DYNAMIC RANGE COMPRESSION OF AUDIO SIGNALS CONSISTENT WITH RECENT TIME-VARYING LOUDNESS MODELS

Ryan J. Cassidy

Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University
Stanford, CA 94305

ABSTRACT

Dynamic range compression may be used to increase the volume of the softer passages of an audio signal relative to its louder portions, thus making the signal better suited to transmission through or storage on a given medium. The level-detection characteristics of typical contemporary dynamic range compressors are analyzed and investigated, thus revealing the shortcomings of such models in light of knowledge about steady-state and time-varying loudness as perceived by the human auditory system. The design of an equal-loudness filter, desired to improve the steady-state properties of compressor level detection, is presented. Finally, the time-varying properties of the level detection scheme presented, configured via attack and release times, are tuned to provide optimal correspondence with a recently proposed model of time-varying loudness.

1. INTRODUCTION

There are many situations when the range of "levels" that a particular signal achieves over time is too large or too small for convenient transmission through or storage on a given medium. Speech signals, for example, may often have portions that are too quiet for transmission across a noisy telephone line (if intelligibility is to be maintained). In a live musical performance, the softer passages may be too quiet to be heard above noise from the audience or air-conditioning system. When listening to music in a car or airplane, noise from the engine or other sources may totally drown out the softer sections of a particular program. It is in situations like these that dynamic range compression may be used to reduce the range of levels in a given signal, hopefully making most or all of the portions of the signal suitable for transmission.

In the 1930s, dynamic range compression was proposed to boost the level of quiet portions of speech for subsequent transmission over a telephone line[1]. Since then, a variety of compressor topologies and designs have evolved (see, for example, [2], [3], [4]), and are in fairly broad use today. In addition to situations in which compression is used to bring softer passages above the noise floor, dynamic range compression is also widely used in the recording industry to enhance the aesthetics of audio signals.

In this paper, we attempt to reconcile contemporary compressor characteristics and parameters with some recently proposed models for time-varying loudness (as perceived by the human auditory system). We also point out that basic compressor topologies as described in the literature [1], [2], [3], [4] use a loudness or level measure that is inconsistent with well-known facts about the loudness of amplitude-invariant or steady sinusoids (again as perceived by humans). To remedy this shortcoming, we present the design of an equal-loudness filter. Finally, we show some examples of this new compressor operating on some basic signals.

Supported by the Natural Sciences and Engineering Research Council of Canada.

2. DYNAMIC RANGE COMPRESSOR ARCHITECTURE

A typical dynamic range compressor architecture is comprised of two main blocks:

1. a level detector, to measure the amplitude envelope of the input signal and thus estimate its level or loudness, and
2. a gain control mechanism based on a static transfer characteristic (which represents the desired output level for a given input level) to control the gain applied to the input signal.

In this paper, we focus primarily on the analysis and design of the level detector. In this way, we examine typical level detection schemes used by contemporary dynamic range processors, and compare their performance to the steady-state and time-varying characteristics of loudness (as perceived by humans).

3. INSTANTANEOUS LEVEL MEASUREMENT AND SOUND LEVEL DETECTION

Many contemporary dynamic range compressors (as described in the literature [1], [2], [3], [4]) effectively derive their sound level measure by first computing the square of the input signal $x(t)$:

$$l_i(t) = (x(t))^2, \tag{1}$$

where $l_i(t)$ is the signal we shall hereafter refer to as the instantaneous sound level.

When the input signal is a "switched-on sinusoid"[1],

$$x(t) = A\sin(2\pi ft)u(t), \tag{3}$$

the instantaneous level is given by

$$l_i(t) = A^2\sin^2(2\pi ft)u(t) = A^2\frac{1 - cos(2\pi(2f)t)}{2}u(t). \tag{4}$$

[1] Note $u(t)$ denotes the unit-step function:

$$u(t) = \begin{cases} 1 & t >= 0 \\ 0 & t < 0 \end{cases}. \tag{2}$$

The instantaneous sound level is next passed through a one-pole lowpass filter, described by the differential equations [4]

$$\frac{\partial l(t)}{\partial t} = \begin{cases} \frac{1}{\tau_a} \cdot (l_i(t) - l(t)) & l_i(t) \geq l(t) \\ \frac{1}{\tau_r} \cdot (l_i(t) - l(t)) & l_i(t) < l(t) \end{cases}, \tag{5}$$

where τ_a is the attack time, τ_r is the release time, and $l(t)$ is the sound level (output) of the detector. A corresponding discrete-time implementation is given by the difference equations

$$l(n) = \begin{cases} \alpha_a l_i(n) + (1-\alpha_a) l(n-1) & l_i(n) \geq l(n) \\ \alpha_r l_i(n) + (1-\alpha_r) l(n-1) & l_i(n) < l(n) \end{cases}, \tag{6}$$

where α_a is the discrete-time attack parameter, α_r is the discrete-time release parameter, and $l_i(n)$ and $l(n)$ are discrete-time analogs of $l_i(t)$ and $l(t)$ described above (let f_s be the sampling frequency in Hz).

The step responses of the level detectors described above are given by

$$s_l(t) = (1 - e^{-t/\tau_a}) u(t), \tag{7}$$

for the continuous-time implementation and

$$s_l(n) = (1 - (1-\alpha_a)^{n+1}) u(n) \tag{8}$$

for the discrete-time implementation. From these equations, it may be seen that the detector output tends monotonically and asymptotically toward input level for a step input.

Under the simplifying assumption that $\tau_r = \tau_a$, the magnitude frequency responses are given by

$$|H(j\omega)| = \frac{\frac{1}{\tau_a}}{\sqrt{\omega^2 + \left(\frac{1}{\tau_a}\right)^2}} \tag{9}$$

for the continuous-time case and

$$|H(e^{j\Omega})| = \frac{\alpha_a}{\sqrt{1 + (1-\alpha_a)^2 - 2(1-\alpha_a)\cos\Omega}}, \tag{10}$$

for the discrete-time case, where ω represents continuous-time angular frequency (in rad/s) and Ω represents discrete-time angular frequency (in rad/sample). The 3-dB frequencies $\omega_{3-\mathrm{dB}}$ and $\Omega_{3-\mathrm{dB}}$ are given by

$$\omega_{3-\mathrm{dB}} = \frac{1}{\tau_a} \tag{11}$$

and

$$\Omega_{3-\mathrm{dB}} = \cos^{-1}\left(\frac{(1-\alpha_a)^2 - 1}{2(1-\alpha_a)}\right). \tag{12}$$

When the detector input is the switched-on sinusoid described by Eq. (3), it may be seen that, as a first approximation, if $2\pi(2f) \gg \omega_{3-\mathrm{dB}}$ (or, in the discrete-time case, $\frac{2\pi(2f)}{f_s} \gg \Omega_{3-\mathrm{dB}}$), the $\cos(2\pi(2f)t)$ term in the squared input will be attenuated such that it may be neglected, and the detector output will exponentially approach one-half the square of the original sinusoid amplitude:

$$\lim_{t\to 0} l(t) \approx \frac{A^2}{2}. \tag{13}$$

If the detector output is subsequently applied to a device which computes the square root of its input, then the output is approximately equal to the RMS value of the sinusoid[2].

[2] This serves to explain the term "RMS detector" used when describing popular dynamic range compressors[5]

If the $\cos(2\pi(2f)t)$ term in Equation (3) is not negligible, then there will be steady-state "ripple" in the detector output, which we denote $l_r(t)$, given by the formula

$$l_r(t) = \frac{-A^2}{2} \cdot \left(|H(j\omega)| \cos(2\pi(2f)t + \angle H(j\omega))\right)\Big|_{\omega=2\pi(2f)}, \tag{14}$$

with a similar expression for the discrete-time case.

4. EQUAL LOUDNESS FILTER DESIGN

It should be emphatically noted that the compression architecture heretofore described derives a sound level that, when the input is a steady sinusoid, is independent of frequency, provided the frequency is sufficiently higher than the 3-dB cut-off of the lowpass filter described in Section 3[3]. It is a well known fact of auditory perception, however, that the sensitivity of the human auditory system to steady sinusoids is decidedly frequency dependent. The first reasonably accurate research detailing this assertion was published by Fletcher and Munson in 1933[6], and was subsequently repeated and revised by Robinson and Dadson[7]. The key results of these studies were the so-called "equal loudness contours"—that is, the frequency-dependent sound intensities (in dB SPL) required to match the loudness a steady pure tone to that of a 1 kHz test tone with a reference sound intensity (again in dB SPL). The equal-loudness contours obtained from the Robinson-Dadson study are shown in Figure 1. Ideally, a dynamic range compressor should derive a frequency-dependent sound level for steady tones in accordance with the equal loudness contours.

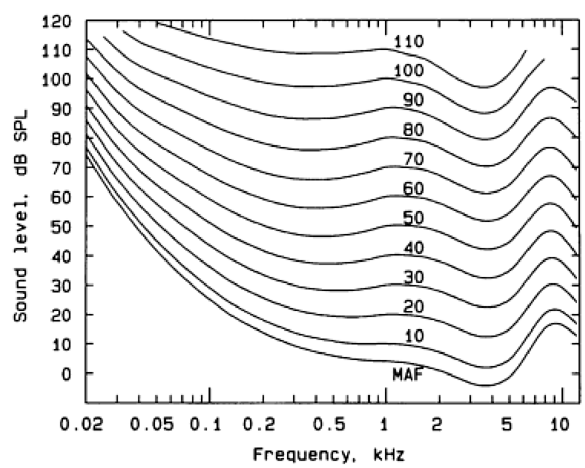

Fig. 1. Robinson-Dadson equal loudness contours [7](a revision of the Fletcher-Munson equal loudness contours).

The goal, then, is to design a filter whose magnitude frequency response approximates the equal loudness contours. If the input signal is passed through this filter prior to the RMS level detection described in Section 3, the loudness of steady tones should be much more accurately estimated by the detector, and the overall compressor performance should be greatly improved.

[3] For typical attack and release time settings, this condition is usually satisfied.

The first requirement is to decide on a "representative curve" whose shape approximates that of the ensemble of equal loudness contours shown in Figure 1. The representative contour, obtained via averaging the equal loudness contours, is shown in Figure 2. For the filter design, we first convert the zero-phase magnitude response to a minimum-phase spectrum obtained using a method described in [8]. Next we obtain an IIR filter using an *equation-error method*[9], conveniently accessible via MATLAB's `invfreqz()` function. Through experimentation with various filter orders, it quickly became apparent that an increasingly accurate approximation could be obtained with increasing filter order (for filter orders up to at least several hundred). Figure 2 shows the approximation obtained with a filter of order 175, which, upon visual inspection, is deemed sufficient. It should be noted that this filter order has been chosen rather arbitrarily, and it is straightforward to design a lower order filter (with a slightly poorer match to the equal-loudness characteristic) using a technique identical to the one described here.

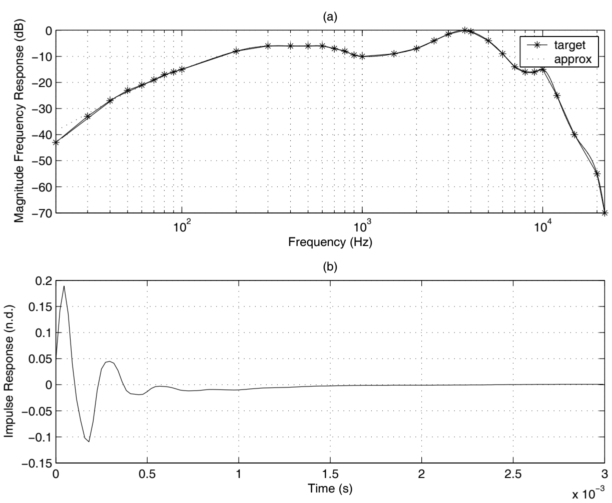

Fig. 2. Equal-loudness filter design: (a) magnitude frequency response of target equal loudness filter (`target`), with the approximate magnitude response (`approx`) obtainined using the MATLAB `invfreqz` function; (b) impulse response of approximation.

5. ATTACK AND RELEASE TIMES WITH SUITABLE CORRESPONDENCE TO TIME-VARYING LOUDNESS MODELS

The previous section focussed on and attempted to remedy the shortcomings of dynamic range compressors with respect to steady-state loudness. In this section, we examine recently compiled knowledge about time-varying loudness, and attempt to tailor our compressor design to accomodate these facts.

In their recently proposed "Model of Loudness Applicable to Time-Varying Sounds", Moore and Glasberg note that "... it is generally agreed that, for a fixed intensity, loudness [of switched-on sinusoids] increases with increasing duration for durations up to 100–200 ms, and then remains roughly constant. For durations less that 100 ms, the loudness increases by roughly 10 phons[4] for each ten-fold increation in duration..."[10]. With this in mind, they proceed to tune the attack time of their proposed loudness model (also based on the 1st-order damping of an instantaneous loudness signal similar to that presented in this paper) to match this time-varying loudness characteristic. The attack and release times proposed by Moore and Glasberg may be calculated as approximately 21.7 ms and 49.5 ms[10]. Thus, with attack and release times similar to these, the time-varying behaviour of the level detector discussed in Section 3 should be consistent with the time-varying loudness characteristics of the human auditory system, thus improving compressor performance.

6. IMPLEMENTATION AND RESULTS

The compressor design presented in this paper has been simulated on a digital computer, using switched-on and "switched-off" sinusoids to verify proper level detector operation. A sample of the results is shown in Figure 3 and Figure 4.

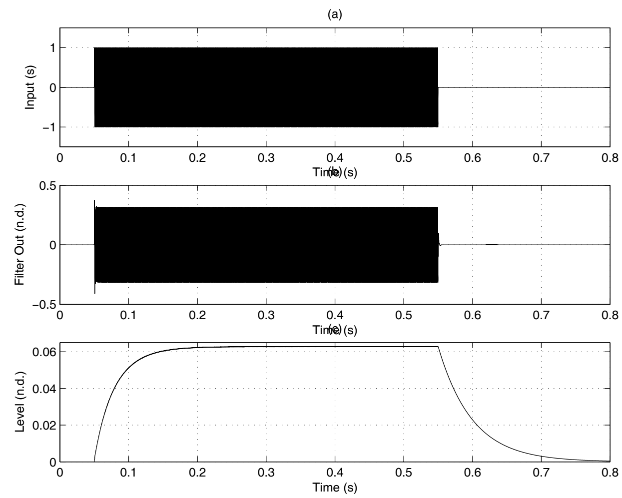

Fig. 3. Dynamic range compressor level detector output in response to a 1 kHz gated sinusoid: (a) input signal (switched on at 50 ms and off at 550 ms), (b) equal-loudness filter output, (c) level detector output.

Part (b) of each figure shows the equal-loudness filter correctly scaling the amplitude of the sinusoid to the magnitude value indicated in Figure 2 (note the 4 kHz sinusoid amplitude in Figure 4 is significantly greater than that of the 1 kHz sinusoid in Figure 3, confirming the increased sensitivity of the human auditory system to tones at this frequency). Note the transient response of the filter to the switched-on sinusoid, as seen in the figures, is reasonably well-behaved. Finally, part (c) of the figures shows the level detector output, with attack and release times set as described in Section 5. Note the release time is slightly longer than the attack time, as desired (more accurate values of the attack and release times have been obtained from the plots, to verify that they match the design values). Also note that the level detector output tends exponentially toward and away from the mean-square of the filtered sinusoid amplitude, again as predicted in Section 3.

[4] The *phon* is a frequency-independent loudness measure, implicitly in decibels. Since 1 kHz is the reference frequency for loudness measures, the loudness of a 1 kHz tone in phons has the same value its the sound intensity level (in dB SPL).

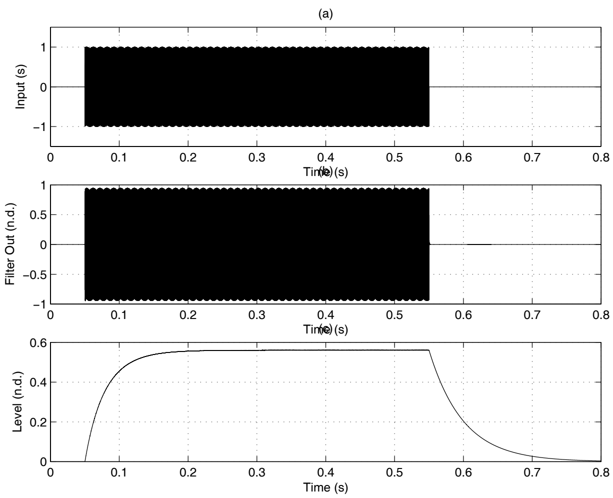

Fig. 4. Dynamic range compressor level detector output in response to a 4 kHz gated sinusoid: (a) input signal (switched on at 50 ms and off at 550 ms), (b) equal-loudness filter output, (c) level detector output.

7. CONCLUSIONS

We have presented a dynamic range compressor design whose steady-state and time-varying level detection characteristics match the loudness characteristics of the human auditory system. Compressors based on such schemes should provide improved dynamic range reduction for a variety of audio signals. Possibilities for future development and investigation include the use of higher order instantaneous level damping systems, and the use of lower-order approximations to the equal-loudness contours for steady tones.

8. REFERENCES

[1] R. C. Mathes and S. B. Wright, "The compandor—an aid against static in radio telephony," *Bell System Technical Journal*, vol. 13, pp. 315–332, 1934.

[2] Fred Floru, "Attack and release time constants in rms-based compressors and limiters," *Proc. 99th A.E.S. Convention, New York*, 1995, preprint 4054 (C-4).

[3] Fred Floru, "Attack and release time constants in rms-based feedback compressors," *Journal of the Audio Engineering Society*, vol. 47, pp. 788–804, 1999.

[4] Jonathan S. Abel and David P. Berners, "On peak-detecting and rms feedback and feedforward compressors," *Proc. 115th A.E.S Convention, New York*, 2003.

[5] THAT 2252, *IC RMS-Level Detector Data Sheet*, THAT Corp., Milford, MA.

[6] Harvey Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *Journal of the Acoustical Society of America*, vol. 5, pp. 82–108, 1933.

[7] D. W. Robinson and R. S. Dadson, "A re-determination of the equal loudness relations for pure tones," *British Journal of Applied Physics*, vol. 7, pp. 166–181, 1956.

[8] Julius O. Smith, *Techniques for Digital Filter Design and System Identification with Application to the Violin*, Ph.D. thesis, Elec. Engineering Dept., Stanford University (CCRMA), June 1983, available as CCRMA Technical Report STAN-M-14.

[9] E. C. Levi, "Complex-curve fitting," *IRE Trans. on Automatic Control*, vol. AC-4, pp. 37–44, 1959.

[10] Brian R. Glasberg and Brian C. J. Moore, "A model of loudness applicable to time-varying sounds," *Journal of the Audio Engineering Society*, vol. 50, pp. 331–342, 2002.