REDS: A NEW ASYMMETRIC ATOM FOR SPARSE AUDIO DECOMPOSITION AND SOUND SYNTHESIS

Julian Neri

SPCL* , CIRMMT†
McGill University, Montréal, Canada
julian.neri@mail.mcgill.ca

Philippe Depalle

SPCL* , CIRMMT†
McGill University, Montréal, Canada
philippe.depalle@mcgill.ca

ABSTRACT

In this paper, we introduce a function designed specifically for sparse audio representations. A progression in the selection of dictionary elements (*atoms*) to sparsely represent audio has occurred: starting with symmetric atoms, then to damped sinusoid and hybrid atoms, and finally to the re-appropriation of the gammatone (GT) and formant-wave-function (FOF) into atoms. These asymmetric atoms have already shown promise in sparse decomposition applications, where they prove to be highly correlated with natural sounds and musical audio, but since neither was originally designed for this application their utility remains limited.

An in-depth comparison of each existing function was conducted based on application specific criteria. A directed design process was completed to create a new atom, the *ramped exponentially damped sinusoid* (REDS), that satisfies all desired properties: the REDS can adapt to a wide range of audio signal features and has good mathematical properties that enable efficient sparse decompositions and synthesis. Moreover, the REDS is proven to be approximately equal to the previous functions under some common conditions.

1. INTRODUCTION

A sparse synthesis model suggests that a signal $\mathbf{s} \in \mathbb{R}^n$ may be represented by a linear combination of a few elements (*atoms*) from dictionary $\mathbf{D} \in \mathbb{R}^{n \times m}$: $\mathbf{s} = \mathbf{D}\mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^m$ is the signal's sparse representation [1] [2]. Decomposing a signal with few elements implies, informally, great meaning is assigned to those elements. On the other hand, creating complex sounds with a few additions provides major efficiency improvements over the alternative (non-sparse) methods. Source-filter synthesis exemplifies a sparse synthesis model because a few waveforms are summed to create a complex sound (typically, a vocal sound) [3]. In either case, a sparse synthesis application requires a dictionary that includes easily controllable atoms capable of representing a wide range of signal content.

Knowledge of salient audio signal features can help guide dictionary design: they are asymmetric in time (short attack and long decay) and usually have time-varying frequency content [4]. Thus, a time-frequency structured signal model that is asymmetric in time (e.g., a damped sinusoid) is appropriate. However, the damped sinusoidal model [5] does not have a smooth attack while real signals almost always do. A compromise involves building a heterogeneous dictionary that includes symmetric atoms (e.g., Gabor atoms) and damped sinusoid atoms. Heterogeneous dictionaries must be indexed by more data, however, because each atom class within the dictionary will have a unique parameter set. More importantly, decomposing asymmetric signal content with a finite number of symmetric atoms will either lead to a non-sparse solution or pre-echo (*dark energy*) [6].

A better approach is to design a homogeneous dictionary (contains a single atom class), wherein the atoms are exponentially damped sinusoids with an attack envelope. Currently, only two functions common in literature have assumed this atomic role: the formant-wave-function [3] [7] (used in audio synthesis) and the gammatone (used in perceptual audio coding) [8] [9]. Matching Pursuit Toolkit (MPTK) supports the use of either function as a dictionary atom [10]. Neither function was designed to be optimized for the task of sparse audio decomposition, though, and they both suffer from limited parameter adaptability.

In this paper, we introduce a new asymmetric atom that is better suited for the sparse synthesis model. A theoretical and practical comparison of existing atom models is presented to consolidate knowledge and highlight their relative strengths and limitations. Our points of comparison reflect the qualities that we seek in a model: ability to match diverse signal behavior (especially transients), and good mathematical properties. Some of the desired mathematical properties include having a concentrated spectrum and an analytic inner product formula. Detailed criteria explanations and justifications reside in a following section.

The paper is structured as follows. Section 2 details the desired atom properties and form. Section 3 includes an analysis and comparison of existing functions. The new atomic model is introduced in Section 4. Section 5 overviews some new atomic model sparse synthesis applications. Fi-

* Sound Processing and Control Laboratory
† Centre for Interdisciplinary Research in Music Media and Technology

nally, in Section 6 we reflect on the future work intended for a sparse audio decomposition system using the new atomic model.

2. ATOM PROPERTIES

2.1. General Form

We generalize the form of a causal asymmetric atom as

$$x[n] = E[n]e^{i\omega_c n}, \tag{1}$$

where

$$E[n] = A[n]e^{-\alpha n}u[n] \tag{2}$$

is the atom's envelope, $\alpha \in \mathbb{R}_{\geq 0}$ is the damping factor, $\omega_c = 2\pi f_c$ is the normalized angular frequency of oscillation ($0 \leq f_c \leq \frac{1}{2}$), $u[n]$ is the unit step function, n is discrete time, and $A[n]$ is an attack envelope that distinguishes each atom ($A[n] \in \mathbb{R}_{\geq 0}\ \forall\ n \in \mathbb{N}$). We introduce the atoms as discrete time signals because, in practice, dictionaries are composed of finite sampled (discretized) atoms. We establish mathematical properties of the atoms (e.g., their derivatives) from their continuous time counterparts.

In the literature, the formant-wave-function and gammatone are typically defined with real sinusoids rather than complex ones. We choose to adopt a complex form for mathematical ease of manipulation and concise representation in transform domains. Moreover, it is necessary to parametrize phase for real valued but not for complex valued atoms: expansion coefficients from a signal decomposition using complex atoms will be complex and will thus provide both the magnitude and phase [5].

2.2. Desired Atom Properties

Our comparison criteria are grouped into three categories: time-frequency properties, control & flexibility, and algorithmic efficiency.

2.2.1. Time-Frequency Properties

A dictionary of atoms with varying degrees of *time and frequency concentration* is important for creating a sparse representation overall. For example, a sustained piano note begins with a short attack, which is best represented with concentrated time (spread frequency) resolution, followed by a long decay, which requires an atom with long time support and a concentrated spectrum. Multi-resolution analysis involves decomposing a signal onto a set of analyzing functions whose time-frequency tiling is non-uniform [11] [12]. We are going one step further by considering that some sounds require excellent time localization in the transient region and concentrated frequency resolution in the decay region. We aim at representing both regions with atoms whose envelopes are closer to those of natural sounds. We quantify concentration in time and frequency by the time spread, T, and frequency spread, B, respectively [13]. The Heisenberg-Gabor inequality states $BT \geq 1$.

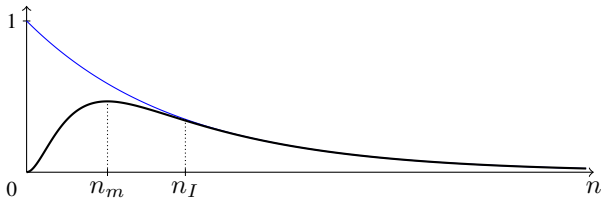

Figure 1: *An example envelope of the form (2) overlaid with an exponential envelope (blue), where n_I is the influence time and n_m is the time location of the envelope maximum.*

Moreover, we prefer an atom that has a *unimodal spectrum*: a spectrum $X(\omega)$ is unimodal if $|X(\omega)|$ is monotonically increasing for $\omega \leq \omega_c$ and monotonically decreasing for $\omega \geq \omega_c$. A function that is truncated in time with a rectangular window admits a non-unimodal spectrum because the truncation is equivalent to convolving the spectrum with a sinc function whose oscillations introduce multiple local maxima/minima [14]. Multiple local maxima/minima in the spectrum can complicate spectral parameter estimation. We prefer an infinitely differentiable atom (i.e., of class C^∞, as defined in [15]) because its spectrum is unimodal.

2.2.2. Control & Flexibility

We modulate the damped sinusoid with A to enhance the atom's adaptability to natural sounds. A damped sinusoid's damping factor α indirectly controls its T and B. Smoothing the damped exponential's initial discontinuity with A concentrates its frequency localization in exchange for a more spread time localization. We want a parametrization of A that enables precise control over its time and frequency characteristics, controllability being an essential aspect of audio synthesis. Furthermore, the attack portion of an audio signal often contains dense spectral content that allows humans to characterize its source [16].

Influence time has a major effect on the atom's overall perceived sound as it controls the degree to which the initial discontinuity is smoothed [16]. We define influence time n_I as the duration that A influences the atom: n_I is the largest value of n for which $e^{-\alpha n}(A[n] - 1) > \delta$ is true (in this paper $\delta = .001$, see Figure 1). The effects of varying influence time are intuitively linked to T and B. In the frequency domain, influence time mostly controls the spectral envelope far from its center frequency (*skirt width* as defined in [3]). Increasing influence time spreads the atom's time localization and concentrates its spectrum.

An important quantity to compare between the atoms is

the time $\Delta_I = n_I - n_m$, where n_m is the time location of E's maximum. n_m is often called a temporal envelope's attack time in sound synthesis [17]. We find n_m by setting E's continuous time derivative equal to zero and solving for n. For a continuous E whose $\alpha > 0$, n_m precedes n_I (i.e., A influences E even after n_m). To compare atoms along this criteria, we equalize their n_m values then compare their Δ_I values. Δ_I indicates the amount of influence that varying the skirt width will have on the bandwidth. We prefer an atom with a small Δ_I value because its 3 dB bandwidth (set through α) is not affected much by the structure of A. An envelope with a small Δ_I also reflects those produced by many acoustic instruments: an exciter increases the system's energy and then releases (at n_I), which results in a freely decaying resonance.

We do not want to complicate the definition of the atom when modulating the damped sinusoid by A either; we encourage *time-domain simplicity*. The damped sinusoid's simple definition enables us to solve for its parameters algebraically. Classic parametric estimation techniques can be used to adapt the damped sinusoid to an arbitrary signal [18]. We want to retain these desirable properties even after introducing A. An atom's time-domain simplicity will depend on how its A marries with the complex damped sinusoid. Finally, after modulating the damped sinusoid with A, we want the atom's envelope to match well with those in actual musical signals.

2.2.3. Algorithmic Efficiency

Fast algorithms are one of the focuses of sparse representations research, as they aim to make sparse decomposition processes more tractable. Amid publications dedicated to creating faster algorithms, some reported techniques have become widely adopted [10]. Specifically, certain analytic formulas are known to increase the algorithm speed because they avoid some of the algorithm's most time consuming numerical calculations (e.g., the inner product).

An envelope shape that enables the inner product of two atoms to be expressed as an analytic formula is required for a fast matching pursuit algorithm [1]. A summary of the relevant algorithm steps are included for justification.

In matching pursuit, a dictionary $\mathbf{D}$ is compared with a signal $\mathbf{s}$ by calculating and storing the inner product $\langle \mathbf{s}, \mathbf{g}_\gamma \rangle$ of each atom $\mathbf{g}_\gamma$ and the signal. The atom that forms the largest inner product $\mathbf{g}_q$ is picked as the signal's best fit. Then $\langle \mathbf{g}_q, \mathbf{g}_\gamma \rangle$ is computed and subtracted from $\langle \mathbf{s}, \mathbf{g}_\gamma \rangle$. This continues until some stopping criteria is met.

Dictionary inner products can be calculated and stored once when the dictionary is static. However, when atom parameters are refined within the iterative loop these inner products cannot be precomputed and, therefore, must be computed at each iteration. Numerical calculations of many inner products at every iteration prohibit speed. Analytic formulas make the process tractable.

Another way to increase the efficiency of a sparse decomposition program is to use parametric atoms, then refine atom parameters using an estimator. Finding a more adapted atom at every iteration may require less iterations overall. Developing parametric estimation techniques sometimes relies on having analytic discrete Fourier transform (DFT) formula. For example, in derivative methods, two spectra are divided to solve for one or more variables [18]. We include each atom's analytic DFT formula in [19] and [20].

Finally, [5] explains how recursion may be exploited to calculate the convolution of damped sinusoidal atoms with a signal: since the impulse response of a complex one-pole filter is a damped complex exponential sinusoid, a recursive filter can efficiently calculate the correlation. We provide each atom's Z-transform to indicate its *causal filter simplicity* and therefore practicality for calculating the correlation. Besides, the Z-transform is useful for source-filter synthesis and auditory filtering.

3. EXISTING FUNCTIONS

3.1. Damped Sinusoid

3.1.1. Background

The damped sinusoid (DS) is essential in audio as it represents a vibrating mode of a resonant structure. The use of a DS model in the context of analysis dates back to Prony's method [21], according to our knowledge, and was the first asymmetric atom used in the context of sparse representations [5].

3.1.2. Properties

Staying with the predefined generic atom expression (1):

$$A_{DS}[n] = 1 \tag{3}$$

and thus $n_m = n_I = 0$. Its continuous-time Fourier transform is well known,

$$X_{DS}(\omega) = \frac{1}{\alpha + i(\omega - \omega_c)} \tag{4}$$

as is the DFT [20], and finally, the Z-transform,

$$X_{DS}[z] = \frac{1}{1 - e^{-\alpha + i\omega_c} z^{-1}} \tag{5}$$

The DS' spectrum is unimodal but not concentrated.

3.2. Gammatone

3.2.1. Background

Auditory filter models are designed to emulate cochlea processing and are central to applications like perceptual audio

coding, where auditory filters are used to determine which sounds should be coded or not according to auditory masking principles. Auditory filter modeling has a variety of applications in bio-mechanics and psychoacoustic research.

The most popular auditory filter model is the gammatone (GT) filter due to its heritage and simple time domain expression. Originally described in 1960 as a fitting function for basilar displacement in the human ear [8], the gammatone filter was later found to precisely describe human auditory filters, as proven from psychoacoustic data [22]. [9] shows that atoms learned optimally from speech and natural sounds resemble gammatones. Designing gammatone filters remains a focus in audio signal processing [23].

More recently, filter models closely related to the gammatone filter have been proposed, such as the all-pass gammatone filter and the cascade family [24]. Added features of these variants do not overlap with our criteria so they are not included for comparison.

3.2.2. Properties

We assign the gammatone as the prototypical auditory filter model. A single variable polynomial envelope function shapes the gammatone:

$$A_{GT}[n] = n^p \tag{6}$$

In literature, $p+1$ is called the filter order. A_{GT} does not converge (its derivative is strictly positive), and thus admits the largest Δ_I in this study, as $n_m = \frac{p}{\alpha}$ and $n_I > 2n_m$. No part of the gammatone is, strictly speaking, a freely decaying sinusoid (excluding when $p = 0$, in which case it is a DS), though it asymptotically approaches a DS as $n \to \infty$.

We demonstrate the filter order's effect by applying the Fourier transform frequency differentiation property to express its spectrum parametrized by p:

$$X_{GT}(\omega) = \frac{p!}{(\alpha + i(\omega - \omega_c))^{p+1}} \tag{7}$$

From its frequency representation, we see that the filter order determines the denominator polynomial order. Finally, referencing the convolution property of the Fourier transform, the gammatone impulse response is a DS convolved with itself p times.

Frequency spread B decreases with respect to the model order, while the time spread T increases. A gammatone of order four ($p = 3$) correlates best with auditory models [23]. The gammatone's spectrum is unimodal and concentrated.

The attack envelope is not parametrized, and therefore cannot be controlled independently of α. After setting p, controlling the atom is solely through α and ω_c. Influence time (or skirt width) is not directly controllable, so one cannot tune the atom to have time concentration in exchange for frequency spread. Thus, the adaptability of this model to a range of sound signal behavior is limited.

We establish an analytic formula for the gammatone's Z-transform that supports an arbitrary integer $p > 0$:

$$X_{GT}[z] = \frac{\sum_{r=1}^{p} \left\langle {p \atop r-1} \right\rangle a^r}{(1-a)^{p+1}} \tag{8}$$

where $a = e^{-\alpha + i\omega_c} z^{-1}$, and the Eulerian number $\left\langle {p \atop r-1} \right\rangle = \sum_{j=0}^{r} (-1)^j \binom{p+1}{j} (r-j)^p$. The gammatone's analytic inner product formula is complicated and described in [20].

3.3. Formant-Wave-Function

3.3.1. Background

In the source-filter model, an output sound signal is considered to be produced by an excitation function sent into a (resonant) filter, referred to as a source-filter pair [3]. Most acoustic instruments involve an exciter, either forced or free, and a resonator [4]. When an instrument's exciter and resonator are independent, or have only a small effect on one another, its sound production mechanism may be described sufficiently with a source-filter model. An example is the voice production system, where excitations produced by glottal pulses are filtered within the vocal tract.

Source-filter synthesis involves sending an excitation function through one or more resonant filters in parallel. The filters are typically one or two pole and defined by their auto-regressive filter coefficients. The excitation function can be an impulse, but is more often an impulse smoothed by a window to emulate natural excitation. The window shape effects the transient portion of the time-domain output from the system, and the skirts of the spectral envelope. The filter coefficients control the shape of the spectral envelope near the resonant peak.

Time-domain formant-wave-function synthesis describes the output of the source-filter model by a single function in the time domain. The amplitude envelope of the function is designed to generically match the output envelope of a source-filter pair: a damped exponential (filter) for which the initial discontinuity is smoothed (excitation). The advantage of this approach is twofold: direct parametrization of the spectral envelope, and efficient synthesis by table lookup [3].

Creating a sustained sound (e.g., a voice) from this model involves filtering a sparse excitation signal made of a (possibly) periodic sequence of short duration signals. Likewise, synthesizing a percussive sound (e.g., a piano) involves summing the output of several resonant filters with comparably long decay times from a single excitation. In the time-domain method, this means that the model synthesizes a signal $\mathbf{s}$ as a linear combination of time-shifted resonant filter impulse responses (i.e., time-frequency atoms). Formally,

we express this as $s[n] = \sum h_\lambda[n-\tau]v_{\lambda,\tau} = \mathbf{Dv}$, where $\mathbf{D}$ is a dictionary of atoms $h_{\lambda,\tau}[n] = h_\lambda[n-\tau]$ that are indexed by λ and time shift τ, and $\mathbf{v}$ contains their amplitude coefficients $v_{\lambda,\tau}$. Thus, the source-filter model is a sparse synthesis representation.

3.3.2. Properties

The formant-wave-function (FOF) is ubiquitous with time-domain wave-function synthesis. It was introduced for its desirable properties: a concentrated spectral envelope that can be controlled rather precisely using two parameters. The FOF's A is defined as:

$$A_{FOF}[n] = \begin{cases} \frac{1}{2}\left(1-\cos(n\beta)\right) & \text{for } 0 \le n \le \frac{\pi}{\beta}, \\ 1 & \text{for } \frac{\pi}{\beta} < n. \end{cases} \tag{9}$$

where $\beta \in \mathbb{R}_{>0}$ controls influence time. Decreasing β increases influence time, $n_I \approx \frac{\pi}{\beta}$, and the time location of the maximum,

$$n_m = \frac{1}{\beta}\cos^{-1}\left(\frac{\alpha^2-\beta^2}{\alpha^2+\beta^2}\right) \tag{10}$$

Δ_I and $\frac{\alpha}{\beta}$ are positively correlated.

A raised cosine is an excellent attack shape in terms of concentration, however, since it is piecewise (its value must be held at one after half of a period) some other design criteria suffer.

$$X_{FOF}(\omega) = \frac{\beta^2}{2}\frac{1+e^{-\frac{\pi}{\beta}(\alpha+i(\omega-\omega_c))}}{(\alpha+i(\omega-\omega_c))((\alpha+i(\omega-\omega_c))^2+\beta^2)} \tag{11}$$

The FOF's spectrum is non-unimodal when the piecewise transition occurs within the window of observation. Moreover, it is difficult to estimate the FOF's parameters and its analytic inner product formula is complicated [7].

We establish the FOF's DFT and Z-transform by converting the cosine function into a sum of complex exponentials and using the linear property:

$$X_{FOF}[z] = \frac{1}{2}\frac{1+a^{N_1}}{1-a} - \frac{1}{4}\left(\frac{1-(ae^{i\beta})^{N_1}}{1-ae^{i\beta}} + \frac{1-(ae^{-i\beta})^{N_1}}{1-ae^{-i\beta}}\right) \tag{12}$$

where $a = e^{-\alpha+i\omega_c}z^{-1}$, and $N_1 = [\frac{\pi}{\beta}]$. From (12), we see that a FOF filter may be implemented as a sum of three complex pole-zero filters. The time-varying input delay complicates controlling attack shape.

3.4. Connecting FOF to Gammatone

Applying the small angle theorem to A_{FOF} reveals a relation between the FOF and GT:

$$\lim_{\beta\to 0}\frac{2}{\beta^2}\left(1-\cos(\beta n)\right) = n^2 \tag{13}$$

We establish that a FOF and gammatone of $p = 2$ are approximately equal when $\beta = \frac{1}{4}\alpha\sqrt{12\epsilon}$, where ϵ is the approximation error (see [20] for proof).

3.5. Recapitulation

Each existing function has several desired properties missing. While the gammatone's unimodal frequency spectrum and time-domain simplicity are appealing, expressing its DFT and inner product is complicated. Most importantly, without a parameter to control influence time, the gammatone is not flexible enough to sparsely represent a variety of signal features. On the other hand, the FOF's attack function enables precise control over its spectral envelope, however, its piecewise construction is problematic: spectral ripples result from a truncation in time, refining its parameters is difficult, and its frequency, Z-transform, and inner product expressions are complicated.

3.6. Towards a New Atom

The starting goal of this paper was to design a $C^\infty A$ that is similar to A_{FOF}. While piecewise construction is the reason for the FOF's shortcomings, approximating the raised cosine with a C^∞ function does not necessarily improve the situation because many functions admit complicated frequency-domain and Z-domain formulas once a unit step is introduced. For example, $e^{-\beta n^2}$ has a compact bell shape that seems to be, at first inspection, a good candidate to replace the raised cosine. However, when a unit step is introduced, it admits a non-algebraic Fourier transform expression (a special function defines the imaginary part). Many bell-shaped functions have the same problem (e.g., $\tanh(\beta n)^2$).

On the other hand, there are A options that are simple but have Δ_I that are large compared to the FOF for equal n_m. In fact, any C^∞ function will have a larger Δ_I than the FOF's for equal n_m. Therefore, our goal became more specific: define a C^∞ A that admits simple mathematical expressions when married with a complex damped exponential, and whose Δ_I is close to that of the FOF's for equal n_m. After an exhaustive search, we resolved that designing a function to satisfy all of the design criteria is difficult.

4. THE NEW ASYMMETRIC ATOM

We have designed a new atom specifically for sparse audio representations. All the aforementioned criteria were in mind when constructing this new atom.

4.1. Background

To reflect generality, we call the new atom the *ramped exponentially damped sinusoid* (REDS). Identically to existing source-filter and auditory filter models, a complex exponentially damped sinusoid defines the atom's decay section. A binomial with one exponential term shapes the atom's attack envelope. By defining the atom as a sum of exponentials (see (16)), all the desired mathematical properties

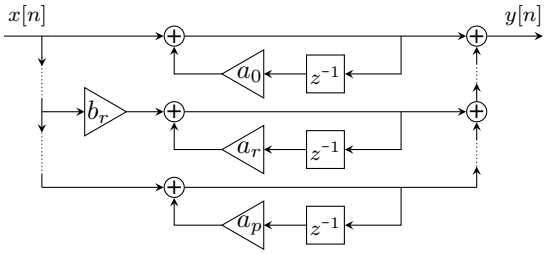

Figure 2: *REDS filter diagram, where* $a_r = e^{-\alpha - r\beta + i\omega_c}$ *and* $b_r = (-1)^r \binom{p}{r}$.

are achieved. The main idea is that the linear property of the Fourier transform and Z-transform can be exploited and each exponential has a transform that is simple and well known.

4.2. Properties

We define REDS concisely in the time-domain by expressing $A_{REDS}[n]$ polynomially as $(1 - e^{-\beta n})^p$:

$$x[n] = \left(1 - e^{-\beta n}\right)^p e^{n(-\alpha + i\omega_c)} u[n] \tag{14}$$

where β controls the influence time (or skirt width) and $p+1$ is the order.

$$n_m = \tfrac{1}{\beta} \log(1 + \tfrac{p\beta}{\alpha}) \tag{15}$$

and $n_I \approx -\frac{1}{\beta} \log(1 - (1 - \delta)^{1/p})$, where δ is the same as in Section 2.2.2.

Like in the gammatone model, order is often constant within an application: we may choose the order, for example, to match with auditory data or to approximate a frame condition [23]. Given that the order is a constant, the number of control parameters and their effect are the same as the FOF. To summarize, the REDS parameter set is a conflation of the source-filter and auditory filter models.

We express the REDS in binomial form to reveal its sum of exponentials construction:

$$x[n] = \sum_{r=0}^{p} (-1)^r \binom{p}{r} e^{n(-\alpha - r\beta + i\omega_c)} u[n] \tag{16}$$

Considering the Fourier transform linear property, we readily find from (16) the Fourier transform of REDS:

$$X_{REDS}(\omega) = \sum_{r=0}^{p} (-1)^r \binom{p}{r} \frac{1}{\alpha + r\beta + i(\omega - \omega_c)} \tag{17}$$

and the analytic DFT [20]. Finally, we apply the linear property to retrieve the Z-transform:

$$X_{REDS}[z] = \sum_{r=0}^{p} (-1)^r \binom{p}{r} \frac{1}{1 - e^{-\alpha - r\beta + i\omega_c} z^{-1}} \tag{18}$$

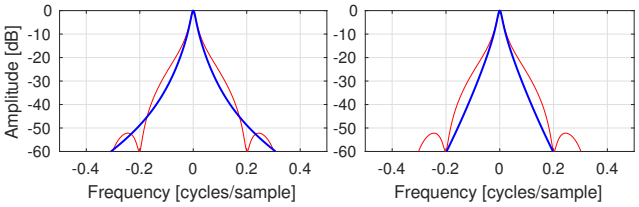

Figure 3: $X_{FOF}[\omega_k]$ *(red) and* $X_{REDS}[\omega_k]$ *(blue) for fixed* β*, where* $p = 2$ *on the left,* $p = 10$ *on the right, and* $\alpha = .05$.

A sum of $p+1$ complex one-pole filters in parallel will thus output a REDS (see Figure 2).

The REDS has a concentrated and unimodal spectrum. Similarly to the FOF, it is possible to precisely control the REDS' spectrum: by varying β one may exchange concentration in time for frequency, and vice versa. The FOF has greater time concentration than the REDS because the raised cosine attack function has a fast uniform transition from zero to one, while the REDS attack envelope is bell-shaped. Formally, $n_{I_{REDS}} > n_{I_{FOF}}$ when $n_{m_{REDS}} = n_{m_{FOF}}$. The REDS' spectral concentration surpasses the FOF's as p increases (see Figure 3).

We established analytic inner product and convolution formulas for two REDS atoms that support the case when atoms have different lengths N (see [20]). Considering that these formulas for Gabor atoms and FOFs provide an efficiency boost in existing programs [7], and the REDS formulas are simpler than those, we assume that using the formulas are more efficient than numerical computations.

4.3. Connection to Existing Functions

A REDS is approximately equal to a GT when β is very small:

$$\lim_{\beta \to 0} \tfrac{1}{\beta^p} \left(1 - e^{-\beta n}\right)^p = n^p \tag{19}$$

We establish that a REDS and a GT are approximately equal when $\beta = \epsilon \alpha p^{-2}$, where ϵ is the approximation error (see [20] for proof). This is important because the REDS filter requires fewer mathematical operations per sample than the gammatone filter. In practice, the perceptual difference between the two is negligible when $\epsilon < .001$, which corresponds to a signal-to-noise ratio between the two atoms greater than 60 dB.

Furthermore, by (13), $A_{REDS}[n] \approx 2A_{FOF}[n]$ when $p = 2$ and their β values are $\frac{1}{4}\epsilon\alpha$.

5. APPLICATIONS

This section includes two sparse synthesis examples: REDS source-filter synthesis, and musical audio decomposition.

5.1. Synthesis

Sparse synthesis via the source-filter model generally involves sending a short excitation periodically through one or more filters, typically one per formant. In this example we checked the ability of the REDS to synthesize a vowel sound /i/ with 5 REDS filters tuned by parameters provided in [3], see Table 1. We set $p = 2$ (see (19)) for better comparison with FOF, which have demonstrated their aptitude for singing voice synthesis. After normalization (see normalization factor in [19] or [20]), a gain G ($0 \leq G \leq 1$) tunes the filter output amplitude. Results show the quality of the synthesized sound as well as the ability to match the spectral envelopes, even in the valleys, mainly controlled by β (see [19]).

Table 1: *REDS filter settings for synthesizing a vowel, where ν_c is center frequency in Hz and $p = 2$.*

ν_c	260	1764	2510	3100	3600
α	.005	.006	.006	.009	.011
β	.018	.059	.034	.011	.008
G	1.0	.501	.447	.316	.056

5.2. Decomposition

We decomposed a set of real audio signals using a standard matching pursuit algorithm[1]. We selected the audio signal set to reflect a range of the source-filter model: it includes a vocal sound (sustained, relatively high damping and smooth attack per atom), a vibraphone (not-sustained, made of low damping and short attack per atom), and a violin (intermediate situation).

We created damped sinusoid dictionaries to fit with each signal's content. Then we made dictionaries for each atom class by modulating the damped sinusoid dictionaries with a set of each atom's $A[n]$. We superimposed each selected atom's Wigner-Ville distribution to show their time-frequency footprint, such as in [6] & [7].

We can represent a signal as time-varying sinusoidal trajectories per the additive model, or as filtered excitation sequences per the source-filter model, by decomposing it onto a dictionary of REDS atoms with constrained damping factors. We chose to demonstrate the ability of the REDS to analyze the signal set from the source-filter viewpoint. For the singing voice, if the dictionary contained atoms with small damping (long time support) then the selected atoms would represent the sinusoidal partials of the signal. We set the damping to be high and in doing so, successfully extracted the excitation sequence of atoms whose spectral content represented vocal formants rather than the sinusoidal partials (see Figure 5). Regarding the vibraphone, we created a dictionary whose damped sinusoids had large time support with low decay rates.

[1]Standard in the sense that the dictionary was static and it did not involve fast algorithms or parameter refinements. We implemented the algorithm based on [1].

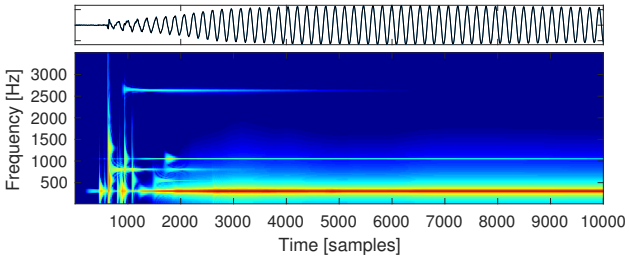

Figure 4: *Sparse representation of a vibraphone (transient part shown) from a decomposition onto* 50 *REDS atoms.*

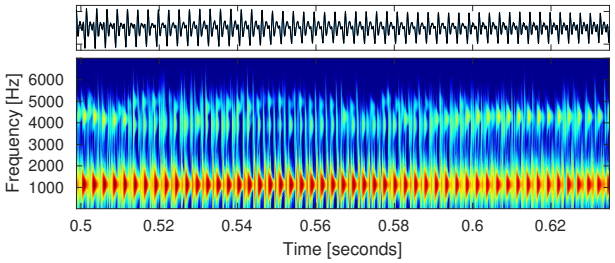

Figure 5: *Source-filter sparse representation of a singing voice. Atom spacing expands/contracts reflecting vibrato.*

Table 2: *Decomposition results. Each audio signal is sampled at* 44100 *Hz. d_s is the signal's duration in seconds.*

				SNR (dB)			
	d_s	Atoms	N	DS	GT	FOF	REDS
Vocal	1.0	10^4	2^8	30.7	35.9	37.8	38.8
Violin	1.6	10^4	2^9	20.0	14.6	27.7	28.0
Vibes	5.5	50	2^{17}	17.6	32.1	36.9	37.1

For each test, the REDS dictionaries provided higher SNR values for the same number of iterations, see Table 2. For the singing voice, the gammatone and REDS were close in performance because the formant time-domain envelopes had very smooth attacks. REDS matched the vibraphone's envelope tightly, while the gammatone caused pre-echo because of its greater amount of symmetry (see [19]). The reconstructed signal from the REDS decomposition had an SNR of 38.8 dB, and consisted of 50 atoms (.04% of the signal length) (see Figure 4). Our companion website includes audio files for each signal approximation and further details the decomposition study results [19].

6. CONCLUSION

We have introduced a new function called REDS that can sparsely represent audio. Through the comparison of functions previously used in sparse representation contexts, we highlighted the most important features for this new function to embody (see Table 3). We have started researching an efficient sparse audio decomposition system that exploits the good properties of the new asymmetric atom.

Since the mathematical properties of the REDS enable efficient implementations of filter banks, source-filter synthesis, and audio coding, the REDS has potential to be used in many audio signal processing fields.

Table 3: *Comparison results.*

Criteria	DS	GT	FOF	REDS
Concentrated Spectrum	–	✓	✓	✓
Unimodal Spectrum	✓	✓	–	✓
Influence Time Control	–	–	✓	✓
Time-Domain Simplicity	✓	✓	–	✓
Causal Filter Simplicity	✓	✓	–	✓
Inner Product Simplicity	✓	–	–	✓

7. REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[2] P. Balazs, M. Doerfler, M. Kowalski, and B. Torresani, "Adapted and adaptive linear time-frequency representations: A synthesis point of view," *IEEE Signal Process. Magazine*, vol. 30, no. 6, pp. 20–31, Nov. 2013.

[3] X. Rodet, *Time-domain formant-wave-function synthesis*, chapter 4-Speech Synthesis, pp. 429–441, J.C. Simon, Ed. New York, 1980.

[4] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*, Springer New York, 2nd edition, 1998.

[5] M. Goodwin, "Matching pursuit with damped sinusoids," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Munich, Germany, Apr. 1997, vol. 3, pp. 2037–2040.

[6] B. L. Sturm, C. Roads, A. McLeran, and J. J. Shynk, "Analysis, visualization, and transformation of audio signals using dictionary-based methods," *J. New Music Research*, vol. 38, no. 4, pp. 325–341, 2009.

[7] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 101–111, Jan. 2003.

[8] J. L. Flanagan, "Models for approximating basilar membrane displacement," *Bell System Technical Journal*, vol. 39, no. 5, pp. 1163–1191, 1960.

[9] E. Smith and M. Lewicki, "Efficient auditory coding," *Nature*, vol. 439, no. 7079, pp. 978–982, Feb. 2006.

[10] S. Krstulovic and R. Gribonval, "MPTK: Matching pursuit made tractable," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2006, vol. 3, pp. 496–499.

[11] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 3rd edition, 2009.

[12] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco, "Theory, implementation and applications of nonstationary gabor frames," *J. Comput. Appl. Math*, vol. 236, no. 6, pp. 1481 – 1496, 2011.

[13] P. Flandrin, *Time-Frequency / Time-Scale Analysis*, vol. 10 of *Wavelet analysis and its applications*, chapter 1, pp. 9 – 47, Academic Press, 1999.

[14] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[15] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer New York, 1983.

[16] A.S. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, MA, 1990.

[17] M. Mathews and J. Miller, *The technology of computer music*, M.I.T. Press, Cambridge, Mass., 1969.

[18] S. Marchand and P. Depalle, "Generalization of the derivative analysis method to non-stationary sinusoidal modeling," in *Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFx-08)*, Espoo, Finland, Sep. 2008.

[19] J. Neri, "REDS website," `http://www.music.mcgill.ca/~julian/dafx17`, 2017.

[20] J. Neri, "Sparse representations of audio signals with asymmetric atoms," M.A. thesis, McGill University, Montréal, Canada, 2017.

[21] G. M. Riche de Prony, "Essai expérimental et analytique sur les lois de la dilatabilité de fluides élastiques," *Journal de l'École Polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[22] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," *Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE*, Dec. 1987.

[23] S. Strahl and A. Mertins, "Analysis and design of gammatone signal models," *J. Acoust. Soc. Am.*, vol. 126, no. 5, pp. 2379–2389, Nov. 2009.

[24] R. Lyon, A. Katsiamis, and E. Drakakis, "History and future of auditory filter models," in *Proc. IEEE Int. Conf. Circuits and Systems (ISCAS)*, Aug. 2010, pp. 3809–3812.