Chapter 8
Digital Audio Effects

8.1 Introduction

Modern processing and circuit technology has made available a number of methods for processing the acoustic signal covering various requirements. Among the different methods, the term *effect* generally refers to the processing of an existing sound in order to make it more suggestive. According to Verfaille [1], the definition of audio effects is the follows.

Definition 8.1. *Digital audio effects (DAFX)* - Digital audio effects are boxes or software tools with input audio signals or sounds which are modified according to some sound control parameters and deliver output signals or sounds.

The possibilities of processing sound for various purposes are practically infinite and any linear, non-linear, stationary and non-stationary transformation that makes a sound or a set of sounds perceptually different from the original can be considered an *effect*. The effect can be inserted on an instrument or on a set of instruments during execution: in this case we will talk, as for the numerical filtering methods studied in Chapter 4, about on-line processing. In postproduction, processing can be done with algorithms operating in non-real time and also in batch mode. In such cases there is a greater chance of processing. Some algorithms, such as the compression/expansion of the time scale (which we will analyze later), are batch type by definition and have no or difficult possibility to be realized online.

In this chapter we present some DASP methodologies for the realization of the most common effects and, given the vastness of the topic that would require other spaces, we want to emphasize, as in other parts of this book, the general and methodological approach.

In the first part we introduce the concept of the simulation of the listening environment illustrating the main methodologies for the realization of artificial reverberators. The second part presents the techniques of dynamic processing of the audio signal. In particular, the compression-expansion techniques and their use in acoustic reproduction are presented. Some effects based on time delay lines modulated variants are, instead, illustrated in the third part of the chapter. In the last part of the chapter the

time-frequency transformations used for time scale change and height translation are introduced. Other effects, such as distortion etc., are not reported for space reasons.

8.2 Room Acoustic Simulation

As is well known, and as previously reported in Chapter 3, that the acoustic environment is a determining part of the subjective quality of listening. Thus, *reverb* strongly characterizes a musical performance. When music is performed in a concert hall, the listener is immersed in a considerable amount of sound reflected from the walls. The perception of these echoes can enrich and make the performance more evocative or, on the contrary, make listening more "tiring" (listening fatigue). For example, in prose the effect of reverberation can decrease the intelligibility of the word, on the contrary a chorus heard in a poorly reverberating environment can be too "dry" and make the performance unattractive.

Recorded music almost always has a certain amount of reverberation due to the environment where the recording is made and/or artificially added in post-production manipulation: *artificial reverberators* are used, therefore, to add reverberation to music recorded in the studio, in movies or to modify the acoustics of a listening room.

For the realization of this effect there is a variety of methodologies almost all based on the use of delay lines appropriately connected to numerical filters implemented according to some criteria of physical modeling of reflection and/or sound propagation.

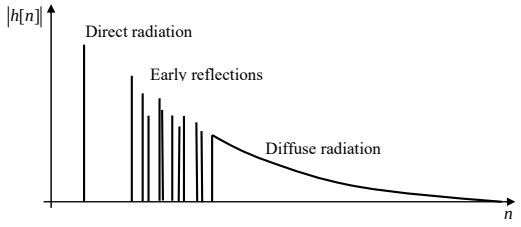

Fig. 8.1 Example of non-linear resistance with piecewise-linear characteristic.

The first work on artificial reverberation dates back to the early 1960s. Schroeder in [2], [3], first proposed the use of *comb* and *all-pass* filters combinations. Since then, a large number of papers have been written on the subject and so far many research contributions have been made in this area. For example, some reverberator models are based on the *Feedback Delay Networks* [9], [11], [13] which represent a multi-channel generalization of Schoreder's models.

As seen in §Chapter 3, in general terms, the impulse response of a listening room can be divided into three parts: direct radiation or *direct signal* (DS); first reflections or *early reflections* (ER); *diffuse radiation* or *subsequent reverberation* (SR) or *late reflections* (LT).

8.2.1 Physical Modeling vs Perceptual Approach

Artificial reverberation can be achieved with different approaches based on even very different philosophies.

8.2.1.1 Convolution with Impulse Response

A first method is to determine the *point-to-point transfer function*(PP-TF) between the source and the listening point. To achieve this goal the reverberated signal is obtained through a convolution operation with an impulse response from a listening room. For each $x_i[n]$ sound source, two impulse responses must be determined for the right and left ear $h_{i,R}[n]$ and $h_{i,L}[n]$ implemented with a FIR filter.

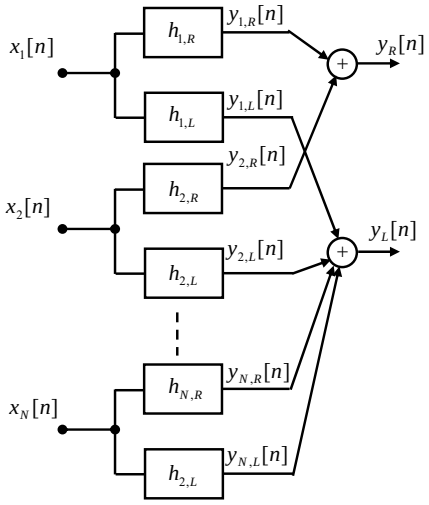

Fig. 8.2 Direct convolution reverberator with *point-to-point transfer function* (PP-TF). Two convolutions must be performed for each source: one for the right ear and the other for the left ear.

The impulse response can be measured in a particular listening room or obtained from a simulator of the type described in §3.

At normal audio sampling rates the impulse response can tens of thousands samples length. In the implementation with FIR filters (2 for each source) TF-PP modeling requires, therefore, rather high computational resources. Another limitation of this method is that it is extremely complex to model the movement of acoustic sources.

For each position of the source (or listener), you need to determine a pair of TF and even if they were previously calculated in memory, a very high memory occupation would be required.

The quality and naturalness of the sound produced with such methodology are however high.

8.2.1.2 Physical modeling

The second approach is based on *physical modeling*: the artificial reverberator is built on a more or less exact model of a real acoustic environment. The listening environ-

ment is modeled with a 3D network (*mesh*) using, for example, a *digital waveguides network* (DWN) or finite element technique. In this case the acoustic response of the room would be available at all points of the space. The listener or source could be moved to any point without changing the underlying model. The approach based on physical modeling allows, moreover, a remarkable naturalness in the positional reconstruction of the 3D sound front.

A rough estimate of the computational cost can be made by simple reasoning. Considering a maximum frequency of 20 kHz at the normal speed of sound in air, the wavelength of the maximum frequency is 17 mm, so we should consider a minimum spatial resolution of about 7.5 mm. For a room of $4 \times 4 \times 3$ m^3 the number of node is about 100 million. For each node it is necessary to consider six directions of propagation and, even in the case of multiplier-less junctions, the number of additions is about 10 for each signal sample. With a sampling frequency of 44.1 kHz we will have 4.41×10^{13} additions which is of the same order of magnitude as an approach based on PP-TF estimation with 3 sources (six convolutions) with a reverberation time of 1s.

8.2.1.3 Perceptual model

A third approach for the realization of an artificial reverberator attempts to reconstruct the acoustic scene considering the perceptual aspects: a reverberant circuit capable of behaving acoustically interesting and perceptually indistinguishable from a natural reverb is synthesized. The objective is to determine an algorithm capable of reproducing the salient characteristics of natural reverberation (reverberation time, reflection density, frequency response, etc.).

This methodology is much more efficient than full physical modeling and the estimation of the PTO. An important aspect of this paradigm is that the perceptual mode is fully parameterizable and easily controllable.

As we will see in this chapter, the perceptual approach allows for the separate implementation of ERs that are usually implemented with a relatively short FIR filter, and LTs that are implemented with appropriately connected comb and all-pass networks.

8.3 Schroeder's Artificial Reverberator

Schroeder was s perhaps one of the first to study reverberators in terms of numerical filters and to propose some architectures that are still widely used today [2], [3], [5]. In the following paragraphs we report some of them.

8.3.1 Schroeder's First Model

The easiest way to simulate an impulse response with exponential decay is to use recursive and all-pass filters §5.2.5.

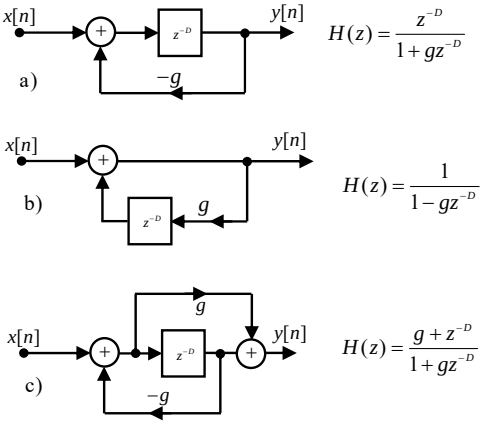

Fig. 8.3 Possible realizations of recursive comb filters (also called plain reverberator): a) Schroeder model; b) Model without delay (see for example[10]); c) All-pass comb filter.

The comb filter, also called *plain reverberator* (PR), is the main element of the reverberator, in fact, this block models the delay with which the reflected wave fronts reach the listener (see Fig. 8.3-a)-b)). The all-pass comb (AP) filter, Fig. 8.3-c) does not model any physical phenomenon but changes the distribution of the peaks of the impulse response that that become "denser". Inserting the AP is just to make the sound more diffuse: the overall effect of the reverb is more realistic.

The reflections obtained with such filters generate, in any case, a rather unnatural sound. Both PR and AP produce a response with impulses that are evenly spaced at a distance equal to the length of the delay line. The comb filter, given its frequency response with resonance and anti-resonance, produces a high coloration of the sound.

The AP filter, while having a flat frequency response, produces a very complex phase distortion. The sound coloration due to the comb and all-pass filters is therefore easily perceptible by a trained listener [7].

A first family of reverberators, suggested by Schroeder in [4], [5], uses two different combinations of PR and AP. Fig. 8.4-a) describes or feedforward structure without any feedback. Fig. 8.4-b) has instead feedback due to PR filters.

In both Schroeder's proposals, a portion of the input signal is mixed directly into the output. The direct sum of the input simulates the proximity of the source to the destination. If the listener, located at a distance r from the source, moves, the reverb level remains constant while the gain of the direct link should increase or decrease with law $1/r^2$ (see §1.7.2). If the listener moves away we will have a point where the reverb level exceeds that of the direct sound and we will have a predominance of the diffuse field (§3.4.5). The coefficient g_0 in Fig. 8.4-b) can then be calculated based on the virtual distance of the source.

However, Schroeder's reverberates require the determination (done more and less empirically) of a number of parameters (gain and delay length). It is very important, as noted by Moorer in [7], to maintain the length of the delay lines prime numebr each other. This choice reduces the possibility of superimposed impulses that could

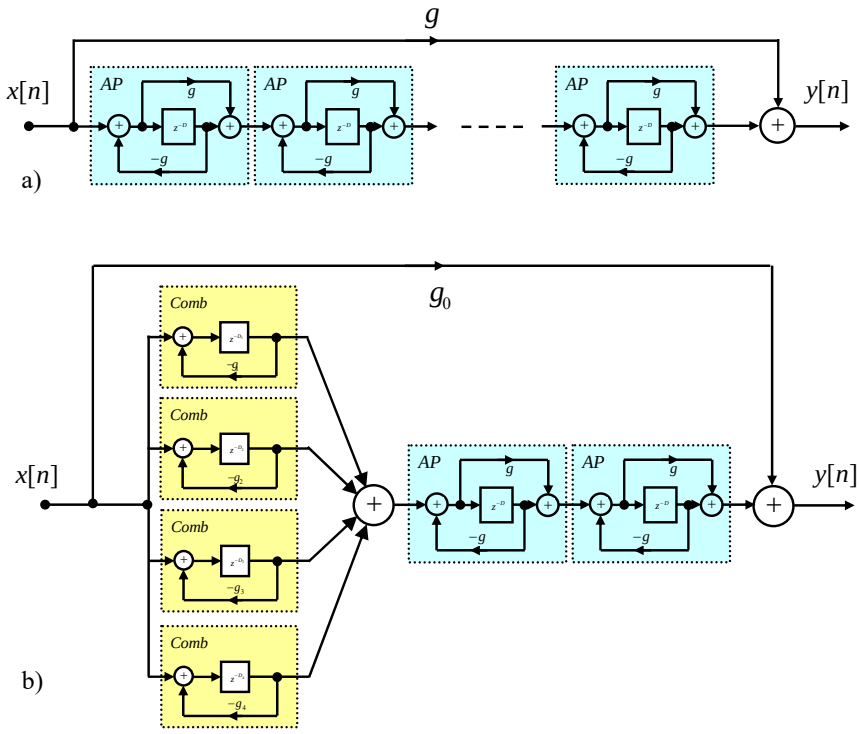

Fig. 8.4 First reverberant structures by Schroeder with comb and all-pass filter combinations suitable for the simulation of late-reflections: a) Cascade allpass filters only; b) Comb and all-pass filters.

generate annoying peaks on the output signal; moreover, it produces a more uniform and dense decay.

8.3.2 The Schroeder-Moorer Model

Again Schroeder in [3] proposed a geometric room simulation model similar to the *ray-tracing* described in §3.6.3.

With this model Moorer calculated the first reflections of the listening room and thus determined the impulse response due to ERs. The proposed model therefore consists of a FIR filter that models the ERs connected to the reverberator in Fig. 8.4-b). The resulting model is shown in Fig. 8.5.

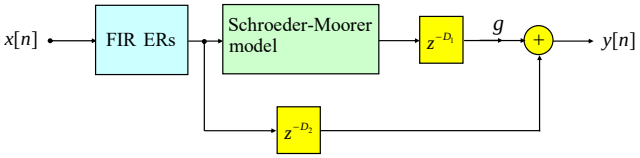

Fig. 8.5 Schroeder-Moorer model with FIR filter for ERs simulation.

8.3.3 The Frequency-Dependent Moorer Model

Schroeder's reverb models in Fig. 8.5 have been and still are widely used. Such models, however, have a not very realistic effect. The main acoustic problems of these models, as pointed out in [7], are:

- The impulse decay response at the beginning is not dense enough and the exponential decay is rather slow. This produces an annoying delay of a few hundred ms.
- The uniformity of the decay curve depends, rather critically, on the choice of parameters (gains and delays) of the base units. Simply changing the length of a delay line can make the sound very "grainy" and unpleasant.
- The reverb tail can be tediously repetitive, due to the imposed periodicity of the delay lines.

To make the effect of Schroeder's reverberator more realistic, Moorer in [7] modifies the basic comb and all-pass structures by inserting simple filters, inside the cycles, that simulate the higher decay at high frequencies due to the propagation of sound in the air and the characteristics of the reflective walls. For example, a $T(z)$ low-pass network function can be inserted in the comb structure in the I-order of type

$$T(z) = \frac{1}{1 + g_1 z^{-D_1}} \tag{8.1}$$

as shown in Fig. 8.6.

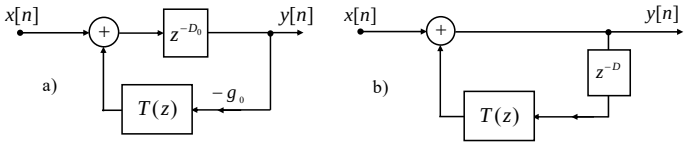

Fig. 8.6 Plain reverberator with TF $T(z)$ inserted within the cycle in order to model the effects of frequency-dependent air attenuation: a) With initial delay D_0; b) With zero initial delay.

The new structure, denoted as *lowpass-feedback comb filter* has a TF defiend as

$$H(z) = \frac{z^{-D_0}}{1 + g_0 T(z) z^{-D_0}} = \frac{z^{-D_0} + g_1 z^{-(D_0 + D_1)}}{1 + g_0 z^{-D_0} + g_1 z^{-D_1}} \tag{8.2}$$

where for the stability $g_0 = k(1-g_0)$, with $0 \leq k < 1$. The presence of the low pass in the feedback causes each replica to disperse more and more, generating a softer and more diffuse reverberator response, in fact each replica is filtered again.

Remark 8.1. To better understand the effect of the $T(z)$ filter in the feedback line consider the reverberator in Fig. 8.6-b), with transfer function equal to

$$H(z) = \frac{1}{1 - g_0 z^{-D} T(z)} \tag{8.3}$$

For $g_0 = 1$ the above equation can be expanded with the geometric series formula. It follows

$$H(z) = 1 + z^{-D}T(z) + \left(z^{-D}T(z)\right)^2 + \left(z^{-D}T(z)\right)^3 + ... \ .$$

Said $t[n]$ the impulse response of the filter $T(z)$, the impulse response $h[n]$ assumes the form

$$h(n) = \delta[n] + t[n-D] + (t*t)[n-2D] + (t*t*t)[n-3D] + ... \ . \tag{8.4}$$

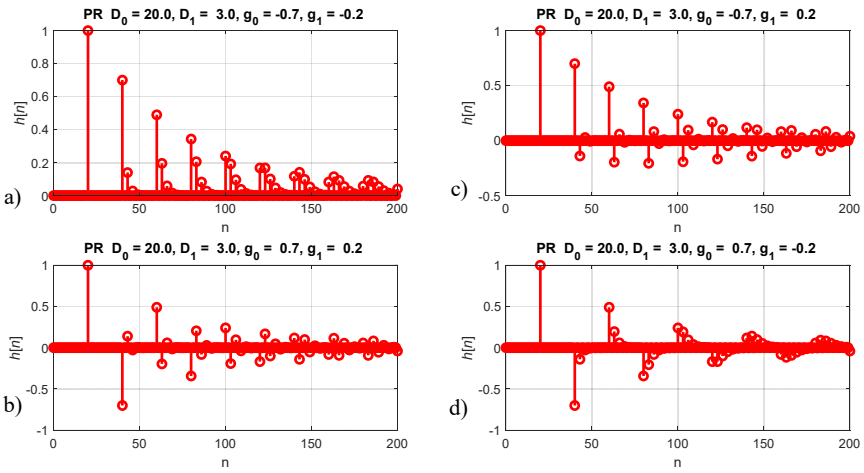

Fig. 8.7 Impulse response for the lowpass-feedback comb filter or plain reverberator with TF (8.2).

From Eqn. (8.4) we can observe that the first echo of the impulse response $h[n]$ for $n = D$ has the form of $t[n]$, for $n = 2D$, instead, it takes the form $t*t$ (which has a longer duration), and so on (see Fig. 8.7). Fig. 8.8, for example, shows the impulse response of the circuit in Figure 9.6-b)for $D = 20$ and $t[n] = \{-1/3, 1/3, -1/3\}$. The density of the impulse response increases, therefore, with time and this behavior is closer to what happens in real istening environments where the number of reflections increases (quadratically) over time.

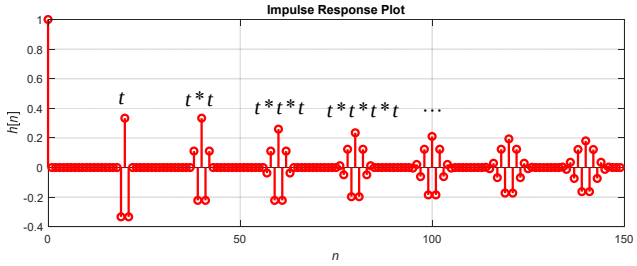

Fig. 8.8 Impulse response from the plain reverberator TF (8.3). The DL with $D = 20$ and the filter $T(z) = -1/3 + 1/3z^{-1} - 1/3z^{-2}$.

Also the basic all-pass structure can be modified by inserting appropriate transfer functions, within the feedforward $F(z)$ and backward $B(z)$ loops, which must be complex conjugated together. This implies that the respective impulse responses are reversed (one is the mirror version of the other).

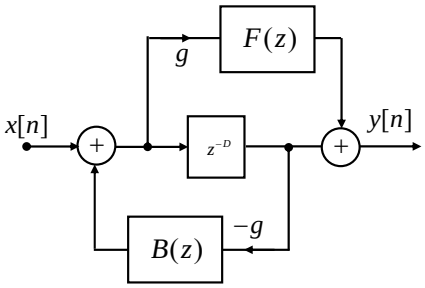

Fig. 8.9 All-pass structure with a network function $B(z)$ and $F(z)$ inserted within the feedback and feedforward cycle respectively.

Usually the $F(z)$ and consequently the $B(z)$, are made with simple FIR filters. This, in fact, makes the problem of stability a simple solution.

8.3.4 Selecting Reverberator Parameters

One of the main problems in making a reverberator is the choice of parameters. In fact, it is not possible to derive certain acoustic characteristics directly from the properties of the filter transfer functions and the lengths of the delay lines. In fact, the perception of reverb quality is not simply related to the typical quantities of network functions (root location, delays, etc.).

As previously observed, one of the (empirical) criteria for the choice of delay lines is to take their prime number lengths from each other. For the choice of the other parameters (for example the coefficients g_1 that control the roll-off of the lowpass-feedback comb filter), as already suggested by Moorer in [7], it is possible to use optimization algorithms with the criterion of minimum squares, a cost function relative to a certain desired response.

Table 8.1 Possible choice of Schroeder reverberator parameters with 6 comb and an 3 all-pass.

	Delay [ms]	25 kHz g_1	50 kHz g_1
Low-pass comb 1	50	0.24	0.46
Low-pass comb 2	56	0.26	0.48
Low-pass comb 3	61	0.28	0.50
Low-pass comb 4	68	0.29	0.52
Low-pass comb 5	72	0.30	0.53
Low-pass comb 6	78	0.32	0.55

	Length [samples]	g
All-pass comb 1	1051	0.7
All-pass comb 2	337	0.7
All-pass comb 3	113	0.7

In Table 8.1 the e parameters of a Schroeder reverberators with 6 parallel lowpass-feedback comb filters with TF (8.2). The parameters g_0 are set as $g_0 = 0.2(1 - g_1)$, and for the 3 all-pass comb lengths equal to prime numbers 113,337 and 1051, have been chosen with $g = 0.7$.

8.4 The Quality of Artificial Reverberation

Before moving on to the study of other more sophisticated architectures to realize artificial reverberators (AR) we see some criteria for the evaluation of their acoustic quality.

As we know the *Quality of a Listening Environment* (QLE) is not a well-defined concept [26]. As we have seen in §3.7, the criteria for defining the goodness of a listening environment is strongly dependent on the type of representation made. In analogy with QLE also the *Artificial Reverberation Quality* (ARQ) cannot be easily defined. In general, however, we can define some subjective criteria, based on perception and supported by some objective measures. In particular the latter, as we will see later, have somehow characterized the choice of more advanced reverberant circuit architectures. It is known that some parameters influence the response of an RA more perceptively. In [22] the authors propose that, in order to be able to optimize the response of an AR, it is necessary to be able to control the following parameters:

FINO QUI

- $T_{60}(f)$ is the *reverberation time* (RT) for each frequency (defined in §3.4.1). For listening rooms it is one of the most used parameters.
- $G_2(f)$ defined as gain for each frequency.
- $C(f)$ defined as *clarity* or ratio between the energy of the impulse response in ER and the energy of the impulse response in LT.
- $\rho(f)$ defined as the interaural correlation coefficient (between the right and left ear).

8.4.1 Energy Decay Curves

As previous explained in Chapter 3, the acoustic characteristics of a room are completely described by the impulse response. In particular, the *energy decay curve* (EDC) proposed by Schroeder and defined §3.4.2 can be used to estimate the RT T_{60}. Ac-

cording with [8] the EDC can be estimated by the backward integration of the squared impulse response using a sliding fixed integration time, $T_0 \approx \frac{1}{5}T_{60}$, using Eqn. (3.16).

So, the EDC can be used also for artificial reverberation. For example, in Fig. 8.10 is reported the echogram and the nomrmalize EDC for the Schroeder reverberator with six low-pass comb and 3 all-pass filter implemented with the parameters in Table 8.1.

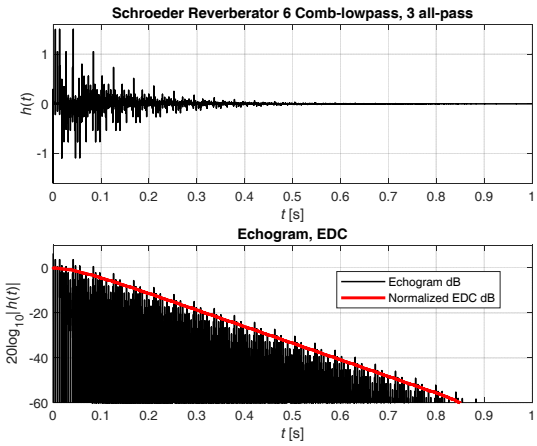

Fig. 8.10 Trend of the Schroeder reverberator impulse response, and the relative echogram, with the low-pass feedback comb parameters proposed by in Table 8.1.

The EDC describes the residual energy of an impulse response from any t moment and normally decreases exponentially with t (i.e. linearly in dB). The RT can be derived from EDC simply by considering the slope of the decay curve (in [dB/s]).

More recently Griesinger [12] and Jot [15] have formulated a variation of EDC(t) by introducing energy decay relief (EDR) defined as

$$\mathrm{EDR}(t,\omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2 . \tag{8.5}$$

The ERD(t,ω), that represents the reverberation decay as a function of time and frequency, can also be used as a qualitative index of artificial reverberators. For example in Fig. 8.11 shown the EDR of the Schroeder reverberator above described.

From a computational point of view it is possible to calculate EDR through *Short Time Fourier Transaform* (STFT) as reported in §3.4.3 (see Eqn.s (3.17) and (3.18)), or with other methods known in literature such as Wigner-Ville distribution, wavelet etc.

Although, as already pointed out, the QRA is not a well-defined concept, there are, however, some objective criteria based on psychophysiological principles of hearing (which are the same that are used for the determination of the QAA). In general, different criteria can be defined for the qualitative assessment of the ER and LR.

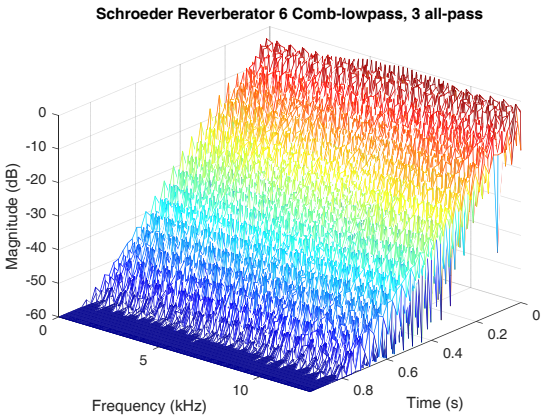

Fig. 8.11 Energy decay relief (EDR) curve of the Schroeder reverberator with the low-pass feedback comb parameters proposed by in Table 8.1.

8.4.2 Characterization of Diffuse Radiation

After the first 60-100 ms, the impulse response measured in a large room can be modeled as a non-stationary Gaussian random process with a time-varying spectral power density [15]. This statistical model assumes a large modal overlap for all frequencies above 100 Hz. For the characterization of diffuse radiation or late reflections (LR), as suggested in [17], the main subjective parameters are:

1. the density of reflections, which in listening rooms increases with the square of time;
2. the similarity between the frequency response of the artificial reverberator and the reference concert hall;
3. the shape of the envelope of the impulse response;
4. the dynamics and the level of distortion;
5. the density of resonant modes (which in listening rooms increases with the square of frequency).

From the previous considerations we can say that LRs can be characterized, starting from the impulse response of the reverberator, only from a perceptual point of view considering the frequency response, the variations of reverberation time with frequency. The main objective parameters are:

1. mode density or frequency density;
2. echo density;
3. unnatural resonances;
4. reverberation time.

For objective parameters it is usual to use the following definitions.

8.4.2.1 Mode Density or Frequency Density

The *frequency density* (or *mode density*) d_f is defined as the number of natural frequencies per Hertz. For example, as reported in [13], for a comb filter you have

$$d_f = DT_s, \quad [1/\text{Hz}]$$

where T_s is the sampling period and D the DL length. It is known, in fact, that a long D comb filter is characterized by $D/2$ resonances between frequencies $[0, 0.5T_s]$, separated by a distance equal to $\Delta f = f_s/D$ (see Fig. 5.5).

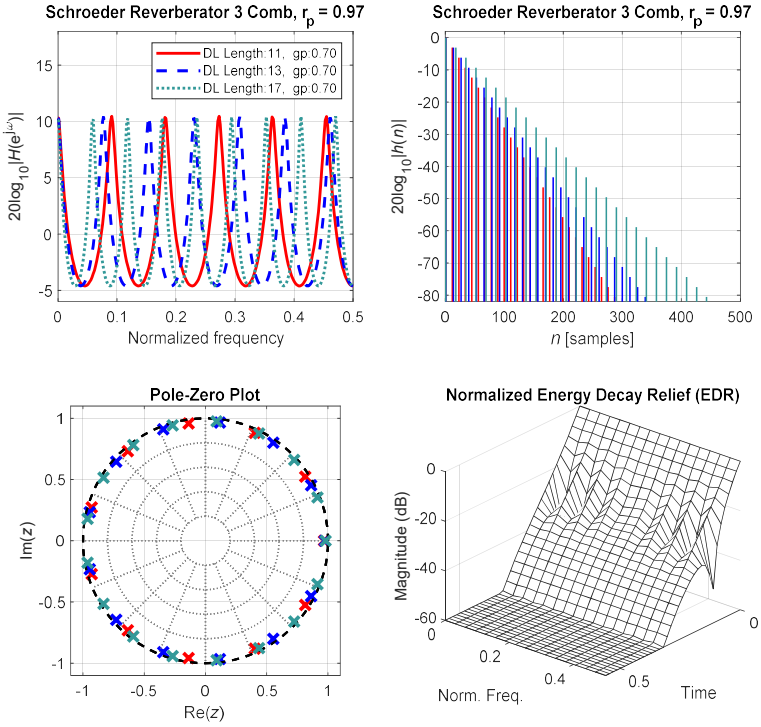

Fig. 8.12 Frequency response of three comb filters in parallel with lengths equal to: 11, 13 and 17. Note that the decay of the comb filters is not uniform: for the same feedback gain $g = 0.7$, the shorter delay line has a higher decay.

The sound of a good reverberator should not have evident colorations; the resonances due to the comb filters should therefore be distributed "quite" uniformely.

Schroeder suggests in [3] that the DL lengths should not be too dissimilar and proposes a maximum ratio between the shorter DL and the longer DL of 1 ./. 1.5 (see Table 8.1). Under these conditions, the frequency density of a Schroeder reverberator with P parallel comb filters, or each of D_p length, is equal to

$$d_f = \sum_{p=1}^{P} D_p T_s = P \cdot \bar{D} \cdot T_s \tag{8.6}$$

where $\bar{D}$ is the average length of the delay lines.

8.4.2.2 Echo density

The echo density is defined as the number of reflections per second. For example, as reported in [13], for a comb filter we have that

$$d_e = \frac{1}{DT_s}, \quad [1/s]$$

and for a parallel bank of P comb filters, we get

$$d_e = \sum_{p=1}^{P} \frac{1}{DT_s} = P\frac{1}{\bar{D}T_s}$$

From the previous expressions we have $P = \sqrt{d_e d_f}$ and $\bar{D} \cdot T_s = \sqrt{d_f / d_e}$.

To obtain a frequency density $d_t = 0.15$ (recommended value in [3]) and an echo density $d_e = 1000$, it is possible to calculate the number of comb filters required $P = \sqrt{0.15 \cdot 1000} \approx 12$ with an average delay line length of $\bar{D}T_S = \sqrt{0.15/1000} \approx 12$ ms.

The echo density value is further increased by cascading a P_A number of all-pass cells (typically 2 or 3). To create a reverberator with sufficient quality, according to Griesinger in [12], an echo density of at least $d_e = 10^4$ and $d_f = 0.45$ (more than 60 comb filters!) is required.

8.4.2.3 Unnatural resonances

In order to avoid unnatural resonances the comb filters should have the same decay rate. So Jot and Chaigne in [13] proposed a simple method to determine the feedback gains of delay lines in order to have uniform decay of all modes.

The same authors suggest that the amplitude of the poles r_p (poles radius on the z-plane) should be calculated according to the criterion

$$r_p = g_p^{1/D_p} = \text{cost} \quad g_p = r_p^{D_p} \tag{8.7}$$

where g_p represents the DL gain. The results are shown in Fig. 8.13).

As already seen in §8.3.2 and illustrated in Fig. 8.14, Moorer has inserted a lowpass filter in the comb filter loop to simulate the frequency-dependent absorption of air and walls. The module of the loop filter, is usually an first order IIR filter with TF: $T(z) = \frac{1}{1-apz^{-1}}$, where a_p depending on the physical characteristics of the absorption. Therefore, the overall TF is

$$H(z) = \frac{1}{1 - g_p T(z) z^{-D}} = \frac{1 - a_p z^{-1}}{1 - a_p z^{-1} - g_p z^{-D}}.$$

The condition (8.7) must therefore be rewritten as

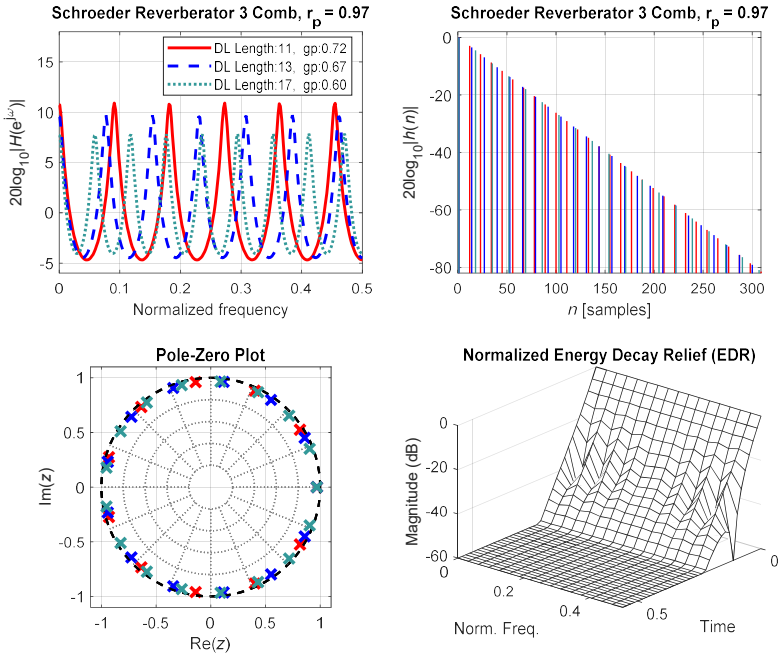

Fig. 8.13 Characteristic curves of a reverberator with three comb filters in parallell with lengths equal to: 11, 13 and 17, with $r_p = 0.97$ in Eqn. (8.7).

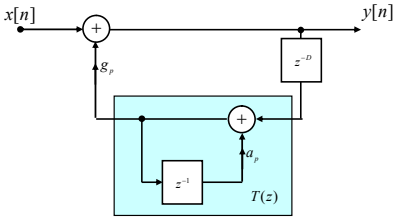

Fig. 8.14 Comb filter with first order IIR low-pass loop filter.

$$r_p(\omega) = \left| g_p T_p(e^{j\omega}) \right|^{1/D_p} = \text{cost}, \quad \forall p \tag{8.8}$$

where, instead of just the p-th DL gain g_p, we consider also the term due to the loop filter $T(z)$.

Fig. 8.15 shows the characteristic curves relative to the example previously described with three parallel comb filters, in which a IIR low-pass first order loop filter, of the type shown in Fig. 8.14, has been inserted. From the figure we can observe that with the same length and gains of the delay lines, the reverberation time is increased. In addition, the modes at low frequencies decay more slowly than those at high frequencies. Finally, note that in this experiment the coefficient a_p has been empirically determined as $a_p = \frac{1}{2}(1 - g_p)$.

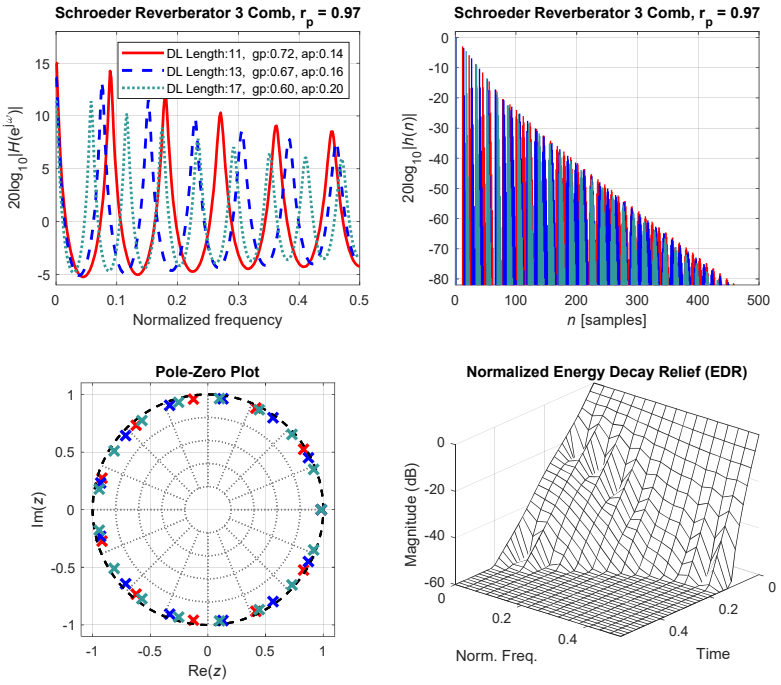

Fig. 8.15 Comb filter with first order IIR loop filter with TF $T(z) = \frac{1}{1-a_p z^{-1}}$.

8.4.2.4 Reverberator with Plug-in All-Pass

As we have seen before, an AR with a good diffuse sound, must be characterized by an impulse response resembling a noisy process with a certain decay. Since white noise is characterized by a white spectrum and a random phase, this suggested the use of all-pass filters: Schroeder proposed the use of all-pass networks in series. Vercoe and Puckette in [31] and Gardner in [32] suggested the use of all-pass nets described in §8.3.2.

However, Gardner [32], pointed out that this type of structure can lead to a colored sound. To avoid this coloration, he suggests inserting additional global feedback on the system with a gain $|g| < 1$. By inserting this feedback, the echo density increases further, making the sound less metallic and more natural.

8.4.3 Early Reflections Characterization

As we know, it is mainly early reflections (ER) that strongly characterize the positional acoustic perception of sound in reverberant environments. The parameters that are normally used for their characterization, defined on the basis of numerous experiments [17]-[20], are for example: 1) the threshold of audibility in relation to the level of the reflected and delayed signal; 2) the level and relative delay that produce the effect of

coloration and/or disturbance; 3) the level and relative delay that produce a sound image (echo).

One of the most accredited methodologies, to define some objective parameters for the characterization of ERs, consists in the analysis of the autocorrelation function of the impulse response $r_{hh}(t)$. Ando in [19] indicates some preference criteria that relate the $r_{hh}(t)$ function and the extent of the first reflections. For example, for a single reflection the subjectively preferred Δt_1 delay is determined by the following reports

$$[\Delta t_1]_d = \tau_d$$

such that

$$|r_{hh}(\tau)| \leq kA^c, \qquad \tau > \tau_d \tag{8.9}$$

where k and c, are constant factors and A represents the reflection amplitude.

In practice, the target delay Δt_1 can be derived from the approxiamtion: $|r_{hh}(\Delta t_1)| = 0.1 \cdot r_{hh}(0)$. The expression (8.4.3) can also be used in the presence of multiple reflections considering the equivalent reflection amplitude calculated as

$$A = \sqrt{\sum_{n=1}^{N} A_n^2}$$

where A_n represents the amplitude of the n-th reflection. The previous relationship say that the perceptually "better reflections" are those with a delay Δt_1 such that Eqn. (8.4.3) is satisfied.

With regard to artificial reverberators, the expression it can be interpreted as reported by Czyzewski in [17], as :

$$|r_{xy}(0)| = \sqrt{\frac{\bar{E}_x}{\bar{E}_y}} \left| \sum_n A_n r_{hh}(\Delta t_n) \right| \tag{8.10}$$

where

- $r_{xy}(0)$ - is the cross correlation between the input and output of the reverberator;
- Δt_n - is the delay of the ER produced by the reverberator;
- $\bar{E}_x$ and $\bar{E}_y$ - represent the average energy of the input and output signal in the analysis window considered.

It indicates that the (8.10) subjective quality of AR's ER can be assessed on the basis of the cross-correlation value.

8.4.3.1 Stereophonic Signal

It is well known that the spatial feel of sound in a listening room can be modelled through ERs. for stereophonic signal, the criteria for the subjective evaluation of the QRA are always defined on the basis of the correlation functions. Ando in [19] proposes a relationship between the desired spatial response and the *interaural crosscorrelation function* (IACC) (see §3.3.3, see Eqn. (3.14)).

In the case of incoherent sounds, where the signal arrives identical to both ears, the IACC takes the form of

$$\text{IACC} = \left| r_{rl}^{(n)}(\tau) \right|_{\max} = \left| \frac{\sum_n A_n^2 R_{rl}^{(n)}(\tau)}{\sqrt{\sum_n A_n^2 R_{ll}^{(n)}(0) \sum_n A_n^2 R_{rr}^{(n)}(0)}} \right|, \qquad |\tau| < 1, \ [\text{ms}] \tag{8.11}$$

where the term $R_{rl}^{(n)}(\tau)$ represents the interaural covariance and $R_{ll}^{(n)}(\tau)$, $R_{rr}^{(n)}(\tau)$ represent the left and right autocovariances respectively of the n-th reflection, measured on the eardrum.

The criterion described by Eqn. (8.11), being measured directly on the eardrum, is not easily adaptable to the qualitative measurement of artificial reverberators. In [17] the author proposes a formula

$$\left| r_{y_r y_l}(0) \right| = \left| \frac{\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f_{y_l}(t) f_{y_r}(t) dt}{\sqrt{\lim \frac{1}{2T} \int_{-T}^{T} f_{y_l}^2(t) dt \cdot \lim \frac{1}{2T} \int_{-T}^{T} f_{y_r}^2(t) dt}} \right|$$

represents the cross-correlation of the reverberator for $t = 0$ and $f_{y_r}(t)$ are the right and left output signals, treated with the A-weighing filter.

8.5 Reverb Model with Feedback Delay Networks

As seen in the previous sections, the realization of an RA requires a number of comb and all-pass filters connected appropriately. For example, to achieve a certain degree of mode density, a number of comb filters must be connected in parallel. To obtain a certain perceivable density of increasing echo over time (as in real rooms) we can connect in series a number of all-pass filters. A first attempt at generalization was suggested by Gerzon in [11] which introduces the concept of unitary multi-channel network. He noticed that the parallel of individual comb filters produces a low quality sound while cross-connecting the feedback lines results in a much higher quality sound.

As proposed by Gerzon [11] and independently from Stautner and Puckette [9], the Feedback Delay Networks model in Fig. 8.16 (already described in §5.2.3) can be seen as a vector extension of the comb filter. In general terms it is possible to say that the FDN model, through the cross coupling between the various channels, produces a much higher echo density and allows a sound with much higher quality than the simple parallel of comb filters.

8.5.1 Stautner and Puckette's Model

The model in Fig. 8.17 is specialized by the same authors [9] with pseudophysical considerations such as a four-input-output network that can be represented with the

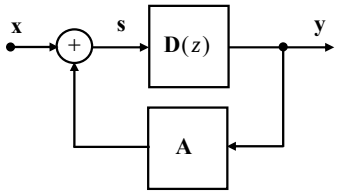

Fig. 8.16 Feedback Delay Networks.

diagram in Fig. 8.16. In particular, an **A** matrix of feedback of the type

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \cdot \mathbf{g}$$

which can be seen as a Hadamard permutation matrix and where the elements of the vector **g** are such that $|g_i| < \frac{1}{\sqrt{2}}$.

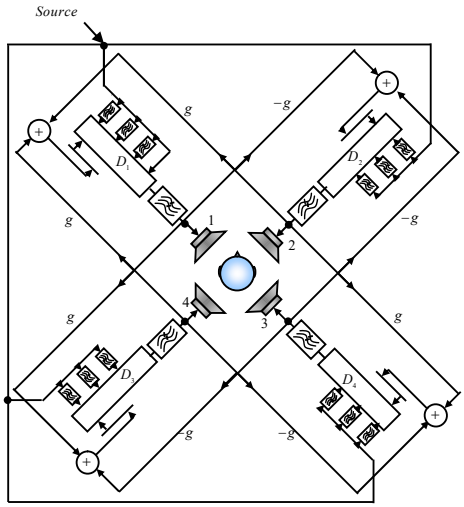

Fig. 8.17 Stautner and Puckette's reverberator with control tone on the input signal [9]

In the Stautner and Puckette model the input signal can be inserted anywhere on the DL. This topology allows to have the outputs (of the loudspeakers) that are mutually incoherent and, as pointed out by many authors [4], [4], [20], this models quite well a diffuse field. The filters at the delay-line input are simple tone controls implemented with shelving filters.

Remark 8.2. One of the main problems of the FDN model lies in the choice of the feedback matrix **A**. In [9], [11] the authors indicate a simple criterion that ensures

stability: matrix A must be obtained as a product of a unit matrix and a vector of gains such that $|g_i| < 1$.

8.5.2 Jot and Chainge Model

A single-input, single-output (SISO) FDN topology, shown in Fig. 8.18-a), was proposed more recently by Jot and Chainge in [13]. This architecture can be seen as an extension of Gerzon's model for monophonic reverberators. Using a vector notation in z domain, we can write

$$\begin{aligned} \mathbf{S}(z) &= \mathbf{D}(z)\left[\mathbf{A}\mathbf{S}(z) + \mathbf{b}X(z)\right] \\ X(z) &= \mathbf{c}^T\mathbf{s}(z) + dX(z). \end{aligned} \tag{8.12}$$

In the case of multi-input, multi-output (MIMO) system in the previous expression, the terms **b** and **c** become matrices. Eliminating $\mathbf{S}(z)$ from the Eqn. (8.12) gives the system's TF

$$H(z) = \mathbf{c}^T\left[\mathbf{D}(z^{-1}) - \mathbf{A}\right]^{-1}\mathbf{b} + d.$$

The poles of $H(z)$ are derived from the equation

$$\det\left|\mathbf{A} - \mathbf{D}(z^{-1})\right| = 0 \tag{8.13}$$

where its solution is simple only for particular choices of matrix **A**.

Remark 8.3. With appropriate choices of matrix **A**, the circuit in Fig. 8.18 can represent any combination of filters. For example, for **A** diagonal the circuit represents a parallel of comb filters. In the case of triangular matrix the equation (8.13) becomes

$$\prod_{i=1}^{N}\left(a_{ip} - z^{D_p}\right) = 0 \tag{8.14}$$

A variant proposed by Jot in [15], of the Jot Chainge model, is the one shown in Fig. 8.18-b).

The insertion of $B_i(z)$ filters allows you to have some control of absorption as a function of frequency. The insertion of the $T(z)$ filter consists instead of a simple tone control.

8.5.3 Choice of Feedback Matrix

As described above, an ideal reverberator should have an exponential LR decay. The impulse response should be that of a modelable diffuse field, i.e., as a time-variant stochastic process [7], [26].

When determining the characteristics of an AR, however, you should refer to a lossless model with infinite reverberation time and make sure that the AR behaves like a "good noise generator".

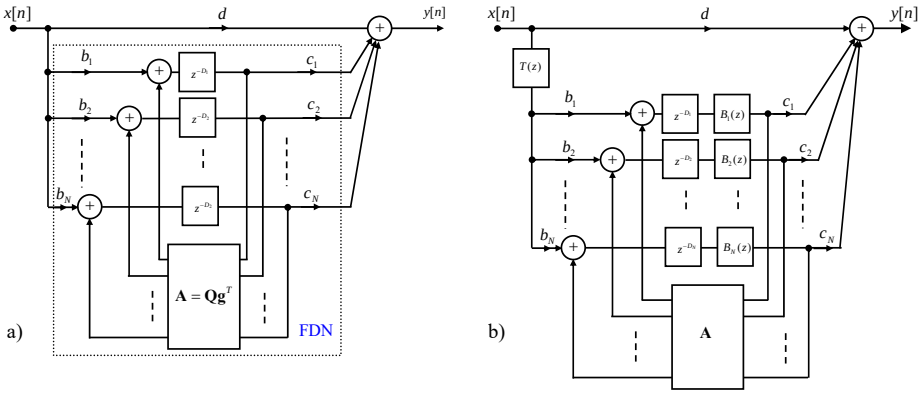

Fig. 8.18 a) Jot Chainge reverberator model. b) Variant with frequency-dependent absorption and shelving filter (tone control).

This modal is defined as *lossless prototype* [4]. After achieving uniformity noise, you can work on getting a certain reverberation time for each frequency band.

It is well known [4] that for FDN models the feature of the lossless prototype strongly depends on the choice of the feedback matrix. Below are some possible choices.

General Method A general criterion for the choice of the feedback matrix derives from the property already described in §8.5.3. The FDN is lossless, if and only if, the eigenvalues of the feedback matrix $\mathbf{A}$, have unit module and corresponding N eigenvectors are linearly independent.

Householder feedback matrix In the case $N \times N$ (N inputs and N outputs) a particularly interesting choice, proposed in [13], consists in the Householder matrix defined as

$$\mathbf{A} = \mathbf{J} - \frac{2}{N}\mathbf{u}_N\mathbf{u}_N^T$$

where $\mathbf{u}_N^T = [1,\ 1,\ ...,\ 1]$ and the identity matrix $\mathbf{J}$ can, be any permutation matrix [23]. The resulting matrix has only two different non zero terms in each column maximizing the echo density.

When N is equal to a power of two, for the calculation of the product $\mathbf{Ax}(z)$, no multiplication operations are necessary but only permutations of the elements of $\mathbf{x}(z)$.

One property of Householder $\mathbf{A}$'s feedback matrix is that for $N \neq 2$ the elements in the matrix are all different from scratch: each delay line feeds all delay lines maximizing echo density.

In the particular case where $\mathbf{J}$ is equal to the identity matrix $\mathbf{I}$, the system behaves as the parallel of comb filters with the highest echo density as shown in Fig. 8.19.

With this configuration Jot notes that an output audible click with a period equal to the sum of the lengths of the delay lines. Jot, suggests that this click can be eliminated by reversing the signs of each element of the vector $\mathbf{c}$.

Another interesting case is for $N = 4$ where the all coefficients of the matrix have the same amplitude.

$$\mathbf{A}_4 = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$$

For $N > 4$ the diagonal elements become larger than the other elements and for very large N the parallel comb bank will be decoupled. In case $N = 16$, an interesting configuration is the one that uses blocks consisting of the $\mathbf{A}_4$ matrix.

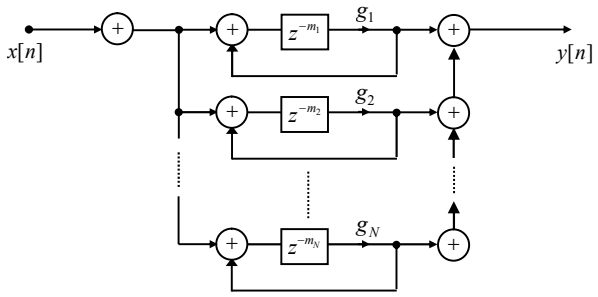

Fig. 8.19 Comb in parallel ($N = 3$) with feedback maximizing echo density.

An interesting method proposed by Gerzon, is to replace each of the four delay lines with an FDN 4×4 of all-pass vectors (which in turn contains 4 delay lines).

Remark 8.4. An interesting choice for the $\mathbf{J}_N$ matrix is the circular permutation matrix. This configuration is equivalent to powering the delay lines in serial mode. This simplifies the problems (hardware and/or software) of memory management.

Triangular feedback matrix Another interesting FDN, proposed by Jot in [25], is the triangular matrix defined as

$$\mathbf{A} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ a & \lambda_2 & 0 \\ b & c & \lambda_3 \end{bmatrix}$$

with this choice in fact the eigenvalues λ_1, λ_2 and λ_3 are placed on the main diagonal whatever the value of a, b and c. However, it is important to note that not all triangular matrices are lossless [4].

Unitary circulating matrix - A further interesting choice of matrix $\mathbf{A}$ is the one proposed by Rocchesso and Smith in [16]. They propose the use of a unitary circulating matrix defined as

$$\mathbf{A} = \begin{bmatrix} a(0) & a(1) & \cdots & a(N-1) \\ a(N-1) & a(0) & \cdots & A(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ a(1) & \cdots & a(N-1) & a(0) \end{bmatrix}.$$

This matrix is characterized by a series of properties such as: if the matrix is circulating then $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$; if the matrix is circulating is lossless it is also unitary. In addition, each circulating matrix can be diagonalized with the DFT transform.

$$\mathbf{A}_d = \mathbf{F}^{-1}\mathbf{A}\mathbf{F} = \frac{1}{N}\mathbf{F}^T\mathbf{A}\mathbf{F}$$

where $\mathbf{F}$ is the DFT matrix (see Eqn. (4.6)). This implies that the eigenvalues of $\mathbf{A}$ can be calculated with the DFT of the first line

$$\{\lambda(\mathbf{A})\} = \text{DFT}\left\{[a(0), a(1), ..., a(N-1)]^T\right\}$$

where $\{\lambda(\mathbf{A})\}$ are the eigenvalues of $\mathbf{A}$. If $\mathbf{A}$ is circulating and unitary all its eigenvalues are arranged in the unitary circle. From the previous properties, it is possible to determine the circulating matrix $\mathbf{A}$ starting from the desired distribution of eigenvalues on the unit circle. Starting, for example, with a desired eigenvalue distribution we can determine the vector of the first line such as

$$[a(0), a(1), ..., a(N-1)]^T = \text{IDFT}\left\{[\lambda(\mathbf{A})]^T\right\} \tag{8.15}$$

and that's true of any lossless matrix. In [16] it is shown, in fact, that given any $\mathbf{A}$ in the form $\mathbf{A} = \mathbf{T}^{-1}\mathbf{D}\mathbf{T}$, it is lossless if d any diagonal matrix with unit module is $\mathbf{T}$ and any invertible matrix.

From a more operational point of view, the use of the unitary circulating feedback matrix reduces the computational cost of the vector matrix product from $O(N^2)$ to $O(Nlog_2N)$ when N is a power of 2.

8.5.4 Other Models

Dattorro's Reverberator - Inspired by Griesinger's work[12], Dattorro's reverberator[39] consists of a number of pre-delay units (one low-pass filter and four all-pass filters) used to decorrelate the input signal and called *diffusers*.

The second section of Dattorro's reverberator, called the *tank*, consists of two different paths, each of which is feedback to the other as illustrated in Fig. 8.20. Each tank path consists of two all-pass two DLs and a low pass filter. The reverberator output is the weighted sum of the tank outputs.

In Dattorro's reverberator, the all-pass filters are of variable length: the length of the DLs is modulated so as to obtain a better diffusion effect.

Direct Convolution Reverberator - As we have already said in the introduction, if we wanted to obtain a very natural sound of an artificial reverberator we could convolve the input signal with the impulse response of a real room. Since convolution done directly in the time domain requires a large number of operations, in practice block methods in the frequency domain [40] or multi-rate methods[41] are used (see §4.6).

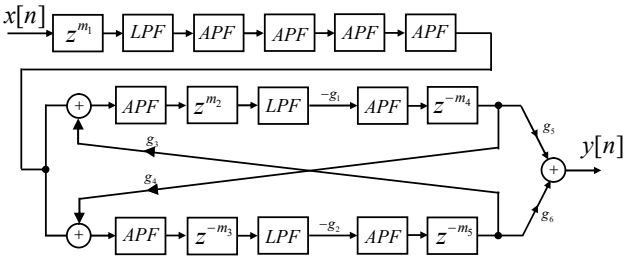

Fig. 8.20 Dattorro's reverberator diagram [39].

FDN with time variants DL For better acoustic performance, Dattorro's reverberator can be realized with time variants delay line From the physical point of view the DL time variant models: 1) the different speed of propagation that occurs in the various layers of air in the presence of temperature gradients; 2) the characteristic time variants of the listening rooms due, for example, to people moving, etc.

This can be easily extended to FDN-based reverberators [33]. The length of the DLs can be modified with different types of modulations such as sinusoidal or pseudo-random sequences which, in order to have a slow variation, can be filtered. To have a very realistic diffusion effect each delay line is modulated with a different seed and/or with different modulation depths. However, care must be taken if the various modulations are all in the same direction: in these cases, in fact, there could be a perceptible effect on the pitch of the note played. In general, these effects are quite small if the modulations are sinusoidal [33].

8.6 Acoustic Modeling with Digital Waveguides Networks

The digital waveguide (DW) paradigm defined by Smith in [27] has long been widely used in the physical modeling of musical instruments. As already mentioned in §6.3, the DW consists of a two-way delay line (or simply two delay lines with opposite directions). A delay line can therefore be considered as a simplified DW and, therefore, the FDN theory can be reformulated in terms of digital waveguides. It is also known that a DW network (DWN) (see also §5.2.3) can simulate wave propagation in any 2D or 3D space direction [4], [28], [29], [35].

In the design of acoustic environment simulators, the main advantage in using DWNs, rather than normal DWs, is that they can model the wavefront explicitly in all directions as in a real situation. With 2D or 3D DWNs, the diffuse field can, in fact, be modeled with some accuracy as well as the growth of echo density and modes that can occur in a very natural way. In particular with DWNs it is possible to model the low frequency modes of the listening room with a certain precision (which is extremely complex with other methods).

It should be noted that the computational cost of DWN is *quite negotiable* and lower than standard techniques based on simulation with finite difference equations especially for coarse-grained DWN. However, the use of coarse-grained DWN allows good physical modeling at low frequencies.

This limitation is acceptable as higher frequency modes cannot be perceived by the ear. Also, since DWN is by definition a lossless circuit, no errors are made in modeling the damping modes. Errors, which can be made by the DWN simulator when modeling an oscillating membrane or acoustic space, are due to mode estimation (tuning) and to the necessarily limited bandwidth. Tuning errors can be due to dispersive phenomena along certain directions [30].

8.6.1 Physical Modeling with Digital Waveguide Networks

A DWN for wave propagation modeling consists of the connection of scattering junctions (SJ) with lines composed of a single delay as shown in Fig. 8.21-a).

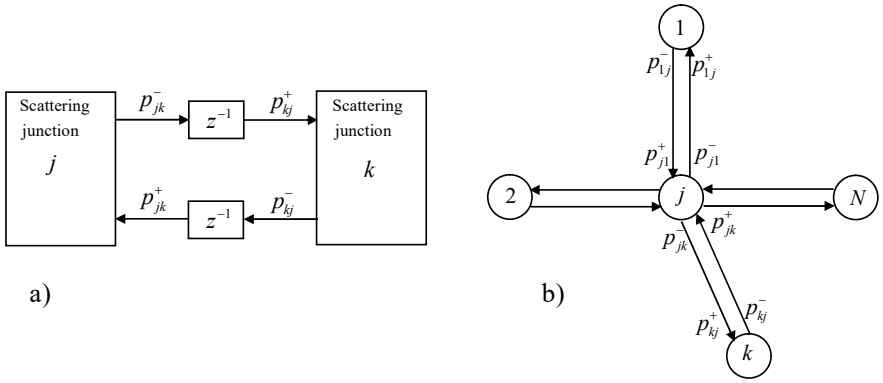

Fig. 8.21 Digital waveguide networks (DWN). a) Scheme of a simple DWN. b) Scattering junction j connected with other DWNs.

A DWN consists of a SJ connection through DW with unit delay. Thus, said p_i the acoustic pressure, u_i the volume velocity and Z_i the acoustic impedance (i.e. $u_i = p_i/Z_i$) of the i-th section of the DW, with reference to Fig. 8.21-b), indicating with p^+_{jk} the signal coming from SJ k toward SJ j, we have

$$p_{jk} = p^+_{jk} - p^-_{jk}$$
$$p^+_{jk} = z^{-1} p^-_{kj}$$

8.6.2 DWN Topologies

The connection of delay elements with SJ can be done in various ways and topology allowing the definition of different models.

To model the wave propagation in a horizontal plane in an enclosed space, you can configure the network (or mesh) in various ways. For example, Fig. 8.22-b) shows a rectangular mesh, while Fig. 8.22-c) shows a triangular mesh.

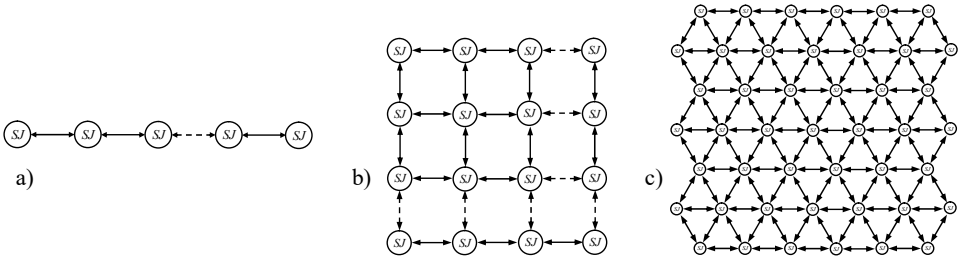

Fig. 8.22 Digital waveguide networks topology. a) Linear DWN: SJ is a two-port (1D) network. b) DWN with rectangular mesh: the SJ is a four-port (2D) network. c) DWN with triangular mesh (SJ is a six-port 2D network)

DWN-2Ds can be used for modeling vibrating membranes such as drums, gongs, ..., together with their excitation models [34]. The sampling frequency of the f_{upd} network is determined by the space between the SJs.

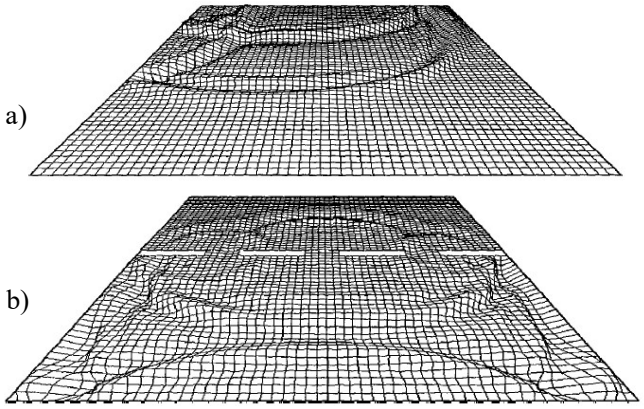

Fig. 8.23 Simulator of a virtual room with 2D DWN with triangular mesh. a) Wave propagation along the mesh. b) Phenomenon of diffraction due to the perforated wall inside the room. The sampling rate is $f_{upd} = 22049$ [Hz] which corresponds to a $d = 0.022$ [m]. (Courtesy of [37]).

Fig. 8.23 [36], [37]) shows, as an example, the simulation of a wave propagation in a 2D virtual room of $(6.6 \times 5.50$ m) with triangular mesh, in which a wall with three openings has been placed. Note how the phenomenon of diffraction is already implicitly contained in the model.

8.7 Dynamic Range Control of Audio Signal

The *automatic gain control* (AGC), realized with a *dynamic range controller* (DRC), is a very important method in the manipulation of the audio signal (see [10], [42], [43]). The devices able to perform AGC, are widely used both for technical reasons,

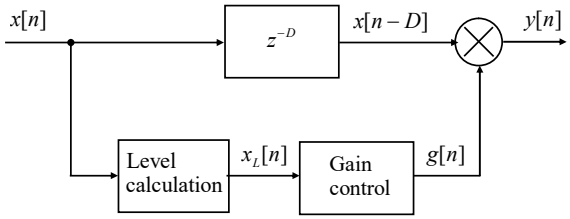

Fig. 8.24 Principle diagram of an automatic gain control (AGC) device which, depending on the control algorithm, can be used as a compressor or dynamics expander.

as for example in the protection of equipment from overloads, and for artistic reasons to realize, for example, particular audio effects.

In applications such as transmission, recording, mixing, playback of audio streams, it is often appropriate to vary the gain of a given channel adaptively to the signal level. Let's make some examples (other applications will be illustrated below) to highlight how such techniques can improve listening quality:

- A voice signal can be affected by large changes in volume due to accidental speaker movement relative to the microphone position. In the case of constant gain, it is difficult to set the correct gain value: if the gain is calibrated on the parts at low volume, it may be too high at the loudest parts (hearing discomfort and/or saturation of electronic devices); if the gain is calibrated to the loudest parts, the low-volume ones will be not amplified and understandable. An adaptive gain allows you to "pull up" the volume when the speaker moves away from the microphone and turns it down when it gets too close, equalizing the signal level.
- During radio or television broadcasting, the broadcasts are carried out by sequencing clips from different sources (music alternating with speech, music from different recordings, intramezzled advertising, etc.); each of these sources may have an intrinsic volume different from the other, so also in these cases it is advisable to vary the gain in an adaptive way so as not to force the listener to turn up and down the volume control of the receiver.
- During a TV talk-show with many microphoned guests in the studio, you should reset the gain corresponding to a given speaker when the speaker is not speaking. Otherwise his microphone would only contribute to the background or ambient noise, which added to that of the other silent guest microphones, could become audible and annoying.

In general we can say that the objective of dynamics signal control is to adapt it "at its best" to the communication channel. The principle diagram for the AGC is shown in Fig. 8.24. The device, commonly called a *compressor*, consists of two branches: one branch with a delay and a multiplier; and the other in which the signal level is measured and appropriately processed. The methods underlying adaptation are: *compression* and *expansion*. With compression the dynamics of the signal is decreased and with expansion it is increased.

Extreme forms of compression/expansion are called *limitation*/*noise-gating*. The control signal $x_L[n]$ consists of a suitably averaged measurement of the input signal level: you can use the peak level $x_{L_{PEAK}}[n]$, the rms (root-mean-square) value $x_{L_{RMS}}[n]$ or a linear combination of them.

8.7.1 DRC Static Curves

The dynamics control of an acoustic signal can be represented by four distinct types of operation: *limiter*, *compressor*, *expander* and *noise-gate*. The characterization of the operating mode of the device is defined by the parameters: *compression ratio*, *intervention threshold* and *gain*. By indicating the values expressed in [dB] with capital letters and natural values with lowercase letters, the *compression ratio* R can be defined as[1],

$$R = \frac{\Delta Y_L}{\Delta X_L} = \frac{Y_L - S}{X_L - S} \tag{8.16}$$

where terms X_L and Y_L indicate the level (peak RMS) of the input and output signal, and the S value, defined as *intervention threshold*, represents a reference value above or below which the AGC device changes its characteristic. In other words, in normalized terms, the Eqn. (8.16) indicates that a dynamic of R dB at the input produces a dynamic of 1 dB at the output.

Fig. 8.25 shows the static input-output characteristics, schematizing the four operating modes defined by the compression ratio value and the intervention threshold:

$$\begin{array}{lll} X_L > S_C & R_C < 1, & \text{compressor} \\ X_L > S_L & R_L \to 0, & \text{limiter} \\ X_L < S_E & R_E < 1, & \text{expander} \\ X_L < S_{NG} & R_{NG} \to \infty, & \text{noise gate.} \end{array}$$

In the graphs we can see the parameters settings that define the static characteristic:

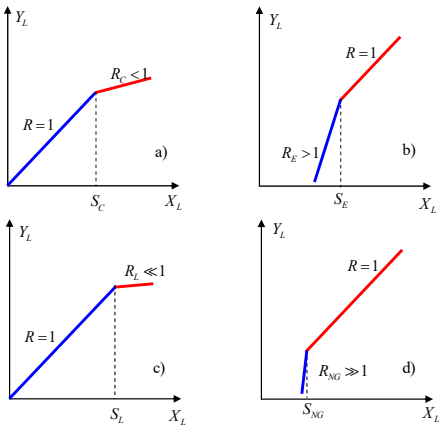

Fig. 8.25 Qualitative static characteristic of a DRC device working as: a) compressor $R_C < 1$ for $X_L > S_C$; b) expander $R_E > 1$ for $X_L < S_E$; c) limiter $R \ll 1$ for $X_L > S_L$; d) noise-gate $R \gg 1$ for $X_L < S_{NG}$.

the compression ratio, the threshold and the gain. The threshold is the knee point while the compression ratio is the slope of the stretch above the threshold.

[1] Some authors, as in [42], [43], use to define the compression ratio as the inverse value of (8.16).

From Eqn. (8.16) it is also possible to derive the relation between the compression ratio R and the slope P, with simple reasoning it is easy to verify that

$$P = 1 - R \tag{8.17}$$

Consider, for example, the typical static characteristic expressed in dB of a compressor shown in Fig. 8.26-a). The DRC device above a certain S_C threshold value, called *compression threshold* ($S_C = -30$ dB in the case described in the graph), has a compression ratio $R = \frac{1}{3}$ and a slope $P_C = \frac{2}{3}$. When the signal has a level below the threshold ($X_L < S_C$), it is not processed while for higher levels the output dynamic is reduced.

Remark 8.5. Note that, we must not make the mistake of interpreting the graph as a saturation curve applied directly to the signal; in this case in abscissa we would have the input sample $x[n]$ and in ordinate the output sample $y[n]$; in this case we would speak of *waveshaping*, an algorithm that has the analogical counterpart in the saturation of an amplification stage, and that introduces considerable harmonic distortion. In fact, in the DRC on abscissae and ordinates we have, respectively, the input and output signal levels both in dB. The level in dB varies more slowly than the signal amplitude, and this poses a first upper limit to the speed with which the gain control varies.

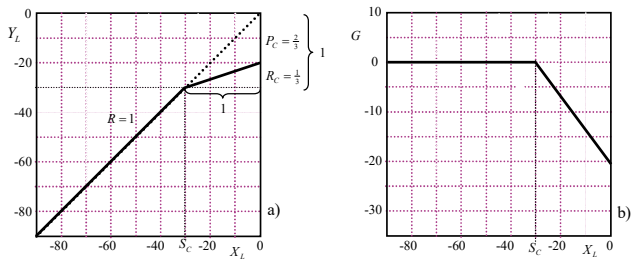

Fig. 8.26 Static characteristic of a DRC device working as an compressor $R_C < 1$: a) static input-output characteristic; b) static input-output characteristic. For the definition of G (control gain) we can observe that the obvious relationship applies.

For the definition of the *control gain* G we can observe that the obvious relationship applies

$$G = Y_L - X_L. \tag{8.18}$$

It is possible, always on the basis (8.18), to define a second curve, called *static gain curve*, shown in Fig. 8.26-b), which expresses the control gain value as a function of the input level $G = F(X_L)$ (in logarithmic domain).

In the linear section the compression ratio and slope are $R_{lin} = 1$ and $P_{lin} = 0$. Above the compressor intervention threshold ($X_L > S_C$), there is $R_C = \frac{1}{3}$ and a slope $P_C = \frac{2}{3}$. It is therefore easy to verify that the trend gain G is equal to

$$G = F(X_L) = \begin{cases} 0 X_L \leq S_C \\ -P_C(X_L - S_C) X_L > S_C \end{cases}$$

The graph represents another form of the static characteristic that carries information completely equivalent to the first. It clearly shows that the more the input level exceeds the threshold, the more the gain is reduced.

Static curves with multiple thresholds The DRC device, in normal use modes, can have different operating modes determined by several trip thresholds. These modes may coexist, i.e. when manipulating a source there may be a need for more than one type of static characteristics to be selected depending on the signal level.

Let's think, for example, of the TV talk-show problem where for low signal levels we may want a noise-gate operation while at higher levels some compression is required. For even higher levels you can think of inserting a limiter that does not allow you to exceed a fixed higher level in any case.

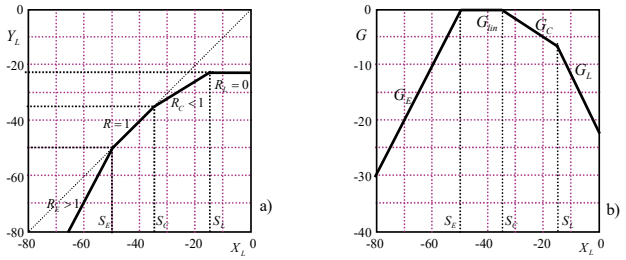

Fig. 8.27 Static characteristic curves. a) Compression ratio R. b) Slope $P = 1 - R$.

Fig.s 8.27-a) and b) show the static characteristics of the compression ratio R and gain G, respectively, of a DRC where the characteristic is selected according to a certain threshold.

By combining several linear segments it is possible to have characteristic curves of any shape.

Table 9.2 shows the operating modes as a function of levels and thresholds, compression ratios, gains related to the various operating modes for a typical multiple threshold device.

Table 8.2 Compression ratios, thresholds and gains of a dynamic range controller with multiple thresholds.

	Thresholds	Input Level	Compression Ratio	Slope or Gain
Limiter	S_L	$X_L \geq S_L$	$R_L \to 0$	$P_L = 1$
Compressor	S_C	$S_C \leq X_L < S_L$	$0 < R_C < 1$	$1 > P_C > 0$
Linear	-	$S_E \leq X_L < S_C$	$R_{lin} = 1$	$P_{lin} = 0$
Expander	S_E	$S_{NG} \leq X_L < S_E$	$1 < R_E < \infty$	$-\infty < P_{lin} < 0$
Noise-Gate	S_{NG}	$X_L < S_{NG}$	$R_{NG} \to \infty$	$P_{NG} \to -\infty$

From a practical point of view it is possible to realize the static gain curve $G = F(X_L)$ with simple geometric considerations on the graph of Fig. 8.27-b) and considering the definitions of Table 8.2 for the various sections of the curve.

The static characteristic function of the gain can be, in fact, calculated with the following relations

$$\begin{array}{lll} X_L \geq S_L, & G_L = -P_L(X_L - S_T) + P_C(S_C - S_L), & \text{limiter} \\ S_C \leq X_L < S_L, & G_C = -P_C(X_L - S_C), & \text{compressor} \\ S_E \leq X_L < S_C, & G_{lin} = 0, & \text{linear} \\ S_{NG} \geq X_L < S_E, & G_E = -P_E(X_L - S_E), & \text{expander} \\ X_L < S_{NG}, & G_{NG} = -P_{NG}(X_L - S_{NG}) + P_E(S_E - S_{NG}), & \text{noise-gate.} \end{array} \tag{8.19}$$

8.7.2 Dynamic Gain Control

The measurement of the input level, as previously indicated, can be related to the maximum or peak value, equivalent to the envelope of the input waveform, or related to the average energy, i.e. the RMS value, of the input signal.

Even if the level value X_L, varies more slowly than the signal variations, this variability may still be too high from a perceptual point of view. In fact, the control gain G by Eqn.s (8.19) gives a gain (converted to natural values) $g[n]$ that varies more slowly than the input signal $x[n]$. Albeit this is inherent in level measurement, this variation may be too abrupt for the acoustic effect to be natural and pleasant. To avoid this problem the signal $g[n]$ is processed by a smoothing filter. Normally a non stationary first-order low-pass filter is used, meaning that the time constant of the filter switches between two values depending on whether the signal level derivative is positive or negative. These two constants are usually called *Attack time* and *Release time*, and are very important in determining the acoustic effect of dynamic treatment.

- *Attack* τ_A is the time constant that intervenes when the gain variation occurs, and therefore determines the velocity with which this variation occurs.
- *Release* τ_R is the time constant that occurs when the gain reduction ceases, and therefore determines how long it takes before the gain is restored to 0 dB.

Fig. 8.28 shows the typical gain trend $g[n]$ as a function of the input signal envelope $X_I[n]$. The action of the AGC is appropriately smoothed and characterized by the time constants τ_A and τ_R.

The choice of τ_A and τ_R values is very important and greatly influences the level of distortion and listening quality. Short Attack and Release values lead to a more effective compression action, while high values ensure a more natural effect, reducing the audibility of the artifice.

For example, in the limiter, where the peak level is used, the attack time must be rather small while the release time is larger. According to [42], [43], the typical values are τ_A = 0.02, 0.04, ..., 10.24 msec, and τ_R= 1, ..., 130, ..., 5000 msec.

From the above, regardless of the type of measurement on the input signal, it is convenient to insert a module for the definition of the time constants τ_A and τ_R. This module, called *gain factor smoothing*, is generally (but not necessarily) inserted after

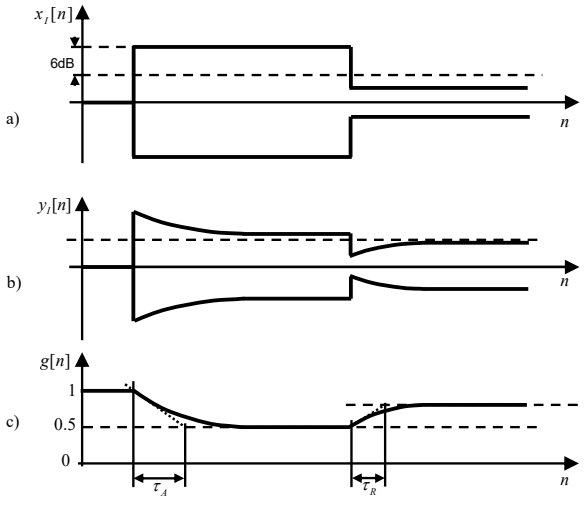

Fig. 8.28 Typical dynamic AGC trend: a) input signal envelope; b) output signal envelope; c) gain trend.

the calculation of gain G and after having converted its value into natural numbers $g[n]$.

$$g[n] = kf[n] + (1-k)g[n-1]$$

where the k parameter assumes different values depending on the level derivative. The switching model for the time constant is therefore of the *hysteretic* type: in the attack phase $k = \tau_A$ is set, while in the release phase $k = \tau_R$ is set. The corresponding network function is therefore

$$H(z) = \frac{k}{1-(1-k)z^{-1}}$$

A schematic diagram that realizes this structure is shown in Fig. 8.29 [42], [44].

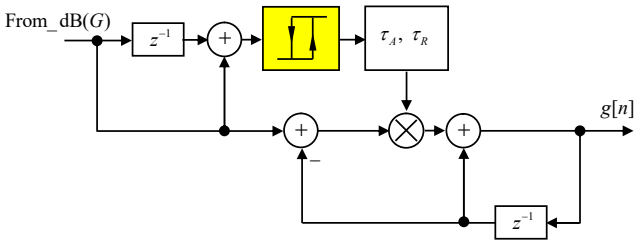

Fig. 8.29 Dynamic control scheme of the hysteretic gain factor smoother. The attack and release time constants are switched according to the level state of. (The `From_dB(G)` function is usually implemented with a LUT).

8.7.3 Signal Level Calculation

The measurement of the signal level, as we have seen above, can be linked to the maximum absolute value (peak) or to the average energy of the signal.

8.7.3.1 Peak level measurement

It is simply a measure of the maximum amplitude reached by the waveform, and because the signals are not stationary, it is a quantity that varies over time. In order to have an appropriate transient behaviour, the peak can be measured by passing the absolute value of the signal through a simple first order low-pass filter by simulating the behaviour of the analogue envelope detector consisting of a diode followed by a *RC* low-pass filter.

As an alternative you can consider the maximum peak in a signal window of a few ms. This sub-sampled information is then smoothed by an IIR or FIR low pass filter to generate an appropriate gain control action. For example, McNally in [42] proposes a method for *peak level measurement* with the following difference equation

$$x_{L_{PEAK}}[n] = \tau_a\,|x[n]| + (1-\tau_a-\tau_r)x_{L_{PEAK}}[n-1]$$

with a TF

$$H(z) = \frac{\tau_a}{1-(1-\tau_a-\tau_r)z^{-1}}$$

whose signal flow graph is shown in 8.30.

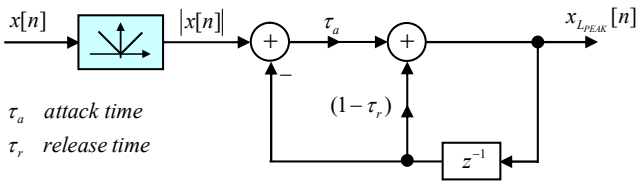

Fig. 8.30 SFG of the input signal peak meter proposed by McNally. The system is an envelope detector followed by a first-order low-pass filter. (Courtesy of [42]).

Remark 8.6. In the diagram, the time constants τ_a and τ_r are related to the dynamics of the level measurement, and should not be confused with the constants τ_A and τ_R defined above, which may have much higher values.

8.7.3.2 RMS level measurement

For the RMS measurement, the information is not the absolute value, but the square of the signal. Moreover, this case larger time constants (i.e. longer averaging times) are generally used, because it is not interesting to find information that follows the instantaneous peak of the signal, but is correlated with the acoustic intensity perceived by the listener. For a segment of N samples, the measurement of the RMS level is by definition given by the square root of the following quantity

$$x_{L_{RMS}}[n] = \frac{1}{N} \sum_{k=n-N+1}^{n} x^2[n] \tag{8.20}$$

however, for reasons of computational simplicity, the square root calculation is avoided and this value is used directly. An alternative way to calculate the approximate RMS value that avoids averaging over N samples is proposed by McNally in [42], and shown in Fig. 8.31.

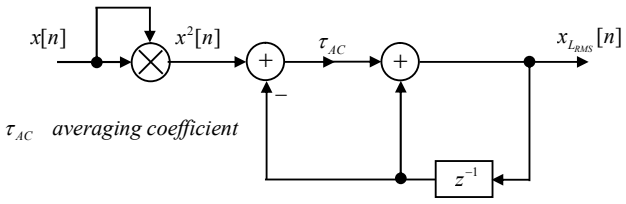

Fig. 8.31 SFG of the RMS input meter proposed by McNally. (Courtesy of [42]).

In the diagram the constant τ_{AC} represents an *average coefficient* that allows to have sufficiently smoothed values. The difference equation results

$$x_{L_{RMS}}[n] = \tau_{AC} x^2[n] + (1 - \tau_{AC}) x_{L_{RMS}}[n-1]$$

with network function equal to

$$H(z) = \frac{\tau_{AC}}{1 - (1 - \tau_{AC}) z^{-1}}.$$

8.7.4 Constructive Considerations of the DRC

Fig. 8.32 shows the diagram of the DRC chain in this case of compressor. The input signal $x[n]$ enters the block for level evaluation x_L (time index n is omitted for simplicity). This value is converted into its logarithmic and multiplied by 0.5 to take into account the square root not evaluated in the RMS calculation with (8.20).

Then the gain G_C is calculated using the expressions (8.19), which is then converted back into natural values.

In order to have adequate attack and release values, the gain signal is lowpass filtered with a structure of the type described in §8.7.2 so as to obtain the $g[n]$ value to be multiplied to the appropriately delayed input signal $x[n-D]$.

For logarithmic conversion, it is convenient to use the logarithm in base 2. In this case the logarithm of the peak value or RMS can be and evaluated as

$$\log_2(x) = \log_2(m \cdot 2^E) = E + \log_2(m)$$

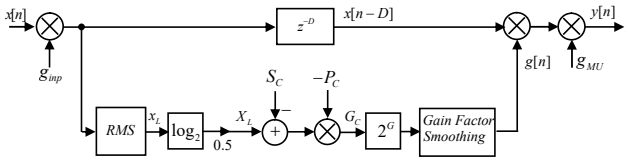

Fig. 8.32 Diagram of a AGC of a compressor.

where E represents the exponent and m the mantissa of the representation of the number x. An evaluation of the computational cost of the calculation can be reduced using a LUT [42].

8.7.4.1 Make-up Gain

Another parameter often present in the device is what is sometimes called make-up gain g_{MU} (sometimes also referred to as g_{out}). It is nothing more than a gain factor cascaded after the compressor to adjust the output level; in fact, since the compressor reduces the level of too intense passages, after its action it is possible to raise the overall level of the signal. This action compensates for the loss of volume suffered

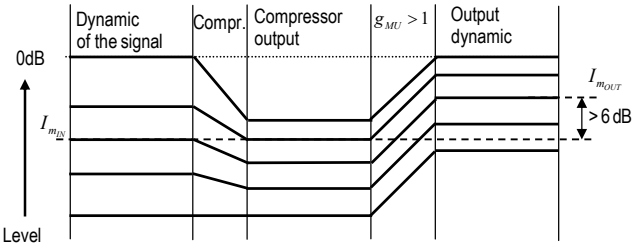

Fig. 8.33 Level diagram. Raising the average level of the input signal by means of a compression and amplification process.

by intense passages and amplifies weak passages, which have not been affected by compression. Fig. 8.33 shows, as an example, the use of the compressor and make-up gain to raise the average level of the input signal. In the figure, I_{mIN} and I_{mOUT} indicate the average input and output levels respectively.

To adjust the input level there may also be an input gain, shown in Fig. 8.32 with g_{inp}.

This technique is often used in the transmission of commercial radio and television messages. In these cases, while not exceeding the maximum sound intensity limits imposed by the regulations, the average level of the audio signal is raised by several dB: the listener perceives the advertising message at a much higher volume than in the previous and subsequent context.

In the case of multi-threshold compressors we will have available many operating modes activated by a selector controlled by the signal level value. For example, Fig. 8.34 shows the complete principle diagram of a DRC device characterized by four operating modes: limiter, compressor, expander and noise-gate.

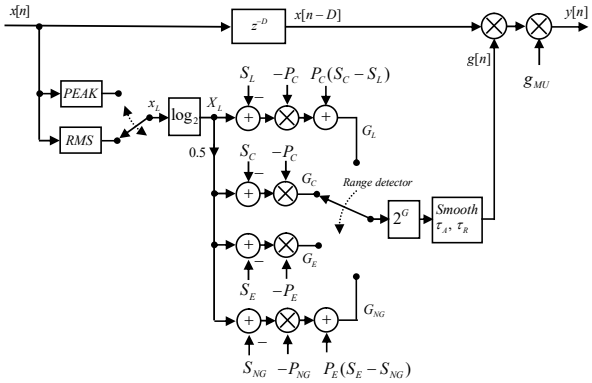

Fig. 8.34 Complete diagram of a device for multiple-threshold DRC: limiter, compressor, expander and noise-gate. (Courtesy of [42]).

In modern playback systems it is necessary to process several channels in parallel as in the case of stereo or multi-channel signals. It is necessary, in these cases, to have a common gain factor: using different gain would have an unbalanced acoustic front. Fig. 8.35 shows an outline of a multi-channel principle. Note that a ticking/interpolation circuit has been inserted in the diagram in order to further reduce the computational cost. This is possible because, as previously pointed out, the variation of the level signals is much smaller than the variation of the input signal.

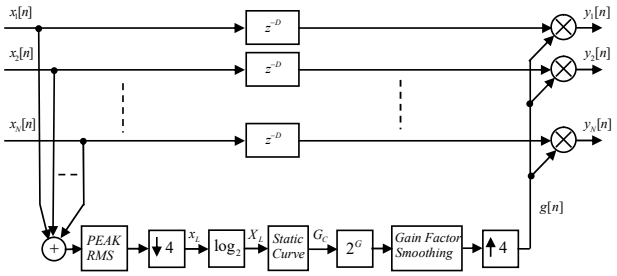

Fig. 8.35 Multi-channel DRC device.

8.7.4.2 Look-ahead time

An important parameter of the DRC concerns the value of the delay, z^{-D} to be applied to the signal before gain adjustment. The value of this delay, often referred to as *look-ahead time*, causes the effect of the dynamic control to occur a few msec before the event that determines it.

To better understand the mechanism of Look-ahead time consider the graphs shown in Fig. 8.35. Fig. 8.35-a) shows the input signal envelope. Fig.s 8.35-b) and c) show the output envelope envelope for a 2:1 (i.e. $R = 1/2$) compressor and S_C = -20 dB with the same values as the attack and release time constants τ_a = 1ms, τ_r = 50ms; for the RMS measurement and τ_A = 30ms τ_R = 500ms; for the gain smoothing filter.

For Fig. 8.36-b) a delay (Lookahead $\tau_{LA} = 0$ has been used while Fig. 8.36-c) is relative to a delay $\tau_{LA} = 30$ms.

From the graph it is easy to see that for $L_A = 0$, the effect of the compressor is almost instantaneous. In this case, in fact, the level change is made immediately after the input signal change (from -20dB to 0dB): the instantaneous peak value, therefore, is not attenuated. Since the maximum value of the signal is unchanged, the g_{MU} output gain cannot be increased.

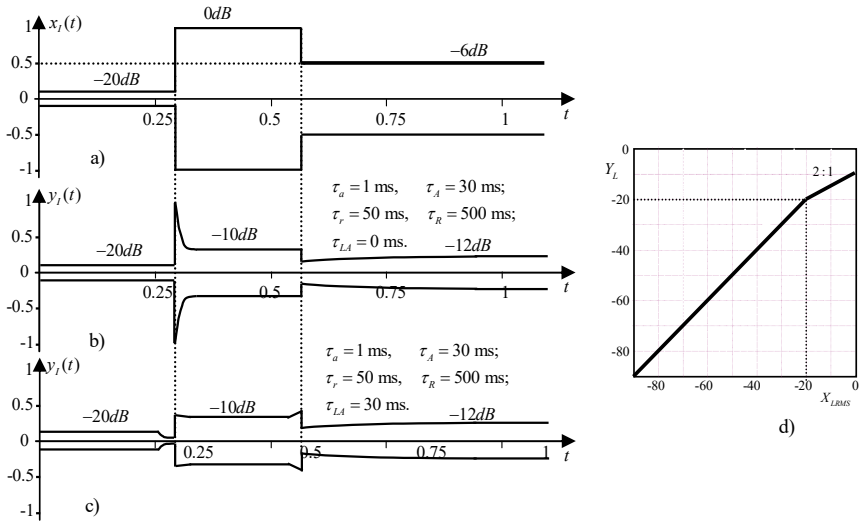

Fig. 8.36 Dynamic signal trend in presence of look-ahead delay: a) input signal envelope; b) compressed signal envelope for $L_A = 0$ms; c) compressed signal envelope for $L_A = 30$ms; d) static compressor characteristic.

In case you have a certain Look-ahead time, as in 8.36-c), where $L_A = 30$ms, the effect of the compressor is earlier than the arrival of the peak. The maximum value, in this case, is decreased.

The decrease of the dynamics already before the arrival of the peak is very important because it safeguards, from the acoustic point of view, the occurrence of the transient phenomenon: the control gain g is decreased before the arrival of the peak. The maximum signal value is attenuated in this case and the output gain cannot take on $g_{MU} > 1$ values.

8.7.4.3 Interpolated spline static characteristic curve

In the more advanced devices the static character istic instead of having a linear trend at times, can have a curved trend. This trend is generally achieved by means of polynomial spline interpolators. The advantage, from a perceptual point of view, is greater in the case of static characteristics with high compression ratios such as limiters.

Fig. 8.37 shows two examples of spline interpolated characteristic curves.

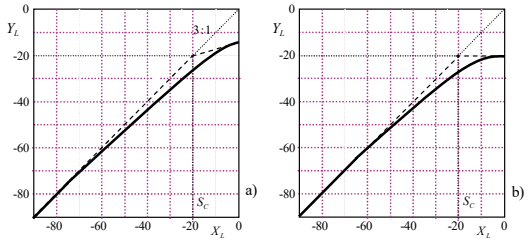

Fig. 8.37 Static interpolated spline characteristics.

8.7.5 DRC with Multiband Approach

The so-called multi-band approach consists in splitting a signal into sub-bands (2 or more) by means of a filter bank and treating the sub-bands separately before they are recombined to reconstruct the signal. The FIR or IIR filter bank can be of uniform or constant-Q or user definable type.

Fig. 8.38 shows the principle scheme of a multi-band DRC. The full band approach presents in some cases a number of problems or side effects arising from DRC. Here are some of them.

Pumping - This term refers to a rather unpleasant acoustic effect generated when DRC is applied to a music signal where several instruments are playing simultaneously, with high compression rate (low threshold, high compression ratio) with relatively fast Attack and Release. In the presence of instruments with high dynamics, which introduces strong and fast variations in the waveform (e.g. drums, bass, vocals, ...), it will trigger the gain reduction. For example, when the kick drum is hit, and DRC is triggered, the other instruments will also suffer the same level reduction. What you hear is a volume of these sounds "changing in time" with the drum or bass drum strokes (or with vocal parts at higher volume). Using a longer Attack and Release (especially the second one) slows down these oscillations making them less audible, but still present.

Breathing - This effect refers to the compression of voice signals. We have talked about how compression leads to the emphasizing of the signal parts at lower volume; well, among these, besides the background noise, there is also the breath of the speaker or singer, which can become annoyingly audible during voice pauses. This problem is more easily solved than pumping, because it may be sufficient to insert a noise-gating module after the compressor.

Advantages of the sub-band approach - The sub-band approach makes it possible to act selectively on the individual channels, compressing more the channels responsible for volume changes and leaving the others unchanged, thereby minimizing (often to zero) the auditory effect of pumping. In addition to having the possibility

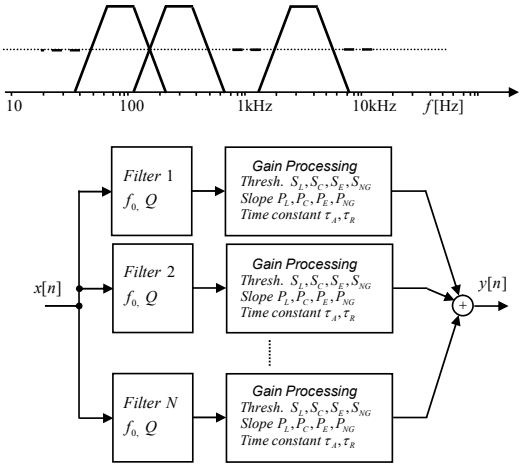

Fig. 8.38 Multiband DRC. Each channel has independent, user-definable settings. 10Hz, 100Hz, 1kHz, 10kHz.

to compress, expand (with Attack and Release aimed at the type of sound material in the band) or leave each band unchanged, the use of different make-up gain for the various channels allows you to perform a real action of equalization of the signal, contextual to the dynamic control. This approach is very powerful in bringing out the more shadowy bands and in blocking those that are too active.

A classic example of multi-band treatment: de-essing - An example of the possibilities offered by a multi-band approach is the "de-esser", a classic and widely used module designed to solve a long-standing problem of recording and vocal performance. For various reasons, ranging from the way the microphone is used to the type of microphone itself, to the use of reverbs of a certain type, to the type of equalization, etc. It often happens that a sung part is affected by overbearing and annoying hissing when the singer pronounces sibilance consonants such as: "s","t", "z", "ch", "j" and "sh". These hisses have a limited band spectrum, approximately between 5 kHz and 10 kHz

To reduce this effect, an equalizer could be used to attenuate the band of interest. This, however, can produce an overall loss of presence and brilliance of the sound, making it muffled.

Full-band compression fails, because sibilant sounds, although audible, have low energy compared to vocal or sound sounds.

The level of sibilants is below the average level of the signal, so it is sufficient to use a compression with a threshold just above the average level in this band, which takes over when the "s" are needed, attenuating them at will. will.

A multi-band N-channel compressor (normally N goes from 2 to 8) is made with a filter bank and N DRC modules containing at least the compressor and the expander modules. Limiter and noise-gate can also be present in each channel, otherwise the limiter can be in one instance after the signal reconstruction and noise-gate before decomposition or after reconstruction.

Limiter and noise-gate can be present in each channel, otherwise the limiter can be inserted after signal reconstruction and noise-gate before decomposition or after reconstruction.

8.7.6 Main applications

Dynamics control is, by definition, a non-linear operation that always produces harmonic output signal distortion [44]. It is obvious, therefore, that DRC devices should be used with great caution and awareness of the operations being carried out on the signal. Nevertheless, compressors are widely used in many audio fields such as recording, transmission and playback.

8.7.6.1 Compression

In compression, the dynamics of the audio signal is decreased to match the playback range, such as in live performances or radio and television broadcasts where the level may be too high to be accurately reproduced or transmitted.

Overload protection - In its extreme form as a limiter, it is used for the protection of equipment such as speakers, modulators etc. In radio transmissions, both in amplitude and frequency modulation, the depth of modulation has very precise limits. Exceeding these limits produces a strong distortion of the transmitted signal.

Improving the quality of microphone footage - To compensate for the change in the output level of a microphone when it is moved away from or near the source as is often the case in broadcast studios due to speaker movements or for the elimination of excessive hissing (de-essing).

Audio effects - DRC is also widely used as an audio effect for individual instruments (drums, bass, guitar etc.) and or used in combination with other effects such as artificial reverberators. In drumming, for example, the insertion of a compressor varies the decay curve making the sound very special and used for certain musical genres.

Increased perceived sound intensity - By reducing the dynamics, the average signal level can be increased through make-up gain (see Fig. 8.33). This can be used to increase the perceived sound level. This application is particularly significant when listening to music from low-power systems such as TVs, radios, keyboards with loudspeakers, etc., where the use of the compressor can increase the perceived sound level by up to 4-6 dB.

8.7.6.2 Expansion

Expansion is the inverse operation of the compression used to increase the dynamics.

Audio effects - Reduced sustain time in musical instruments.

Noise gating - Is an extreme form of expansion generally adopted to eliminate background noise.

Increased S/N - The compression and expansion used in conjunction with emphasis and de-emphasis networks can be used to increase S/N in low dynamic audio media. Widely used in the past for reproduction with magnetic tapes or audiocassettes, they were used as a noise reduction system (Dolby system etc.).

In case the level measurement is the RMS, the typical parameters for operation both as compressor and expander are $\tau_a = 5$ msec and $\tau_r = 130$ msec. As for the value of the time constants for gain smoothing, typical values for Attack are $\tau_A = 0.16, ..., 5, ..., 2600$ msec, while for Release they are $\tau_R = 1, ..., 130, ..., 5000$ msec.

Compander noise-reduction devices The compander is a noise-reduction device mainly used in analog recording based on compression and expansion. The operating principle is quite simple.

As shown in Fig. 8.39, the signal, before being recorded or transmitted, is compressed so that the low-level part is moved away from the background noise of the recording or transmission equipment. During the listening phase the signal is expanded, with a characteristic curve complementary to the compression, so that it is brought back to the original dynamics. Since the background noise of the tape (or transmission equipment) is at a lower level than the useful signal, the expansion operation decreases the level even more by increasing the S/N. Noise reduction systems require source coding before storage and decoding after reading. With this methodology it is possible to obtain an improvement in the S/N even greater than 10 dB.

In commercial devices based on the compander principle (e.g. Dolby, DBX, etc.) filters (weighing) with response curves derived from psychoacoustic models that tend to increase the S/N in the frequencies where the ear is more sensitive are inserted in the coding/decoding phase.

More sophisticated compenders are multi-bands, for example the Dolby A system is implemented in 4 sub-bands.

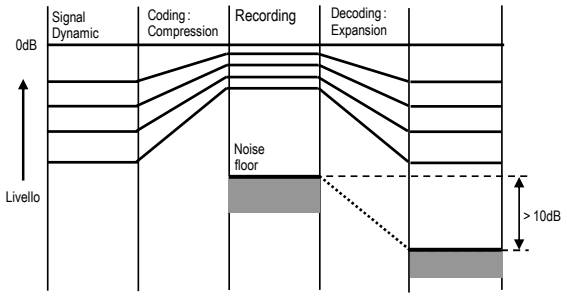

Fig. 8.39 Multiband DRC. Each channel has independent, user-definable settings. 10 100 1kHz 10kHz Hz.

8.7.6.3 Loudness control

Automatic level adjustment to reduce the difference between weak and strong parts. It may be necessary to increase the level of weaker sounds in order to increase the perceived S/N. In cars, for example, where weaker sounds would be covered by background noise, or to "move away" from the noise floor due to the recording equipment.

Remark 8.7. Note that, in television broadcasting, there are very specific standards in transmission levels. In addition, inconsistent levels can be deeply annoying to viewers; therefore this issue was, and still is, considered a very critical technical aspect to deal with when managing the large variety of program genres typically handled by broadcasters nowadays [53].

XXX
ESPANDERE MONO MULTI CANALE FIGS
XXX

8.8 Effects Based Time-Variants Fractional-Delay Lines

The *time variant fractional delay lines* (TV-FDLs) (see §5.5.5), represent the fundamental element for the realization of audio effects and sound synthesis algorithms. They also form the basis for very important acoustic models.

The main effects based on time delay lines variants are: *Vibrato*, *Flanging*, *Chorus*, *Phasing*, *Leslie*, etc. In the next paragraphs these effects will be defined and the basic algorithms for their realization will be presented.

8.8.1 Angular Modulation with TV-FDL and Vibrato

Before analyzing the various effects, let's briefly see how with a TV-FDL it is possible to realize phase and frequency angular modulations.

In the case of discrete-time (DT) signals, called $x_m[n]$ the modulating signal and $x[n]$ the modulated or carrier signal, the generic modulation can be expressed by a non-linear law of the type

$$y[n] = f(x[n], x_m[n])$$

where the function $f(\cdot)$ is defined as the *modulation law* of and is such that there is always a function $g(\cdot)$ for which the following applies

$$x_m[n] = g(x[n], y[n])$$

where the function $g(\cdot)$ expresses the *demodulation law*.

Phase modulation (PM), for example, can be expressed as[2]

$$y[n] = x_{PM}[n] = x[n - k_p x_m[n]] \tag{8.21}$$

with k_p defined as phase modulation constant. If in Eqn. (8.21) we set $D[n] = k_p x_m[n]$, it is easy to observe how phase modulation can be obtained very simply by means of a time variable delay line.

[2] The phase modulation can also be expressed as a convolution between the carrier $x[n]$ and an time-variant impulse response defiend as $h[n] = \delta[n - x_m[n]]$; therefore we have that $x_{PM}[n] = x[n] \star \delta[n - x_m[n]] = x[n - k_p x_m[n]]$.

Since $D[n]$ is generally a continues function, it can be expressed as a whole part plus a fractional part i.e. $D[n]$ is TV-FDL. The modulating signal can be realized by means of a so called *Low Frequency Oscillator* (LFO) as shown in Fig. 8.40-a).

The LFO oscillator can generally switch to various waveforms, whether deterministic: sinusoidal, triangular, linear sweeps, exponential sweeps, etc. both stochastic: low-pass filtered noise and so on.

To understand more deeply the problems of angular modulations let's consider the simple case of sinusoidal carrier i.e.

$$x[n] = \cos(\omega_0 n) = \cos\phi[n]$$

with $\omega_0 = \pi f_0$ and $\phi[n]$ defined as *instantaneous phase*.

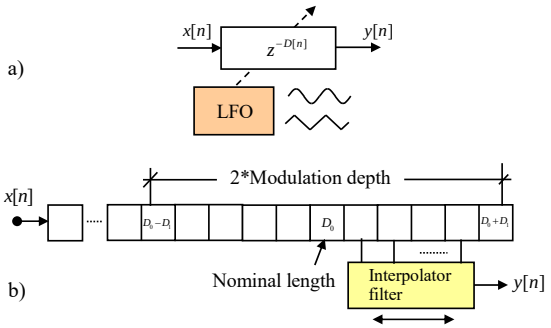

Fig. 8.40 Time-variant fractional delay-line (TV-DFL) which realizes a phase modulation: a) Low Frequency Oscillator (LFO) modulator; b) Realization scheme.

8.8.1.1 Phase modulation

In the PM we have, by definition, that the instantaneous phase is proportional to the modulating signal, i.e.

$$\psi[n] = \psi_{PM}[n] = 2\pi f_0 n + k_p x_m[n]$$

the expression of phase modulation then turns out to be

$$x_{PM}[n] = \cos(2\pi f_0 n + k_p x_m[n]) \tag{8.22}$$

Defining the *instantaneous frequency* $\omega[n]$ as a phase change[3], in the case of discrete-time, indicating with ∇ the differentiation operator, this applies to

$$\omega[n] = \omega_{PM}[n] = \omega_0[n] + k_p \nabla\{x_m[n]\}. \tag{8.23}$$

The above expression can be interpreted in terms of *frequency deviation* Δf_{PM} around f_0 . It turns out then

$$\Delta f_{PM} = \frac{k_p \nabla\{x_m[n]\}}{2\pi}. \tag{8.24}$$

[3] In the case of continuous time the angular or pulsation velocity is equal to $\omega(t) = d\phi(t)dt$.

8.8.1.2 Frequency modulation

In frequency modulation (FM) we have, by definition, that the instantaneous pulse is proportional to the modulating signal; i.e.

$$\omega[n] = \omega_{FM}[n] = \omega_0 + k_f x_m[n] \tag{8.25}$$

so the instantaneous phase is equal to

$$\phi[n] = \varphi_{FM}[n] = \omega_0 n + k_f \sum_{k=-\infty}^{n} x_m[k]$$

Therefore, the expression of the FM signal is

$$x_{FM}[n] = \cos(2\pi f_0 n + k_f \sum_{k=-\infty}^{n} x_m[k]) \tag{8.26}$$

From Eqn. (8.25) frequency deviation Δf around FM the f_0

$$\Delta f_{FM} = \frac{k_f x_m[n]}{2\pi}. \tag{8.27}$$

The term $k_f x_m[n]$ represents the instant pulsation deviation.

Definition 8.2. The quantity m, defined by the ratio between the maximum frequency deviation and the frequency of the modulating signal:

$$m = \frac{f_{\max} - f_{\min}}{f_0} = \frac{\Delta f}{f_0} \tag{8.28}$$

is called a modulation index.

In case the modulating signal is also sinusoidal, i.e. $x_m[n] = \cos(\omega_m n)$ taking into account the definition of modulation index, it Eqn. (8.26) can be expressed as

$$x_{FM}[n] = \cos(\omega_0 n + m\sin(\omega_m n)). \tag{8.29}$$

Remark 8.8. Note that the Eqn. (8.29) can be rewritten as

$$x_{FM}[n] = \cos(m\sin(\omega_m n)) \cdot \cos(\omega_0 n) - \sin(m\sin(\omega_m n)) \cdot \sin(\omega_0 n)$$

from which, through the serial development of Bessel's terms e $\cos(m\sin(\omega_m n))$ and $\sin(m\sin(\omega_m n))$, it is possible to derive the expression of the FM signal spectrum.

As an example let's consider the angular modulations shown in Figure 9.47, implemented with the scheme of Figure 9.46-b), in which the carrier signal $x[n]$ is sinusoidal while $D[n] = kx_m[n]$ is triangular.

From the figure it can be seen that in the case of phase modulation, when the reading velocity increases, the output frequency is higher and is proportional to the slope of the curve.

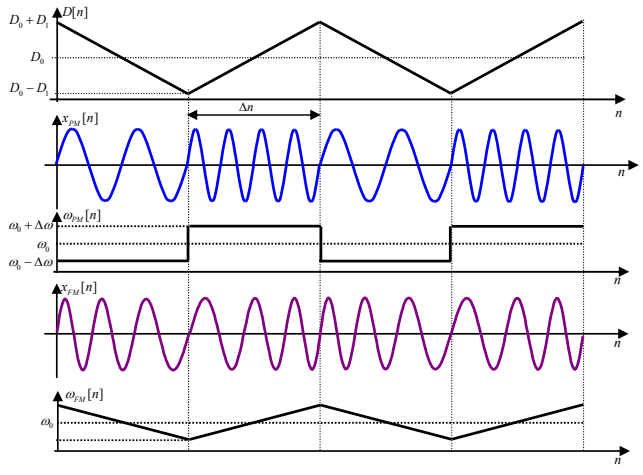

Fig. 8.41 Example of phase and frequency modulation with sinusoidal carrier and triangular modulating signal ($\omega_0 = 2\pi\frac{3}{100}$); $D_1 = 10; \Delta 1000$.)

In particular in this case we have that

$$\Delta\omega = \frac{\Delta D[n]}{\Delta n} = \frac{D_1}{\Delta n}$$

In FM, the frequency trend is identical to the $D[n]$ signal.

8.8.1.3 The Vibrato

The *Vibrato effect* is realized with a simple low frequency modulation using the TV-FDL scheme in Fig. 8.40.

In practice, the effect is obtained with a sinusoidal modulating signal. Sometimes other waveforms can be used to obtain particular effects: triangular, square, etc. or even random signals.

For the evaluation of the modulation index m, necessary to obtain a certain depth of effect, it is appropriate to express frequency deviation in terms of fractions of octave b_w where, by definition, the distance between two frequencies is expressed in terms of the ratio

$$f_1/f_0 = 2^{b_w}.$$

The maximum and minimum frequencies of the modulated signal will then be $f_{\max} = f_0 2^{b_w}$, $f_{\min} = f_0/2^{b_w}$. From the definition Eqn. (8.28) it is easy to express the modulation index m as a function of b_w as

$$m = \frac{f_{\max} - f_{\min}}{f_0} = 2^{b_w} - \frac{1}{2^{b_w}}$$

if we want to obtain a vibrato that oscillates of k semitones around a certain frequency[4] f_0 we will have that the band, in fractions of an octave, will be equal to

[4] The distance of a semitone corresponds to a ratio of $b_w = 1/12$ of octave. That is, the ratio between two frequencies one semitone apart is equal to $f_1/f_0 = 2^{1/12}$.

$b_w = k/12$; that is to say

$$m = 2^{k/12} - \frac{1}{2^{k/12}}$$

For example, for $k = 1$ (plus or minus one semitone) the modulation index is equal to $m = 0.1156$. It follows, with reference to Fig. 8.40, that the nominal length of the delay line D_1 will be proportional to m to the sampling rate f_s with the law

$$2D_1 = m\frac{f_s}{f_0}.$$

In the example above considering $f_s = 44.1$ kHz, $f_0 = 20$ Hz the modulation depth is $D_1 = \frac{44.1 \cdot 10^3}{2 \cdot 20} 0.1156 = 127.45$. In this case the nominal length of the line could be $D_0 = 128$ while its length in terms of memory occupation will be 256 samples.

Note that the length of the delay line should be dimensioned on the basis of the maximum modulation index on the minimum frequency of the signal.

Remark 8.9. For higher frequencies the length, in terms of samples, of the modulation depth is very small. If in the example above we consider an $f_0 = 2$ kHz, we have that $D_1 = 1.2745$. It is very important in these cases to have an interpolator filter with order 5 or 7 at least.

8.8.1.4 The Flanging

The *Flanging effect* is achieved by mixing the signal with its delayed copy with a time-varying delay[5].

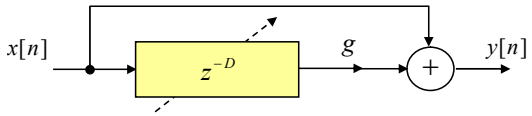

Fig. 8.42 Simple flanger scheme made with a non-recursive comb filter.

The length of the delay line generally varies from 1 to 10 ms, delays of more than $50-70$ ms would be audible as an echo, therefore you do not hear an echo but a timbre modulation due to the variation in frequency response introduced by the delay modulation.

The diagram in Fig. 8.42 is in fact equivalent to that of the comb filter already described in §5.2, which has a number of zeros (notches) equal to $D/2$. The output of the flanger is given by

$$y[n] = x[n] + gx[n - D[n]]$$

[5] It is said that the *flanging effect* was originally discovered by The Beatles during the production of an album. A tape recorder was used to make a delay (probably to be used as an echo) when someone touched the flange of the reel (hence the name *flanging*) by changing the pitch. The characteristic sound is therefore due to the mixing of the original signal with its delayed and variable pitch version.

where $D[n] = M[n] - \alpha[n]$ it is given by the sum of an integer part and a fractional part. In the case of delay line modulation, the variation law generally can be expressed as $D[n] = D_0\,(1 + m_D f_D[n])$ (see §5.5.5, Eqn. (5.66)). In the case of sinusoidal modulation we then have

$$D[n] = D_0\,[1 + m_D \sin(2\pi f_{FL} n)] = D_0 + D_1 \sin(2\pi f_{FL} n).$$

The frequency f_{FL} defines the rate of spectral variation and generally has a frequency in the order of Hz or lower. D_0 represents the average density of the notch filters or the average length of the delay line (see Fig. 8.40-b). The parameter $D_1 = m_D D_0$, (defined in [47] as `CHORUS_WIDTH`) limits the maximum excursion of the *Depth* line, defined as

$$D_{epth} = [D_{max} - D_{min}] = 2D_1.$$

The frequency response of the flanger, shown in Figure 9.49-a), is typical of the comb filter with $D/2$ zeros that are evenly spaced on the frequency axis.

8.8.1.5 Flanging reversed

In case $g < 0$, the frequency response, shown in Fig. 8.43-b), is inverted and has resonances (or peaks). In this case the effect is more of a "high pass" type. For $g = -1$, in fact, we have that $y[n] = x[n] - x\,[n - D]$ that for $D = 1$, corresponds to a first order differentiator circuit.

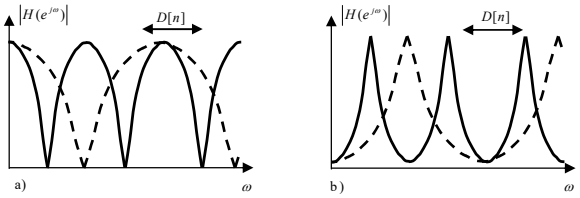

Fig. 8.43 Flanger spectrum variation due to modulation of the delay line length: a) with $g > 0$; b) with $g < 0$.

8.8.1.6 General effect scheme based on delay line

A variant to the basic scheme previously discussed is the one proposed by Dattorro in [47] and shown in Fig. 8.44. In this case a signal is taken from the delay line and fed back to the input through a feedback gain g_{fb}. The feedback signal is taken at a fixed delay of length K which is usually equal to the average value of the variation of the line length; therefore $K = D_0$. Note that, in [47] the term K, which corresponds to the central tap of the delay line, is defined as `NOMINAL_DELAY`.

The other parameters of the structure are the *feed forward gain* g_{ff} and the *blend gain* g_{bl}, such that the network function of the structure described in Fig. 8.44 results to be

$$H(z) = \frac{g_{bl} + g_{ff} z^{-D[n]}}{1 + g_{fb} z^{-K}}.$$

The sound you get, for high gfb values, is metallic and intense. In this case, attention must also be paid to the stability of the system, which, if not satisfied, would lead to overflow and/or clipping errors.

The structure of Fig. 8.44 is of a general type and can be used for the realization of various effects.

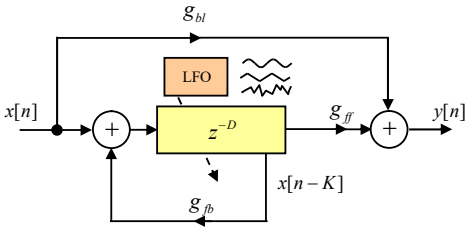

Fig. 8.44 Standard effect of commercial type with variable length fractional delay line. The LFO signal can be deterministically shaped or simply low-pass filtered noise.

8.8.1.7 Stereo Flanging

Another variant of the flange effect is the extension to the stereophonic case. If the source is already stereophonic (or multi-channel) in nature, the effect can be applied independently on each channel.

Very often, as in the case of some musical instruments, the source is monophonic and you want to obtain a stereophonic effect. Thus, a simple variant, easily implemented, consists in modulating two independent flangers through the same LFO signal but with signals phased by 90° as shown in Fig. 8.45.

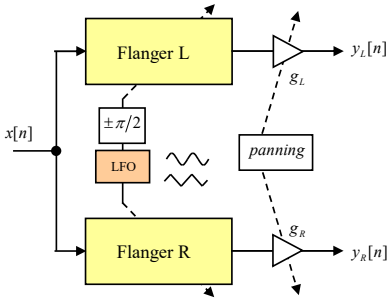

Fig. 8.45 Flanger with monophonic input and stereo output.

The quadrature modulation, taking advantage of the *precedence effect* (see §2.4.1.3), dynamically alters the positioning of the source on the stereo front in a pleasant way and, in the case of a monophonic source, an artificial stereo field is generated.

For these reasons it is generally useful to have additional control over the output levels of the effect, such as *stereo panning* or *pan-pot*, and the ability to reverse the modulation phase.

8.8.1.8 The Chorus

The *Chorus effect* consists of the sum of at least two voices "singing" in unison. Each of the voices will be affected by random microvariations (delay, pitch, amplitude, etc.). As a consequence of the beats due to random modulations, the sound is richer and more evocative.

The basic structure to realize the chorus effect, illustrated in Fig. 8.46, is very similar to that of the flanger but repeated for each voice.

In case of only one delay line, and without feedback ($g_{fb} = 0$), for a nominal delay around $D_0 = 20$ms, the corresponding algorithm is known as *Duobling effect* and consists in duplicating the voice.

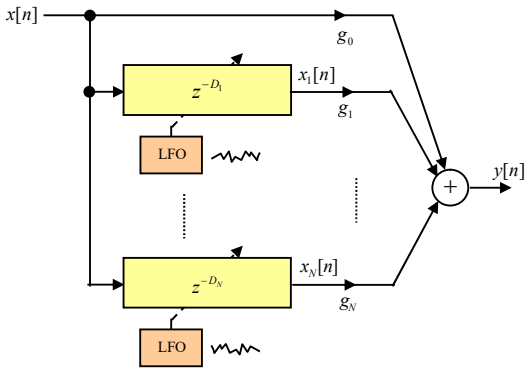

Fig. 8.46 Principle diagram of the Chorus effect. For a richer and "bigger sound" it is usual to use random modulating signals

As shown in Table 8.3, which shows the typical settings of some effects that can be achieved with delay lines, the Chorus parameters are different from those of Flanging where the delay lines are generally longer and the greater the modulation depth.

For a good chorus effect, for example, Dattorro in [47] suggests for a sampling frequency $f_s = 44.1$ kHz, a length or `NOMINAL_DELAY`, equal to 400 ($D_0 \approx 9$ ms), an excursion or `CHORUS_WIDTH`, equal to 350 samples ($D_1 \approx 8$ ms) and a modulation frequency of about 0.15 Hz.

The Chorus effect adds richness, presence and thickness to the sound that seems to be obtained by superimposing several voices or several instruments played together

Table 8.3 Typical parameter values for effects with varying time delay lines.

	g_{bl}	g_{ff}	g_{fb}	Onset	Depth	Mod
Vibrato	0.0	1.0	0.0	0ms	0 - 5ms	0.1 - 5 Hz sinus
Flanger	0.707	0.707	–0.707	0ms	1 - 10ms	0.1 - 1 Hz sinus
Chorus	1.0	0.707	0.0	1 - 30ms	5 - 30ms	Lowpass noise
White Chorus	0.707	1.0	0.707	1 - 30ms	5 - 30ms	Lowpass noise
Doubling	0.707	0.707	0.0	10 - 100ms	1 - 100ms	Lowpass noise
Echo	1.0	≤ 1.0	<1.0	50 - ∞ ms	80 - ∞ ms	-

For the *Doubling effect*, we mean a superimposition of the same voice made by means of the overdubbing technique. The singer records a second voice by superimposing it on the first (usually listening to the first voice on headphones as a reference). The two recordings, even if made by the same singer and in the same key, are affected by micro variations that to produce a stronger or "bigger sound" than can be obtained with a single voice or instrument. The resulting effect is known as Doubling effect.

Remark 8.10. When designing the Chorus effect, great attention must be paid to the type of interpolator circuit. A linear interpolator, in fact, in addition to being time variant, has distinct low-pass characteristics making a "closed sound". An all-pass interpolation, as discussed in [48], could also be critical: the distortions introduced would be audible as "lack of transparency". In general, however, when all-pass interpolators and negative feedback are used, the Chorus is white.

Remark 8.11. Note that, before the introduction of delay lines (analog or digital) the choral and vibrato effect was introduced in the early 1940's in Hammond organs using the so called "vibrato scanner".

The vibrato scanner consists of two elements. The first called *vibrato phase-shift line* is composed of a cascade 2-port *LC* low-pass filter sections, that form an analog delay line. The second one is a rotating switch, called *scanning pick-up*, that travel along the line and thus encounter waves increasingly delayed in phase at each successive tap, and the signal it picks up will continuously change in phase creating a particularly warm and rich vibrato. The rate at which this phase shift occurs will depend on how many line sections are scanned each second.

8.8.1.9 Phasing

The Phasing effect is closely related to the Flanging effect as it is also based on the shifting of the frequency of a notch filter. Very often, in fact, there is confusion between the two terms in commercial devices.

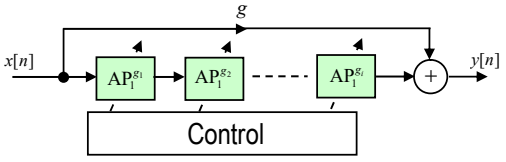

Fig. 8.47 Phaser circuit schematic diagram.

Unlike the flanger where the notches are evenly spaced (by effect of the comb filter), the phaser effect consists of modulating some notch filters that are not evenly spaced. To obtain separate control of the position of the filter poles, instead of the simple delay line of the comb filter, it is possible to replace the z^{-1} element with an all-pass cell of the type

$$z^{-1} \rightarrow \frac{g_i + z^{-1}}{1 + g_i z^{-1}}.$$

The resulting structure therefore consists of a chain of $1st$- or $2nd$-order all-pass sections as shown in Fig. 8.47.

The overall delay due to the all-pass delay line is not constant but dependent on frequency. As a result of the frequency and phase response of the all-pass cells we have that at the output some bands of frequency (or components) will have a phase (and consequently a delay) greater than others, hence the name "Phasing". Using a second order filter you can act more selectively on the group delay.

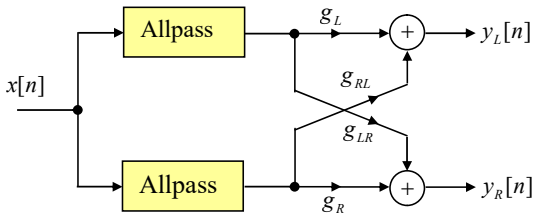

Fig. 8.48 Principle diagram of the stereo phaser circuitry.

As with Flanging, and generally for all other effects, you can generalize phasing to multi-channel audio. For the two-channel case a possible scheme, widely used in practice, is the one shown in Fig. 8.48.

8.8.2 Amplification Systems with Rotating Speakers (Leslie)

One of the most popular effects, used mainly with electronic and electro-mechanical organs, is made by means of an amplifier equipped with "Rotary Speakers". This device, also called "Leslie" from the name of its inventor (Don Leslie), is a real electromechanical audio processors able to modulate acoustically in a very characteristic and suggestive way the sounds of the Hammond organ and more rarely of other instruments (guitars, electric piano, etc.). Thus, according to [48], the Leslie it is not a "Hi-fi" speaker, but rather a part of a musical instrument.

The Leslie consists of a tube amplifier and a two-way speaker with a crossover frequency around 800 Hz. The schematic diagram of one of the most popular models (Model 122) is shown in Fig. 8.49. For low frequencies, a closed box speaker and a cylinder rotating around its axis are used. The cylinder, usually made of wood, has a side opening to diffuse the sound with an horn-section. The high frequencies are diffused by a rotating horn. Due to the high directivity of the horn, compared to a fixed listener, the sound obtained is affected by a considerable variation in volume.

Its rotation speed also determines a Doppler effect. For an accurate formulation of the Rotary Speaker model, a few considerations should be made:

- The sound comes out of one horn, the other is closed and serves only as a counterweight to cancel the centrifugal force;
- The characteristic Rotary Speaker sound is achieved by appropriately placing microphones near the horns. Usually two microphones are used for the high frequencies and only one for the low frequencies.

With reference to Fig. 8.49 where is shown a schematic model with a single pickup microphone, neglecting for simplicity the phenomena of reflection and refraction, the rotation of the horn determines at the ends of the microphone an amplitude modulated signal - the signal will be maximum when the microphone is aligned with the horn $\psi = 0$ - and a frequency modulation, by Doppler effect, whose deviation will be maximum at the maximum relative speed of the horn with respect to the microphone or for $\psi = \pm\frac{\pi}{2}$.

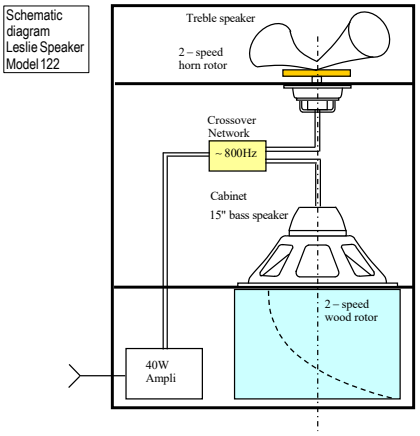

Fig. 8.49 Schematic diagram of the Leslie Mod. 122 amplifier designed exclusively for amplifying Hammond organs.

The Leslie has two rotation speeds denoted as *choral* and *tremolo*. In choral mode the rotation speed is low 15-120 rpm in tremolo mode you have a speed of 300-500 rpm. The musician can switch these speeds by means of a command. The rotating horn speed is 0.35-2.8 m/s in choral mode and 7-11.7 m/s in tremolo mode.

8.8.2.1 Doppler effect simulation with delay line

The Doppler effect (see §1.8.4.4) consists in the variation of pitch perceived by a fixed observer (listener) in the case of moving acoustic sources according to the following relation

$$f_a = f_s \frac{1}{1 \pm \frac{v_s}{c}} \tag{8.30}$$

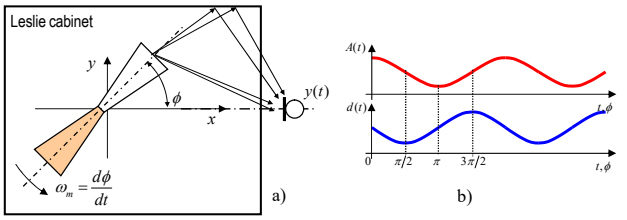

Fig. 8.50 Principle diagram for determining the Leslie model with a single microphone: a) model geometry; b) amplitude envelope and frequency deviation trend.

where f_s and f_a represent respectively frequency of the source and the perceived frequency by the fixed listener, v_s the speed of the source and c the propagation speed of the acoustic wave. The "+" sign is valid in case the source approaches the listener (a higher frequency is perceived), the "-" sign in case it moves away. Considering for example a note at 800 Hz at a speed of 11.7 m/s we will have a frequency deviation of about 55 Hz which corresponds to a frequency modulation index $m = 0.068$.

To simulate the Doppler effect using a delay line (see [4] and [50] for more details) simply change the line readout pointer in accordance with the expression (8.30) and (8.24) so the change in the delay line readout pointer is proportional to the Doppler frequency deviation. For which we have that

$$\nabla\{D[n]\} \propto \frac{f_s - f_a}{f_a} = \pm\frac{v_s}{c}$$

Usually three microphones, two for the horns and one for the bass rotor, are used for the Leslie, arranged as shown in Fig. 8.51.

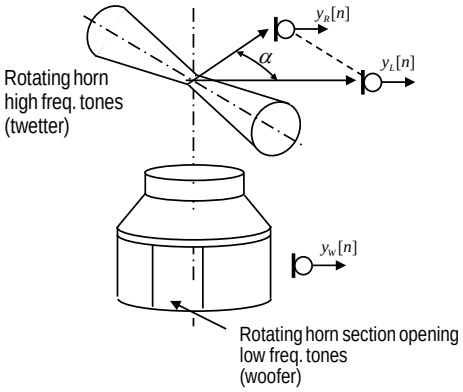

Fig. 8.51 Typical three-microphone shooting technique for Leslie with two rotors: two for the horn and one for the bass rotor.

Only one microphone is used for the rotor because the low frequencies are less directional and the Doppler effect is less evident.

The diagram for the determination of the model for the twitter-horn only is shown in Fig. 8.52-a while Fig. 8.52-b) shows the amplitude and frequency deviation acquired by the two microphones.

To simulate a Leslie effect, a frequency modulator that simulates the Doppler effect, made with a TV-TDL, and an AM amplitude modulation with a modulating signal of the same frequency as the FM modulator but delayed by an angle of $\pi/2$, must be inserted for each channel.

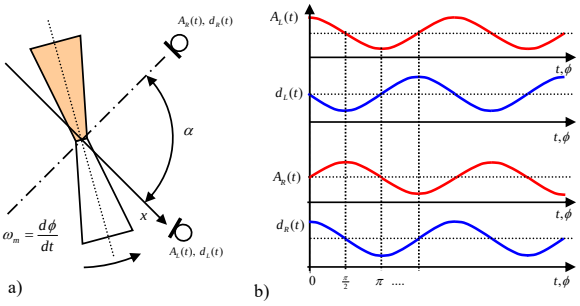

Fig. 8.52 Principle diagram for the determination of the Leslie model with two microphones with angle α: a) geometric model; b) envelope and frequency deviations trend.

A possible simplified scheme for creating the Leslie effect is shown in Figure 9.59. Note that in the circuit in the figure the modulating signal is equal to $\cos(\omega_m n)$ with ω_m angular velocity of the horn. The TV-TDL is in fact phase modulated so the frequency deviation is for the (8.24) proportional to its derivative. Between the FM modulating and the AM modulating there is then an angle equal to $\pi/2$.

For a more accurate model it is possible to replace the modulators with time-variants TFs (with the law of the modulating signal) that take into account the sound propagation model (at least the first reflections) that comes out of the horn and propagates inside the cabinet. This model could, for example, be calculated using one of the simulation techniques for listening rooms described in Chapter 3, or obtained with real measurements such as those shown in Fig. 8.54 for a Leslie Model [50].

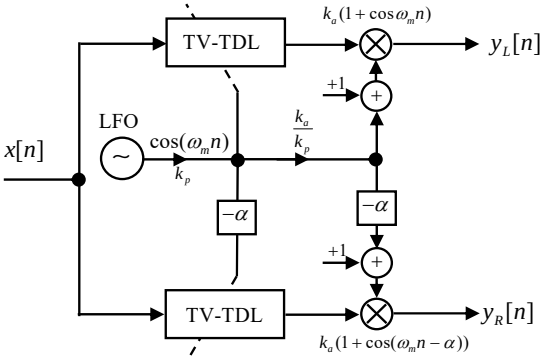

Fig. 8.53 Simplified diagram of the model that implements the Leslie amplifier's rotary speaker effect.

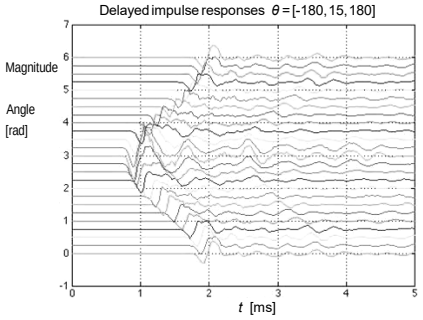

Fig. 8.54 Free field impulse responses of the Leslie Model 600 measured for twetter-horn angles $\theta \in [0\ 2\pi]$ with step $\pi/12$. (Courtesy of [50]).

As for the bass section, not reported for simplicity, it is sufficient to consider only amplitude modulation.

Another important aspect for the determination of the Leslie's timbre concerns the rotation speeds of the horn, the cylinder and the duration of the transient relative to the change of speed between the choral and tremolo effect.

Fig. 8.55 shows a principle diagram of the simulator complete with rotor speed control device.

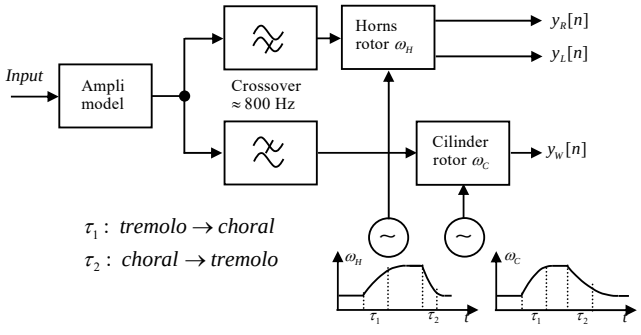

Fig. 8.55 Scheme of rotary speakers simulator with speed and time constant control.

Remark 8.12. Rotary speaker systems have been widely used in the past and many models were produced and with numerous variations (shape, size, speaker rotation technique, number of ways, etc.) by various companies around the world.

It is obvious that the diagrams presented above can be modified according to the device to be simulated. In particular, in order to obtain the simulation of a specific model it is necessary to take into account: the crossover cut-off frequency (for 2-way loudspeakers); the type of cabinet; the distance and position of the microphones; the distortion model of the amplifier (the "sound" has very different characteristics in the case of tube or transistor technology); the rotor and horn speeds; the duration of the transients in the speed change, etc.

8.9 Effects Based on Time-Frequency Transformations

The need to change key is a very common and simple solution in modern Western musical notation: raising or lowering the key of a song simply means moving the positions of the notes up or down on the pentagram of the desired amount. The performer, whether real or virtual (e.g. a MIDI file), can play the song more or less quickly regardless of the key used.

In the case of signals, the transposition operation, called pitch-shifting, leaving the length unchanged, is a difficult problem to formalize and does not provide simple solutions in closed form. The dual problem consists in the possibility of varying the time scale of a signal by compressing or expanding it without changing the timbre (or spectral) structure of the sound. This type of processing is called time transposition or, in jargon, time-stretching. Thus we can consider the following definitions

- *Time-stretching* - is the operation of changing the length of a signal, without affecting its spectral content,
- *Pitch-shifting* - is the operation of raising or lowering the original pitch of a sound without affecting its length.

Changing the pitch of a sound without changing its duration or the possibility to compress/expand time without changing its pitch may be necessary in many situations. In post-productions, for example, you may need to insert a voice comment or a soundtrack into a specific time slot. Think for example of commercials, movie voiceovers, dubbing phases, etc.

The possibility of time-stretching and/or pitch-shifting, in recent years, has encouraged the expansion of new frontiers in musical composition. Think of remixed songs where we have collages of sampled song parts from different authors. Each part will be characterized by a key and a time duration: the composition of the various songs can be performed acoustically and artistically interesting only if you make a normalization of the keys and juxtapose the various parts that must have a precise rhythmic structure.

Time-frequency transformations (TFT) are also widely used in studio recorded music post-productions. These techniques, in fact, can be useful to perform many operations such as, for example, correction of errors of musicians or singers, insertion of special effects, etc..

The frequency and time transposition algorithms are very numerous and different methodologies are available in the scientific literature. Here, while not providing an exhaustive view of the problem, we want to highlight some of the fundamental problems inherent to this topic.

The main problem in building a height transporter is to determine exactly what height it is. The height, in fact, is a quantity that cannot be precisely defined by context. For example, the fundamental frequency of an organ pipe can be 20 Hz: that is the height. A right-hand pianist can play a note 20 times per second: that is tempo. Both these events are represented by a fundamental frequency of 20 Hz, however one is considered pitch and should be translated, the other is considered time and should not be modified by the algorithm. This is why a height transposer must essentially choose a frequency that arbitrarily divides height from time.

The main problem in building a height transposer is to determine exactly what height it is. The height, in fact, is a quantity that cannot be precisely defined within the context. For example, the fundamental frequency of an organ pipe can be 20 Hz: that is the height. A right-hand pianist can play a note 20 times per second: that is tempo. Both these events are represented by a fundamental frequency of 20 Hz, however one is considered pitch and should be transposed, the other is considered time and should not be modified by the transposition algorithm. So a height transposer must essentially choose what is the height, and what is the time.

Another problem encountered in the design of a height transposer involves the duality time-frequency of audio signals. To determine a frequency of a signal, this must be observed for at least one period of the waveform. So it is impossible to determine whether a signal contains precisely that frequency at that precise time. The more accurately the frequency is known, the less accurately the time in which it occurs is known, and vice versa: this is the fundamental *principle of uncertainty* that governs the time-frequency relationship.

When a tone is played at half speed, its pitch is shifted downwards by a factor of 2 (i.e. an octave). In order to re-establish the height, the transposer must adjust the frequencies in a complementary way; in this case, the frequencies must be multiplied by 2. The height translation process is a process in which frequencies are multiplied by a constant height translation factor. Multiplying the frequencies by 2 will shift one octave up, dividing by 2 will shift one octave down. A factor of 21/12 will translate them by one semitone (12 semitones make an octave). It should also be noted that multiplication by a constant factor is the only way frequency ratios are preserved and therefore is the only possible way to leave the harmonic structure unchanged during a height shift.

8.9.1 Frequency Translation and Compression/Expansion of the Time Scale: Linear and Stationary Case

In the Signal Theory, with the term *frequency translation* is generally understood the analogous theorem (sometimes also called modulation theorem) related to the Fourier Transform [51] (FT)[6] and defined as follows

$$\Im\{x(t)\} = X(j\omega) \Leftrightarrow \Im\left\{e^{j\omega_0 t}x(t)\right\} = X(j(\omega - \omega_0))$$

Multiplication of the $x(t)$ signal by the complex exponential $e^{j\omega_0}$ is equivalent to shifting its spectrum by a factor of ω_0.

To understand the effect of frequency translation on a sound, let us consider Fig. 8.56. From the figure we can see that the signal spectrum is moved up "as is" by the quantity ω_0.

[6] Remind the reader that the FT of a signal, indicated as $X(j\omega) = \Im\{x(t)\}$, is defined as: $X(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$. The antitransformed, $x(t) = \Im^{-1}\{X(j\omega)\}$, is defined as $x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t}d\omega$.

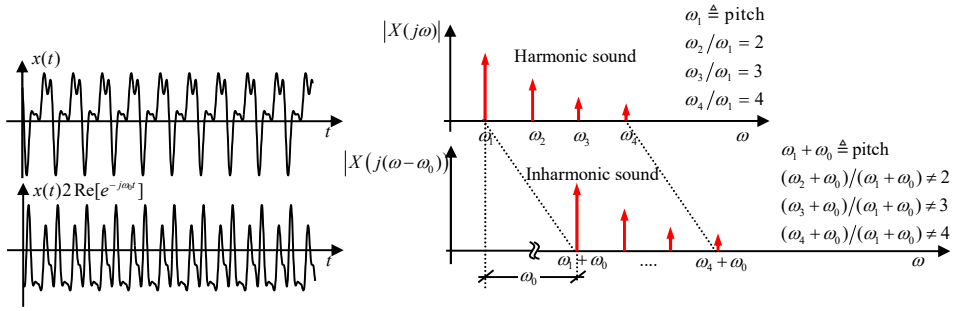

Fig. 8.56 Graphic representation of the frequency translation theorem. In the translated signal the proportion between the fundamental frequency and its harmonics is not respected.

The information contained in the signal is not degraded because the transformation is reversible and it is possible to return the signal spectrum to its original position or, in the Electrical Communications jargon, to the *base band*. This transformation is in fact used in amplitude modulation transmissions: a) where the signal is shifted in frequency to make it compatible with the band of the transmission channel; b) in the so-called multiplexing technique which allows several signals to be combined on the same channel.

In these cases the $x(t)$ signal is defined as modulating while the signal $e^{\pm j\omega_0}$ is defined as carrier (in practical cases it is considered the only real part of the carrier).

However, in the case of musical signals, this transformation cannot be used to raise or lower the key of a piece of music; in fact, the ratio between the fundamental frequency and its harmonics is not respected in the translated signal. In the case of a musical signal this means that the timbral structure of the sound is completely distorted and the resulting sound is said to be “inharmonic". In order to have a correct acoustic structure of the translated sound completely similar to the original sound, the ratio between the harmonics must be unchanged.

To increase the pitch of a piece of music while respecting the harmonic structure, simply play it faster. For example, to increase the key by half-tone you need to increase the sampling rate by a factor of $\sqrt[12]{2}$. In this case the harmonic structure would be unchanged and the timbre of the sound would be the same.

From a theoretical point of view the justification is given by the theorem of the variation of the Fourier transform time scale defined as

$$\Im\{x(t/\omega_0)\} = \omega_0 X(j\omega \cdot \omega_0) \tag{8.31}$$

From the previous expression it is easy to observe how a compression in the time domain corresponds to an enlargement in the frequency domain.

The (8.31) can be interpreted, in fact, as the modification of the reading speed of a musical signal: reading faster, the signal rises in pitch and its duration is reduced.

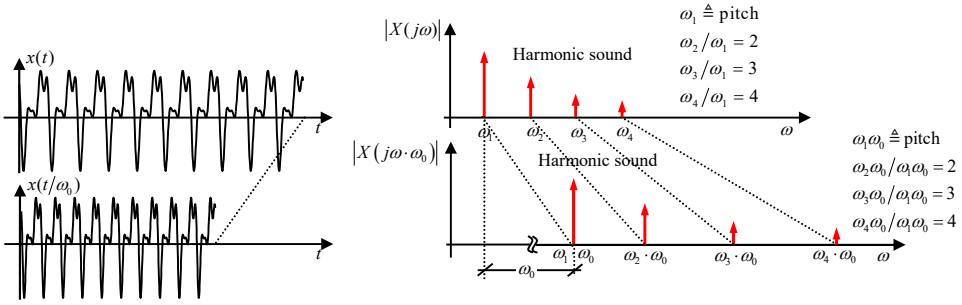

Fig. 8.57 Graphic representation of the time scale theorem. By changing the signal time scale the ratio between the fundamental frequency and its harmonics is identical to the original sound.

8.9.2 Classification of TFT algorithms

Given the ill-posed nature of the TFT problem, very different philosophies are available in literature for the determination of time-stretching and pitch-shifting methodologies; consequently, even the classification of algorithms may not be simple and intuitive. A possible classification, then, can be thought on the basis of the type of implementation used.

- *Off-line* - In signal editing (acoustic corrections etc.) the processing can be done on previously recorded material and the processing time or type of algorithm (batch or on-line) are not important because the implementation is out of time.
- *Real-time* - In cases where pitch transposition is used as an effect during an execution, it is obvious that it is necessary to work in real time with very strict constraints on processing time and group delay. The choice of algorithms is therefore very limited.

In the case of real-time realization it is possible to think about a further classification of the algorithms.

- *Batch and mini batch algorithms* - The most frequent situation is the use of buffers that collect data in batches before providing them for processing. In this way a better processing quality can undoubtedly be obtained, but, on the other hand, the use of buffers implies considerable computational resources due to the possible iterations of the algorithm; moreover, each iteration involves a window delay which, although minimal, multiplied by the number of iterations, is unacceptable for audio effects such as chorusing and harmonization.
- *On-line algorithms* - This type of implementation requires direct communication between the signal source and the computer. Any kind of influential delay is eliminated, so these algorithms are better suited for chorusing and harmonizing changes. Another great advantage of this mode is, in general, a lower computational cost which is reflected only in the indispensable transformation operations.

The time scale transformation algorithms cannot work in real time, having an input duration different from the output duration, but can only be realized in off-line mode. To change the time scale, it is necessary to work on a buffer in which the signal is

inserted and then read again with a higher or lower sampling rate. So, if used in real-time this buffer would empty or go into overflow

On the other hand, a height (or pitch) transposer can also work in real time because the duration of the input is identical to the duration of the output. In this case, however, the problem is to give an exact definition of what pitch (or height) is. As already indicated at the beginning of this paragraph (see also Chapter 2), the perceived pitch of a sound is in fact a non-objective but perceptual quantity that strongly depends on the context in which it is to be defined.

From the point of view of algorithm implementation, the time scale variation can, theoretically, be realized with two blocks in cascade. The first block performs the variation of the sampling frequency and the second one performs a pitch transposition bringing the signal back to its original tone.

As shown in Fig. 8.58, the first block resamples the signal and changes its duration. The change of the time scales can be done with simple interpolation and decimation operations and, in case of non-rational ratios, with an appropriate interpolation system as described in Chapter 5.

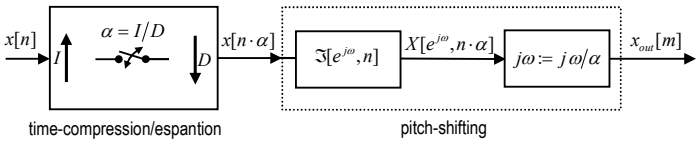

Fig. 8.58 Principle scheme of a time-stretching algorithm. The first block performs the variation of the time scale by resampling; in the second block the sound is adjusted to its original pitch.

8.9.3 Time Domain FTF Algorithms

If you want to expand the time scale without changing the pitch, you must stretch the sound by inserting parts that do not exist on the original signal. The most intuitive method, which is the basis of many algorithms for stretching, is to repeat even multiple signal segments. On the contrary, in the compression problem, parts of the message must be cut acoustically tolerable.

The compression/expansion scheme in Fig. 8.58, in which the expansion or compression of the time scale occurs completely asynchronously with respect to the signal, although consistent from the theoretical point of view, can produce in certain situations results that are not acoustically acceptable or in any case, that can be improved.

To understand some issues related to the most appropriate choice of segments to be eliminated or added in the compression or expansion of the time scale, think about the need to elongate a pattern of a percussive instrument such as drums. The best thing to do in this case is to insert pause segments between two successive hits thus emulating a drummer who accelerates or slows down his performance. With this mode the duration and timbre characteristic of the single drum beat is not modified.

The emulation of the drummer could be done with an algorithm that analyzes the signal in a synchronous way, inserting null amplitude signal strokes between two successive hits. It is obvious that in this case, as shown in Fig. 8.59-b), the structure of the single percussive sound would not be modified, as follows

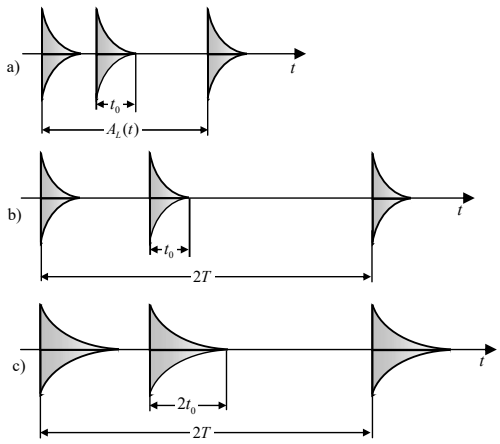

Fig. 8.59 Stretching of a percussive sound: a) original duration; b) synchronous expansion with pause insertion; c) asynchronous extension carried out with the diagram in Fig. 8.58. In this case the duration of the single percussion sound is doubled.

The stretching that can be obtained with the diagram in Fig. 8.58, which is completely asynchronous to the signal, would also lengthen the transient duration of the single shot (see Fig. 8.59-c)). In percussive sounds this phenomenon would make the characteristic of the instrument very distorted and this modification of the signal structure would not be acoustically admissible.

Generally the most complex sounds are not separated by distinct pauses and it is impossible to insert pause stretches to lengthen their duration.

In these situations the stretching of the signal must be done in a more or less "psycho-acoustically" admissible way. In practice, as shown in Fig. 8.60, the algorithms control the periodicity zones within a segment and the elongation is obtained by replicating periodic stretches leaving, as much as possible, the content of transient phenomena for which the auditory system is much more sensitive.

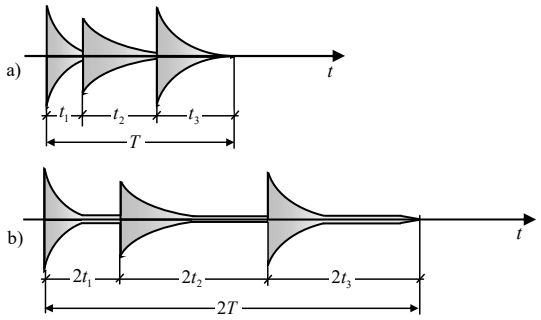

Fig. 8.60 Stretching of a complex sound: a) original duration; b) synchronous stretching with insertion of a periodic repetition tail at the transient end.

8.9.3.1 Rotating read head method with a tape recorder

An analog method for performing frequency time transformations is based on tape recorders equipped with multiple read heads placed on a rotating drum as shown in Fig. 8.61 [55].

The duration of the output signal depends on the absolute speed of the tape. The playback speed depends on the relative speed between the head and the tape. The pitch is shifted in the tape segments "seen" by the individual head during drum rotation. The head outputs are mixed together to produce the height shifted signal.

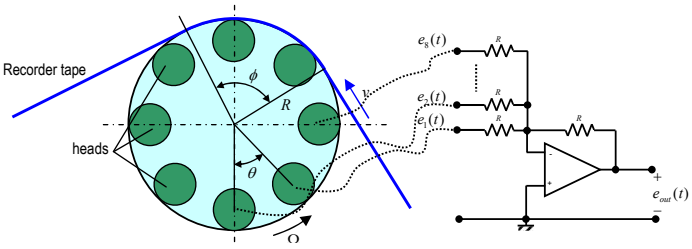

Fig. 8.61 Diagram of an analogue tape recorder with rotating heads. Each head reads the same tape "window" several times.

Time scale reduction is obtained by adjusting the absolute speed of the tape while pitch transposition is obtained by varying the angular velocity of the drum. To have a more regular output the head contact signals are usually mixed together.

In the case of a numerical implementation, it is easy to hypothesize the tape running on a drum as a delay-line where the signal "runs" (running window) and the heads as pointers that repeatedly read the segment present at a different speed from the line feed itself.

8.9.3.2 Periodicity-Detection Algorithm

In [52] Dattorro describes a simple and commercially available method of frequency transposition and time scale compression/expansion (LEXICON model 2400).

The compression expansion of the time scale is obtained simply by changing the sampling rate of the signal while the device works as a pitch-shifter.

Dattorro's algorithm represents, in practice, a numerical version similar to the rotating heads method.

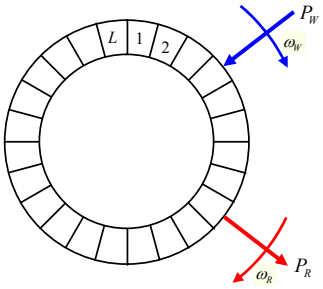

Fig. 8.62 Delay line with circular buffer and two pointers.

The splicing algorithm is realized by means of a delay line with circular buffer (see §5.4) and two pointers as shown in Fig. 8.62. The data writing pointer indicated with P_W rotates with angular speed ω_W, while the other one, indicated with P_R and rotating at one speed ω_R, is used to read the data present on the buffer.

The writing velocity ω_W is regulated by the clock which also controls the A/D and in the realization of Dattorro this varies from 36 kHz to 64 kHz. The ω_R reading velocity, which is also that of the D/A, is kept fixed and equal to 48kHz. Since the speed of the two pointers is different, they will cross periodically. To prevent acoustic discontinuity at the output, before the pointers cross each other, the reading pointer jumps forwards or backwards (for time compression or expansion respectively) by a precise amount of time $T_{R_{1-2}}$.

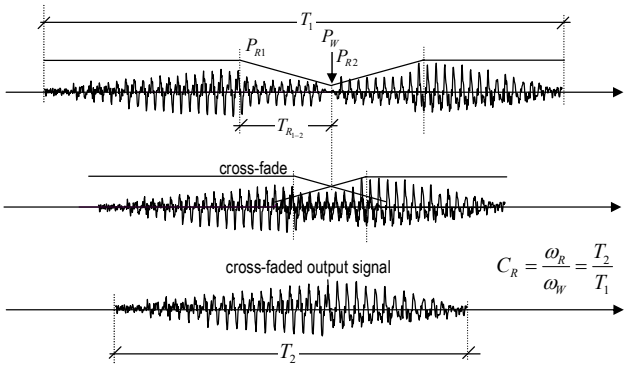

Fig. 8.63 Representation of the cross-fade operation for the reduction of artifacts.

As shown in Fig. 8.63, the new output is mixed (cross-faded) with the old one to filter out any possible discontinuity. The cross-fade curve is adjusted proportionally to the jump distance in order to minimize the discontinuity of the output sound.

The compression ratio C_R is defined by the ratio between the two pointer speeds as $C_R = \omega_R/\omega_W$ and can vary in the range (0.750, 1.333).

To avoid artifacts as much as possible, the amount of the reading pointer jump is not determined arbitrarily. For this purpose a special algorithm, called *Periodicity-Detection Algorithm*, is used, which "looks ahead" for the detection of signal periodicity and optimal mixing.

Time stretching or shrinking, without pitch-shifting is just an illusion, as the original audio is always available at the same speed. Small parts of the input signal, in the order of msec, are in fact repeated in case of expansion, or cut off in case of compression.

8.9.3.3 Synchronous Overlap and Add Algorithm

Designed mainly for the vocal signal [62], the Synchronous Overlap and Add (SOLA) method is based on the segmentation of the input signal into sections of identical length A_i: the technique consists in the overlapping-addition of the various sections that are processed one by one by lengthening or shortening them in order to obtain an output signal of greater or lesser length.

The duration of the synthesis tract is regulated by the relationship:

$$S_i = \alpha A_i$$

where α indicates the scale change factor: for $\alpha < 1$ you get a time scale compression and for $\alpha > 1$ an expansion.

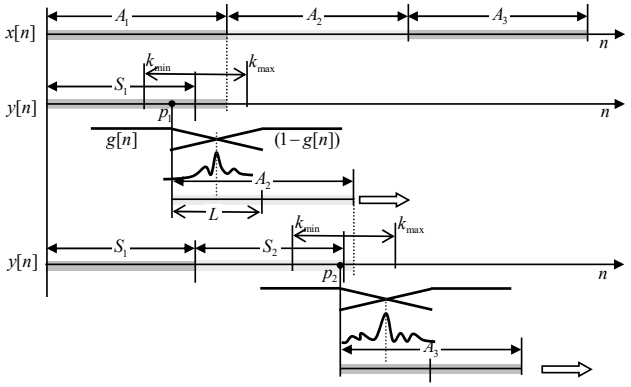

Fig. 8.64 Synchronous Overlap and Add (SOLA) method: time scale compression $\alpha < 1$.

During the compression phase where $\alpha < 1$, as shown in Fig. 8.64 [63], the input signal $x[n]$ is divided into blocks of αN samples A_1, A_2, A_3, ... The first S_1 synthesis block of the output signal $y[n]$ contains the first samples duplicated by A_1, the remaining n samples of A_1 are duplicated in $y[n]$ beyond the S_1 block.

An overlap region ($k_{\min}$, $k_{\max}$) of length L equal to the length of the sequence formed by the remaining n samples is now chosen. The second analysis block A_2 is scrolled through this region in order to find the best alignment point p_1. This point

is determined by calculating the correlation on the overlap region between the output signal $y[n]$ and the analysis block in question.

The block length L, starting at p_1 and extending beyond S_2, is obtained by fading between $y[n]$ and A_2. The new output signal $y[n]$ will then have the block S_2 added. It will be followed by a new sequence of n remaining samples, duplicated by A_2. It continues in the same way for the other blocks

In case the best alignment point is beyond the output signal $y[n]$ (and therefore in the L sequence of remaining samples), the new synthesis block S_n will present some duplicate samples from the A_{n-1} analysis block, a part of samples obtained by fading between $y[n]$ and A_n, and a part of samples from the A_n block. The remaining samples from A_n will be duplicated beyond S_n.

During the time scale expansion, where $\alpha > 1$, as shown in Fig. 8.65 [63], the input signal $x[n]$ is again divided into analysis blocks of N samples A_1, A_2, A_3, ... The first S_1 synthesis block of the $y[n]$ output signal contains the first αN samples duplicated by $x[n]$, including samples belonging to A_1, A_2, and possibly the following blocks, depending on the expansion ratio. Again, the remaining samples of $x[n]$ are duplicated in addition to the synthesis blocks.

The overlap-add procedure is similar to that performed in the case of time scale compression.

The normalized cross-correlation function of the m-th block can be expressed as [63]

$$r_{xy}[n] = \frac{\sum_{k=0}^{L-1} x[mA_i + k]y[mS_i + n + k]}{\sqrt{\sum_{k=0}^{L-1} x^2[mA_i + k] \sum_{k=0}^{L-1} y^2[mS_i + n + k]}} \tag{8.32}$$

where mS_i is the $k_{\min}$ position, L is the length of the overlap region, n is the time index that varies from $k_{\min}$ to $k_{\max}$.

The best alignment point $(n = p)$ is determined by searching for the maximum cross-correlation function $r_{xy}[n]$ in the region between $k_{\min}$ and $k_{\max}$. The output signal $y[n]$ at the m-th block, using the overlap-added function, is

$$y[mS_i + p + n] = g[n]x[mA_i + n] + [1 - g[n]]\, y[mS_i + p + n], \quad 0 \le n \le L - 1 \tag{8.33}$$

the remaining samples are duplicated

$$y\,[mS_i + p + n] = x\,[mA_i + n]\,, \quad 0 \le n \le L - 1 \tag{8.34}$$

Note that the input and output fading function $g[n] = n/L$ and $(1 - g[n])$ respectively, is such that the average gain in the overlap region L is unitary.

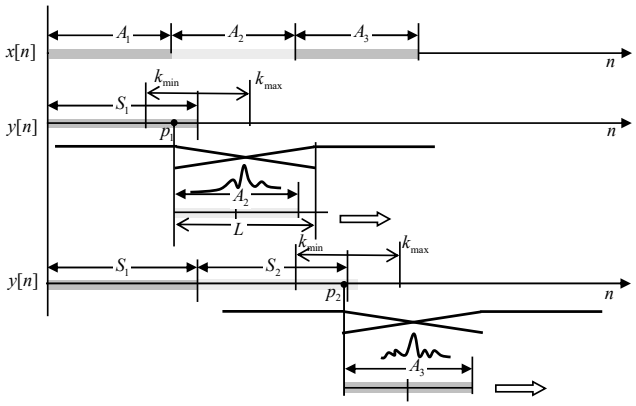

Fig. 8.65 Synchronous Overlap and Add (SOLA) method: time scale compression $\alpha > 1$.

8.9.3.4 Time stretching and pitch transposition with Pitch-Synchronous Overlap and Add (P-SOLA) algorithm

A variant of the SOLA methodology, particularly suitable for vocal signal for small variations in pitch translation, is the one proposed in [63] called *Pitch Synchronous OverLapp-Add* (PSOLA).

The methodology proposed in this algorithm consists in the exact determination of the fundamental frequency of the sound, which is usually identified with pitch, and overlap synchronously with the pitch itself.

For our purpose we assume that the sequence of the pitch period[7] $P[n]$, estimated with a certain algorithm, is evaluated at regular intervals uniformly spaced throughout the duration of the signal $x[n]$ (in intervals equal to the detected fundamental period). Furthermore, these markers, as for example shown in Fig. 8.66, should be positioned at the points where the signal assumes a maximum value. These two constraints are often in conflict, especially because the assumption that the fundamental period is constant for the entire window is not entirely true.

With this technique it is possible to make a time expansion, a pitch translation, the combination of the two alterations with consequent scaling of the formant frequencies (typical characteristic of the vocal signal).

The time extension algorithm consists of two phases: the first phase analyzes and divides the input sound into segments (pitch labeling); the second phase synthesizes an expanded version in time by superimposing and adding the segments extracted from the analysis algorithm.

In the analysis phase, the period of the fundamental frequency $P[n]$ of the input signal and the time instants known as pitch markers or height markers are first determined. The pitch markers are at maximum amplitude at a synchronous rate during the periodic part of the sound, and at a constant rate during the non-vocal parts. In practice, $P[n]$ is therefore considered constant over the time interval (n_i, n_{i+1}). Next, we extract the segment centered at each n_i time mark, using a Hanning window of

[7] The evaluation of pitch extraction techniques is beyond the scope of this paragraph and reference should be made to specific literature such as, for example, [54].

length $L_i = 2P[n_i]$ (two pitch periods), to fade in the additions at the beginning and end of each segment.

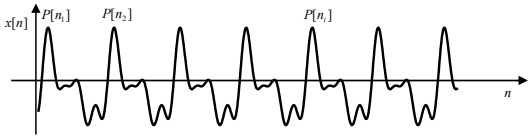

Fig. 8.66 Pitch labeling of the segnal $x[n]$.

8.9.4 Algorithms Operating in the Time-Frequency Domain

In the previous chapter we introduced some algorithms to perform time-frequency transformations operating on the signal directly in the time domain. The fundamental idea of most of these algorithms is in fact based on the techniques of overlap-and-add on stretches of signal addressed by means of pointers and delay lines.

Even if, as previously pointed out, an exact classification of TFF algorithms is not possible, we see in this paragraph some methods to operate in the time-frequency domain.

8.9.4.1 Phase Vocoder

Introduced by Flanagan in 1966 [56], the *Phase Vocoder* (PV) represents one of the oldest methods for frequency time analysis of the vocal signal. Used mainly for voice signal coding [57]-[61], the basic idea to perform the time-frequency transformation is to consider the signal as non-stationary.

In this case the signal spectrum is a function of two variables: time, intended as an index of the time sample n, and frequency represented by the variable ω; and is defined by a "shape" transformation.

$$X(n, e^{j\omega}) = \sum_{m=-\infty}^{\infty} h[n-m]x[m]e^{-j\omega m} \tag{8.35}$$

It is possible, then, to process the signal in the two-dimensional domain (n, ω). Therefore, its implementation can be made considering the following dual approaches.

- *Filter bank PV* - As originally proposed by Flanagan [56], the input signal is decomposed through a filter bank, so each filter sets a frequency and the signals evolve over time. Thus, we can write $X(n, e^{j\omega}) \Rightarrow X_{\omega_k}[n]$.
- *STFT PV*- Proposed in 1976 by Portnoff [59] and defined by Eqn. 8.35) every analysis window fixes the time variable n. So we have that $X(n, e^{j\omega}) \Rightarrow X_n(e^{j\omega})$.

In the case of filters bank for each channel centered around the frequency ω_k, the real and imaginary parts, $a_{\omega_k}[n]$, $b_{\omega_k}[n]$, are extracted. Subsequently these are transformed into a signal $\left|X_{\omega_k}[n]\right|$ calculated as

$$\left|X_{\omega_k}[n]\right| = \sqrt{a^2_{\omega_k}[n] + b^2_{\omega_k}[n]} \tag{8.36}$$

representing the spectral envelope of the frequency ω_k and a signal representing the derivative of the phase $\dot{\theta}_{\omega_k}$ calculated as

$$\dot{\theta}_{\omega_k}[n] = \frac{b_{\omega_k}[n]\dot{a}_{\omega_k}[n] - a_{\omega_k}[n]\dot{b}_{\omega_k}[n]}{a^2_{\omega_k}[n] + b^2_{\omega_k}[n]} \tag{8.37}$$

The typical filter bank PV scheme is shown in Fig. 8.67.

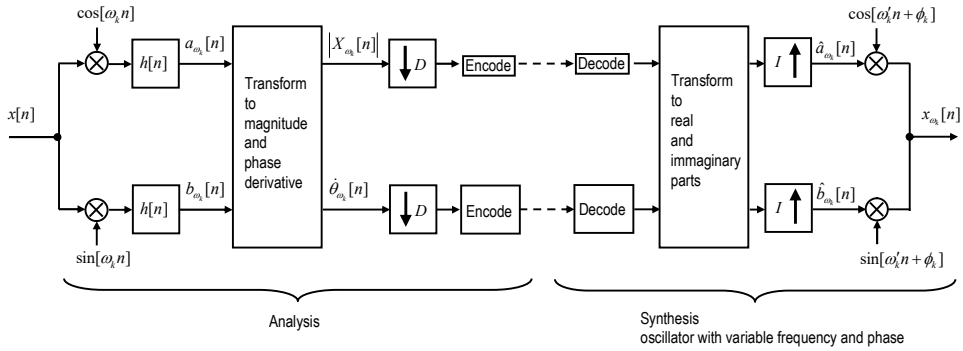

Fig. 8.67 Filters bank Phase Vocoder.

Remark 8.13. Note that, for the phase calculation we prefer to use the derivative quantity rather than the direct one: the direct calculation of the phase as θ_{ω_k}=atan($b_{\omega_k}/a_{\omega_k}$) produces a signal with numerous discontinuities related to the jumps due to the non-monodromicity of the phase. In addition, other numerical problems are possible in cases where $a_{\omega_k} = 0$ (the denominator assumes zero values).

By adjusting the decimation-interpolation rates and frequencies of the synthesis oscillators it is possible to perform both compression-expansion and frequency transformations.

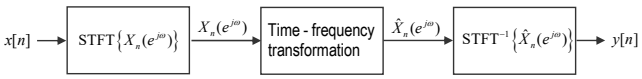

Fig. 8.68 Phase Vocoder implemented with STFT technique.

By fixing the time variable you can implement the Phase Vocoder with algorithms operating directly in frequency domain. In this case, "windowing" a short portion of the signal will fix the time variable and calculate (for that portion) the FFT. This technique, derived directly from the definition of Short-Time Fourier Transform

(8.37), can be used for TFTs according to the scheme shown in Fig. 8.68. In this case each time interval n will be characterized by an instantaneous frequency and phase. For details see the literature on the topic [57]-[61].

8.9.4.2 Subband Analysis Synchronous Overlap-and-Add Algoritm

A modification of the SOLA algorithm, which can also be seen as a PV modification, called *Subband Analysis Synchronous Overlap-and-Add* (SASOLA), proposed in [63], consists of processing the input signal in sub-bands.

Changing the time scale of voice signals using the SOLA technique can produce excellent results, but is considered unsuitable for musical audio signals due to the complexity of the audio waveform in the audible bandwidth between 20 Hz and 20 kHz.

The best alignment point in the overlap region L obtained from the cross correlation function expressed by the Eqn. (8.32) may not always be ideal. This is mainly due to the lack of a single "pitch period" in most audio signals. The SASOLA method solves this problem by dividing the entire signal bandwidth into smaller bands using the sub-band filtering technique. The analysis filters bank divides the wideband audio signal into smaller sub-bands before performing the time scale change. The modified sub-band signals are then reconstructed at the output using the synthesis filters bank.

A schematic diagram of the SASOLA algorithm for changing the time scale is shown in Fig. 8.69.

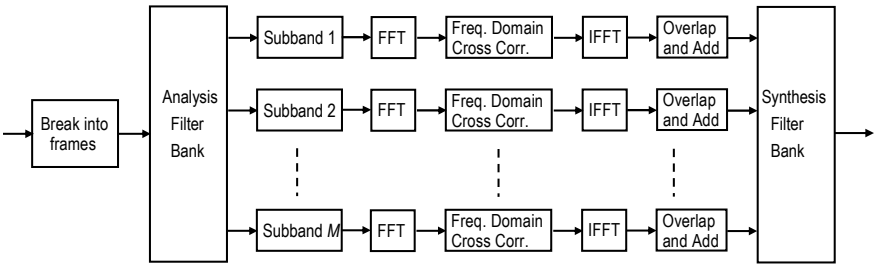

Fig. 8.69 Subband Analysis Synchronous Overlap-and-Add (SASOLA).

In SASOLA the entire bandwidth of the audio or voice signal between 20 Hz and 20 kHz is first divided into smaller sub-bands. Subsequently, the SOLA function is applied to each sub-band signal. The sub-band components are then added together, following the filtering synthesis process, to regenerate the required output signal on a modified time scale.

Remark 8.14. Recent changes aimed at reducing the artifacts of the SOLA method has been proposed in [68] and [69].

Röbel in [68] has proposed a new method for shape invariant real-time modification of speech signals. The method can be seen as SOLA algorithm that is using the standard PV algorithm for phase synchronization. This method, denoted PVSOLA,

provide an improved time synchronization during overlap add and, consequently an improved quality of the transformed speech signals.

In addition in [69], some improvements have been proposed for the treatment of polyphonic signals, the distinction between sinusoidal and noisy frequency components and the treatment of transients. The proposed algorithm is also characterized by a lower latency that makes it more suitable for real-time implementations.

Many methods for TFTs are available in the specialist literature and an extended treatment of such techniques would require a separate text. Among the emerging techniques we want to emphasize the use of wavelet or Gabor transformations, which for space reasons we do not report [1], [64]-[69].

References

1. U. Zölzer (editor), "DAFX Digital Audio Effects (Second Edition)", ISBN: 978-0-470-66599-2 John Wiley & Sons, 2011.
2. M.R. Schroeder, "Improved quasi-stereophony and colorless artificial reverberation", J. of Audio Engineering Society, Vol. 10, No. 3, pp. 229-223, 1962.
3. M.R. Schroeder, "Natural sounding artificial reverberation", Journal of Acoustic Soc. of America, 33:1061, 1962.
4. Julius O. Smith, "Physical Audio Signal Processing: Digital Waveguide Modeling of Musical Instruments and Audio Effects", Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Web published at http://ccrma.stanford.edu/j̃os/pasp/, May 2004.
5. M.R. Schroeder, "Digital simulation of sound transmission in reverberant spaces. Part 1", Journal of Acoustic Soc. of America, Vol.47, No. 2, pp. 424-431, 1970.
6. M.A. Gerzon,"The Design of Distance Panpots", Proc. 92nd AES Convention, Preprint No. 3308, Vienna 1992.
7. J.A. Moorer, "About this reverberation business", Computer Music Journal, Vol. 3, No. 2, pp 13-18, 1979.
8. ISO 3382-1:2009, "Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces, " International Organization for Standardization, Genéve, 2009.
9. J. Stautner, M. Puckette, "Designing Multi-Channel Reverberetors", Computer Music Journal, Vol. 6, No. 1, pp. 52-65, 1982.
10. S.J. Orfanidis, "Introduction to Signal Processing", Prentice Hall, 1996.
11. M.A Gerzon, "Unitary (energy preserving) multichannel networks with feedback", Electronics Letters, Vol. 12, No. 3, pp. 13-18, 1976.
12. D. Griesinger, "Practical Processors and Programs for Digital Reverberation", Proc. AES 7th Int. Conf., pp. 187-195, Toronto, 1989.
13. J.M. Jot, A. Chaigne, "Digital Delay Networks for Designing Artificial Reverberations", Proc. 94th, AES Convetion, Preprint No. 3030, 1991.
14. J.M. Jot," Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality", ICMC: International Computer Music Conference, Thessaloniki, Greece, Septembre 1997.
15. J.M. Jot, "An analysis/synthesis approach to real-time artificial reverberation", in Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc, Vol. 2, pp. 221-224, 1992
16. D. Rocchesso, J.O. Smith, "Circulant and elliptic feedback delay networks for articial reverberation", IEEE Transactions on Speech and Audio Processing, Vol. 5, No. 1, pp. 51-60, Jan. 1997.
17. A. Czyzewski, "A Method of Artificial Reverberation Quality Testing", J. Audio Eng. Soc., Vol. 38, No. 3, March 1990.
18. M. Barron, "The Subjective Effects of First Reflections in Concert Hall - The Need for Lateral Reflections", J. Sound Vib., Vol. 15, pp. 475-494, 1971.
19. Y. Ando, "Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo", J. of Acoust. Soc. Am., Vol. 62, pp.1436-1441, Dec. 1977.
20. Y. Ando, "Concert Hall Acoustics", Springer-Verlag, 1985.
21. W.V. Keet, "The influence of Early Reflections on Spatial Impression", in Proc. 6th Int. Cong. on Acoustic, pp E-2-4, Tokyo JP, 1969.
22. J.P. Jullien, E. Kahle, M. Marin, O. Warusfel, G. Bloch, and J.-M. Jot, "Spatializer: A perceptual approach," AES 94th Convention, vol. Preprint 3465 (B1-5), pp. 1-13, March 16-19 1993.
23. W.G. Gardner, "Reverberation algorithms", in Applications of Digital Signal Processing to Audio and Acoustics, M. Kahrs and K. Brandenburg, Eds., pp. 85-131. Kluwier Academic Publishers, Boston/Dordrecht/London, 1999.
24. J.O. Smith and D. Rocchesso, "Connections between feedback delay networks and waveguide networks for digital reverberation", in Proceedings of the 1994 International Computer Music Conference, Århus. 1995, pp. 376-377, Computer Music Association.

25. J.M. Jot, “Etude et Réalization d’un Spatialisatuer de Sons par Modèles Physiques et Perceptifs”, PhD thesis, French Telecom, Paris, Paris 92 E 019, 1992.
26. H. Kuttruff, “Room Acoustic”, Fourth edition, Elsevier, 2000.
27. J.O. Smith, “Physical modeling using digital waveguides”, Computer Music Journal, vol. 16, no. 4, pp. 74-91, Winter 1992, Special issue: Physical Modeling of Musical Instruments, Part I.
28. S.A. Van Duyne and J.O. Smith, “Physical modeling with the 2-D digital waveguide mesh”, Proceedings of the 1993 International Computer Music Conference, Tokyo. 1993, pp. 40-47, Computer Music Association.
29. L. Savioja, J. Backman, A. Järvinen, and T. Takala, “Waveguide mesh method for low-frequency simulation of room acoustics”, Proceedings of the 15th International Conference on Acoustics (ICA-95), Trondheim, Norway, pp. 637-640, June 1995.
30. L. Savioja and V. Välimäki, “Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques”, IEEE Transactions on Speech and Audio Processing, pp. 184-194, March 2000.
31. Barry Vercoe and Miller Puckette. “Synthetic Spaces - Artificial Acoustic Ambience from Active Boundary Computation,” NSF proposal (1985). Available from Music and Cognition office at MIT Media Lab.
32. William Grant Gardner, “The Virtual Acoustic Room”, Thesis S.B., Computer Science and Engine, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1982.
33. J. Frenette, “Reducing Artificial Reverberation Requirements using Time-Variant Feedback Delay Networks”. MsThesis, University of Miami, Dec. 2000.
34. A. Scott, S.A. Van Duyne, J.R. Pierce, “Travelling Wave Implementation of a Lossless Mode-Coupling Filter and The Wave Digital Hammer”, Proceedings of the International Computer Music Conf., Aarthus, Denmark, pp. 411-418, Sept. 1994.
35. L. Savioja, T. Rinne and T. Takala, “Simulation of room acoustics with a 3-D finite difference mesh,” Proceedings of the International Computer Music Conf., Aarthus, Denmark, pp. 463-466, Sept. 1994.
36. D.T. Murphy, D.M. Howard, A.M. Tyrrell, “Multi-channel reverberation for computer music applications,” 1998 IEEE WORKSHOP ON SIGNAL PROCESSING SYSTEMS-SIPS, 1998.
37. D. T. Murphy, D.M.;Howard, “Digital waveguide modelling of room acoustics: comparing mesh topologies,” Proceedings. 25th EUROMICRO Conference, 1999., Vol. 2, pp. 82-89, Sept. 1999.
38. S.A. Van Duyne, J.O. Smith, ”The tetrahedral digital waveguide mesh”, Applications of Signal Processing to Audio and Acoustics, 1995, IEEE ASSP Workshop on, 15-18 pp 234-237, Oct. 1995.
39. J. Dattorro, “Effect Design Part 1: Reverberator and Other Filters,” J. Audio Eng. Soc., vol. 45, n.9, pp. 660-684, Sept. 1997.
40. W.G. Gardner, “Efficient Convolution without Input-Output Delay,” J. Audio Eng. Soc., vol. 43, pp. 127-136, March 1995.
41. U. Zölzer, N. Fleige, M. Schonle, and M. Schusdziara, “Multirate Digital Reverberation System,” in Proc. Audio Eng. Conv., 1990.
42. G.W. McNally, “Dynamic Range Control of Digital Audio Signals”, J. Audio Eng. Soc., Vol. 32, No., 5 pp. 316-327, May 1984.
43. Udo Zölzer, “Digital Audio Signal Processing”, J. Wiley, ISBN 0 471 97226 6, England, 1997.
44. P. Dutilleux, U. Zölzer, “Nonlinear Processing”, in DAFX Digital Audio Effects in Zolzer (edt), ISBN 0 471 49078 4, John Wiley & Sons, 2002.
45. D. Mapes-Riordan, W.M. Leach, “The Designof a Digital Signal Peak Limiter for Audio Signal Processing”, J. Audio Eng. Soc.,Volume 36 Number 7/8 pp. 562 · 574; July 1988.
46. P. Dutilleux, U. Zölzer, “Delays”, in DAFX Digital Audio Effects in Zolzer (edt), ISBN 0 471 49078 4, John Wiley & Sons, 2002.
47. J. Dattorro, “Effect Design Part 2: Delay-line modulation and chorus,” J. Audio Eng. Soc., vol. 45, n. 10, pp. 764-788, Oct. 1997.
48. C.A. Henricksen, “Unearthing the mysteries of the Leslie cabinet,” Recording Engineer Producer magazine, April 1981, articolo disponibile in Internet http://www.theatreorgans.com/hammond/faq/mystery/mystery.html.

49. S. Disch and U. Zölzer, "Modulation and Delay Line Based Digital Audio Effects", Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99), NTNU, Trondheim, December 9-11, 1999.
50. J.O. Smith, S. Serafin, J. Abel, and D. Berners, "Doppler simulation and the leslie", in Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, pp. 13-20, September 26 2002.
51. A. Papoulis, "The Fourier Integral and Its Applications," McGraw-Hill, New York, 1962.
52. J. Dattorro, "Using Digital Signal Processor Chips in a Stereo Audio Time Compressor/-Expander", Presented at the 88rd Convention of the AES, preprint 2500 (M-6), October 1987.
53. ITU-R.BS1770-2 "Algorithms to measure audio programme loudness and true-peak audio level," 2011.
54. W. Hess, "Pitch Determination of Speech Signals", Springer Verlag, New York, 1983.
55. C.J. Roehrig , "Time and Pitch Scaling of Audio Signals", AES 89th Convention, preprint n. 2954 (E1), September 21-25, 1990.
56. J. L. Flanagan and R. M. Golden, "Phase vocoder," Bell System Technical Journal, 1966.
57. M. R. Portnoff, "Time-scale modifications of speech based on short-time Fourier analysis," IEEE Trans. Acoust., Speech, Signal Processing, vol. 29, pp. 374-390, 1981.
58. J. Laroche, M. Dolson, "Improved Phase Vocoder Time-Scale Modification of Audio", IEEE Transactions on Speech and Audio Processing, vol 7, no. 3, May 1999.
59. M.R. Portnoff, "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform", IEEE Transactions on Acoustics, Speech and Signal Processing ASSP-24:243-248, 1976.
60. Portnoff, M.R., "Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis", IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-28(1):55-69, 1980.
61. M.R. Portnoff, "Short-Time Fourier Analysis of Sampled Speech", IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-29(3):364-373, 1981.
62. S. Roucos and A. M. Wilgus, "High Quality Time-Scale Modification for Speech", in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 493-496, March 1985.
63. R.K.C. Tan, A.H.J. Lin, "A time-scale modification algorithm based on the subband time-domain technique for broad-band signal applications", J.Audio Eng.Soc., vol.48, no.5, May 2000.
64. E. Moulines and F. Charpentier, "Pitch Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones," Speech Commun. (EUROSPEECH'89), vol. 9, no. 5/6, pp. 453-467, 1990.
65. P. De Gersem, B. De Moor, M. Moonen, "Applications of the continuous wavelet transform in the processing of music", K.U.Leuven, ESAT/SISTA, 1997.
66. Z. Pruša and P. Rajmic, "Toward high-quality real-time signal reconstruction from STFT magnitude," IEEE Signal Processing Letters, vol. 24, no. 6, June 2017
67. E.S. Ottosen and M. Dorfler, "A Phase Vocoder based on Nonstationary Gabor Frames," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume: 25 , Issue: 11 , Nov. 2017.
68. A. Röbel, "A shape-invariant phase vocoder for speech transformation," in Proc. 13th Int. Conf. on Digital Audio Effects, pp. 1-8, 2010
69. S. Kraft, M. Holters, A. von dem Knesebeck, and U. Zölzer, "Improved PVSOLA time-stretching and pitch-shifting for polyphonic audio," in Proc. Int. Conf. Digital Audio Effects (DAFx–12), Sep 2012