Grid-based decimation for invertible wavelet transforms in audio processing

Nicki Holighaus, Günther Koliander, Clara Hollomey, and Friedrich Pillichshammer

***Abstract*—The constant center frequency to bandwidth ratio (Q-factor) of wavelet transforms provides a very natural representation for audio data. So far, invertible wavelet transforms have either required non-uniform decimation—leading to irregular data structures that are cumbersome to work with—or require excessively high oversampling with unacceptable computational overhead. Here, we present a novel decimation strategy for wavelet transforms that leads to invertible representations, i.e., representations that allow perfect reconstruction, with minimal oversampling and uniform decimation. Numerical evidence shows that the resulting representation is highly stable in the sense of frame theory. The obtained wavelet coefficients can be stored in a time-frequency matrix with a natural interpretation of columns as time frames and rows as frequency channels. In a special case, the employed decimation strategy corresponds to a time-frequency lattice related to certain one-dimensional low discrepancy sequences. The matrix structure of the wavelet coefficients immediately grants access to a large number of algorithms that have been successfully used in time-frequency audio processing, but could not previously be used jointly with invertible wavelet transforms. We demonstrate the application of our method in processing based on non-negative matrix factorization and phaseless reconstruction.**

***Index Terms*—wavelet transform, quasi-random sequences, sampling, perfect reconstruction, shift-invariant systems, uniform decimation**

I. Introduction

The wavelet transform is one of the most important and well-studied time-frequency filter banks, only rivaled by the short-time Fourier transform. Due to its constant center frequency to bandwidth ratio, or *Q-factor* [1], wavelets provide a natural and physically meaningful representation of audio. Consequently, constant-Q filter banks in general [2]–[4] and wavelet transforms in particular [5]–[8] have been used to great success in the analysis of speech and audio signals. For the longest time, however, constant-Q filter banks have been inaccessible, or at least inconvenient, for processing audio at sufficient fidelity: Wavelet bases [9] and undecimated, discrete wavelet systems [10], [11] possess a very low Q-factor that is not suitable for processing audio. Modern implementations of constant-Q filter banks allow for the tuning of the Q-factor [12], [13] and the oversampling rate, but computationally efficient, numerically stable, and invertible constant-Q filter banks have only recently been constructed [13], [14], by means of mathematical frame theory [15]. Although the availability of invertible representations with appropriate frequency resolution, i.e., a sufficiently large Q-factor, presents an important step towards making constant-Q filter banks more attractive for audio processing, the adaptation of successful processing schemes from the ubiquitous short-time Fourier transform (STFT) to the constant-Q setting remains challenging. One main obstruction is given by the geometry of established wavelet sampling sets: Since the bandwidth of constant-Q filters is proportional to their center frequency, it is natural to consider frequency bands spaced evenly on a logarithmic frequency axis, i.e., a fixed number B of channels per octave, or equivalently $B/12$ channels per semitone on a musical scale. However, to construct an invertible filter bank with such a distribution of channels, it is necessary to either abstain from in-channel decimation entirely, resulting in prohibitively high oversampling, or to choose the decimation factors to be inversely proportional to the channel center frequency (as commonly done for wavelet bases), leading to a nonuniform array of filter bank coefficients that are not aligned in time. In light of the number of successful time-frequency processing schemes that rely on time-alignment of the short-time Fourier coefficients and their segmentation into *time frames*, the nonuniform structure of the constant-Q coefficients is more than just a minor inconvenience. Furthermore, filter banks with non-uniform decimation only admit efficient, direct algorithms for reconstruction from the filter bank coefficients under restrictive conditions. If these conditions are not satisfied, even efficient reconstruction algorithms involve costly iteration [16], [17].

Contribution: Previously, grid-based decimation strategies were considered ill-suited for the continuous wavelet transforms and not expected to provide invertible signal representations. For discrete signals, known constructions did not admit perfect reconstruction at moderate, or even low, oversampling rates. In this paper, we propose the first grid-based decimation strategies for wavelet transforms with a tunable Q-factor that allow perfect reconstruction at oversampling rates close to 1 and that provide excellent numerical stability at moderate oversampling rates, ranging from 2 to 8, as commonly used for audio processing with the STFT. To achieve this, we

Manuscript received XXX; revised August XXX.

N. Holighaus (corresponding author) and C. Hollomey are with the Acoustics Research Institute (ARI), Austrian Academy of Sciences, Wohllebengasse 12–14, 1040 Vienna, Austria. G. Koliander is with ARI and the Faculty of Mathematics, University of Vienna, Austria. F. Pillichshammer is with the Institute of Financial Mathematics and Applied Number Theory, Johannes Kepler University Linz, Austria. e-mail: {nicki.holighaus,guenther.koliander,clara.hollomey}@oeaw.ac.at, friedrich.pillichshammer@jku.ac.at

Extended results, audio files and code for reproducing the presented experiments is available at: http://ltfat.org/notes/057

This work is supported by the Austrian Science Fund (FWF): I 3067–N30 (N.H.), Y 1199 “Time-Frequency Analysis, Randomness and Sampling.” (G.K.), and F5509-N26, which is a part of the Special Research Program “Quasi-Monte Carlo Methods: Theory and Applications.” (F.P.).

combine shift-invariant systems [18], [19] with ideas from quasi-random sampling using low discrepancy sequences [20], [21]. Additionally, similar to wavelet bases and prior invertible constant-Q implementations, we allow an arbitrarily small lowpass region to be covered by a complementary set of lowpass filters. Our construction is validated numerically by computing frame bounds and accumulated spectrograms of the decimated wavelet systems across a range of system parameters covering variations of the mother wavelet, the number of frequency channels, the decimation factor, and the oversampling rate. As a proof of concept, we apply the proposed wavelet decimation to several audio applications. We replicate an experiment on signal enhancement based on Itakura-Saito non-negative matrix factorization (NMF) as proposed by Févotte et al. [22], which relies heavily on the natural interpretation of the representation coefficients as a *time-frequency matrix*. We further evaluate the suitability of the proposed method for phaseless reconstruction with the fast Griffin-Lim algorithm [23], in relation to STFT and standard, irregular wavelet decimation.

Random and quasi-random sampling of time-frequency representations: This work is not the first to consider (quasi-) randomized sampling of time-frequency representations or more general integral transforms. In a series of recent works, Levie et al. consider Monte Carlo [24], [25] and Quasi-Monte Carlo [26] style random sampling of time-frequency integral operators. Their work is concerned with the approximation of continuous domain time-frequency processing by means of Monte Carlo integration with (quasi-) random time-frequency samples. By invoking prior results on Monte Carlo integration, the authors demonstrate that the approximation error can be controlled, provided that the considered time-frequency representation is *linear volume discretizable*. They further show that this property is satisfied by the STFT, the wavelet transform and a custom blending of the two, referred to as localizing time-frequency transform. The invertibility or stability of the sampled representation is not investigated, however. In the context of random sampling, these properties are considered in the literature on *relevant sampling*, introduced by Bass and Gröchenig for bandlimited functions [27] and later generalized to various settings [28]–[30], including time-frequency representations [31]. Relevant sampling provides a probabilistic framework for stable sampling of functions that are localized in a domain of finite volume, e.g., bandlimited signals that have only negligible energy outside a finite interval. Our work differs from these prior works in multiple ways: The considered sampling sets are not fully (quasi-) random, but correspond to a uniform time-frequency grid, up to the introduction of a small, quasi-random delay in every wavelet channel. Further, we consider perfect reconstruction of arbitrary signals, without localization assumptions.

Paper structure: A short introduction to wavelet systems and quasi-random sequences is given in Section II, before presenting the proposed decimation scheme in Section III. Our construction is evaluated numerically (Section IV) and applied in two illustrating experiments in audio processing (Section V). Specifically, we consider the decomposition of a signal with non-negative matrix factorization and phaseless reconstruction from time-frequency coefficients based on the fast Griffin-Lim algorithm. A short discussion of the provided code implementation is found in Section VI. The paper concludes with a summary of the results and an outlook towards related future work (Section VII).

II. Technical Background

Before introducing our novel decimation strategy, we review some basics of wavelet systems and quasi-random sequences.

A. The wavelet transform

A wavelet system is a collection of functions (or vectors) generated from a single prototype, the *mother wavelet*, by translation and dilation. Since we are interested in real-valued signals, audio signals in particular, we consider a mother wavelet ψ such that its Fourier transform vanishes for negative frequencies, i.e., $\hat{\psi}(\xi) = 0$ for $\xi \in (-\infty, 0]$. Such mother wavelets are often called *analytic*, although the terminology *analytic wavelet transform* has been used in at least two different manners in the past [32], [33]. The continuous wavelet system is generated via dilation by $s > 0$ and translation by $x \in \mathbb{R}$ of the mother wavelet

$$\psi_{(x,s)} := s^{-1/2}\psi\left(\frac{\bullet - x}{s}\right).$$

A signal f can now be filtered using this system resulting in the continuous wavelet transform

$$\begin{aligned} W_\psi f(x,s) &= \frac{1}{\sqrt{s}} \int f(t)\overline{\psi\left(\frac{t-x}{s}\right)}\,\mathrm{d}t \\ &= \langle f, \psi_{(x,s)} \rangle_{L_2} \\ &= \left[f * \overline{\psi\left(-\bullet/s\right)}\right](x)\,. \end{aligned} \tag{1}$$

The mathematical study of wavelet transforms usually considers $W_\psi f$ in terms of the inner product representation of the transform coefficients, but the final equality justifies the interpretation as a filter bank. Importantly, the wavelet transform is shift-invariant, i.e., the wavelet transform of a delayed signal $f(\bullet - y)$ equals the delayed (in the first variable) wavelet transform of f. Also note that the scale s is inversely proportional to the filter center frequency: If ψ has its passband around frequency ξ_1, then $\psi(\bullet/s)$ has its passband around $\xi_s := \xi_1/s$.

In applications, only a discrete subset of the continuous wavelet system can be considered; the system, or equivalently the transform, is *decimated*. It is desirable that any function f can be stably recovered from the decimated transform. Mathematically, these properties correspond to the discrete subset of the continuous wavelet system constituting a frame [15]. The ratio B/A of the so-called upper (B) and lower (A) frame bounds quantifies how well the decimated transform preserves signal energy, i.e., how stable it is in numerical computation.

Commonly, the discrete subset $\big((x_{l,j}, s_j)\big)_{l\in\mathbb{Z}, j\in I} \subset \mathbb{R} \times [0,\infty)$ of translation-dilation pairs is generated by straightforward decimation rules. We denote the corresponding decimated wavelet system by $\{\psi_{l,j}\}_{l\in\mathbb{Z}, j\in I}$ with $\psi_{l,j} := \psi_{(x_{l,j}, s_j)}$. Regarding the choice of the scales s_j, it is essential to note

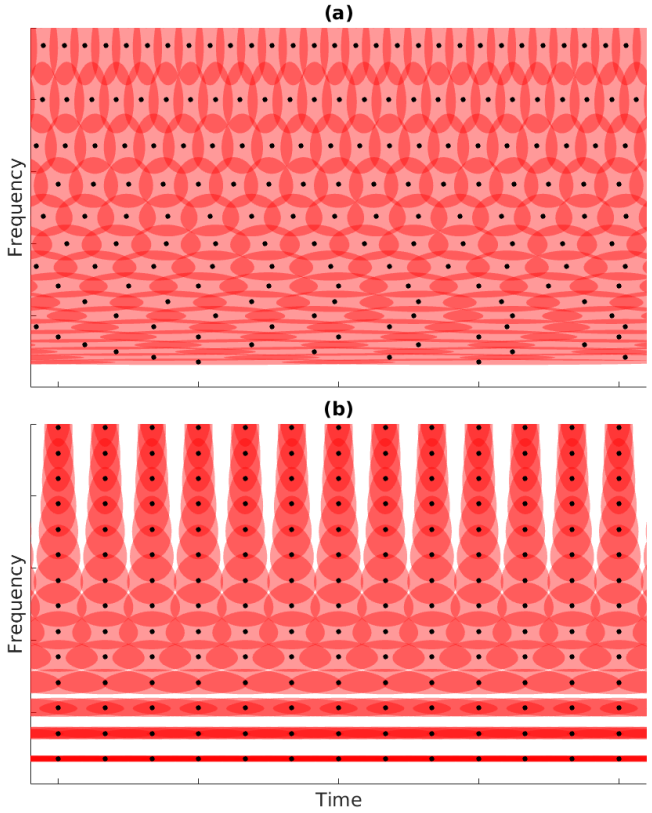

Figure 1. Time-frequency geometry of decimated wavelet systems: (a) The classical nonuniform wavelet decimation provides a mostly even covering of time-frequency space, at the cost of introducing nonuniform decimation. (b) Wavelet decimation on a uniform grid results in an increasingly uneven covering away from a small frequency strip.

that the bandwidth of wavelet filters increases linearly with their center frequency ξ_s, which is inversely proportional to their scale s. Thus, it is natural to decimate the frequency channels by taking integer powers of a fixed *base scale* $a > 0$, i.e., $s_j = a^{-j}$. Conversely, the width of the wavelet impulse responses is proportional to the scale, suggesting in-channel decimation according to $x_{l,j} = a^{-j} \cdot lb$ for some fixed $b > 0$, see also [34, Chapter 3]. Intuitively, the time-frequency region covered by the individual $\psi_{(x_{l,j},s_j)}$ corresponds roughly to ellipses of constant area that grow narrower as j increases, or, more accurately, to hyperbolic circles of constant radius, centered at $(x_{l,j}, \xi_1/s_j)$, see Fig. 1(a).

This decimation strategy choosing $(x_{l,j}, s_j) = (a^{-j} \cdot lb, a^{-j})$, with geometric frequency spacing and channel-dependent decimation factor (hop size) $d_j = a^{-j} \cdot b$, is customary and well-studied for wavelet bases, overcomplete wavelet systems, and constant-Q transforms, e.g., [13], [34]. We will refer to this convention as *classical wavelet decimation*. However, the dependence of the decimation factor d_j on the scale leads to filter bank coefficients that are not time-aligned and, in many applications, encumber the workflow. In contrast to the STFT, the time-frequency geometry of the wavelet transform described above is ill-suited for decimation on a uniform, rectangular grid, independent of whether the center frequencies are spaced linearly or logarithmically, see Fig. 1(b). It is easy to see and, in fact, straightforward to prove, that such a decimation strategy cannot lead to a numerically stable system with perfect reconstruction. Considering that many successful processing schemes for the STFT rely on both perfect reconstruction and the representation of the coefficients as a time-frequency matrix, this presents a notable obstruction to the adoption of wavelets in (audio) signal processing.

On the other hand, perfect reconstruction from filter banks with non-uniform decimation, in general, introduces a non-negligible computational burden compared to uniform decimation, see [17], [35], and requires costly iteration. Although this iteration can sometimes be avoided [13], [36], this imposes severe restrictions on the considered filters and decimation scheme.

B. Quasi-random sequences

Quasi-random sequences, also known as low discrepancy sequences [20], are deterministic sequences of numbers or k-dimensional coordinates that share some properties of uniformly distributed random numbers and can be used as a replacement for the latter in various applications, most prominently the quasi-Monte Carlo method for numerical integration [20]. The term *discrepancy* refers to one of several related measures that quantify how uniformly distributed a set of points is. This property is important, e.g., to obtain error bounds in quasi-Monte Carlo integration. While most modern theoretical studies are concerned with quasi-random sequences of high dimensionality, we only require 1-dimensional sequences. Specifically, any N consecutive elements of a low discrepancy sequence in dimension $D = 1$ provides N numbers that are particularly well distributed in the unit interval.

Decimation strategies for wavelets can be derived from any low discrepancy sequence. Here, we focus on one carefully chosen example each out of two prominent classes of quasi-random sequences.

Kronecker sequences: A deceptively simple construction are so-called Kronecker sequences which are of the form $(\{\alpha l\})_{l\geq 0} = (0, \{\alpha\}, \{2\alpha\}, \ldots)$ for some real α, where $\{x\} := x - \lfloor x \rfloor$ denotes the fractional part of a real x. It is well-known that Kronecker sequences are particularly well distributed if α is a *badly approximable number*. These are irrational numbers that are particularly poorly approximated by rationals. Formally, a number $\alpha \in \mathbb{R}$ is badly approximable if there is a constant $c > 0$, such that

$$|\alpha - p/q| > c/q^2,$$

for all nonzero integers $p, q \in \mathbb{Z} \setminus \{0\}$. It is known that an irrational number is badly approximable if and only if the coefficients of its continued fraction expansion are bounded. Among all the badly approximable numbers, the *golden ratio* $\phi = \frac{1+\sqrt{5}}{2}$ maximizes the optimal constant c. Since this property is shared exactly with all so-called equivalent numbers of the form $\frac{a\phi+b}{c\phi+d}$, for integers a, b, c, d with $ad - bc = \pm 1$, we may likewise consider the Kronecker sequence with $\alpha = 1 - 1/\phi$, i.e., $a, c = 1$, $b = -1$, and $d = 0$. For more information on Kronecker sequences we refer to the books [37], [38].

Digital $(0,1)$*-sequences over* $\mathbb{Z}_2$*:* A $(0,1)$-sequence in base 2 is an infinite sequence $(x_l)_{l\geq 0} = (x_0, x_1, \dots)$ in the unit-interval $[0,1)$ with the following property: for every $m \in \mathbb{N}_0$ and every $k \in \{0, 1, \dots, 2^m - 1\}$, the elementary interval $[\frac{k}{2^m}, \frac{k+1}{2^m})$ contains exactly one element of the point set

$$\{x_l \ : \ p2^m \leq l \leq (p+1)2^m - 1\} \quad \text{for every } p \in \mathbb{N}_0.$$

One example for such a sequence is the well-known van der Corput sequence [39] (see also [40]), which is related to the bit-reversal permutation. Typically, $(0,1)$-sequences in base 2 are constructed by the so-called digital method over the finite field $\mathbb{Z}_2$ of order two, i.e., the integers modulo 2. To this end, let $\mathbf{C} = (c_{r,k})_{r,k\geq 1}$ be an $\infty \times \infty$-matrix over $\mathbb{Z}_2$, i.e., with entries $c_{r,k}$ from $\{0,1\}$, such that for every $m \in \mathbb{N}$ the left-upper $m \times m$ sub-matrix is non-singular. Then the sequence $(x_l)_{l\geq 0}$ defined by

$$x_l = \sum_{r=0}^{\infty} \frac{\eta_{l,r}}{2^{r+1}} \quad \text{where} \quad \eta_{l,r} = \sum_{k=0}^{\infty} c_{r+1,k+1} l_k \pmod 2,$$

where $l_k \in \{0,1\}$ are the binary digits of the index l, i.e., $l = l_0 + l_1 2 + l_2 2^2 + \cdots$ (which are obviously 0 from a certain index on), is a $(0,1)$-sequence in base 2. In this context, the van der Corput sequence is covered by choosing the identity matrix for $\mathbf{C}$. We will use the specific digital $(0,1)$-sequence which is obtained from the infinite matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 1 & 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

For more information on (digital) $(0,1)$-sequences we refer to the books [20], [21] or the survey article [40, Section 3.2].

III. A New Convention for Wavelet Sampling

Due to the geometric intuition outlined in Section II-A, linear spacing of wavelet center frequencies $\xi_j \sim s_j^{-1}$ has hardly been considered in the literature. For the same reason, uniform decimation, i.e., $d_j = d$ for some fixed $d > 0$, is usually disregarded. We will now describe a decimation scheme that uses both of these conventions and which can be used to construct numerically stable, perfect reconstruction wavelet transforms with moderate, or even marginal, oversampling.

Given a largest scale of interest $b > 0$ (or equivalently a minimal frequency of interest $\xi_{\min} > 0$), a decimation factor $d > 0$, a third parameter $q > 0$ that determines the step size in the frequency direction, and finally a low discrepancy sequence $(\delta_0, \delta_1, \delta_2, \dots)$ determining channel specific delays, we select the translation-dilation pairs

$$(x_{l,j}, s_j) = \left(d(l + \delta_j), \frac{1}{b^{-1} + q^{-1} j} \right), \tag{2}$$

for all integers $l \in \mathbb{Z}$ and non-negative integers $j \in \mathbb{N}_0$. In other words, we consider the discrete wavelet system $\{\psi_{l,j}\}_{l\in\mathbb{Z}, j\in\mathbb{N}_0}$, with

$$\psi_{l,j}(t) = \sqrt{\tfrac{1}{b} + \tfrac{j}{q}} \cdot \psi\left((\tfrac{1}{b} + \tfrac{j}{q}) \cdot (t - d(l + \delta_j)) \right). \tag{3}$$

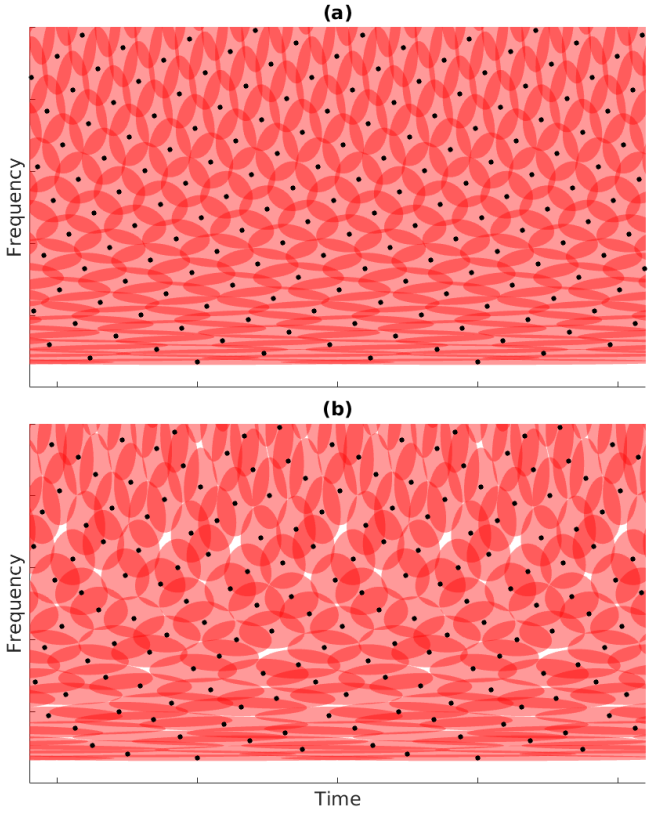

Figure 2. Time-frequency geometry of decimated wavelet systems with quasi-random delay: (a) Decimation on a uniform grid with delays chosen according to the *golden* Kronecker sequence provides an even covering comparable to the classic wavelet decimation (Fig. 1(a)). (b) Deriving the delays instead from the digital sequence described in Section II-B, we obtain a covering that is slightly less even. In both cases, the resulting covering will become uneven below a certain frequency, similar to the bottom region in Fig. 1(b).

In contrast to the rectangular grid shown in Fig. 1(b), which can be generated by the above construction with delays $\delta_j = 0$ for all j, choosing the δ_j as the elements of a quasi-random sequence provides a wavelet system that covers the time-frequency plane surprisingly evenly, see Fig. 2. Intriguingly, if we choose the δ_j as the elements of a $\{l\alpha\}$-sequence, the points $\{(x_{l,j}, \xi_j)\}_{l\in\mathbb{Z}, j\in\mathbb{N}_0}$ form a uniform (skewed) time-frequency grid.

A. Numerically stable wavelet transforms with perfect reconstruction and uniform decimation

As indicated by Fig. 2, decimating the wavelet system according to a time-frequency grid with quasi-random channel delays yields a surprisingly uniform covering of the time-frequency plane. However, due to the uniform spacing in frequency, this is only true as long as the scale of the wavelets is small enough (relative to the frequency step parameter q). Since the wavelet bandwidth is proportional to its center frequency, we find that the lower frequency region is insufficiently covered, similar to Fig. 1(b), for arbitrarily large base scale b. The size of this region depends on the chosen wavelet and the frequency step parameter q. To compensate for this

lack of coverage, we introduce the additional lowpass elements

$$\psi_{l,j}(t) = \frac{1}{\sqrt{b}}\psi\left(\frac{t-d(l+\delta_j)}{b}\right)e^{2\pi i\xi_1\cdot\frac{j(t-d(l+\delta_j))}{q}}, \quad (4)$$

for negative integers $j \geq -q/b$. In words, we demodulate the wavelet at the base scale b in uniform steps for as long as the center frequency remains non-negative. Note that the construction above implies that the index of the quasi-random sequence of delays is shifted accordingly, i.e., we denote the j-th element by $\delta_{j-\lfloor q/b\rfloor}$ instead of δ_j. Clearly, this is only one of many possibilities for the construction of suitable lowpass filters, chosen here because it preserves the filter prototype, or mother wavelet, and the uniform decimation of the system in time and frequency.

The full system $\{\psi_{l,j}\}_{l\in\mathbb{Z},j\in\mathbb{Z}\cap[-q/b,\infty)}$ remains uniformly decimated and its properties, in particular the perfect reconstruction property and numerical stability can be studied using frame theory for uniform filter banks [19] or, equivalently, shift-invariant systems [18], which also provides highly efficient algorithms for reconstruction from the filter bank coefficients. We observe that, in practice, choosing the decimation parameters d and $1/q$ small enough is sufficient to obtain an invertible, numerically stable filter bank. In Section IV, we validate this assessment through a numerical study.

B. Notes on implementation

Although the parametrization of $\{\psi_{l,j}\}_{l,j}$ introduced above, namely, in terms of $b > 0$ and $q > 0$, arises naturally, it is inconvenient for practical construction. By restricting to the case $q/b \in \mathbb{N}$, we can instead construct the proposed decimated wavelet system by selecting a mother wavelet ψ, the desired number $M+1$ of frequency channels equidistantly spaced at center frequencies in $[0, \xi_s/2]$, where ξ_s is the sampling rate, and the number M_{LP} of lowpass/demodulated filters. The relation to the parametrization introduced above is given by $M_{\text{LP}} = q/b$ and $\xi_1/q = \xi_s/(2M)$. In the provided implementation, the wavelet is automatically scaled such that the center frequency of $\psi_{l,0}$ equals $\xi_s \cdot M_{\text{LP}}/(2M)$. The redundancy of the resulting system is then controlled by selecting the decimation factor d, similar to the common practice for the STFT.

IV. Numerical Evaluation

To validate that the proposed decimation strategy indeed leads to invertible, numerically stable wavelet systems, we compute the *optimal frame bound ratio* R_{FB} for various wavelet filter banks so decimated. For a filter bank acting on finite sequences, the optimal frame bound ratio equals the condition number of the matrix that realizes the composition of filter bank analysis and synthesis. The filter bank is invertible if and only if this condition number is finite and it is perfectly stable if $R_{\text{FB}} = 1$. To compute this number efficiently, we use the *Large Time-Frequency Analysis Toolbox* (LTFAT, ltfat.org), which also implements the wavelet filter banks used throughout this work. The evaluation presented here can be reproduced using the code available at: http://ltfat.org/notes/057.

In a pre-test for the proposed evaluation, we noticed that for fixed M, the frame bound ratio R_{FB} decreases monotonically with increasing M_{LP}, up to a certain value of M_{LP} after which a further increase yields no benefit. Furthermore, we found that this value of M_{LP} does not depend on the choice of M. Increasing M_{LP} stabilizes the frequency response of the filter bank, which equals $\Psi = \sum_j |\widehat{\psi_{0,j}}|^2$ up to a positive multiplicative constant, especially in the lowpass region. Strong fluctuations of Ψ have detrimental effect on R_{FB}, and increasing M_{LP} reduces these fluctuation, explaining the first observed effect. The lack of a dependence on M can be explained by the fact that the frequency response of two such filter banks that only differ in the choice of M are equal up to a dilation and a positive multiplicative constant. Altogether, these observations allow us to greatly reduce the number of tested configurations.

With the above considerations in mind, we perform an optimization of the frame bounds by first determining, for a fixed mother wavelet and oversampling rate, the smallest value of M_{LP} that yields optimal R_{FB}. We then proceed to determine the value $M \in \{128, 256, 384, 512, 640, 768, 1024, 1280, 1536, 2048\}$ that optimizes R_{FB}, with the additional restriction that $M \geq 50M_{\text{LP}}$, by an exhaustive search. Finally, we further refine the choice of M by a divide and conquer approach starting from the determined optimizer and its two neighbors in the sequence of tested values for M. The restriction $M_{\text{LP}} \leq M/50$ ensures that the center frequency of $\psi_{l,0}$ is no larger than $\xi_s/100$, e.g., 441 Hz for a sampling rate of 44.1 kHz.

Results: In Table I, we show the obtained optimal frame bounds for oversampling rates $1.2,\ 2,\ 4$, and 8. As mother wavelet, we consider Cauchy wavelets [41], defined (up to a normalization constant) by $\widehat{\psi}(\xi) = \xi^{\frac{\alpha-1}{2}}e^{-2\pi\xi}$, for α equal to $100,\ 300,\ 900$, and 2700, where higher α implies a higher Q-factor. As an example of a compactly supported wavelet, we consider the modulated fourth order B-spline defined (up to a normalization constant) by $\widehat{\psi}(\xi) = \sin(\pi(\xi-\xi_{\text{fm}}))^4/(\pi(\xi-\xi_{\text{fm}}))^4$, with ξ_{fm} equal to $3,\ 6$, and 10, implying Q-factors that are roughly equivalent to Cauchy wavelets with α equal to $257,\ 1024$ and 2842, respectively, see [42].

As is to be expected, the optimal frame bound ratio decreases with higher oversampling rate. For fixed oversampling, however, there is only limited dependence on the mother wavelet and Q-factor. Specifically, we obtain R_{FB} close to $14,\ 3,\ 1.65$, and 1.3 for oversampling rates $1.2,\ 2,\ 4$, and 8 across almost all conditions with Kronecker-sequence based decimation. For redundancy 4 and below, the results for the decimation based on a digital-(0,1) sequence yield slightly worse stability, i.e., larger R_{FB}. Interestingly, for factor 8 oversampling, R_{FB} is improved by using this decimation scheme, for all but one conditions, indicating that a decimation scheme based on digital-(0,1) sequences is beneficial when large oversampling rates are used.

Note that our divide and conquer refinement procedure is only guaranteed to find the global minimum of R_{FB} as a function in M if this function is convex. This is usually not the case as illustrated in Fig. 3, where we show the function

Table I
OPTIMIZED FRAME BOUND RATIOS FOR THE PROPOSED WAVELET SYSTEMS. TABLE VALUES ARE R_{FB} (M_{LP}, M), I.E., THE OPTIMAL FRAME BOUND RATIO R_{FB} IS ACHIEVED WITH M_{LP} LOWPASS CHANNELS AND M TOTAL CHANNELS.

Kronecker-sequence delays							
	Cauchy				B-Spline		
Oversampling	$\alpha = 100$	$\alpha = 300$	$\alpha = 900$	$\alpha = 2700$	$\xi_{\text{fm}} = 3$	$\xi_{\text{fm}} = 6$	$\xi_{\text{fm}} = 10$
1.2	15.06 (2, 102)	14.17 (4, 202)	13.61 (6, 307)	13.74 (11, 550)	15.04 (4, 202)	14.08 (7, 449)	14.08 (12, 620)
2	3.22 (3, 448)	2.98 (5, 253)	2.92 (9, 501)	2.94 (15, 764)	3.01 (5, 253)	2.93 (9, 511)	2.97 (15, 764)
4	1.72 (4, 768)	1.62 (7, 350)	1.60 (12, 768)	1.59 (20, 1012)	1.60 (6, 363)	1.59 (12, 768)	1.59 (20, 1024)
8	1.31 (4, 214)	1.25 (8, 404)	1.23 (14, 702)	1.24 (26, 1306)	1.25 (7, 473)	1.24 (15, 757)	1.25 (25, 1250)
Digital (0,1)-sequence delays							
	Cauchy				B-Spline		
Oversampling	$\alpha = 100$	$\alpha = 300$	$\alpha = 900$	$\alpha = 2700$	$\xi_{\text{fm}} = 3$	$\xi_{\text{fm}} = 6$	$\xi_{\text{fm}} = 10$
1.2	20.38 (2, 127)	19.27 (4, 255)	17.48 (6, 384)	16.02 (11, 1023)	20.67 (4, 256)	17.26 (7, 384)	16.98 (12, 640)
2	3.80 (3, 260)	3.73 (5, 384)	3.66 (8, 656)	3.57 (14, 771)	3.72 (5, 388)	3.62 (9, 771)	3.62 (14, 771)
4	1.74 (4, 255)	1.67 (7, 383)	1.63 (12, 639)	1.62 (21, 1279)	1.78 (6, 640)	1.62 (12, 639)	1.60 (20, 1023)
8	1.27 (5, 256)	1.20 (9, 511)	1.20 (16, 1535)	1.21 (27, 1791)	1.21 (8, 511)	1.27 (16, 1408)	1.22 (27, 1791)

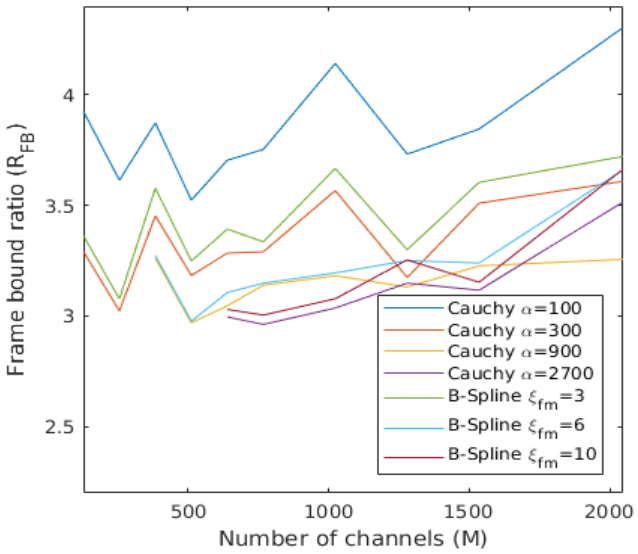

Figure 3. Examples for the dependence of R_{FB} on M for Kronecker-sequence delays and oversampling rate 2.

for a fixed oversampling rate. Thus, it may be possible to further improve the reported bounds. Furthermore, this is an explanation why the reported optimal parameter M does not depend monotonically on the Q-factor of the mother wavelet which is larger for larger choices of α or ξ_{fm}.

To validate that the wavelet coverings illustrated in Figs. 1 and 2 above do indeed conform to the time-frequency domain being well covered by the proposed wavelet systems, we calculated accumulated spectrograms [43] of these systems (see Fig. 4). More specifically, we calculated the spectrogram for each wavelet in the system using a short-time Fourier transform with a Gaussian window and summed all of them. Since the spectrogram of a signal is a representation of its time-frequency energy localization, this sum illustrates the time-frequency area that is well represented by the different wavelet systems. Our simulations, illustrated by Fig. 4, essentially confirm the findings of Fig. 1 and 2. The classical wavelet system in (a) and the delay shifted system using the *golden* Kronecker sequence (c) show the best uniformity. Clearly, the uniform system without delays (b) on the one hand overemphasizes certain time-frequency regions while completely missing the area in between. The uniform system with delays based on a digital $(0, 1)$-sequence (d) does not cover the area quite as uniformly. However, the accumulated spectrograms presented here were obtained with an oversampling rate of approximately 2 such that this finding matches the computed frame bounds: At low oversampling rates, the Kronecker sequence delays are superior to the digital $(0, 1)$-sequence delays.

V. EXPERIMENTS IN AUDIO PROCESSING

We present two applications of the proposed wavelet decimation scheme. In the first application, we replicate experiments from prior work on NMF-based signal decomposition [22], which originally relied on the STFT. We simply substitute the wavelet transform for the STFT in a plug-and-play manner, leaving all other parameters unchanged. Only the parameters of the wavelet transform, i.e., the mother wavelet ψ, the number of channels M and lowpass channels M_{LP}, were adjusted. Given that the considered processing scheme was originally conceived for the STFT, and not adapted in any way, it is not expected that using the wavelet transform will outperform the original methods using the STFT. Instead, the goal is to demonstrate that even a naive plug-and-play approach can achieve comparable results.

In a second application, we compare the performance of phaseless reconstruction with the fast Griffin-Lim algorithm (FGLA) [23], [44] between the proposed decimation scheme and classical decimation. Results on phaseless reconstruction with FGLA from STFT coefficients are provided as reference.

Accompanying audio files and code for reproducing the presented experiments is available at: http://ltfat.org/notes/057.

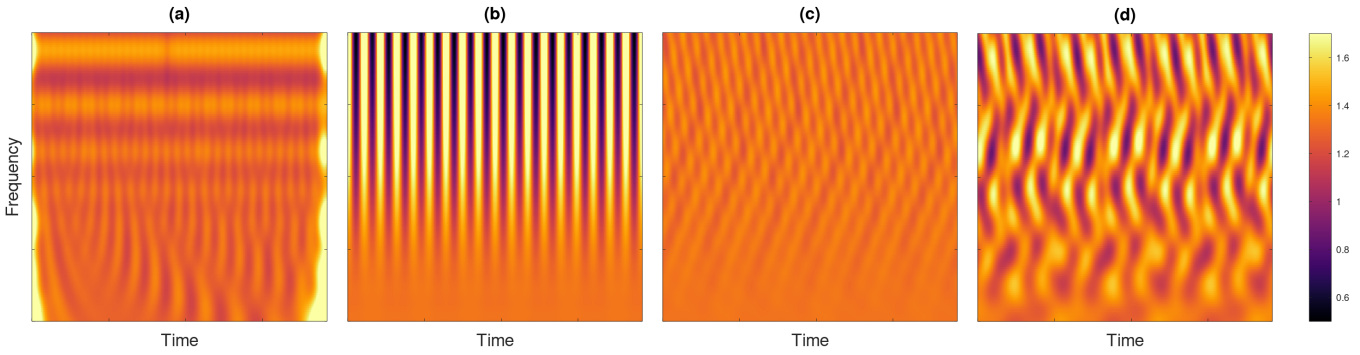

Figure 4. Accumulated spectrograms of decimated wavelet systems: (a) classical nonuniform decimation, (b) uniform grid, (c) uniform grid with delays chosen according to the *golden* Kronecker sequence, (d) delays derived from the digital sequence described in Section II-B.

A. Signal decomposition with non-negative matrix factorization

In [22], Févotte et al. showed that non-negative matrix factorization with the Itakura-Saito cost function achieves a meaningful decomposition of STFT spectrograms of audio data as $S = W \cdot H$, with W and H being non-negative component and activation matrices, respectively. The authors presented the decomposition, denoising, and upmixing of a 1920s recording of "My Heart (Will Always Lead Me Back to You)" by Luis Armstrong and His Hot Five. According to the subjective analysis in [22], the recording contains trumpet, clarinet, trombone, double bass, and piano tracks, as well as significant hiss noise and crackling. The signal used was an excerpt of length $108\,\mathrm{s}$, sampled at $\xi_s = 11\,025\,\mathrm{Hz}$, with a total length of $L = 1\,191\,735$ samples. In the original contribution, an Itakura-Saito NMF decomposition with 10 components was computed from an STFT spectrogram with $M = 129$ channels from the zero to Nyquist frequencies and an oversampling rate of 2, using an inverse Gamma prior to regularize the component activation matrix.

In our experiment, we adapted the code provided with the follow-up contribution by Févotte [45], which uses a different smoothing prior, but is otherwise identical. As in that contribution, the regularization parameter was set to $\lambda = 25$. When replicating the experiment, we used a wavelet transform with a Cauchy wavelet [32], [41], with hyperparameter $\alpha = 450$. Decimation was based on the Kronecker sequence and the oversampling rate was set to 2. The parameters were optimized[1] as in Section IV, leading to a total of $M + 1 = 449$ channels and $M_{\mathrm{LP}} = 6$ lowpass channels, i.e., the center frequency of $\psi_{l,0}$ is at $74\,\mathrm{Hz}$. To achieve the desired oversampling rate, the decimation factor was set to $d = 448$. The resulting NMF components and signal decomposition are shown in Fig. 5.

Results: Comparing to the results shown in [22], we see that the harmonic structure of the components in W is not as pronounced when the wavelet transform is used, but still present. This is not entirely surprising, considering the large bandwidth of high frequency wavelets and the non-aligned phase space covering of the Kronecker sequence based decimation. Nonetheless, the decomposition achieves a separation into lead, accompaniment and noise comparable to the results presented for the STFT in [22], [45]. The main difference is that the trombone track is not as clearly separated, but mostly mixed with the lead track. Audio examples, including individual components, as well as denoised and upmixed versions of the original signal, are provided on the website.

B. Phaseless reconstruction with the fast Griffin-Lim algorithm

In this experiment, we evaluate phaseless reconstruction, i.e., the reconstruction of an audio signal from magnitude-only time-frequency coefficients in our proposed wavelet decimation scheme. The Griffin-Lim algorithm [44] (GLA) remains the most popular iterative method for phaseless reconstruction from STFT or general time-frequency spectrograms. Here, we consider the fast Griffin-Lim (FGLA) variant proposed by Perraudin et al. [23], which introduces a Nesterov-like acceleration term. The reconstruction error is measured in spectral SNR, often referred to as spectral convergence and given by

$$\mathbf{SNR}_{\mathbf{MS}}(f, f_r) = 10\ \log_{10} \frac{\|W_\psi f\|^2}{\|\,|W_\psi f_r| - |W_\psi f|\,\|^2}, \qquad (5)$$

with the target signal f and the reconstructed signal f_r. In the literature, it is not always clear with respect to which time-frequency representation the quantity $\mathbf{SNR}_{\mathbf{MS}}$ is computed. Since we compare results across different representations and parameter choices, we fix the representation for computing $\mathbf{SNR}_{\mathbf{MS}}$, as in [46]. Here, we choose a highly oversampled wavelet transform using a Cauchy-type mother wavelet with $\alpha = 1000$, with $M + 1 = 181$ geometrically spaced frequency channels and $d = 3$, i.e., approximately 60-fold oversampling, as reference representation.

The experimental setup is similar to [42, Section 4.2]: We consider the same 15 signals from the EBU SQAM dataset[2] and test Cauchy-type mother wavelets with $\alpha = 1000$, at

[1]Due to memory constraints, we restricted the optimization to choose no more than 769 channels.

[2]The first 5 seconds of signals 01, 02, 04, 14, 15, 16, 27, 39, 49, 50, 51, 52, 53, 54, and 70.

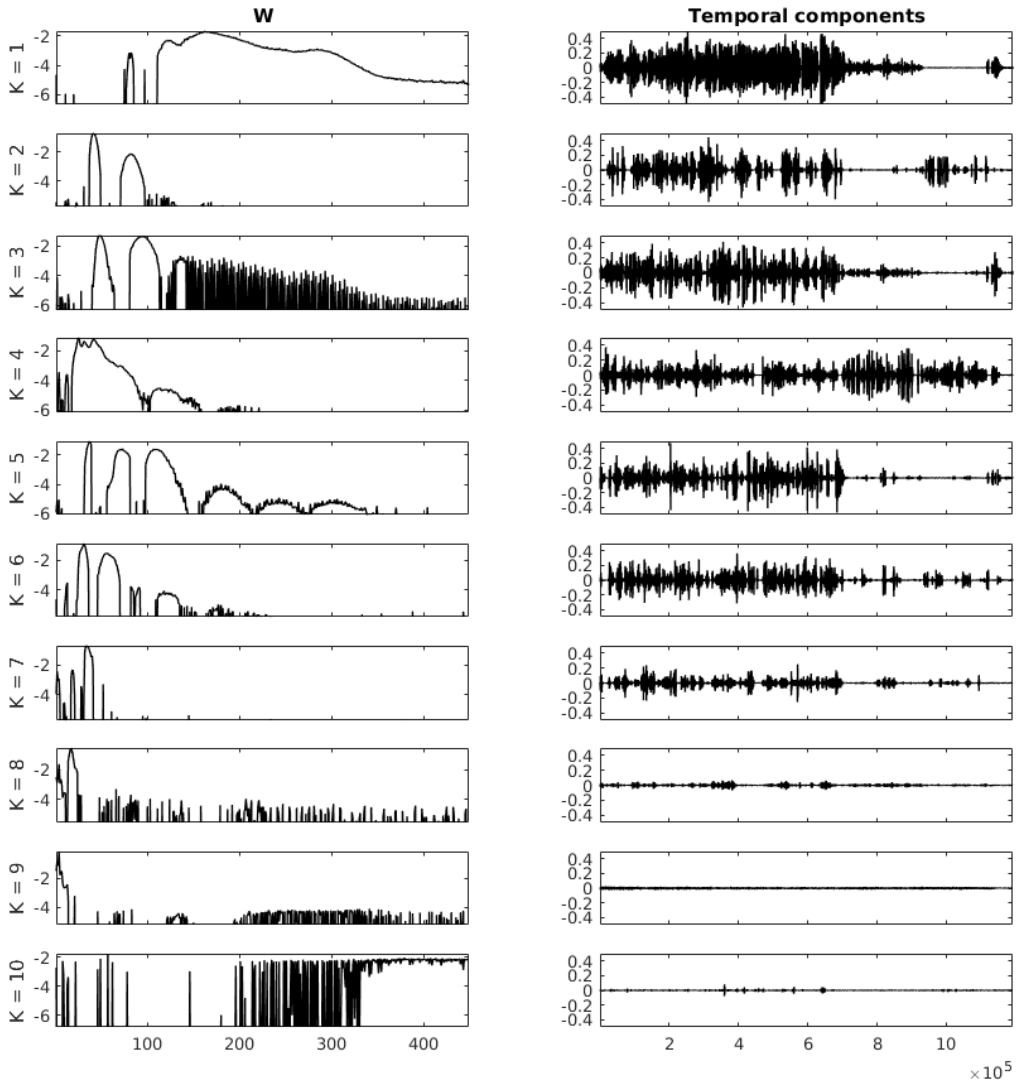

Figure 5. NMF decomposition of music excerpt. (Left) dB-scaled columns of W, i.e., NMF components. (Right) Reconstructed signal components. Components 4, and $6-8$ capture most of the piano and double bass tracks. Components 9, 10 characterize large parts of the hiss and crackling noise. The remaining components contain most of the trumpet and clarinet lead track, as well as the trombone.

oversampling rates[3] of 3, 5, and 10. For the proposed method, we chose $M = 750$, 968, and 1369 for low, medium, and high oversampling, respectively. We further chose 11, 13, and 17 low pass channels. The number of lowpass channels is determined as in Section IV. After promising results were obtained for $M = 750$ at oversampling rate 3, the parameters for higher rates were chosen by isotropic scaling of the sampling grid, i.e., the product dM remained constant[4]. As references, we first consider wavelet transforms at the same oversampling rates, but with geometric frequency spacing. Since we found that the uniform decimation scheme used in [42, Section 4.2] leads to unstable systems at low and medium oversampling, we further use channel-dependent decimation as discussed in Section II-A. The number of channels is set to $M = 90$, 125, and 180 for low, medium, and high oversampling, respectively. Furthermore, we consider STFTs with a Hann window at the same redundancies with $M = 1536$, 1920, and 2880 channels and uniform decimation. To prevent issues with poor initialization, we first computed 20 FGLA iterations for initial phase 0 and for five random uniformly distributed phase initializations. The best of these six candidates is chosen to compute the final solution by applying another 130 FGLA iterations.

Results: In Fig. 6, our results are summarized. Overall, we see a clear trend that an increase in oversampling improves performance, matching the results obtained in [46]. On average, the classical wavelet decimation is superior to the STFT at medium and high oversampling and to the proposed method at high oversampling. At medium oversampling both wavelet methods are roughly on par. At low oversampling, the performance of the proposed method is superior to both reference methods. On the associated website, we provide

[3]Note that the value M/a considered in [42, Section 4.2] corresponds to roughly half the oversampling rate, considering that coefficients are complex-valued.

[4]Note that the requirement that $d, M \in \mathbb{N}$ leads to round-off errors. Moreover, the implementation we use chooses the largest value d such that $M/(2d)$ is at least as large as the desired oversampling rate. Hence, the actual oversampling rate can be slightly larger.

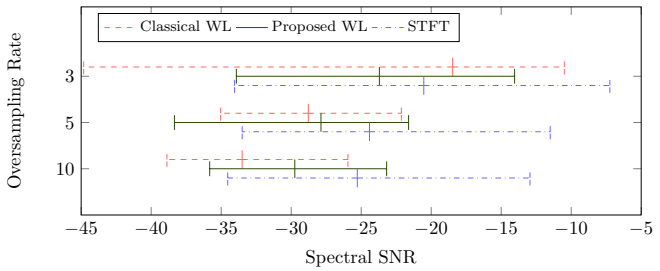

Figure 6. Whisker plots showing the minimal, median, and maximal spectral SNR after phaseless reconstruction for 15 signals and three transforms, namely, the classical wavelet transform, the proposed wavelet transform with Kronecker sequence based decimation, and the STFT. The different oversampling rates are arranged vertically.

audio examples and additional results obtained when computing $\mathbf{SNR}_{\mathrm{MS}}$ with respect to a reference STFT. As observed in [46], $\mathbf{SNR}_{\mathrm{MS}}$ is biased towards representations that are *similar* to the reference, and we found that $\mathbf{SNR}_{\mathrm{MS}}$ with STFT reference, when compared to Fig. 6, favors the STFT over the wavelet transform in general, and Kronecker decimation over classical wavelet decimation in particular. However, the overall findings are similar, such that we do not include these results here.

VI. Wavelets in the Large Time-Frequency Analysis Toolbox

The wavelet filter banks used for the experiments are implemented in the Large Time-Frequency Analysis Toolbox (LTFAT[5]) [47] through the function `waveletfilters` in LTFAT's *filterbank module*. `waveletfilters` accepts the filter bank length L and a vector of wavelet scales as input arguments and returns the wavelet filters corresponding to the specified scales, their corresponding decimation factors and center frequencies.

By default, `waveletfilters` selects a Cauchy-type mother wavelet with $\alpha = 300$ and the unit scale $s = 1$ corresponds to a center frequency of $\xi_s/20$. Further, non-uniform, integer decimation factors adapted to the wavelet bandwidth are automatically chosen by default. Various optional input parameters may be used to control the mother wavelet, oversampling rate, decimation scheme, and other characteristics of the wavelet filter bank. In particular, `waveletfilters` covers all configurations used in [32], [42] and in the present work: The wavelet filters are specified in the frequency domain according to [32]. Currently, the Cauchy [48], Morse [49], Morlet, frequency B-spline [50], analytic spline [51], and complex spline wavelets are implemented for positive and negative scales. The individual filters are generated by the separate function `freqwavelet`, which enables the straightforward addition of further wavelet prototypes in the future. In addition to integer decimation as considered here, `waveletfilters` supports fractional and non-uniform decimation for later use. To construct perfect reconstruction filter banks, the frequency range from 0 Hz to the center frequency of the largest wavelet scale can be covered either by a single lowpass or several frequency-shifted, scaled copies of that wavelet as described in Section III-A.

To compute the filter bank coefficients, the LTFAT function `filterbank` accepts a target signal and the wavelet filters and decimation factors output by `waveletfilters` as input arguments. `filterbank` instantiates the wavelet filter bank and uses fast FFT-based convolution to retrieve the coefficients. The output of `waveletfilters`, and the resulting filter bank coefficients are fully compatible with the LTFAT filterbank module, streamlining further processing with LTFAT, e.g., visualization via `plotfilterbank`, time-frequency reassignment [52], and the phaseless reconstruction [23], [53]. Synthesis from filter bank coefficients is realized by `ifilterbank`. If analysis-synthesis filter bank pairs with perfect reconstruction are desired, a dual filter bank can be obtained by applying `filterbankdual` to the output of `waveletfilters`, respectively `filterbankrealdual` for filter banks that cover only positive frequencies. Note that this is only possible for uniform filter banks or under strict conditions on non-uniform filter banks. Generally, perfect reconstruction can be achieved with an iterative method, implemented in `ifilterbankiter`, provided that R_{FB} is finite.

[5]They can be found in the current development version and will be integrated into future releases, from LTFAT 2.5.0 onwards.

VII. Conclusion and Outlook

We proposed a novel, uniform wavelet decimation scheme with quasi-random delays. Conceptually, and through numerical evaluation, we demonstrated that the proposed scheme is suitable for constructing wavelet systems with the perfect reconstruction property, even at oversampling rates close to 1. In an audio decomposition application using Itakura-Saito NMF, we have shown that processing schemes previously proposed for the STFT are easily adapted to our setting by interpreting the filter bank coefficients as a time-frequency matrix, in contrast to classical, non-uniform wavelet decimation. Finally, we observed promising performance of our method in phaseless reconstruction for diverse audio signals.

We expect that the proposed decimation strategy will become a valuable asset for future work in different applications in audio and beyond. Furthermore, whereas the analysis of the proposed method in this work is restricted to a numerical investigation, a rigorous study of its formal, mathematical properties in the context of frame and function space theory is in preparation.

Acknowledgments

We would like to thank Cédric Févotte for kindly providing their code for Itakura-Saito NMF and some guidance on its use. We further thank Georg Tauböck for fruitful discussion on potential applications of the proposed method.

References

[1] J. C. Brown and M. S. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *The Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2698–2701, 1992.

[2] C. Schörkhuber, A. Klapuri, and A. Sontacchi, "Audio pitch shifting using the constant-Q transform," *J. Audio Eng. Soc*, vol. 61, no. 7/8, pp. 562–572, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16871

[3] B. Fuentes, A. Liutkus, R. Badeau, and G. Richard, "Probabilistic model for main melody extraction using Constant-Q transform," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 5357–5360.

[4] M. Todisco, H. Delgado, and N. Evans, "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," *Computer Speech & Language*, vol. 45, pp. 516–535, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230816303114

[5] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494618305799

[6] S. Kadambe and G. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 917–924, 1992.

[7] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[8] C.-C. Lin, S.-H. Chen, T.-K. Truong, and Y. Chang, "Audio classification and categorization based on wavelets and support vector machine," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 644–651, 2005.

[9] S. Mallat, *A Wavelet Tour of Signal Processing (Third Edition)*. Boston: Academic Press, 2009.

[10] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets*, J.-M. Combes, A. Grossmann, and P. Tchamitchian, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 286–297.

[11] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets*, J.-M. Combes, A. Grossmann, and P. Tchamitchian, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 298–304.

[12] I. W. Selesnick, "Wavelet transform with tunable q-factor," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3560–3575, 2011.

[13] N. Holighaus, M. Dörfler, G. A. Velasco, and T. Grill, "A framework for invertible, real-time constant-Q transforms," *IEEE Audio, Speech, Language Process.*, vol. 21, no. 4, pp. 775–785, Apr. 2013.

[14] C. Schörkhuber, A. Klapuri, N. Holighaus, and M. Dörfler, "A matlab toolbox for efficient perfect reconstruction time-frequency transforms with log-frequency resolution," in *Proceedings of the 53rd International Audio Engineering Society Conference: Semantic Audio*, Jan 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17112

[15] O. Christensen, *An Introduction to Frames and Riesz Bases*, ser. Applied and Numerical Harmonic Analysis. Cham: Springer International Publishing, 2016.

[16] T. Necciari, N. Holighaus, P. Balazs, Z. Průša, P. Majdak, and O. Derrien, "Audlet filter banks: A versatile analysis/synthesis framework using auditory frequency scales," *Applied Sciences*, vol. 8, no. 1:96, 2018.

[17] T. Necciari, P. Balazs, N. Holighaus, and P. L. Søndergaard, "The erblet transform: An auditory-based time-frequency representation with perfect reconstruction," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 498–502.

[18] A. J. E. M. Janssen, *The duality condition for Weyl-Heisenberg frames*. Boston, MA: Birkhäuser Boston, 1998, pp. 33–84.

[19] H. Bölcskei, F. Hlawatsch, and H. Feichtinger, "Frame-theoretic analysis of oversampled filter banks," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3256–3268, 1998.

[20] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611970081

[21] J. Dick and F. Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi–Monte Carlo Integration*. Cambridge University Press, 2010.

[22] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[23] N. Perraudin, P. Balazs, and P. L. Sondergaard, "A fast Griffin-Lim algorithm," in *Proc. IEEE Appl. Sig. Process. Audio Acoustics*, New Paltz, NY, USA, Oct. 2013.

[24] R. Levie and H. Avron, "Randomized signal processing with continuous frames," *J Fourier Anal Appl*, vol. 28, no. 5, 2021.

[25] ——, "Randomized continuous frames in time-frequency analysis," *arXiv preprint arXiv:2009.10525 [math.NA]*, 2021.

[26] R. Levie, H. Avron, and G. Kutyniok, "Quasi Monte Carlo time-frequency analysis," *arXiv preprint arXiv:2011.02025 [math.NA]*, 2021.

[27] R. F. Bass and K. Gröchenig, "Relevant sampling of band-limited functions," *Illinois Journal of Mathematics*, vol. 57, no. 1, pp. 43 – 58, 2013. [Online]. Available: https://doi.org/10.1215/ijm/1403534485

[28] H. Führ and J. Xian, "Relevant sampling in finitely generated shift-invariant spaces," *Journal of Approximation Theory*, vol. 240, pp. 1–15, 2019.

[29] D. Patel and S. Sampath, "Random sampling in reproducing kernel subspaces of $\mathbf{l}^p(\mathbb{R}^n)$," *Journal of Mathematical Analysis and Applications*, vol. 491, no. 1, p. 124270, 2020.

[30] P. Goyal, D. Patel, and S. Sampath, "Random sampling in reproducing kernel subspace of mixed lebesgue spaces," 2021.

[31] G. A. Velasco, "Relevant sampling of the short-time fourier transform of time-frequency localized functions," 2017.

[32] N. Holighaus, G. Koliander, Z. Průša, and L. D. Abreu, "Characterization of analytic wavelet transforms and a new phaseless reconstruction algorithm," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 3894–3908, 2019.

[33] J. M. Lilly and S. C. Olhede, "On the analytic wavelet transform," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 4135–4156, 2010.

[34] I. Daubechies, *Ten lectures on wavelets.* SIAM, 1992.

[35] T. Necciari, N. Holighaus, P. Balazs, Z. Průša, P. Majdak, and O. Derrien, "Audlet filter banks: A versatile analysis/synthesis framework using auditory frequency scales," *Appl. Sci.*, vol. 8, no. 1(96), Jan. 2018.

[36] 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, Oct. 2011.

[37] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*. Springer Berlin Heidelberg, 1997.

[38] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. John Wiley, 1974.

[39] J. G. van der Corput, "Verteilungsfunktionen I-II," *Proc. Akad. Wet. Amsterdam*, vol. 38, pp. 813–821, 1058–1066, 1935.

[40] H. Faure, P. Kritzer, and F. Pillichshammer, "From van der Corput to modern constructions of sequences for quasi-Monte Carlo rules," *Indagationes Mathematicae*, vol. 26, no. 5, pp. 760–822, 2015, in memoriam J.G. van der Corput (1890–1975).

[41] I. Daubechies and T. Paul, "Time-frequency localisation operators—a geometric phase space approach: II The use of dilations," *Inverse Prob.*, vol. 4, no. 3, pp. 661–680, Aug. 1988.

[42] N. Holighaus, G. Koliander, Z. Průša, and L. Abreu, "Non-iterative phaseless reconstruction from wavelet transform magnitude," in *Proceedings of the International Conference on Digital Audio Effects 2019 (DAFx19)*, Sept 2019. [Online]. Available: http://dafx.de/paper-archive/2019/DAFx2019_paper_23.pdf

[43] L. D. Abreu, K. Gröchenig, and J. L. Romero, "On accumulated spectrograms," *Transactions of the American Mathematical Society*, vol. 368, pp. 3629–3649, 01 2016.

[44] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr. 1984.

[45] C. Févotte, "Majorization-minimization algorithm for smooth itakura-saito nonnegative matrix factorization," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1980–1983.

[46] A. Marafioti, N. Holighaus, and P. Majdak, "Time-frequency phase retrieval for audio—the effect of transform parameters," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3585–3596, 2021.

[47] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The large time-frequency analysis toolbox 2.0," in *Sound, Music, and Motion*, M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, Eds. Cham, Switzerland: Springer, 2014, pp. 419–442.

[48] I. Daubechies and T. Paul, "Time-frequency localisation operators—a geometric phase space approach: II The use of dilations," *Inverse Problems*, vol. 4, no. 3, pp. 661—-680, 1988.

[49] S. C. Olhede and A. T. Walden, "Generalized Morse wavelets," *IEEE Trans. Sig. Process.*, vol. 50, no. 11, pp. 2661–2670, Nov. 2002.

[50] R. X. Gao and R. Yan, *Wavelets: Theory and applications for manufacturing*. Springer Science & Business Media, 2010.

[51] K. N. Chaudhury and M. Unser, "Construction of Hilbert Transform Pairs of Wavelet Bases and Gabor-Like Transforms," *IEEE Trans. Sig. Process.*, vol. 57, no. 9, pp. 3411—-3425, 2009.

[52] N. Holighaus, Z. Průša, and P. L. S. ndergaard, "Reassignment and synchrosqueezing for general time-frequency filter banks, subsampling and processing," *Signal Processing*, vol. 125, no. Supplement C, pp. 1 – 8, 2016.
[53] Z. Průša and N. Holighaus, "Non-iterative filter bank phase (re)construction," in *Proc. 25th European Signal Processing Conference (EUSIPCO–2017)*, Aug 2017, pp. 952–956.