype. The transformed prototype h'_{LP} is given by

$$(9) \quad h'_{LP}(n) = h_{LP}(n) \cdot [1+am \cdot \cos(2\pi f_a n)]$$

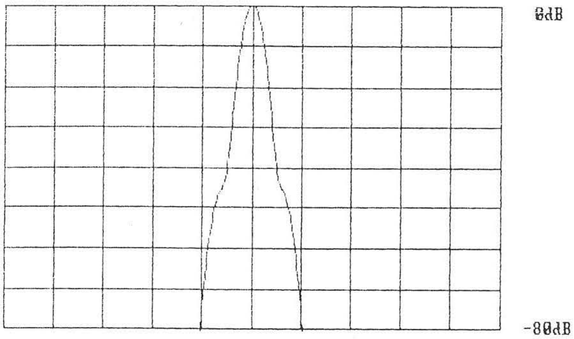

Figure 3

An other useful technique is to phase modulate the shifting cosine in (2), even if it produces distortion in phase linearity. The impulse response of the resulting bandpass filter is

$$(11) \quad h_{BP}(n) = h_{LP}(n) \cdot \cos(2\pi f_c n + ø(n))$$

The figure 4 shows the effect of using a gaussian window as modulating function.

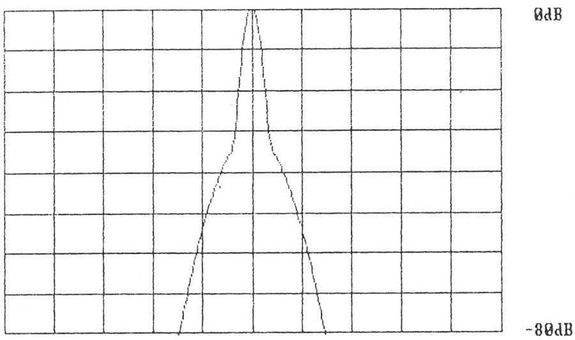

Figure 4

6. RESULTS AND DISCUSSION

We implemented the parallel formant synthesizer (1 channel, 4 formants) in real time on a TMS 32020 microprocessor system and now we are testing its performance. The two implementation models have been realized using various prototype filters. The gaussian window resulted turned out to be one of the best for acoustic resonance simulation. It is stored in a table of 1024 values. Working at a 44 kHz sampling rate , a minimum bandwidth of 40 Hz can be obtained. In the implementation of phase modulation, instead of putting in the table the phase distorting function ø(n), we preferred to use the differences Δø(n), which give the step increments for the cosine table look-up.

In conclusion, the proposed model allows a high quality synthesis of instrumental and vocal sounds, with a reduced number of perceptually relevant control parameters. In particular the global model is computationally more efficient. In the case of slow variation with respect to the impulse response duration, the intermodulation, introduced by the discrete updating of control parameters, is negligible. The additive model is more versatile and can produce a great variety of sounds. More freedom is allowed in parameter control: for example, the pitches of different formants need not be equal. However, the intermodulation problem may occur. Finally, it is sufficient to specify the amplitude envelopes of every formant by defining 8-16 breakpoints and using linear interpolation.

REFERENCES

[1] Holmes J. "Formant Synthesizers: Cascade or Parallel?", Speech Communication, vol. 2, p.251-273, 1983

[2] Gabor D., "Theory of Communication", J. of the IEE, vol 93, p.429, 1946

[3] Gabor D.,"Acoustical quanta and the theory of hearing", Nature,159,1044,p. 591-594, 1947

[4] Roads C., " Granular synthesis of sound", Comp. Music J. vol. 2, n. 2, p. 61-62, 1978.

[5] Cavaliere S., Piccialli A.,"Phase modulation with interpolated time functions: synthesis by formants", Proc. ICMC 86, p. 293-297, 1986.

[6] De Poli G., Piccialli A., "Forme d' onda per la sintesi granulare sincrona", Proc. CIM 88, 1988

[7] Kaegi W., Tempelaars S., "VOSIM: A new sound synthesis system", J. Audio Eng. Soc., vol 26, n.6, 1978

[8] Bass S.C., Goeddel T.W., "The efficient implementation of subtractive synthesis", IEEE Micro, vol. 1,n. 3, p. 24-37, 1981

[9] De Poli G., Longo G.,Mian G.A., "An effective software tool for digital filter design", Proc. IASTED Int Symp. MSPC, p. 237-242, 1986.

[10] Rodet X., Time domain formant wave function synthesis, in: Simon J.G.(ed.), Spoken Language Generatinon and Understanding, (Dordrecht, 1980)

[11] Meyer Y., Jaffard S., Rioul O., "L' analyse par ondelettes", Pour la Science, n. 119, p. 28-37, 1987

[12] Lienard J.,"Speech analysis and reconstruction usiing short-time elementary waveform", Proc. ICASSP 87, p. 948-951, 1987

[13] Harris F.J., "On the use of window for harmonic analysis with discrete Fourier transform", Proc. IEEE, vol. 66, n. 1, p. 51-85, jan. 1978.

[14] De Poli G.,"A tutorial on digital sound synthesis techniques", Comp. Music J., vol. 7, n. 4, P. 8-26, 1983