Audio Dynamic Range Compression For Minimum Perceived Distortion

BARRY A. BLESSER, Student Member, IEEE
Elektromesstechnik Wilhelm Franz KG
Lahr/Schwarzwald, West Germany
and
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Mass. 02139

Abstract

When manual dynamic range compression is replaced by an automatic compression system, the sound engineer must still be able to choose some characteristics of the compression. An automatic system and the theory motivating its design are described. We have built a device (EMT 156) that enables the sound engineer to adjust any of the six static parameters; also, the system automatically chooses its own dynamic properties as a function of the input program signal.

Manuscript received July 17, 1968; revised December 3, 1968.

The experimental research and development of the EMT 156 compression system was performed at Elektromesstechnik Wilhelm Franz KG. The theoretical exploration was undertaken at the Research Laboratory of Electronics, M.I.T., and was supported in part by the National Institutes of Health (Grants 5 PO1 GM-14940-02 and 1 PO1 GM-15006-01) and the Joint Services Electronics Program (Contact DA 28-043-AMC-02536(E)).

I. Introduction

Establishing an effective dynamic range, like many of the tasks facing the sound engineer, is an art form. The problem is to pass an original program signal with a large dynamic range through a channel with a reduced dynamic range. The original dynamic range, that is, the ratio of the largest to the smallest program signals, of classical symphonic music can be as high as 90 dB. Once the material has been recorded, the dynamic range has been reduced to no more than 70 dB. One should, however, contrast this with a 40-dB dynamic range of an FM station, or a 20-dB range of an AM station. The fundamental limitations of all channels, including tape recordings, record disk platters, motion picture films, and radio broadcast stations, mean that the reproduced program differs from the original. Given these limitations, the sound engineer can either obscure that part of the program which is below the noise level, distort the high-level peaks, or introduce distortion that results from amplifying the low-level signals more than the high-level signals. The last technique is called *compression and limiting*. Because the channel cannot reproduce the input exactly, some kind of distortion must be introduced. Compression, if performed properly, is the least objectionable kind of distortion.

Originally, compression resulted when the sound engineer adjusted the gain manually to produce the most pleasing effect. But the response time of a sound engineer to the sudden appearance of a loud passage is relatively long; a delay of 1 second would be considered fast. During this second, the listener would perceive distortion as the signal exceeded the maximum level of the channel. Previewing the program material so that the gain could be reduced before the actual appearance of the high-level passage would permit the sound engineer to reduce the program level before the peak occurs. Under most conditions, however, previewing is neither practical nor economical.

This leads one directly to the idea of automatic gain control. An automated system reduces the gain when the program material is excessively loud, and increases the gain when the program material is too low. Once the sound engineer decides to employ some kind of automated gain control, the quality of the dynamic range compression becomes dependent on the characteristics of the equipment, and not on his own artistic sense and reflexes. With an automated compression system, adjusting the gain of the program before it enters the compression system would have little effect on the program signal that is leaving. Thus, the compression system judges when the signal is too loud or too soft. If it decides on the wrong gain, or changes its gain at the wrong time, there is nothing that the sound engineer can do to correct it. The sound engineer has lost control of the dynamic properties of the program material. Before the sound engineer gives control of the dynamic range to the system, he should convince himself that the compressor is doing exactly what he would do if he could adjust the gain as quickly as the automated compression system. One may also argue that the automated system is more effective since it responds faster and it examines more aspects of the signal than a sound engineer can.

The nature of the compression that a sound engineer would employ for classical music and popular music is radically different. Also, the dynamic range characteristics of an AM radio station are extremely different from those of an FM station or a disk-recording system. When the sound engineer manually controls the dynamic range compression, he takes these factors into account. With an automated system, he must be permitted to adjust the characteristics of the system.

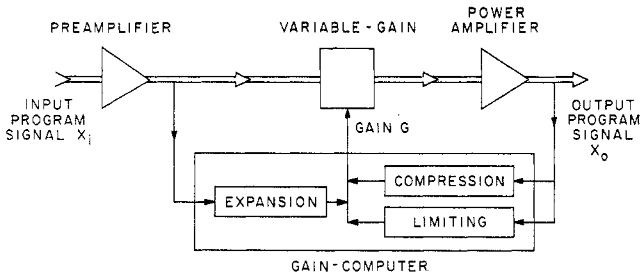


Fig. 1. Model of generalized compression system.

Generalized Compression Systems

Almost all compression systems can be modeled as shown in Fig. 1. The program signal at the input passes through the preamplifier, the variable-gain section, and the output power amplifier. The gain-computer section examines the input and output program signals and determines what the gain should be. With the exception of this section, the technical characteristics of the entire system are easily specified. The input impedance should be high, the output impedance low,[1] the audio channel linear, the frequency response flat over the desired range, and the signal-to-noise ratio high. These specifications are purely technical and are generally of no interest to the sound engineer if they meet the minimum requirements for his audio system.

The model in Fig. 1 shows the gain-computer section monitoring the input and output signals, although either one of them contains sufficient information to determine the same gain. Let us define a set of functions for the three cases when the gain is controlled by input, output, or both:

$$G = F_1(X_o)$$
$$G = F_2(X_i)$$
$$G = F_3(X_o, X_i).$$

If the input and output are measured in dB, then

$$X_o = X_i + G.$$

For each of the three possible cases, the gain is the same for a given input (and hence the outputs are the same) if the 3 functions are related in the following way:

$$F_1(X) = F_2(X - F_1(X)) = F_3(X, X - F_1(X)).$$

If each of these functions exists and contains no singularities, then a compression system could be built using either the input, output, or both for controlling the gain. In practice, however, it is much easier to implement a feed-forward system for compression ratios less than one, and a feedback system for ratios greater than one. For this reason, the compression and limiting section are shown with the output signal controlling the gain, and the expansion section is shown with the input signal controlling the gain. Being able to implement an arbitrary gain function, however, does not help to determine which gain function should be used.

The characteristics of the gain-computer section determine the quality of the compression and are the most difficult to formulate. The designer of this section must ask himself, "If I were the gain-computing section and knew the input and output signals, what gain would I decide upon?" This is a difficult question. The answer must take into account the nature of auditory perception if one is to fool the listener into thinking that the program signal is almost the same as the original.

In essence, the gain-computer section is a specialized analog computer whose parameters are controlled externally by the sound engineer and internally by the program signal itself. For the purposes of discussion, these parameters are divided into two classes: static and dynamic. The static parameters specify the performance of the system to periodic, although not necessarily sinusoidal, waveforms; and the dynamic parameters specify the performance for changing input signal level. It is desirable to find a set of parameters that have some psychological significance. Most common audio measuring techniques have been chosen for their usefulness and simplicity in particular applications, but they are not relevant to perception. Unfortunately, our knowledge of acoustical psychophysics is so meager that the problem of choosing a good set of variables remains unsolved.

Conventional compression systems make very little attempt to use a complex set of measuring techniques; rather, they simply compute some sort of value based on peak and average. Instead of having fixed parameters, this compression system allows the sound engineer to adjust and experiment with the relevant parameters that we have found. We have attempted, with partial success, to generate a set of orthogonal adjustments that operate on distinct characteristics of the program signal. This allows each parameter to be changed independently of the settings on the other controls. For example, it would be

[1] Alternatively, the input and output impedances are made to be 600 ohms.

advantageous to control the peak level output independently from the loudness. This means that the parameter affecting the peak must compensate for its effect in such a way that it does not change subjective loudness.

Our understanding of channel characteristics is, fortunately, much greater. The range of a channel is limited by the maximum level, above which large amounts of distortion are introduced, and by the inherent additive noise level of the channel electronics. In the same way, the range of the program is the ratio of the highest level signals to either the minimum desired signal or the inherent noise level. The incoming program signal must be compressed if the channel range is less than the program range. The difference between the dynamic range of the channel and program must be equal to the difference in amplification for high-level signals and low-level signals. Although this amplification factor is specified by the channel and program characteristics, the sound engineer, or the designer of the gain-computer section, must still decide how this gain variation should be distributed over the range of input signals. For example, peak signals may be left untouched and the low-level signals amplified; or peaks may be attenuated and low-level signals only somewhat amplified. Furthermore, all of the gain variation can be concentrated into a small part of the range, or spread equally throughout. In our system, the shape of the compression curve is controlled by the static parameters and is adjusted externally by the sound engineer.

Once having satisfied the maximum and minimum level requirements of the channel program, all remaining choices, and there are many, are based on subjective judgments. With one set of dynamic characteristics, classical music may sound "muddy" or "full"; yet with the same set, popular music may appear pleasing. We are of the opinion that "optimum" performance for popular music will require adjustments that are very different from those of classical music. The listener of classical music is sensitive to aspects of the program signal that are irrelevant to popular music. The issue of optimization is tied to what the listener expects to hear. Part of the expectation is adaptive, in the sense that the listener expects a different reverberation when he listens in a small room. He does not wish to find, however, that the orchestral balance between instruments has been changed. If he were listening to the same music in his car, however, he would prefer that the balance be changed if the original balance prevented him from hearing some of the quieter instruments. We must emphasize the point that varying the task of a compression system will vary the "optimum" settings; or in other words, there is no single overall solution to the compressor problem.

Even without the hope of a solution, it is necessary to formulate the problem. This requires the discovery of the perceptually relevant characteristics of the program and compression system. We have found the following terms to be useful for describing compression systems.

1) *Average loudness* is the general sensation of loudness experienced by the listener. In many cases, this is approximated by a short-term average of the full-wave rectified audio signal. For complex waveforms, however, this is a poor approximation. Zwicker [1] and others have actually formulated a sophisticated algorithm that separates the signal into frequency bands. Each of these bands, and the interaction of the bands, is processed independently. But, even their measure of loudness does not include the fact that commercial messages appear louder than the corresponding news reports.

2) *Short-term dynamic range* is the ratio of maximum to minimum average program level during a short interval, for example, 1 second. This parameter may correspond to the sensation of "fullness." Reducing the short-term dynamic range is accompanied by various notes modulating each other, thereby giving an overly "muddy" or distorted sound. A piano solo has an extremely large short-term dynamic range, whereas an aria from an opera has a very small range.

3) *Long-term dynamic range* is the ratio of maximum to minimum average program level during a long interval, for example, approximately 30 seconds. This parameter may correspond to the sensation of "mood." Large changes in program level are often accompanied by a change in the mood created by the music. Excessive compression of the long-term dynamic range results in a sound that is "dull" or uninteresting. Orchestral music is likely to have a very large range, whereas popular music has no significant range.

4) *Envelope-attack time* is the time that it takes the loudness or level to increase. This parameter may correspond to the sensation of "crispness," is heavily dependent on the type of instrument and the recording technique. The effects of the recording hall, microphone placement, artificial processing, acoustic reflection, etc., however, increase the attack-time considerably. A bass sound has a very long attack time, whereas a harpsichord has a very short attack time.

5) *Envelope-decay time* is the time it takes the level to decrease and is almost completely dependent on the reverberation time. In a large hall it can approach 2 seconds, whereas in an acoustically treated studio it can be 20 ms. A compression system that increases the decay time may increase the sensation of "largeness."

To our knowledge, no detailed psychological studies have ever been performed on these kinds of parameters. We do not have an overall perspective for the relationship between the subjective effects and the compression variables. Controlled experiments with musicians as subjects should be performed in order to gain some structured insight into the compression phenomena. Most of the fixed parameters in the EMT-156 compression system were determined by trial and error; and most of the critical parameters were left variable so that the sound engineer could experiment in his particular application.

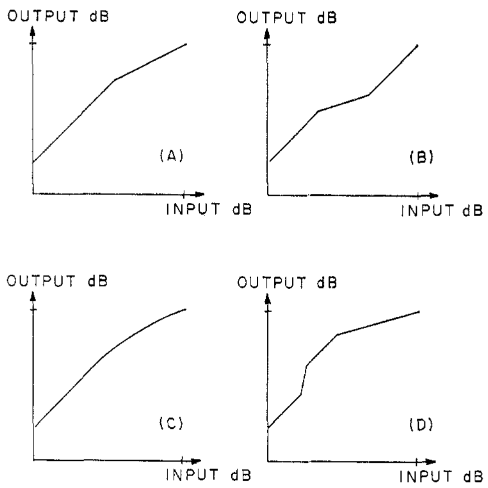


Fig. 2. Static characteristics of four compression systems with the same compression factor.

Static Characteristics

As we have stated, dynamic range compression results when high-level signals are amplified less than low-level signals. The static input-output curves for 4 different types of compression systems are shown in Fig. 2. In each case the total dynamic range compression is identical, but the quality of the sound is very different.

Each is useful in a different application. For the compression part of a compander system the curve in Fig. 2(B) might be used. The curves in Fig. 2(A) and (C) are very similar, except that the compression ratio for the former remains constant and for the latter gradually increases with increasing signal level. Many conventional compressors have characteristics represented by Fig. 2(C). A compression system with the curve of Fig. 2(D) produces the greatest increase in loudness, whereas a system with the curve of Fig. 2(B) leaves the effective loudness unchanged.

A straight-line compression curve using a dB scale has the advantage that the effect of compression is level independent. That is, increasing the input level does not change the output program quality. With the gradually changing curve of Fig. 2(C) increasing the input level increases the compression ratio and decreases the dynamic range. With a complex system it is better to have level independence. It has also been argued that a linear compression curve corresponds to the ear's loudness characteristics and is, therefore, a natural parameter to control, but we will now show that this is not true.

For a sine wave or white noise, the compression ratio is proportional to the increase in the level at the input divided by the increase in the level at the output (in dB). The relationship between signal level and loudness, as shown by Stevens [2], is approximately

$$L = k(S)^{0.6}, \tag{1}$$

where L is the loudness in sones, and S is the signal level. This can be rewritten as

$$\log L = \log k + 0.6 \log S. \tag{2}$$

The compression ratio CR is defined as the change (in dB) in input signal necessary to produce a 1-dB change in the output signal. For the curve in Fig. 2(A), this can be written

$$\mathrm{CR} \cdot \log_{10} (S_o/S_r) = \log_{10} (S_i/S_r), \tag{3}$$

where S_o is the output signal, S_i is the input signal, and $S_r = 100$ percent modulation level. The loudness of the input and output signals with (2) is

$$\begin{aligned} \log L_i &= \log k + 0.6 \log S_i, \\ \log L_o &= \log k + 0.6 \log S_o. \end{aligned} \tag{4}$$

Combining (3) and (4) gives

$$\mathrm{CR} \cdot \log L_o = \log L_i + (\mathrm{CR} - 1)(\log k + 0.6 \log_{10} S_r). \tag{5}$$

Consider the case when the loudness of the input increases by a factor A and the output then increases by a factor B. Rewriting (5) gives

$$\begin{aligned} \mathrm{CR} \cdot \log (L_o \cdot B) = \log (L_i \cdot A) + (\mathrm{CR} - 1) \\ \cdot (\log k + 0.6 \log_{10} S_r). \end{aligned} \tag{6}$$

Subtracting (5) from (6) gives

$$\mathrm{CR} = \log A / \log B. \tag{7}$$

The results of (7) show that within a given program, loudness ratios do not scale proportionally. When a linear compression curve is transformed to a loudness scale, it is no longer linear. The subjective loudness balance within the program is destroyed. For three loudness levels with ratios $1:(1+k):(1+3k)$ entering the compression system, the output ratios are not $1:2k:4k$, rather, they are $1:2k:4k^2$. Clearly, we could find another compression curve, other than a linear one, which when transformed to loudness would yield a linear loudness curve. Unfortunately, such a curve requires a knowledge of the acoustic reproduction level in order to be accurate. If the sound engineer knew this level, he could improve the subjective balance. Obviously this is impossible. The acoustic reproduction level varies from listener to listener, and furthermore, such a system would not be level independent.

Using a compression system with the characteristics shown in Fig. 2(A) reduces the dynamic range of the program, but also reduces the signal-to-noise ratio by the same amount. The amount (defined as the compression factor) is equal to the difference in amplification when the input is normal program and when it is noise alone. It is very large if the compression region is extensive. This should suggest that the additional gain used to produce compression be removed when noise is present. In this way, the gain with normal program, and the gain with noise, would be the same; there would be no degradation in signal-to-noise. The distinction between low-level program and noise can often be made on the basis of level. Useful signals generally remain at least 10 dB above the background noise. If this margin is not great enough, spectral and temporal difference would distinguish the two classes. The part of the system that operates in this region is called the expansion section, since the CR is less than one in the transition region between noise and low-level signal.

Because the compression and expansion sections use short-term averages of the input and output signals for making their decisions, the instantaneous dynamic range is not affected. That is, the peak signal at the output may exceed the maximum level of the channel unless some other peak measuring section is used. With the addition of a limiter section, the gain is reduced whenever the output attempts to exceed a preset threshold. The compression ratio for the limiter is usually above 20:1.

The complete static characteristics for the EMT 156 compression system are shown in Fig. 3. Region A in Fig. 3 is the expansion region. In this region a decrease in the input signal results in a reduction in gain. The minimum gain is zero and the maximum gain is determined by compression-factor adjustment. The transition is gradual rather than sharp to minimize "pumping" and objectionable noise modulation. As the input signal passes through this region, the output appears to fade more rapidly.

Region B in Fig. 3 is the compression region. In this region any increase in input signal results in a decrease in gain and a corresponding decrease in dynamic range. Input signals above the rotation point are attenuated, and signals below are amplified. As is the case with the expander, the maximum gain for the compression region is determined by the compression-factor adjustment. For large signals, the limiter wrests control from the compression section and monitors the peaks rather than the loudness.

Region C in Fig. 3 is the limiter region. In this region an increase in input signal results in no increase in peak output signal. The gain is instantaneously reduced by the same amount that the input peak increases. After initial limiter transients, the limiter does not produce any distortion.

For steady-state input signals, there is no change in gain as determined by any of the 3 sections and the output is an exact replica of the input. A steady-state input signal is one that has a constant average, rms, and peak value. It does not have to be periodic.

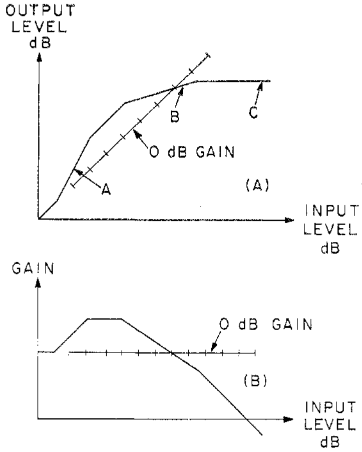
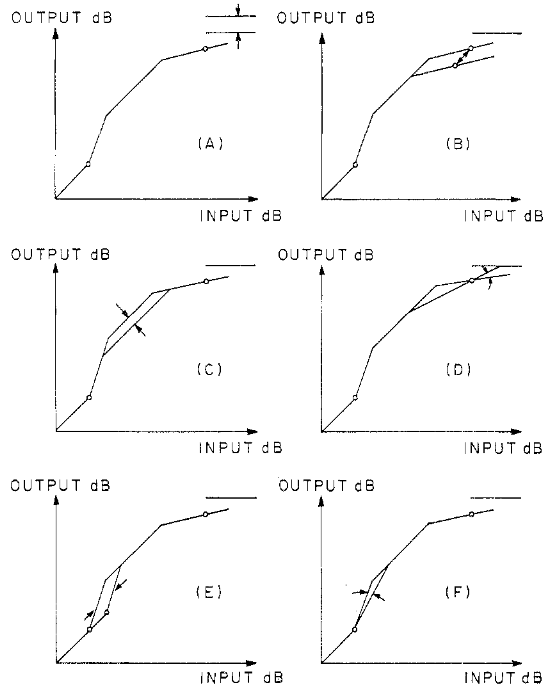


Fig. 3. Static characteristics of the EMT 156 compression system.

Fig. 4. Effect of changing static adjustments on the static characteristics of the EMT 156 compression system.



Adjustment of Static Characteristics

A typical set of static characteristics which might be used in a compression system is shown in Fig. 3. There are 6 independent adjustments.

1) The limiter-threshold adjustment controls the output level for signals above the threshold for more than 100 μs. Although the amount and frequency of limiting is determined by the input signal level, it is also affected by the compression rotation point, the compression ratio, and the compression factor. See Fig. 4(A).

2) The compression rotation-point adjustment controls the output signal level for a given input signal when the signal is in the compression region. It effectively shifts the compression region up and down relative to the expansion and limiting regions. The rotation point is defined as the intersection of the compression curve with the 0 dB line. See Fig. 4(B).

3) The compression-factor adjustment controls the maximum gain of the system when the input signal is

between the compression and expansion regions. A large compression factor produces a louder output signal, as well as a reduced signal-to-noise ratio, for signals between the compression and expansion regions. See Fig. 4(C).

4) The compression-ratio adjustment rotates the compression curve about the compression rotation point, and thereby controls the dynamic range for high-level signals. Because it rotates about a 0 dB gain point, it does not affect the over-all loudness for normal programs with a 0 dB level. A large compression ratio produces a sound that is "dense." See Fig. 4(D).

5) The expansion rotation-point adjustment controls the level at which the system returns to 0 dB gain for low-level signals. Like the compression rotation point, it moves the expansion region relative to the compression and limiting regions. The expansion rotation point is adjusted so that it is just slightly higher than the noise level of the program. See Fig. 4(E).

6) The expansion-ratio switch rotates the expansion curve about the expansion rotation point, and thereby controls dynamic range expansion for low-level signals. See Fig. 4(F).

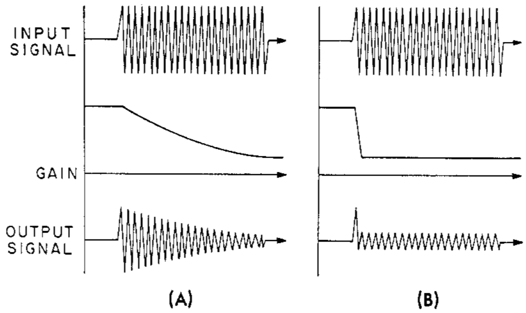
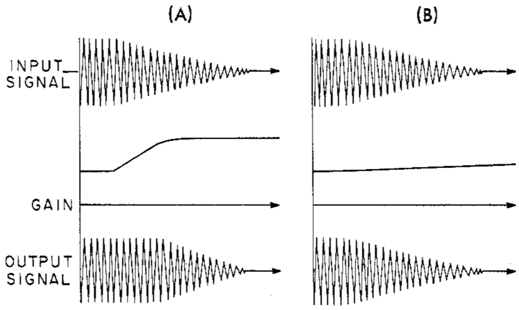


Fig. 5. Response of the compression system to tone burst with (A) long attack time, and (B) short attack time.

Fig. 6. Response of the compression system to damped tone burst with (A) short release time, and (B) long release time.



Dynamic Characteristics of Compression

The dynamic properties of a compression system may be thought of as a relationship between the gain variations and the complex program signals. To define this relationship we must answer four questions.

1) When should the gain change?
2) How rapidly should the gain increase?
3) How rapidly should the gain decrease?
4) How should the limiting, compression, and expansion sections interact?

First, let us consider the response of the system to an input transient. The time that it takes the system gain to go from its initial value to 63 percent of its final value is called the *attack time*. The value of the attack time usually ranges from 100 μs to 100 ms. Fig. 5 shows the effect of both long and short attack times on a sine-wave burst. Notice that the output resembles the input signal more exactly when the attack time is short. When it is long, the initial part of the input transient is accentuated. This overemphasis is perceived at the beginning of all musical notes and speech syllables. To avoid overemphasis, a compression system should have an attack time no greater than 10 ms. A transient that decays in this amount of time is generally not perceived.

The disadvantage of using a very short attack time is that the gain tends to be determined by the peaks of the program and not by the energy or loudness. Under some conditions this can change the effective compression ratio. When the system is operating in the compression region, an increase in loudness at the input should result in a smaller increase in loudness at the output. Consider the case when a musical note with a high peak-to-average ratio suddenly appears, as for example, when a harpsichord note is added to a background symphony. The peaks of the program signal with the harpsichord note may be 10 dB greater than the peaks without it, but the actual increase in loudness may be only 3 dB. If the compression ratio is 2:1 and the attack time is short, the gain is reduced by 5 dB and the loudness of the output signal is reduced by 2 dB. This produces a very "flat" sound with an apparently negative compression ratio. This phenomenon is directly related to the fact that the ear responds to the loudness and that signals with high-level peaks are not necessarily loud. For this reason, the attack-time is not made too short and a compromise is sought. We have chosen 5 ms for our system. The compromise attempts to minimize the effects of accentuating initial transients and to prevent the peaks from controlling the gain.

The release time may be defined as the time that it takes the gain to reach 37 percent of its initial value when the input signal decreases. The value of the release time usually varies from 100 ms to 20 seconds. Fig. 6 shows the effect of long and short release times; the decreasing part of the signal, or reverberation, is affected. In the same way that a fast attack time results in a faithful reproduction of the initial part of the transient, a long release time insures that the final transients are not accentuated. With a long release time and short attack time the gain can decrease rapidly, but it recovers slowly. Under this condition, a single high-level note can create a "hole" in the program. For example, a single piano note played with other music in the background causes all other instruments to be reduced in loudness for as long as 10 seconds after the piano note occurs. Alternatively, if the release time is short, the gain varies rapidly and often. To the listener it appears that the individual notes are modulating each other. In fact, individual notes are being distorted and the music sounds "muddy."

Because the complex program signals are not steady-state, the dynamic release characteristics of the gain computer have a much greater effect on the nature of the compression than do the static characteristics. Only the long-term dynamic range compression is unaffected by the release-time function. The short-term dynamic range compression depends on the rapidity with which the gain recovers to its previous value after a peak transient.

The following example illustrates the point. Consider a compression system adjusted for both a very high compression ratio and a long release time. The long release time prevents the gain from changing rapidly after peaks, and for normal programs the gain remains constant. Thus, even though there is a high compression ratio, there is no change in the dynamic range.

The program following a large loudness peak retains the same dynamic relationship to the peak if the gain has not had enough time to change. If it has recovered, then the program is treated independently from the previous peak. Consider a hypothetical loudness curve for some program, as shown in Fig. 7(A). The dynamic range, or difference between the loudness minima and maxima, is constant throughout. For some particular release time, the output loudness curve might be that of Fig. 7(B). In region A, the peaks are sufficiently close together to prevent the gain from recovering. The output program is identical to the input program and there is no short-term dynamic range compression. In interval B, however, there is enough time for partial recovery of the gain during the interval between peaks. As a result, there is some short-term compression. The original dynamic range x_1 has been reduced to x_2. The effective short-term compression ratio is x_1/x_2. In region C, the loudness changes sufficiently slowly that the peaks have no effect on the program in between them. The gain has enough time to recover completely. The original dynamic range has been reduced to x_3, and the effective compression ratio is x_1/x_3. This long-term compression ratio is independent of the release time and is determined by the static compression-ratio adjustment. Notice that the long-term ratio is always greater than the short-term ratio.

Now, let us define a sound package as the part of the program signal containing a long series of peaks all of which are of about the same amplitude. Fig. 7(C) shows the amount of short-term compression for the package as a function of the time interval between peaks T and the release time t_r. The effective compression ratio CR_e can be shown to be related to the static compression ratio CR_s and long-term release time t_r as follows:

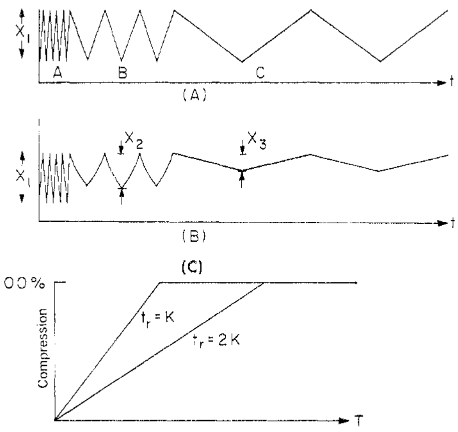


Fig. 7. Effect of release-time function on short-term dynamic range compression: (A) hypothetical input program, (B) output program with the input of (A) and single release-time function, (C) relationship between short-term compression and interval between peaks in a sound package. (100 percent corresponds to maximum static compressions x_1/x_3.)

$$\mathrm{CR}_e = 1 + (\mathrm{CR}_s - 1)T/k \cdot t_r \quad \text{for } T < k \cdot t_r,$$

where k is some constant. The importance of the release time function should now be clear.

The release-time function just described has only one degree of freedom—one parameter determines all of its properties. If it were used in the compression section, we would find that no matter what value it had, the output program would be of poor quality. With a long release time, the listener would judge the program as being "flat," "empty," or "hollow"; with a short release time, he would judge it as being "distorted," "muddy," or "full." A single release-time function does not take into account the psychophysical characteristics of hearing.

In order to prevent the listener from perceiving the effects produced by changing gain, the function must meet several requirements. 1) The gain must not change during the duration of a single note to insure that it is a faithful reproduction of the original. 2) Transients should not dominate the program. 3) The subjective balance between different musical notes should appear the same, even if the actual balance is changed. Subjective balance can be accomplished if the long- and short-term dynamic range compression is a function of the program. If there is a large discrepancy between the ratio of the short- and long-term dynamic range of the input to that of the output, the music does not sound natural. Our goal is to find a release-time function that satisfies these requirements.

As a first approximation, we divide our release-time function into two separate time constants. In this way, the function can consider each peak to be both an isolated transient and a contribution to the over-all loudness. The short-term release time controls the recovery of a percentage, P_s, of the gain change caused by the last peak. The recovery of the remaining percentage of gain change is controlled by the long-term release time.

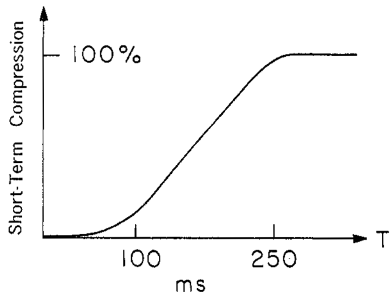
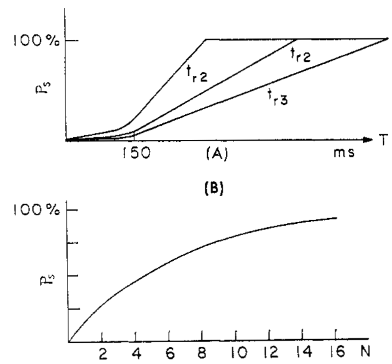


Fig. 8. Relationship between short-term compression and interval between peaks in a sound package. (100 percent short-term compression corresponds to maximum compression of gain change assigned to short-term release time.)

Fig. 9. Percentage of gain reduction, P_s, assigned to short-term release time as a function of (A) interval between peaks in a sound package, and (B) number of peaks in a 1-second sound package.



The short-term release time compensates for part of the undesirable effects of a fast attack time. One must recall that the short-attack time does not suppress low-energy peaks. Thus, if the release-time function did not allow the gain change caused by one of these peaks to recover quickly, it would blank the remaining program. With the rapid recovery, a gain change produced by a random click or pop does not modulate the program, as is the case with many compression systems. A typical value for short release time is 150 ms.

The gain also recovers, however, following the initial peak of a musical note. By changing the gain during the note, rather than after it, the system distorts it. Clearly, the only way to avoid this is to postpone recovery until much of the note has passed. To do this, we have included a holding time in the short-term release function. That is, after a peak has reduced the gain, the new gain is held at that value for approximately 100 ms before it is allowed to recover. Often a new peak arrives during the holding time and recovery is again postponed. In effect, the holding time gives the compression section time to decide if it really wants to recover. The rapid but small gain variations that were present without the holding time are no longer there. The holding time guarantees that during this period there is no dynamic range compression.

To determine the dynamic range compression for our release-time function, let us again consider a sound package with peaks separated by an interval of T. Fig. 8 shows the short-term compression as a function of T. In this case, the maximum short-term compression is equal to the maximum total compression multiplied by the P_s. The amount of long-term compression as a function of T is shown in Fig. 7(C), except that the maximum long-term compression is equal to the maximum compression multiplied by the factor $1-P_s$. The essence of good dynamic range compression is exemplified by proper behavior of the factor P_s. If the compression section can correctly decide how much gain variation should be temporary and how much should be permanent, then the listener will not perceive the changing gain. Experimentally, we have arrived at the following function. Fig. 9(A) shows the relationship between P_s and the interval T in a sound package. Notice that the relationship is made a function of the long-term release time. Thus, increasing the long-term release time also increases the necessary distance between peaks in order for the peaks to be considered isolated events. In other words, it reduced the percentage of the gain reduction assigned to the short-term release. This, in effect, reduces the short-term dynamic range compression. Thus increasing the long-term release time decreases the long-term dynamic range compression, but it also decreases the short-term dynamic range compression proportionally. If the sound package contains only a few peaks, the system behaves slightly differently. Figure 9(B) shows the relationship between P_s and the number of peaks that have occurred during the last second.

Perhaps the operation of the entire release-time function can be clarified by a visual analogy. Consider a cube in space. Because the front and back faces are at a different distance from the viewer they should appear to have different sizes. The viewer compensates for the difference in distance so that the cube appears to be a cube. Now, when the cube is moved closer, the ratio of the distances between the front and back faces from the viewer changes. For our analogy, the distance between the two faces of the cube is the short-term dynamic range, the distance from the center of the cube to the viewer is the long-term dynamic range, and moving the cube closer is compression. In the case of the cube, the viewer changes the compensation in order for the cube to remain a cube; however, the listener does not know that we have performed compression, so his auditory system does not produce the correct compensation. Or more probably, the auditory system cannot produce compensation of this kind, since compression is not a naturally occurring phenomenon. The compression system, in addition to changing the dynamic range, must also change the perspective. The way the system accomplishes this is through the release-time function.

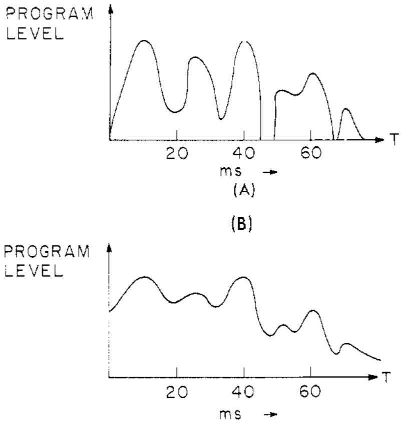


Fig. 10. Two loudness waveforms of input program signals each having the same loudness peaks with (A) high short-term dynamic range, and (B) low short-term dynamic range.

We conclude from the previous discussion that the long-term release time controls the amount of dynamic range compression and also adjusts the short-term range for proper perspective. The one question that remains is "How much compression is needed?" If the sound engineer knows the answer for his particular application, he can adjust the manual release-time controls. For any specific setting, however, the effective compression ratio is fixed. Even if the input program has a small dynamic range, it would be compressed further. Fig. 10 shows the loudness contours of two different program signals. One has almost no dynamic range, while the other has a very large range. Yet, they are both treated equally by the compression system, because the compression section computes the gain based on the peaks of the 5-ms average of the output program. The two signals shown in Fig. 10 both have the same peak loudness.

We solve this problem by allowing the program's peak-to-average loudness ratio to control the release-time function. In this way the original dynamic range determines how much compression should take place. Programs that have a small dynamic range, or are precompressed, are not subject to as high an effective compression ratio as programs with an initially large dynamic range. Having this feature solves the last problem created by the fast-attack time.

The compression section computes a measure of level by performing a 5-ms averaging on the output signal. This average treats all spectral components above 200 Hz equally. The frequency response of the average ear is relatively flat at high levels, yet at low levels it is 60 dB more sensitive to the midrange than to the low frequencies. Consider the case of a 50-Hz organ note playing with a 1-kHz flute. If the level of the organ is 10 dB higher than the flute, it dominates the gain computer. The gain variations produced by the organ note modulate the loudness of the flute. The listener perceives that the flute is modulated by an unknown source, since he is so much more sensitive to the flute note that he hardly notices the organ note. This problem can be solved by adding a variable-frequency filter option to the compression section. The filter modifies its frequency response to match the ear's. In this way the compression section is using an approximation to loudness rather than energy to control the gain. But matching the ear's characteristics requires that we know the level at which the program is being reproduced. If the listener reproduces the program at high levels, there is no problem. At low levels, however, such an option would be necessary. This is the same problem that occurred with the compression curve. Namely, to effectively take into account the ear's characteristics, one needs to know the acoustic reproduction level.

Limiting Characteristics

The function of the limiting section is to protect the channel from being overmodulated by high-level peak transients. Since the compression section attack time is only 5 ms, all transients with a duration shorter than this value retain their original dynamic range. The limiter can compress these only if its attack time is extremely small. We have achieved attack times of 50–100 μs by optimizing the feedback limiter [3]. In general, limiting is an undesirable form of compression, since it usually affects the program quality adversely. The infinite compression ratio of the limiting threshold combined with the extremely small attack time means that the gain is determined solely by peaks rather than average energy or loudness. Because peak level has no perceptual correlate, the gain change appears to be unrelated to the program content, and large limiting gain changes are sometimes perceived as clicks. Clearly, the way to avoid these problems is to prevent peak signals from exceeding the threshold frequently. To accomplish this, we created another threshold 2 dB below the limiting level. The frequency with which peaks enter this window determines the limiter release time. Thus, for a single peak above the limiting threshold, the system reduces the gain and then immediately returns it to the previous value. Since the gain is reduced for only a short amount of time, nothing is perceived. If, on the other hand, the peaks are repetitive and an integral part of the program, the system reduces the gain for the first peak, and observes that the succeeding peaks lie within the window. The release time then approaches infinity and the gain remains constant. The constant gain affects the program quality, since there is no longer any short-term compression. This is a small price to pay for no distortion. The ability to vary the limiter's release time from 150 ms to more than 30 seconds as a function of the program gives all of the advantages of a limiter with few of the disadvantages.

The variable release-time feature allows the program peaks to approach the absolute maximum permissible level without introducing distortion. The sound engineer need not be concerned if the compression section increases the gain until limiting begins. When this happens, the limiter overrides the compression section if the peaks are repetitive, or returns control to the compression section if the peaks are isolated events. The frequency of limiting is determined by the compression-section adjustments, in particular, the compression rotation-point control. The compression section has no knowledge of the peaks, since it computes a 5-ms average when deciding on some gain. If the compression section calculates a gain that results in an average output program level of -8 dB relative to the limiting level, then limiting occurs only when the peak-to-average ratio becomes greater than 8 dB. If this margin is too small for the particular application, the rotation point should be lowered.

For program signals with a high peak-to-average ratio, the compression section pushes the average level up and the limiting section holds the peaks down. It is quite reasonable for them to alternate control. With both sections operating together, the density of the output program tends to be maximum. When the compression system tries to increase the average loudness, the peaks activate the limiter. The limiter reduces the gain but, with seldomly occurring peaks, allows it to return quickly. Thus, the loudness is unaffected. The advantage of combining the limiting and compressing function in one system should now be clear. Also, when the two sections are alternating control, the compression section is continuously monitoring the true output even if it is not controlling the gain. Thus, when the limiter relinquishes control, the compression section is prepared to resume operation.

Expansion Characteristics

The function of the expansion section is to preserve the effective signal-to-noise ratio. The increase in gain that accompanies the low-level program also amplifies the noise level. Clearly, this is inherent in compression systems, but it is not necessary to amplify the noise when no useful input program is present. The gain is set to 0 dB when noise alone is present. Implementing this requires that the expansion section distinguish between low-level program and noise. The discrimination is made on the basis of average level if one can assume that the noise level is at least 10 dB below the minimum useful signal level. It is also possible to use the spectral differences, since a program tends to have much more energy in the low frequencies and noise tends to have a much more uniform spectrum. As the program is fading below a threshold, the expansion section decreases the gain from the compression-factor to 0 dB.

When the program level is near the noise level, the resulting gain variations become extremely displeasing. Gain variations at high levels are masked by the complex program. The listener perceives the gain variations at low levels by the modulation of the background noise. Modulated noise is much more noticeable than uniform noise of the same level. Also, the random-noise fluctuations present with the program vary the gain, which in turn modulates the program. Having noise-modulated program and program-modulated noise is intolerable. Only by preventing the gain from varying very rapidly, can one avoid this situation. The expansion attack and release times are made large and both the rate and amount of gain variation are reduced.

A long attack time, however, prevents the expansion section from quickly releasing control to the compression section when high-level program appears. Under this condition, the expansion section still thinks the program is noise and holds the gain at 0 dB. A short time later, it finally releases control and the listener perceives a sharp change in loudness. If the attack times of both the compression and expansion sections are equal, the transition is graceful and unnoticed. Both the low-level and high-level conditions can be satisfied if the attack time is program-controlled. In this system, the attack time is proportional to the logarithm of the input program level. With program signals sufficiently large to activate the compression section, the expansion attack time is approximately equal to the compression attack time.

Similarly, when normal program stops, leaving only noise, the transition from compression to expansion sections must also be smooth and unnoticed. That is, the release time of the two sections should match. If the expansion release time is larger, the compression section allows the gain to rise by the compression factor before the expansion section reduces it back to zero. If the expansion release time is smaller the expansion section reduces the gain by the compression factor before the compression section increases the gain to zero. A temporary increase or decrease in the gain during the transition between sections can be very conspicuous. There is no simple way to match the release times, since the compression release time is a complex function of the program. But, if the expansion section is properly coupled to the compression section, the expansion release time automatically adapts to the compression release time.

The importance of this cross-coupling can be appreciated by considering the following illustration. Suppose the program consists of an announcer talking in a noisy environment. Without the coupling, the noise would vary with each syllable and pause. This noise "pumping" is a common indication that the program is compressed. With the coupling, however, the gain does not change (with normal speech at 0 dB). Each time there is a pause, control is switched from compression to expansion; following the pause, control is returned to the compression section. The level of music program signals, in contrast to speech, does not decrease suddenly. During a pause, the reverberation of the last notes fades until the level is low enough to activate the expansion section. The expansion section reduces the gain and causes the reverberation to fade more rapidly. In this case, the gain was more than 0 dB when the expansion section was activated. The cross coupling insures that the gain returns to 0 dB smoothly.

Summary

Dynamic range compression is necessary in order to pass an input program signal through a channel that has less dynamic range than the program. Under normal conditions, the sound engineer either adjusts the gain manually, or employs an automated compression system. If he uses the automated system, he must still be able to choose the characteristics of compression that are appropriate for his application. This system allows for the 6 static parameters to be adjusted by the user; it also adjusts its own dynamic characteristics as a function of the program signal. Thus, the gain variations are distributed in such a way that the listener does not perceive a sense of unnaturalness.

The operating region is divided into three sections: expansion, compression, and limiting. The expansion section processes low-level signals to distinguish low-level program from noise alone. The compression section, for high-level signals, compresses the long- and short-term dynamic range. The limiting section is operational whenever peaks would exceed the maximum signal limitations of the channel.

The expansion section allows the gain, for low-level program, to be equal to the compression factor; yet, it reduces the gain to zero when the input signal is noise alone. Long attack and release times are necessary in order to keep low-level gain variations at minimum. For large signals, however, the attack time must become short enough to match the compression section's attack time; and the release time must track the release time of the compression section.

The compression section is most important, since all of the major gain variations are caused by this section. The amount of long- and short-term dynamic range compression is controlled by the release-time function. This function is divided into two sections: long-term and short-term. Thus, a given gain decrease caused by a loudness transient is divided into these two classes. The percentage assigned to long-term compression recovers slowly; the remainder, assigned to short-term compression, recovers quickly. Control of the short-term compression is achieved by varying the percentage of the gain change assigned to that class, rather than by varying the short-term release time. The percentage, however, is controlled by the long-term release time. In this way, a balance between long- and short-term dynamic range compression is achieved. This parameter, in the automatic mode, is controlled by the density, or dynamic range, of the original program.

The limiting section functions only to prevent peak signals from overloading the channels. Therefore, it has a high compression ratio, and a short attack time. The undesirable aspects of limiting require that it not happen too often. Thus, for frequent peaks, the gain is reduced but not allowed to return until the peaks are longer near the limiting threshold. For seldomly occurring peaks, the release time is short. In this way, frequent peaks do not produce frequent limiting, and randomly occurring peaks do not blank out the program as a result of a long release time.

REFERENCES

[1] E. Zwicker, "Ein Verfahren zur Berechnung der Lautstaeker," *Acustica*, vol. 10, no. 1, p. 304, 1960.

[2] S. S. Stevens, "The psychophysics of sensory function," in *Sensory Communication*, W. A. Rosenblith, Ed. Cambridge, Mass.: M.I.T. Press, 1961.

[3] B. Blesser and A. Kent, "Analysis of a feedback-controlled limiter using a logarithmic measuring scale," *IEEE Trans. Audio and Electroacoustics*, vol. AU-16, pp. 481–485, December 1968.

Barry A. Blesser (S'68) was born in Brooklyn, N. Y., on April 3, 1943. He received the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1964 and 1965, respectively. As an undergraduate, he participated in the cooperative program with Raytheon Company's Advanced Development Laboratory, Wayland, Mass., where he did research for his Master's degree on "Radar Video Data Processor for Air Traffic Control."

While he was Technical Manager of the M.I.T. radio station, WTBS, he became interested in audio frequency communications systems with an emphasis on the psychology of perception. This later led to the development of a dynamic range compression system which is now being manufactured by Elektromesstechnik Wilhelm Franz KG, Lahr, West Germany. During the last few years, he has acted as a Consultant for special problems in audio frequency systems. Since 1966, he has been an Instructor of Electrical Engineering and a member of the Research Laboratory of Electronics at M.I.T., where he has been engaged in research on the perception of spectrally rotated speech as a means of studying the complexity of human cognition. At present, he is finishing his doctoral dissertation and is teaching a communications project laboratory course.

Mr. Blesser is a member of Tau Beta Pi and Eta Kappa Nu.