Modeling the effects of dynamic range compression on signals in noise

Ryan M. Corey[a] and Andrew C. Singer[b]

Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA

ABSTRACT:

Hearing aids use dynamic range compression (DRC), a form of automatic gain control, to make quiet sounds louder and loud sounds quieter. Compression can improve listening comfort, but it can also cause unwanted distortion in noisy environments. It has been widely reported that DRC performs poorly in noise, but there has been little mathematical analysis of these noise-induced distortion effects. This work introduces a mathematical model to study the behavior of DRC in noise. By making simplifying assumptions about the signal envelopes, we define an effective compression function that models the compression applied to one signal in the presence of another. Using the properties of concave functions, we prove results about DRC that have been previously observed experimentally: that the effective compression applied to each sound in a mixture is weaker than it would have been for the signal alone; that uncorrelated signal envelopes become negatively correlated when compressed as a mixture; and that compression can reduce the long-term signal-to-noise ratio in certain conditions. These theoretical results are supported by software experiments using recorded speech signals. © *2021 Acoustical Society of America.*

https://doi.org/10.1121/10.0005314

I. INTRODUCTION

Hearing aids often perform poorly in noisy environments, where people with hearing loss need help most. One challenge for hearing aids in noise is a nonlinear processing technique known as dynamic range compression (DRC), which improves audibility and comfort by making quiet sounds louder and loud sounds quieter (Allen, 2003; Kates, 2005; Souza, 2002; Villchur, 1973). Compression is used in all modern hearing aids, but it can cause unwanted distortion when applied to multiple overlapping sounds. For example, a sudden noise can reduce the gain applied to speech sounds. This effect is well documented empirically but has been little studied mathematically. To better understand DRC in noisy environments, this work applies tools from signal processing theory to model the effects of DRC on sound mixtures.

The auditory systems of people with hearing loss often have reduced dynamic range: Quiet sounds need to be amplified in order to be audible, but loud sounds can cause discomfort. Hearing aids with DRC apply level-dependent amplification so that the output signal has a smaller dynamic range than the input signal. A typical DRC system is shown in Fig. 1. An envelope detector tracks the level of the input signal over time in one or more frequency bands while a compression function adjusts the amplification to keep the output level within a comfortable range. Both the envelope detector and the compression function are nonlinear processes, so when the input contains sounds from multiple

sources, changes in one component signal can affect the processing applied to the others.

This interaction between signals can be difficult to measure, but hearing researchers have found three quantifiable effects. First, noise can reduce the effect of a compressor, especially at low signal-to-noise ratios (SNR) (Souza *et al.*, 2006). The DRC system applies gain based on the stronger signal and has little effect on the dynamic range of the weaker signal. This effect can measured by comparing the overall dynamic ranges of the input and output signals (Braida *et al.*, 1982; Stone and Moore, 1992). Second, fluctuations in the input level of one component signal vary the output levels of other components. This interaction has been called across-source modulation (Stone and Moore, 2007) and can be measured using the correlation coefficient between output envelopes. Finally, at high SNR, compressors tend to amplify low-level noise more strongly than the higher-level signal of interest, which can reduce the long-term average SNR (Alexander and Masterson, 2015; Hagerman and Olofsson, 2004; Rhebergen *et al.*, 2009; Souza *et al.*, 2006).

The adverse effects of noise on DRC systems have been well documented empirically, but the problem has received little formal mathematical analysis. While experimental work is useful for studying the consequences of these effects, especially on human listeners, theoretical results can help to understand their causes. This work applies signal processing research methods to the DRC distortion problem: First, we make simplifying assumptions to develop a tractable mathematical model of a complex system. Next, we use that model to prove theorems that explain the behavior of the system.

[a]Electronic mail: corey1@illinois.edu, ORCID: 0000-0002-3122-124X.
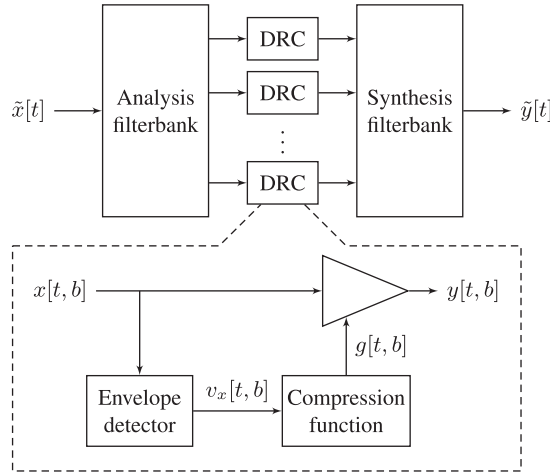[b]ORCID: 0000-0001-9926-7036.

FIG. 1. A typical DRC system performs automatic gain control in each of several frequency bands or channels.

Finally, we validate those assumption-based theoretical results using experiments with a realistic system.

Compression systems are difficult to analyze because of the complex interactions between the envelope detector and compression function, both of which are nonlinear. Using the simplifying assumption that envelopes are additive in signal mixtures, we can separate the effects of the envelope detector from those of the compression function. To characterize the interaction between signals in a mixture, we introduce the effective compression function (ECF), which relates the input and output levels of one signal in the presence of another. The ECF is used to explain the three effects described above: that noise reduces the effect of compression (Sec. IV), that compression induces negative correlation between signal envelopes (Sec. V), and that compression can reduce long-term average SNR in certain conditions (Sec. VI).

Each section includes a theorem about the effect and simulation experiments that illustrate it. The theorems rely on the concavity, or downward curvature, of the compression function and, like many results in signal processing theory, take the form of inequalities. The experiments illustrate the predictions of each theorem and show how the results change when the assumptions are violated.

II. DRC

Because most modern hearing aids are digital, we formulate the DRC system in discrete time. Let the sequence $\tilde{x}[t]$ be a sampled audio signal at the input of the DRC system, where t is the sample index. Let $\tilde{y}[t]$ be the output of the system.

A. Filterbank and envelope detector

Compression is often performed separately in several frequency bands. A filterbank splits the signal into B channels corresponding to different bands, which may be linearly or nonlinearly spaced and may or may not overlap. Let $x[t, b]$ and $y[t, b]$ be the filterbank representations of $\tilde{x}[t]$ and $\tilde{y}[t]$, respectively, in channels $b = 1, \ldots, B$.

The gain applied by DRC is calculated from the signal envelope, which tracks the signal level over time. Level is typically defined in terms of either magnitude ($|x|$) or power (x^2); this work uses power. Let the non-negative signal $v_x[t, b]$ be the envelope of the input signal $x[t, b]$ at time index t and channel b. In the theoretical analysis presented here, the envelope is an abstract property of a signal, such as a statistical variance. In real DRC systems, the envelope is estimated from the observed signal, typically using a moving average.

Most DRC systems respond faster to increases in signal level (attack mode) than to decreases in signal level (release mode) in order to suppress sudden loud sounds. There are many ways of implementing an envelope detector (Giannoulis *et al.*, 2012). A representative detector is the nonlinear recursive filter (Kates, 2008),

$$
v_x[t, b] = \begin{cases} \beta_\mathrm{a} v_x[t-1, b] + (1 - \beta_\mathrm{a})|x[t, b]|^2, \\ \qquad \text{if } |x[t, b]|^2 \geq v_x[t-1, b] \\ \beta_\mathrm{r} v_x[t-1, b] + (1 - \beta_\mathrm{r})|x[t, b]|^2, \text{otherwise}, \end{cases}
$$

$$(1)$$

for $b = 1, \ldots, B$, and where β_a and β_r are constants that determine the attack and release times.

Because envelope detection is a nonlinear process, it contributes to the distortion effects of DRC systems. The theorems in this work do not depend on the filterbank structure or the choice of attack and release time, but these parameters do affect the rate of fluctuation of the measured envelopes and therefore the distribution of envelope samples. Many nonlinear interaction effects are more severe for fast-acting and many-channel compression than for slow-acting and few-channel compression (Alexander and Masterson, 2015; Alexander and Rallapalli, 2017; Naylor and Johannesson, 2009; Plomp, 1988; Rallapalli and Alexander, 2019; Reinhart *et al.*, 2017), though these parameters do not necessarily impact speech intelligibility (Salorio-Corbetto *et al.*, 2020).

B. Compression function

A compression function $\mathcal{C}_b$ determines the instantaneous mapping between input level and target output level in each channel,

$$
v_y[t, b] = \mathcal{C}_b(v_x[t, b]), \quad b = 1, \ldots, B, \tag{2}
$$

where $v_y[t, b]$ is the target output level. The amplification applied in each channel is then

$$
g[t, b] = \sqrt{\frac{v_y[t, b]}{v_x[t, b]}}, \quad b = 1, \ldots, B, \tag{3}
$$

so that the output is the product

$$
y[t, b] = g[t, b]x[t, b], \quad b = 1, \ldots, B. \tag{4}
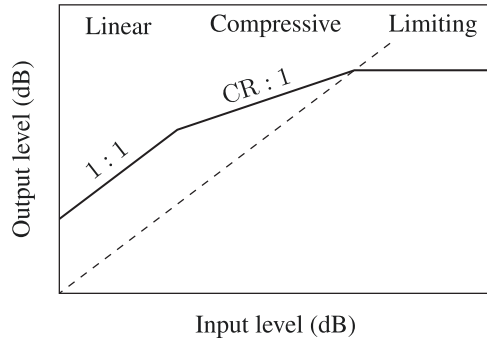$$

Ryan M. Corey and Andrew C. Singer

FIG. 2. A compression function $\mathcal{C}_b$, shown here on a logarithmic scale, maps input levels to output levels.

Note that the target output level $v_y[t, b]$ is not necessarily equal to the measured envelope of $y[t, b]$ because the envelope is a moving average. Longer release times cause gains to lag behind short-term signal levels, especially for dynamic signals such as speech (Braida *et al.*, 1982; Stone and Moore, 1992).

Although compression functions are defined here in terms of input and output level (i.e., power), they are often visualized and described on a logarithmic scale, such as in decibels (dB). A typical "knee-shaped" compression function is shown in Fig. 2: It features a linear region in which gain is constant, a compressive region where the output level increases by less than the input level, and and a limiting region that prevents the output from exceeding a maximum safe level.

The strength of compression can be characterized by the compression ratio (CR), which is the inverse of the slope of the compression function on a log-log scale, as shown in Fig. 2. For example, in a 3:1 compressor, the output increases by 1 dB for every 3 dB increase in the input. For a constant CR, the compression function is given by the power-law relationship

$$\mathcal{C}_b(v) = G_0[b]v^{(1/\mathrm{CR})}, \tag{5}$$

where $G_0[b]$ is a constant power gain factor. Thus, for a 3:1 compressor, the output level is proportional the cube root of the input level. In limiters, $\mathcal{C}_b(v)$ is constant and so the CR is infinite.

While most compressors reported in the literature use some combination of linear, power-law, and limiting compression functions, many others are possible. To make our analysis as general as possible, we allow the compression function to be any mapping between non-negative numbers such that the output level grows no faster than the input level. More precisely, we require it to be a concave function.

Definition 1. A function $\mathcal{C}_b(v)$ is a *compression function* if it is concave, non-negative, and nondecreasing for all $v > 0$.

In mathematics, a function $f(x)$ is said to be concave if for any $\lambda \in [0, 1]$ and any x_1 and x_2,

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2). \tag{6}$$

Note that Definition 1 includes non-differentiable functions such as knee-shaped compression curves. It excludes dynamic range expanders, which some hearing aids apply at low signal levels to reduce noise. The proofs in this work will also involve convex functions that satisfy Eq. (6) with the inequality reversed. Convex and concave functions are widely used to prove inequalities in signal processing and information theory (Cover and Thomas, 2006).

To describe how much a compression function reduces the dynamic range of a signal, we could compute its CR. Because the CR can be infinite, however, it is more convenient to work with its inverse, the compression slope.

Definition 2. For all points v at which a compression function $\mathcal{C}_b(v)$ is differentiable, the *compression slope* $\mathrm{CS}_b(v)$ is the slope of $\mathcal{C}_b(v)$ on a log-log scale,

$$\mathrm{CS}_b(v) = \frac{\mathrm{d}}{\mathrm{d}u} \ln \mathcal{C}_b(e^u)|_{u=\ln v}, \tag{7}$$

$$= \frac{\mathcal{C}_b'(v)}{\mathcal{C}_b(v)} v. \tag{8}$$

For example, if $\mathcal{C}_b(v) = G_0[b]v^\alpha$, then $\mathrm{CS}_b(v) = \alpha$ for all v. The smaller the compression slope, the more the dynamic range of the signal is reduced.

C. Experimental methods

To validate the predictions of the mathematical model, each section of this work includes experiments using speech recordings and a software DRC system. The theoretical results in this work rely on simplifying assumptions, but the simulation experiments are more realistic and therefore illustrate the limitations of the model. Wherever possible, the experiments use methods and performance metrics from prior work in the literature.

Although the compression function varies with each experiment, all simulations in this work use the same envelope detector. The input is first processed by a short-time Fourier transform with 8 ms windows and 50% overlap. A frequency-domain filterbank splits the signals into 6 Mel-spaced bands from 0 to 8 kHz, which are roughly linearly spaced at lower frequencies and exponentially spaced at higher frequencies. Within each band, the envelopes are computed using the nonlinear recursive filter (1) with an attack time of 10 ms and a release time of 50 ms as defined by ANSI S3.22–1996 (ANSI, 1996). All speech signals are 60-s clips derived from the Voice Cloning Toolkit (VCTK) dataset of quasi-anechoic read speech (Veaux *et al.*, 2017). The figures in this work use logarithmic scales for envelope level. These levels are given in dB relative to the mean wideband signal level. That is, each speech signal has a mean level of 0 dB across channels.

III. MODELING COMPRESSION OF SOUND MIXTURES

Hearing aids are often used in noisy environments with several simultaneous sound sources. The interactions between multiple signals are difficult to analyze because DRC involves two nonlinear operations: envelope detection

J. Acoust. Soc. Am. **150** (1), July 2021

Ryan M. Corey and Andrew C. Singer   161

and level-dependent amplification. To create a tractable model for sound mixtures, we make a simplifying assumption about the signal envelopes that allows us to separate the effects of these two nonlinearities, as shown in Fig. 3. Under this model, the filterbank and envelope detector determine the relationship between input signals and envelope values; they act independently on each component signal. Meanwhile, the compression function determines the output levels from these envelopes; it acts independently at each time index and within each channel. In this work, we focus on the compression function.

A. Envelope model

Suppose that the input to the system is $\tilde{x}[t] = \tilde{s}_1[t] + \tilde{s}_2[t]$, where $\tilde{s}_1[t]$ and $\tilde{s}_2[t]$ are two discrete-time signals. For example, $\tilde{s}_1$ and $\tilde{s}_2$ could be two speech signals as captured at the listening device microphone, including any reverberation effects. Because a filterbank is a linear system, the filterbank representation of the input is

$$x[t, b] = s_1[t, b] + s_2[t, b], \quad b = 1, \ldots, B, \qquad (9)$$

where $s_1[t, b]$ and $s_2[t, b]$ are the filterbank representations of $\tilde{s}_1[t]$ and $\tilde{s}_2[t]$, respectively.

Because envelope detection is a nonlinear process, the additivity property of Eq. (9) does not hold in general for the signal envelopes measured by practical envelope detectors. However, to simplify our analysis, the signal envelopes can be *modeled* as obeying additivity.

Assumption 1. *The envelopes* $v_{s_1}[t, b]$, $v_{s_2}[t, b]$, *and* $v_x[t, b]$ *of* $s_1[t, b]$, $s_2[t, b]$, *and* $x[t, b]$, *respectively, satisfy*

$$v_x[t, b] = v_{s_1}[t, b] + v_{s_2}[t, b], \quad b = 1, \ldots, B. \qquad (10)$$

This assumption is justified if we think of the envelopes as abstract properties of signals, such as parameters of a process that generates them, rather than as measurements. For example, suppose that $s_1[t, b]$ and $s_2[t, b]$ are sample functions of random processes that are uncorrelated with each other (Hajek, 2015). Then the variance of the mixture is given by $\mathrm{Var}(x[t, b]) = \mathrm{Var}(s_1[t, b]) + \mathrm{Var}(s_2[t, b])$. If

$v_x[t, b]$ were any linear transformation of the sequence $\mathrm{Var}(x[t, b])$, then the envelopes would satisfy Assumption 1. Because the variance is an ensemble average, not a time average, the processes need not be stationary or ergodic.

Of course, real compression systems cannot observe the underlying variance of a random process; they derive envelopes from recorded samples. The accuracy of the additive envelope model depends on the signals: Two sinusoids would strongly violate the assumption (Ludvigsen, 1993), while two signals that are disjoint across time and channels would satisfy it exactly. To test the accuracy of the assumption for a realistic envelope detector, we applied the software envelope detector described in Sec. II C to a mixture of two speech signals and compared the envelope of the mixture, $v_x[t, b]$, to the sum of the envelopes of the component signals, $v_{s_1}[t, b] + v_{s_2}[t, b]$. Figure 4 shows a set of envelope samples drawn from different time frames and frequency channels plotted on a decibel scale. In this experiment, the assumption is accurate to within 1 dB for 93% of samples.

B. Output model

Care is also required in analyzing the components of the output of a nonlinear system. Let $\tilde{y}[t] = \tilde{r}_1[t] + \tilde{r}_2[t]$, where $\tilde{r}_1[t]$ is the component of the output corresponding to $\tilde{s}_1[t]$ and $\tilde{r}_2[t]$ is the component corresponding to $\tilde{s}_2[t]$. For systems with the additivity property, like linear filters, these components can be calculated by applying the same system to $\tilde{s}_1$ and $\tilde{s}_2$. For nonlinear systems like DRC, each component of the output depends on all components of the input. In general, nonlinear distortion artifacts cannot be clearly attributed to one input signal or the other, and they cannot be easily classified as helpful or harmful to intelligibility (Ludvigsen, 1993). For the relatively mild compression used in hearing aids—compared to aggressive compression-based effects in electronic music, for example—a reasonable approach is to treat the nonlinear system as a time-varying linear system.

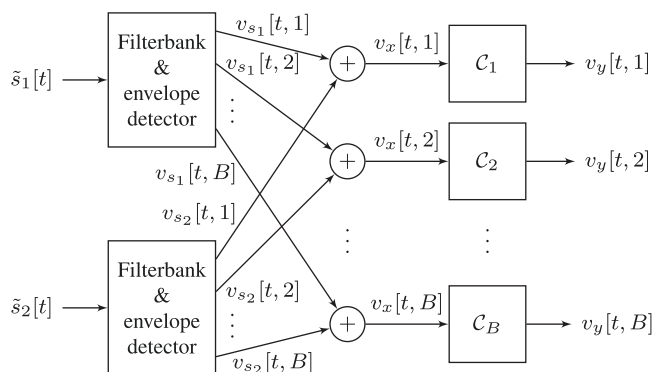In this work, the output components are determined by calculating the level-dependent amplification sequence

FIG. 3. A simplified model separates the effects of the filterbank and envelope detector from those of the compression functions $\mathcal{C}_1, \ldots, \mathcal{C}_B$. The former act independently across signals, while the latter act independently across time and channels.
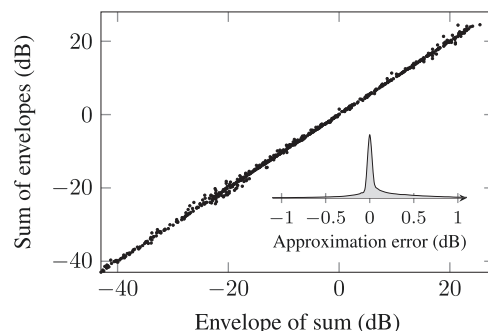
FIG. 4. Empirical evaluation of Assumption 1 with a mixture of two speech signals. The plotted points are samples of the sum of the envelopes, $v_{s_1}[t, b] + v_{s_2}[t, b]$, and the envelope of the sum, $v_x[t, b]$. The inset plot shows a histogram of the difference $10 \log_{10}(v_{s_1} + v_{s_2}) - 10 \log_{10} v_x$.

Ryan M. Corey and Andrew C. Singer

$g[t, b]$ based on the mixture $x[t, b]$, then applying it to each component,

$$y[t, b] = g[t, b]x[t, b] \tag{11}$$

$$= g[t, b](s_1[t, b] + s_2[t, b]) \tag{12}$$

$$= \underbrace{g[t, b]s_1[t, b]}_{r_1[t,b]} + \underbrace{g[t, b]s_2[t, b]}_{r_2[t,b]}, \tag{13}$$

for all time indices t and channels $b = 1, \ldots, B$. This definition of the output components is used in the mathematical analysis below. Similarly, in the software simulations, the two input signals are stored in memory alongside their mixture and the amplification sequence is applied separately to each, allowing the output components to be computed exactly. This time-varying linear approach to computing output coefficients is conceptually related to the phase inversion technique of Hagerman and Olofsson (2004), which is often used in laboratory experiments with real hearing aids where the time-varying amplification sequence cannot be observed directly.

C. Effective compression function

The additive models for the input envelopes and output signal components, while imperfect, allow us to study the dominant source of nonlinearity in a DRC system: the compression function. Although the signals $s_1[t, b]$ and $s_2[t, b]$ may have different levels, the amplification $g[t, b]$ applied to both of them is the same and is computed from the overall level of the input signal,

$$g[t, b] = \sqrt{\frac{\mathcal{C}_b(v_x[t, b])}{v_x[t, b]}}. \tag{14}$$

Under Assumption 1, the amplification is

$$g[t, b] = \sqrt{\frac{\mathcal{C}_b(v_{s_1}[t, b] + v_{s_2}[t, b])}{v_{s_1}[t, b] + v_{s_2}[t, b]}}, \tag{15}$$

resulting in the output levels

$$v_{r_1}[t, b] = \frac{\mathcal{C}_b(v_{s_1}[t, b] + v_{s_2}[t, b])}{v_{s_1}[t, b] + v_{s_2}[t, b]} v_{s_1}[t, b], \tag{16}$$

$$v_{r_2}[t, b] = \frac{\mathcal{C}_b(v_{s_1}[t, b] + v_{s_2}[t, b])}{v_{s_1}[t, b] + v_{s_2}[t, b]} v_{s_2}[t, b], \tag{17}$$

for channels $b = 1, \ldots, B$. The gain and therefore the output levels are functions of both input signal levels, as illustrated in Fig. 5. The gain applied to $s_1[t, b]$ in the presence of $s_2[t, b]$ is weaker than it would have been for $s_1[t, b]$ alone. To characterize this effect, we can define an effective compression function (ECF) that relates the input and output levels of one signal in the presence of another.

Definition 3. The ECF $\hat{\mathcal{C}}_b(v_1|v_2)$ applied to a signal with level $v_1 > 0$ in the presence of a signal with level $v_2 \geq 0$ is given by

$$\hat{\mathcal{C}}_b(v_1|v_2) = \frac{\mathcal{C}_b(v_1 + v_2)}{v_1 + v_2} v_1, \tag{18}$$

where $\mathcal{C}_b(v)$ is the compression function applied to the mixture level $v_1 + v_2$.

Using this definition, Eqs. (16) and (17) become

$$v_{r_1}[t, b] = \hat{\mathcal{C}}_b(v_{s_1}[t, b]|v_{s_2}[t, b]), \tag{19}$$

$$v_{r_2}[t, b] = \hat{\mathcal{C}}_b(v_{s_2}[t, b]|v_{s_1}[t, b]), \tag{20}$$

for $b = 1, \ldots, B$. The ECF expresses the dependence between the levels of the two signal components. The ECF can be used to mathematically characterize the nonlinear interactions between signals in DRC systems, including the effective CR, the across-source modulation effect, and the SNR.

IV. EFFECTIVE COMPRESSION PERFORMANCE

When DRC is applied to a mixture of multiple signals, it has a weaker effect on the dynamic range of each component signal than it would if they were processed independently. Intuitively, if a signal of interest is weaker than a noise source, then the noise level will determine the gain applied to both signals and the target signal will not be compressed. Even when the target signal has a higher level, the noise will cause the gain to decrease less than it should with respect to the target level.

To quantify the effect of noise on compression performance, we can measure the change in the output level of the target signal in response to a change in its input level and compare that relationship to the nominal CR. Even without noise, the long-term effective compression ratio (ECR) is generally lower than the nominal ratio because of the time-averaging effects of the envelope detector (Braida *et al.*, 1982; Stone and Moore, 1992). However, it has been
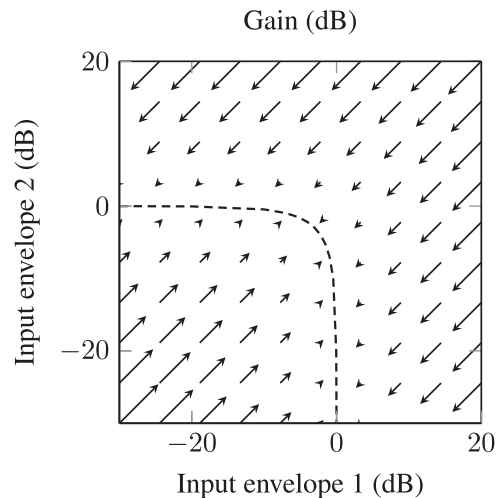

FIG. 5. Gain applied to a mixture signal as a function of the signal envelopes $v_{s_1}[t, b]$ and $v_{s_2}[t, b]$ for $\mathcal{C}_b(v) = v^{1/3}$ under Assumption 1. The length of the arrows is proportional to the power gain $g^2[t, b]$ in dB and the dashed curve shows the equilibrium mixture level $\mathcal{C}_b(v_{s_1} + v_{s_2}) = v_{s_1} + v_{s_2}$.

J. Acoust. Soc. Am. **150** (1), July 2021

Ryan M. Corey and Andrew C. Singer   163

observed that the ECR of a DRC system is further reduced in the presence of noise (Souza *et al.*, 2006). While this noise-induced reduction in CR has been previously measured as a long-term average for particular signals, here we show that it is a short-term effect caused by the concavity of the compression function. Although the magnitude of the reduction depends on the compression function and the signal characteristics, the effect occurs for every compression function and at every SNR.

Because the instantaneous CR can be infinite, we will instead use the effective compression slope, defined as the log-log slope of the ECF.

Definition 4. If $\hat{\mathcal{C}}_b(v_1|v_2)$ is differentiable with respect to v_1, then the *effective compression slope* $\hat{CS}_b(v_1|v_2)$ is given by

$$\hat{CS}_b(v_1|v_2) = \frac{\partial}{\partial u}\hat{\mathcal{C}}_b(e^u|v_2)|_{u=\ln v_1} \tag{21}$$

$$= \frac{\frac{\partial}{\partial v_1}\hat{\mathcal{C}}_b(v_1|v_2)}{\hat{\mathcal{C}}_b(v_1|v_2)}v_1. \tag{22}$$

Note that the effective compression slope defined here is a function of the two signal levels, so it provides a more complete description of the compression system than the long-term ECR. Furthermore, it only measures the instantaneous effects of interaction between signals, not the time-averaging effects of the envelope detector. The simplified envelope model allows us to analyze these two compression-weakening mechanisms separately.

A. Noise reduces compression performance

Using the properties of the ECF, it can be shown that the effective compression slope from Definition 4 is always larger than the nominal compression slope from Definition 2—equivalently, the instantaneous ECR is always smaller than the nominal CR—meaning that, when applied to a mixture, the system is less compressive on each component signal than it would be if applied to the components separately,

$$\hat{CS}_b(v_{s_1}|v_{s_2}) \geq CS_b(v_{s_1}+v_{s_2}), \quad b = 1, \dots, B. \tag{23}$$

Notably, this result applies to any pair of signal levels v_{s_1} and v_{s_2}. Whereas the across-source modulation result of Sec. V relies on probabilistic averaging and the SNR result of Sec. VI uses time averaging, Eq. (23) holds for each individual envelope sample.

The proof relies on concavity. Because the lemmas and theorems in this work follow from the properties of compression functions, which act independently across time and frequency, the time and channel indices $[t, b]$ are omitted in their statements and proofs.

Theorem 1. *If a compression function $\mathcal{C}(v)$ is differentiable at $v_x = v_1 + v_2$, then its effective compression slope satisfies*

$$\hat{CS}(v_1|v_2) \geq CS(v_x), \tag{24}$$

with equality if $\mathcal{C}(v)$ is linear or if $v_2 = 0$.

Proof. Because $\mathcal{C}(v)$ is defined to be concave and non-negative for $v > 0$, it follows that

$$\mathcal{C}(v) - v\mathcal{C}'(v) \geq 0 \tag{25}$$

for all v at which $\mathcal{C}$ is differentiable, with equality if $\mathcal{C}$ is linear. The effective compression slope is given by

$$\hat{CS}(v_1|v_2) = \frac{\frac{\partial}{\partial v_1}\hat{\mathcal{C}}(v_1|v_2)}{\hat{\mathcal{C}}(v_1|v_2)}v_1 \tag{26}$$

$$= \frac{v_x}{\mathcal{C}(v_x)}\left(\frac{\mathcal{C}'(v_x)v_1 + \mathcal{C}(v_x)}{v_x} - \frac{\mathcal{C}(v_x)v_1}{v_x^2}\right) \tag{27}$$

$$= \frac{\mathcal{C}'(v_x)}{\mathcal{C}(v_x)}v_1 + 1 - \frac{v_1}{v_x} \tag{28}$$

$$= \frac{\mathcal{C}'(v_x)}{\mathcal{C}(v_x)}v_x - \frac{\mathcal{C}'(v_x)}{\mathcal{C}(v_x)}v_2 + \frac{v_2}{v_x} \tag{29}$$

$$= CS(v_x) + \frac{v_2}{v_x\mathcal{C}(v_x)}\left(\mathcal{C}(v_x) - v_x\mathcal{C}'(v_x)\right) \tag{30}$$

$$\geq CS(v_x), \tag{31}$$

with equality if $\mathcal{C}$ is linear or if $v_2 = 0$. $\square$

Suppose that v_1 corresponds to a sound source of interest and v_2 is the level of unwanted noise. The proof illustrates that the effective compression slope for the target increases with the level of the interfering signal. For example, in the limit as v_1/v_x approaches 0, the slope from Eq. (28) approaches 1, so that the system applies linear gain to the target signal. At low SNR, the gain applied to both signals is determined by the noise. The theorem shows, however, that even at high SNR, the compression effect is slightly weaker.

B. Experiments

Theorem 1 shows that under the simplified envelope model, noise always reduces the effect of compression on a signal of interest. To verify this result experimentally in a realistic system, the software DRC system described in Sec. II C was applied to a mixture of speech at a wideband level of 0 dB and varying levels of white Gaussian noise with a nominal CR of 3:1.

Figure 6 shows the effective compression performance of the system for three wideband SNRs. The dashed line shows the nominal compression function $\mathcal{C}_b(v) = v^{1/3}$ for all b. The solid curves are the ECFs $\hat{\mathcal{C}}_b(v_{s_1}|v_{s_2})$ predicted by the model for constant noise power v_{s_2} equal to the variance of the Gaussian noise. The plotted points show speech input envelope samples and their corresponding output levels computed using the time-varying gain of the software DRC system. The curves align closely with the nominal compression function when the speech has higher level than the noise, but they are nearly linear when the noise has higher level.
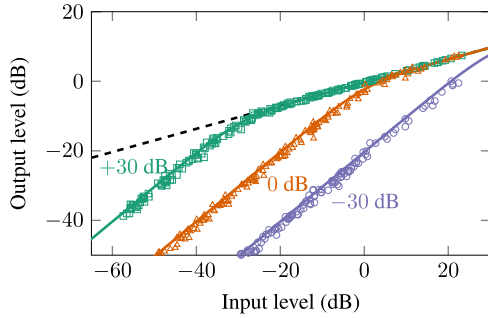
Ryan M. Corey and Andrew C. Singer

FIG. 6. (Color online) ECF for speech and white noise at different wide-band input SNRs. The dashed line shows the nominal compression function, the curves show the ECFs evaluated with constant noise level, and the plotted points show speech envelope samples.

The long-term ECR depends on the distribution of envelope samples. For target signals whose envelopes are usually above the noise level, the long-term ECR will be close to the nominal ratio. When the noise is usually more intense, as in the rightmost curve of Fig. 6, the long-term ECR will be close to unity. Using the method of Souza *et al.* (2006), which measures dynamic range between the 5th and 95th percentiles of input and output envelope samples, and averaging across signal bands, the long-term ECRs from the experiments here were 1.01 at –30 dB SNR, 1.17 at 0 dB, and 1.75 at +30 dB.

V. ACROSS-SOURCE MODULATION DISTORTION

DRC creates distortion in mixtures because the presence of one signal alters the gain applied to another signal. It has been observed experimentally (Alexander and Masterson, 2015; Stone and Moore, 2004, 2007, 2008) that when two signals are mixed together and passed through a compressor, their output envelopes become negatively correlated: As one sound becomes louder, the other sound becomes quieter. The across-source modulation coefficient, a measure of this negative correlation, was found to be correlated with reduced speech intelligibility (Stone and Moore, 2007, 2008).

A. Output levels are anticorrelated

The ECF can be used to show that if the input envelopes $v_{s_1}[t, b]$ and $v_{s_2}[t, b]$ are independent random processes, then the covariance between the output levels in each channel is negative,

$$\text{Cov}(v_{r_1}[t, b], v_{r_2}[t, b]) \leq 0, \quad b = 1, \ldots, B. \quad (32)$$

The covariance is an ensemble mean over the distributions of the envelope samples $v_{r_1}[t, b]$ and $v_{r_2}[t, b]$. Although the covariance is often measured empirically using a time average, our mathematical analysis applies to each time index and channel independently.

We first show that the ECF is nondecreasing in one envelope and nonincreasing in the other.

Lemma 1. Any ECF $\hat{\mathcal{C}}(v_1|v_2)$ is nondecreasing in v_1 and nonincreasing in v_2 for $v_1, v_2 \geq 0$.

Proof. Because $\mathcal{C}(v)$ is nondecreasing and v_2 is non-negative, $\hat{\mathcal{C}}(v_1|v_2) = \mathcal{C}(v_1 + v_2)(v_1/v_1 + v_2)$ is the product of two nondecreasing functions of v_1 and is therefore nondecreasing. Because $\mathcal{C}(v)$ is concave and non-negative, $\mathcal{C}(v)/v$ is nonincreasing for $v > 0$. Then $\mathcal{C}(v_1 + v_2)/(v_1 + v_2)$ is nonincreasing in v_2. □

Next, we will need the following result about functions of random variables. Let $\mathbb{E}$ denote the expectation of a random variable, that is, its probabilistic mean.

Lemma 2. If $f(x)$ is nondecreasing, $g(x)$ is nonincreasing, X is a random variable, and $\mathbb{E}[f(X)]$, $\mathbb{E}[g(X)]$, and $\mathbb{E}[f(X)g(X)]$ exist, then

$$\mathbb{E}[f(X)g(X)] \leq \mathbb{E}[f(X)]\mathbb{E}[g(X)]. \quad (33)$$

Proof. See Appendix A. □

We can now prove that independent envelopes become negatively correlated when compressed.

Theorem 2. If $\hat{\mathcal{C}}(v_1|v_2)$ is an ECF and V_1 and V_2 are independent random variables, then

$$\text{Cov}\left(\hat{\mathcal{C}}(V_1|V_2), \hat{\mathcal{C}}(V_2|V_1)\right) \leq 0. \quad (34)$$

Proof. Because $\text{Cov}(\hat{\mathcal{C}}(V_1|V_2), \hat{\mathcal{C}}(V_2|V_1)) = \mathbb{E}[\hat{\mathcal{C}}(V_1|V_2)\hat{\mathcal{C}}(V_2|V_1)] - \mathbb{E}[\hat{\mathcal{C}}(V_1|V_2)]\mathbb{E}[\hat{\mathcal{C}}(V_2|V_1)]$, it is sufficient to show that

$$\mathbb{E}\left[\hat{\mathcal{C}}(V_1|V_2)\hat{\mathcal{C}}(V_2|V_1)\right] \leq \mathbb{E}\left[\hat{\mathcal{C}}(V_1|V_2)\right]\mathbb{E}\left[\hat{\mathcal{C}}(V_2|V_1)\right]. \quad (35)$$

From Lemma 1, $\hat{\mathcal{C}}(v_1|v_2)$ is a nondecreasing function of v_1 and a nonincreasing function of v_2. Let $\mathbb{E}[X|Y]$ denote the conditional expectation of X given Y. From iterated expectation and application of Lemma 2, we have

$$\mathbb{E}\left[\hat{\mathcal{C}}(V_1|V_2)\hat{\mathcal{C}}(V_2|V_1)\right]$$
$$= \mathbb{E}_{V_2}\left[\mathbb{E}_{V_1}\left[\hat{\mathcal{C}}(V_1|V_2)\hat{\mathcal{C}}(V_2|V_1)|V_2\right]\right] \quad (36)$$
$$\leq \mathbb{E}_{V_2}\left[\mathbb{E}_{V_1}\left[\hat{\mathcal{C}}(V_1|V_2)|V_2\right]\mathbb{E}_{V_1}\left[\hat{\mathcal{C}}(V_2|V_1)|V_2\right]\right]. \quad (37)$$

Now, because V_1 and V_2 are independent, $\mathbb{E}_{V_1}[\hat{\mathcal{C}}(V_1|V_2)|V_2]$ is a nonincreasing function of V_2 and $\mathbb{E}_{V_1}[\hat{\mathcal{C}}(V_2|V_1)|V_2]$ is a nondecreasing function of V_2. Applying Lemma 2 once more,

$$\mathbb{E}\left[\hat{\mathcal{C}}(V_1|V_2)\hat{\mathcal{C}}(V_2|V_1)\right] \leq \mathbb{E}_{V_2}\left[\mathbb{E}_{V_1}\left[\hat{\mathcal{C}}(V_1|V_2)|V_2\right]\right]$$
$$\cdot \mathbb{E}_{V_2}\left[\mathbb{E}_{V_1}\left[\hat{\mathcal{C}}(V_2|V_1)|V_2\right]\right] \quad (38)$$
$$= \mathbb{E}\left[\hat{\mathcal{C}}(V_1|V_2)\right]\mathbb{E}\left[\hat{\mathcal{C}}(V_2|V_1)\right]. \quad (39)$$
□

For linear gain, the theorem holds with equality because $\hat{\mathcal{C}}(v_1|v_2)$ does not depend on v_2. The magnitude of the negative correlation depends on the compression function: Stronger

compression causes the ECFs and the conditional expectations to increase or decrease more quickly, resulting in a stronger negative correlation. The channel structure and time constants of the envelope detector affect the correlation indirectly by altering the distributions of V_1 and V_2.

B. Experiments

To illustrate the negative correlation effect with a realistic envelope detector, the software DRC system from Sec. II C was applied to mixtures of speech and white Gaussian noise with a 3:1 CR. Each plot in Fig. 7 shows pairs of measured envelope samples for two signals: the input mixtures $(v_{s_1}[t,b], v_{s_2}[t,b])$ on the left and the output mixtures $(v_{r_1}[t,b], v_{r_2}[t,b])$ on the right. The top plots are for speech and white noise and the bottom plots are for two speech signals. The correlation coefficient ρ is computed on a linear scale and averaged across channels. The dashed curve shows the equilibrium level $v_{r_1} + v_{r_2} = 1$, which indicates perfect negative correlation ($\rho = -1$).

The input levels are mostly uncorrelated between the two component signals, but DRC shifts the levels according to a vector field like that in Fig. 5, producing correlated output levels. Because the white noise has nearly constant envelope, the effect of DRC is most visible at high speech levels: When the speech signal is strong, both speech and noise are attenuated, bending the distribution of level pairs downward and producing a negative correlation. When the interfering signal is speech, which has a wide dynamic range, the signal components interact at all levels. At low instantaneous SNR, the weaker speech signal of interest is modulated according to the level of the stronger interfering speech.

VI. OUTPUT SNR

Of the three nonlinear interaction effects discussed in this work, the impact of DRC on long-term SNR is both the
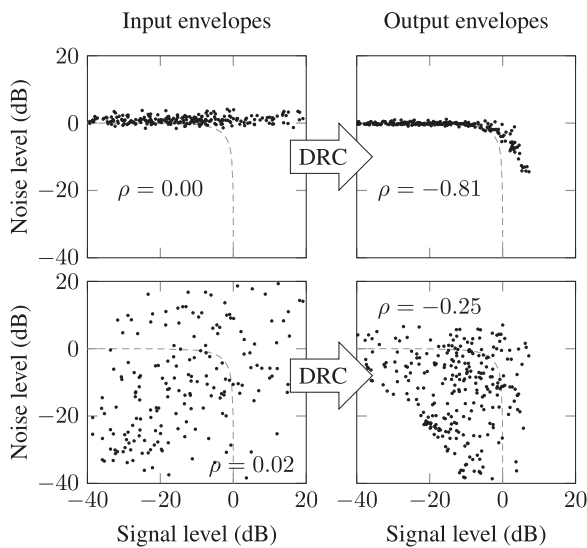

FIG. 7. Sample input and output envelope pairs for mixtures of two signals in a DRC system. Top: Speech and white noise. Bottom: Speech and speech. The dashed curve shows the equilibrium level of the compression function as in Fig. 5.

most studied empirically and the most challenging to analyze mathematically. Hagerman and Olofsson (2004) showed that fast-acting compression improved the average output SNR of speech in babble noise at negative average input SNR but made it worse for positive input SNR. Souza et al. (2006) found that DRC reduced the SNR of speech in speech-shaped noise. Naylor and Johannesson (2009) showed that this effect depends on the type of noise, filter-bank structure, and envelope time constants. Brons et al. (2015) and Miller et al. (2017) demonstrated the effect in hearing aids that include nonlinear noise reduction, which can interact with DRC in complex ways (Kortlang et al., 2018). SNR changes appear to be more severe for fast compression (Alexander and Masterson, 2015; May et al., 2018) and less severe with reverberation (Reinhart et al., 2017).

It is important to remember that long-term SNR is not the same as intelligibility. Listening tests suggest that DRC can improve intelligibility with some types of noise but not others (Kowalewski et al., 2018; Rhebergen et al., 2017; Rhebergen et al., 2009; Yund and Buckles, 1995).

A. Output SNR for constant-envelope noise

While it is difficult to say much in general about the effect of compression on output SNR, we can prove a result for an important special case: a target signal with a time-varying envelope and a noise signal with constant envelope. Stationary white noise, for example, has constant variance, and its measured envelope fluctuates only slightly over time. Meanwhile, information-rich signals such as speech tend to vary rapidly. Many classic speech enhancement algorithms, such as spectral subtraction, assume that the noise spectrum is constant while the speech level varies (Loizou, 2013). When the mixture level is high, it is assumed that speech is present and the gain is increased, while at lower levels the output is attenuated to remove noise. Because these speech enhancement systems amplify high-level signals and attenuate low-level signals, they act as dynamic range *expanders*.

If a dynamic range expander can improve long-term SNR, it stands to reason that a compressor might make it worse. To see why, let us analyze the effect of compression on the average SNR over time. Because the envelope is proportional to the power of a signal component, the average SNR at the input is given by

$$\text{SNR}_{\text{in}}[b] = \frac{\text{mean}_t v_{s_1}[t,b]}{\text{mean}_t v_{s_2}[t,b]}, \quad b = 1, \dots, B, \quad (40)$$

and the average SNR at the output is

$$\text{SNR}_{\text{out}}[b] = \frac{\text{mean}_t v_{r_1}[t,b]}{\text{mean}_t v_{r_2}[t,b]} \quad (41)$$

$$= \frac{\text{mean}_t \hat{\mathcal{C}}_b(v_{s_1}[t,b]|v_{s_2}[t,b])}{\text{mean}_t \hat{\mathcal{C}}_b(v_{s_2}[t,b]|v_{s_1}[t,b])}. \quad (42)$$

If the compression function were linear, then the input and output SNRs would be identical. For a concave

compression function with convex gain, it can be shown that, if the noise envelope is constant, then the average output SNR is lower than the average input SNR,

$$\text{SNR}_{\text{out}}[b] \leq \text{SNR}_{\text{in}}[b], \quad b = 1, \ldots, B. \tag{43}$$

Unlike in Sec. V, here the envelopes are not modeled as random processes and the quantities of interest are time averages, not ensemble averages.

The proofs in this section rely on an additional technical condition on the compression function. Not only must $\mathcal{C}_b(v)$ be non-negative and concave, the gain function $\mathcal{C}_b(v)/v$ must be convex. This condition is satisfied for many smooth compression functions, including linear, power-law, and logarithmic, but not for some functions with corners like that in Fig. 2. This condition ensures that the ECF is concave in its first argument and convex in its second.

Lemma 3. If $\mathcal{C}(v)$ is a compression function and $\mathcal{C}(v)/v$ is convex for all $v > 0$, then the ECF $\hat{\mathcal{C}}(v_1|v_2)$ is concave in v_1 and convex in v_2.

Proof. See Appendix B. □

This property lets us take advantage of Jensen's inequality (Cover and Thomas, 2006), one form of which states that for any convex function $f(x)$,

$$\text{mean}_t f(x[t]) \geq f(\text{mean}_t x[t]), \tag{44}$$

with equality if $f(x)$ is linear or $x[t]$ is constant. The same property holds with the inequality reversed if $f(x)$ is a concave function. Jensen's inequality allows us to prove that the average output SNR is no larger than the average input SNR.

Theorem 3. *If $\mathcal{C}(v)$ is a compression function and $\mathcal{C}(v)/v$ is convex for all $v > 0$, $v_1[t] > 0$ for all t, and $v_2[t] = \bar{v}_2 > 0$ for all t, then*

$$\text{SNR}_{\text{out}} \leq \text{SNR}_{\text{in}} \tag{45}$$

with equality if $v_1[t]$ is constant or $\mathcal{C}$ is linear.

Proof. Since $v_2[t]$ is fixed, the output SNR can be written

$$\text{SNR}_{\text{out}} = \frac{\text{mean}_t \hat{\mathcal{C}}(v_1[t]|\bar{v}_2)}{\text{mean}_t \hat{\mathcal{C}}(\bar{v}_2|v_1[t])}. \tag{46}$$

The numerator is the mean over t of a concave function of $v_1[t]$. By Jensen's inequality,

$$\text{mean}_t \hat{\mathcal{C}}(v_1[t]|\bar{v}_2) \leq \hat{\mathcal{C}}(\text{mean}_t v_1[t]|\bar{v}_2), \tag{47}$$

with equality when $\mathcal{C}$ is linear or $v_1[t]$ is constant. Similarly, the denominator is the mean over t of a convex function of $v_1[t]$. Again applying Jensen's inequality,

$$\text{mean}_t \hat{\mathcal{C}}(\bar{v}_2|v_1[t]) \geq \hat{\mathcal{C}}(\bar{v}_2|\text{mean}_t v_1[t]), \tag{48}$$

with equality when $\mathcal{C}$ is linear or $v_1[t]$ is constant. Let $\bar{v}_1 = \text{mean}_t v_1[t]$. Since the numerator and denominator of Eq. (46) are both positive, we have

$$\text{SNR}_{\text{out}} \leq \frac{\hat{\mathcal{C}}(\bar{v}_1|\bar{v}_2)}{\hat{\mathcal{C}}(\bar{v}_2|\bar{v}_1)} \tag{49}$$

$$= \frac{\bar{v}_1 \mathcal{C}(\bar{v}_1 + \bar{v}_2)/(\bar{v}_1 + \bar{v}_2)}{\bar{v}_2 \mathcal{C}(\bar{v}_1 + \bar{v}_2)/(\bar{v}_1 + \bar{v}_2)} \tag{50}$$

$$= \frac{\bar{v}_1}{\bar{v}_2} \tag{51}$$

$$= \text{SNR}_{\text{in}}, \tag{52}$$

with equality when $\mathcal{C}$ is linear or $v_1[t]$ is constant. □

B. Experiments

Because Theorem 3 requires stronger assumptions than the other theorems in this work, it is especially important to validate its predictions experimentally. Figure 8 compares the input and output SNRs for speech in white Gaussian noise—which has a relatively steady but not constant envelope—at different CRs. For this section, the software simulations use a knee-shaped compression function like that in Fig. 2, which is commonly used in hearing aids but violates the technical condition required for Lemma 3. The knee point is 40 dB below the wideband average speech level. Although the assumptions are violated, the results still show the behavior predicted by the theorem. The SNR-reducing effect is greatest at high input SNRs; at low input SNRs, the noise level determines the gain and the ECF is linear, so the SNR is not affected.

As an inequality, Theorem 3 does not predict the magnitude of SNR reduction, but the equality condition suggests that the effects are smaller for more linear compression functions. Indeed, the experiments show that higher CRs have stronger effects on SNR, which is consistent with results in the literature (Naylor and Johannesson, 2009; Rhebergen *et al.*, 2009).

Theorem 3 applies only to constant-envelope noise. When the target and noise signals both vary strongly with time, the weaker signal will be amplified more and the stronger signal less, pushing their average output levels closer together. Figure 9 shows the results of the SNR experiment with 3:1 knee-shaped compression and different noise types. With white noise, the long-term SNR is always reduced, as predicted by Theorem 3. With speech babble, generated by
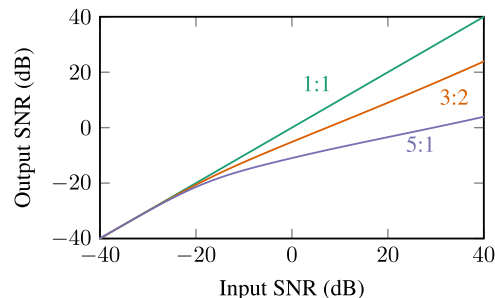

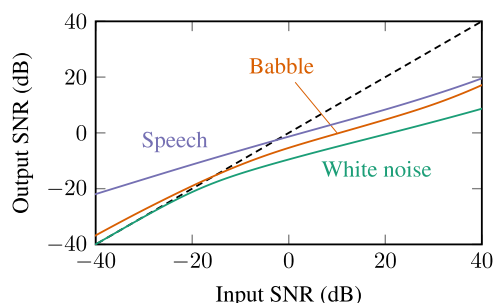FIG. 8. (Color online) Effect of DRC with different CRs on long-term SNR of speech in white noise.

FIG. 9. (Color online) Effect of compression on long-term SNR of mixtures of speech with different types of noise.

mixing 14 VCTK speech clips, the SNR is slightly increased at low input SNRs. When the target and interference signals are both single-talker speech signals, the long-term SNR is improved when it is negative but made worse when it is positive.

These results align well with those in the literature. Hagerman and Olofsson (2004) showed that fast-acting compression can improve negative SNRs but worsen SNRs near zero for speech in babble noise. Naylor and Johannesson (2009) found that output SNR is always reduced for speech in unmodulated noise, greatly reduced at positive input SNR and slightly reduced at negative input SNR for speech in modulated noise, and symmetrically increased at negative input SNR and decreased at positive input SNR for a mixture of two speech signals. Reinhart *et al.* (2017) performed experiments with different numbers of talkers and found that the SNR improvement at negative input SNRs declined with each additional interfering talker, consistent with the results for speech babble here.

VII. DISCUSSION

The mathematical analysis presented here confirms the empirical evidence from the hearing literature that DRC causes unintended distortion in noise. The effects of this distortion depend on the characteristics of the signals, especially their relative levels. At low SNR, the ECF for the target signal becomes nearly linear and the dynamic range of that signal is not changed. At high SNR, the signal of interest is amplified by less than the noise, reducing average SNR. At all SNRs, the signal components modulate each other, compressing the weaker signal according to the level of the stronger component.

The theorems in this work apply to ideal envelopes that obey the additivity assumption. In that sense, they are optimistic predictions. Real DRC systems that use measured envelopes would exhibit even stronger interactions between signals. Further theoretial work is required to model distortion within the envelope detector, predict the effects of filterbank structure and envelope time constants, and show how these nonlinearities interact with those of the compression function.

Can anything be done to improve the performance of DRC systems in noise? The analysis shows that all these effects are caused by the concave curvature of the compression function, which is also what makes the system compressive. The results for effective compression performance and across-source modulation hold instantaneously, not just over time, and apply to any compression function and any combination of signals, even hypothetical ideal signals that have independent envelopes. It seems, then, that nonlinear interactions are inevitable whenever signals are compressed as a mixture.

A possible solution is to compress the component signals of a mixture independently, as music producers do when mixing instrumental and vocal recordings. Listening tests have shown improved intelligibility when signals are compressed before rather than after mixing (Rhebergen *et al.*, 2009; Stone and Moore, 2008). Of course, real hearing aids do not have access to the unmixed source signals, so a practical multisource compression system must perform source separation. Hassager *et al.* (2017) used a single-microphone classification method to separate direct from reverberant signal components, helping to preserve spatial cues that can be distorted by DRC. May *et al.* (2018) proposed a single-microphone separation system that applies fast-acting compression to speech components and slow-acting compression to noise components; listening experiments with an ideal separation algorithm improved both quality and intelligibility (Kowalewski *et al.*, 2020). Corey and Singer (2017) used a multimicrophone separation method to apply separate compression functions to each of several competing speech signals. The output exhibited better measures of across-source modulation distortion, effective compression performance, and SNR compared to a conventional system. The modeling framework described here could be applied to analyze the performance of these multisource compression systems and to devise new ones.

VIII. CONCLUSIONS

The mathematical tools introduced in this work can help researchers to understand the distortion effects of conventional DRC systems in noise and to devise new approaches to nonlinear processing for mixtures of multiple signals. The additive envelope model allows the envelope detector and compression function to be analyzed independently, greatly reducing the complexity of the system. The ECF models interactions between signal envelopes at the input and output of any compression function, characterizing system behavior across all signal levels. It can be used to analyze instantaneous interactions or integrated into long-term or probabilistic models to study average effects.

Like the human auditory system itself, DRC is a complex nonlinear system that defies simple analysis. By modeling how DRC systems behave in the presence of noise, we can develop and analyze new strategies for nonlinear signal processing in the most challenging environments.

Ryan M. Corey and Andrew C. Singer

APPENDIX A: PROOF OF LEMMA 2

Lemma 2. If $f(x)$ is nondecreasing, $g(x)$ is nonincreasing, X is a random variable, and $\mathbb{E}[f(X)]$, $\mathbb{E}[g(X)]$, and $\mathbb{E}[f(X)g(X)]$ exist, then

$$\mathbb{E}[f(X)g(X)] \leq \mathbb{E}[f(X)]\mathbb{E}[g(X)]. \tag{A1}$$

Proof. Because $f(x)$ is nondecreasing and $g(x)$ is nonincreasing, for every x and y we have

$$[f(x) - f(y)][g(x) - g(y)] \leq 0. \tag{A2}$$

It is sufficient to show that $\mathbb{E}[f(X)g(X)] - \mathbb{E}[f(X)]\mathbb{E}[g(X)] \leq 0$. If X has cumulative distribution function $P(x)$, then

$$\mathbb{E}[f(X)g(X)] - \mathbb{E}[f(X)]\mathbb{E}[g(X)]$$
$$= \int_x f(x)g(x)\mathrm{d}P(x) - \int_x f(x)\mathrm{d}P(x)\int_y g(y)\mathrm{d}P(y) \tag{A3}$$

$$= \int_x\int_y f(x)[g(x) - g(y)]\,\mathrm{d}P(y)\mathrm{d}P(x) \tag{A4}$$

$$= \int_x\int_{y<x} f(x)[g(x) - g(y)]\,\mathrm{d}P(y)\mathrm{d}P(x)$$
$$+ \int_y\int_{x<y} f(x)[g(x) - g(y)]\,\mathrm{d}P(x)\mathrm{d}P(y) \tag{A5}$$

$$= \int_x\int_{y<x} f(x)[g(x) - g(y)]\,\mathrm{d}P(y)\mathrm{d}P(x)$$
$$+ \int_x\int_{y<x} f(y)[g(y) - g(x)]\,\mathrm{d}P(y)\mathrm{d}P(x) \tag{A6}$$

$$= \int_x\int_{y<x} [f(x) - f(y)][g(x) - g(y)]\,\mathrm{d}P(y)\mathrm{d}P(x) \tag{A7}$$

$$\leq 0. \tag{A8}$$

Equation (A5) swaps the order of integration using Fubini's theorem (Knapp, 2005) and line Eq. (A6) exchanges the integration variables x and y. □

APPENDIX B: PROOF OF LEMMA 3

Lemma 3. If $\mathcal{C}(v)$ is a compression function and $\mathcal{C}(v)/v$ is convex for all $v > 0$, then the ECF $\hat{\mathcal{C}}(v_1|v_2)$ is concave in v_1 and convex in v_2.

Proof. Starting with Definition 3 and letting $v_1 = \lambda p + (1 - \lambda)q$,

$$\hat{\mathcal{C}}(v_1|v_2) = \frac{\mathcal{C}(\lambda p + (1-\lambda)q + v_2)}{\lambda p + (1-\lambda)q + v_2}(\lambda p + (1-\lambda)q) \tag{B1}$$

$$= \mathcal{C}(\lambda(p + v_2) + (1-\lambda)(q + v_2)) \tag{B2}$$

$$- v_2\frac{\mathcal{C}(\lambda(p + v_2) + (1-\lambda)(q + v_2))}{\lambda(p + v_2) + (1-\lambda)(q + v_2)}. \tag{B3}$$

Because $\mathcal{C}(v)$ is concave and $\mathcal{C}(v)/v$ is convex,

$$\hat{\mathcal{C}}(v_1|v_2) \geq \lambda\mathcal{C}(p + v_2) + (1-\lambda)\mathcal{C}(q + v_2) \tag{B4}$$

$$- v_2\left(\lambda\frac{\mathcal{C}(p + v_2)}{p + v_2} + (1-\lambda)\frac{\mathcal{C}(q + v_2)}{q + v_2}\right) \tag{B5}$$

$$= \lambda\frac{\mathcal{C}(p + v_2)}{p + v_2}p + (1-\lambda)\frac{\mathcal{C}(q + v_2)}{q + v_2} \tag{B6}$$

$$= \lambda\hat{\mathcal{C}}(p|v_2) + (1-\lambda)\hat{\mathcal{C}}(q|v_2). \tag{B7}$$

Therefore, $\hat{\mathcal{C}}(v_1|v_2)$ is concave in v_1.

Similarly, letting $v_2 = \lambda p + (1 - \lambda)q$,

$$\hat{\mathcal{C}}(v_1|v_2) = \frac{\mathcal{C}(v_1 + \lambda p + (1-\lambda)q)}{v_1 + \lambda p + (1-\lambda)q}v_1 \tag{B8}$$

$$= \frac{\mathcal{C}(\lambda(v_1 + p) + (1-\lambda)(v_1 + q))}{\lambda(v_1 + p) + (1-\lambda)(v_1 + q)}v_1 \tag{B9}$$

$$\leq \lambda\frac{\mathcal{C}(v_1 + p)}{v_1 + p}v_1 + (1-\lambda)\frac{\mathcal{C}(v_1 + q)}{v_1 + q}v_1 \tag{B10}$$

$$= \lambda\hat{\mathcal{C}}(v_1|p) + (1-\lambda)\hat{\mathcal{C}}(v_1|q). \tag{B11}$$

Therefore, $\hat{\mathcal{C}}(v_1|v_2)$ is convex in v_2. □

Alexander, J. M., and Masterson, K. (**2015**). "Effects of WDRC release time and number of channels on output SNR and speech recognition," Ear Hear. **36**(2), e35–e49.

Alexander, J. M., and Rallapalli, V. (**2017**). "Acoustic and perceptual effects of amplitude and frequency compression on high-frequency speech," J. Acoust. Soc. Am. **142**(2), 908–923.

Allen, J. B. (**2003**). "Amplitude compression in hearing aids," in *MIT Encyclopedia of Communication Disorders*, edited by R. Kent (MIT Press, Cambridge, MA), pp. 413–423.

ANSI (**1996**). ANSI S3.22-1996, *Specification of Hearing Aid Characteristics* (ANSI, New York).

Braida, L., Durlach, N., De Gennaro, S., Peterson, P., Bustamante, D., Studebaker, G., and Bess, F. (**1982**). "Review of recent research on multi-band amplitude compression for the hearing impaired," in *The Vanderbilt Hearing Aid Report*, edited by G. Studebaker and F. H. Bess (York Press, London).

Brons, I., Houben, R., and Dreschler, W. A. (**2015**). "Acoustical and perceptual comparison of noise reduction and compression in hearing aids," J. Speech Lang. Hear. Res. **58**(4), 1363–1376.

Corey, R. M., and Singer, A. C. (**2017**). "Dynamic range compression for noisy mixtures using source separation and beamforming," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 15–18, New Paltz, NY.

Cover, T. M., and Thomas, J. A. (**2006**). *Elements of Information Theory* (Wiley, New York).

Giannoulis, D., Massberg, M., and Reiss, J. D. (**2012**). "Digital dynamic range compressor design—A tutorial and analysis," J. Audio Eng. Soc. **60**(6), 399–408.

J. Acoust. Soc. Am. **150** (1), July 2021

Ryan M. Corey and Andrew C. Singer   169

Hagerman, B., and Olofsson, Å. (**2004**). "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise," Acta Acust. united Ac. **90**(2), 356–361.

Hajek, B. (**2015**). *Random Processes for Engineers* (Cambridge University Press, Cambridge, UK).

Hassager, H. G., May, T., Wiinberg, A., and Dau, T. (**2017**). "Preserving spatial perception in rooms using direct-sound driven dynamic range compression," J. Acoust. Soc. Am. **141**(6), 4556–4566.

Kates, J. M. (**2005**). "Principles of digital dynamic-range compression," Trends Amplif. **9**(2), 45–76.

Kates, J. M. (**2008**). *Digital Hearing Aids* (Plural Publishing, London).

Knapp, A. W. (**2005**). *Basic Real Analysis* (Birkhäuser, Basel, Switzerland).

Kortlang, S., Chen, Z., Gerkmann, T., Kollmeier, B., Hohmann, V., and Ewert, S. D. (**2018**). "Evaluation of combined dynamic compression and single channel noise reduction for hearing aid applications," Int. J. Audiol. **57**, S43–S54.

Kowalewski, B., Dau, T., and May, T. (**2020**). "Perceptual evaluation of signal-to-noise-ratio-aware dynamic range compression in hearing aids," Trends Hear. **24**, 233121652093053–233121652093014.

Kowalewski, B., Zaar, J., Fereczkowski, M., MacDonald, E. N., Strelcyk, O., May, T., and Dau, T. (**2018**). "Effects of slow-and fast-acting compression on hearing-impaired listeners' consonant–vowel identification in interrupted noise," Trends Hear. **22**, 233121651880087–233121651880012.

Loizou, P. C. (**2013**). *Speech Enhancement: Theory and Practice* (CRC Press, Boca Raton, FL).

Ludvigsen, C. (**1993**). "The use of objective methods to predict the intelligibility of hearing aid processed speech," in *Proceedings of the* 15th Danavox Symposium, December 15, 1993, Kolding, Denmark, pp. 81–94.

May, T., Kowalewski, B., and Dau, T. (**2018**). "Signal-to-noise-ratio-aware dynamic range compression in hearing aids," Trends Hear. **22**, 233121651879090–233121651879012.

Miller, C. W., Bentler, R. A., Wu, Y.-H., Lewis, J., and Tremblay, K. (**2017**). "Output signal-to-noise ratio and speech perception in noise: Effects of Algorithm," Int. J. Audiol. **56**(8), 568–579.

Naylor, G., and Johannesson, R. B. (**2009**). "Long-term signal-to-noise ratio at the input and output of amplitude-compression systems," J. Am. Acad. Audiol. **20**(3), 161–171.

Plomp, R. (**1988**). "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function," J. Acoust. Soc. Am. **83**(6), 2322–2327.

Rallapalli, V. H., and Alexander, J. M. (**2019**). "Effects of noise and reverberation on speech recognition with variants of a multichannel adaptive dynamic range compression scheme," Int. J. Audiol. **58**(10), 661–669.

Reinhart, P., Zahorik, P., and Souza, P. E. (**2017**). "Effects of reverberation, background talker number, and compression release time on signal-to-noise ratio," J. Acoust. Soc. Am. **142**(1), EL130–EL135.

Rhebergen, K. S., Maalderink, T. H., and Dreschler, W. A. (**2017**). "Characterizing speech intelligibility in noise after wide dynamic range compression," Ear Hear. **38**(2), 194–204.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2009**). "The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise," J. Acoust. Soc. Am. **126**(6), 3236–3245.

Salorio-Corbetto, M., Baer, T., Stone, M. A., and Moore, B. C. (**2020**). "Effect of the number of amplitude-compression channels and compression speed on speech recognition by listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **147**(3), 1344–1358.

Souza, P. E. (**2002**). "Effects of compression on speech acoustics, intelligibility, and sound quality," Trends Amplif. **6**(4), 131–165.

Souza, P. E., Jenstad, L. M., and Boike, K. T. (**2006**). "Measuring the acoustic effects of compression amplification on speech in noise," J. Acoust. Soc. Am. **119**(1), 41–44.

Stone, M. A., and Moore, B. C. (**1992**). "Syllabic compression: Effective compression ratios for signals modulated at different rates," British J. Audiol. **26**(6), 351–361.

Stone, M. A., and Moore, B. C. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am. **116**(4), 2311–2323.

Stone, M. A., and Moore, B. C. (**2007**). "Quantifying the effects of fast-acting compression on the envelope of speech," J. Acoust. Soc. Am. **121**(3), 1654–1664.

Stone, M. A., and Moore, B. C. (**2008**). "Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **123**(2), 1063–1076.

Veaux, C., Yamagishi, J., and MacDonald, K. (**2019**). "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit (version 0.92)," University of Edinburgh Centre for Speech Technology Research, https://datashare.ed.ac.uk/handle/10283/3443.

Villchur, E. (**1973**). "Signal processing to improve speech intelligibility in perceptive deafness," J. Acoust. Soc. Am. **53**(6), 1646–1657.

Yund, E. W., and Buckles, K. M. (**1995**). "Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids," J. Acoust. Soc. Am. **97**(2), 1224–1240.

170   J. Acoust. Soc. Am. **150** (1), July 2021

Ryan M. Corey and Andrew C. Singer