2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)

# Acceleration of Audio Inpainting by Support Restriction

Pavel Rajmic*, Hana Bartlová†, Zdeněk Průša‡ and Nicki Holighaus‡
* SPLab, Brno University of Technology, Czech Republic, rajmic@feec.vutbr.cz
† Faculty of Mechanical Engineering, Brno University of Technology, Czech Republic
‡ Acoustics Research Institute, Austrian Academy of Sciences, Austria

***Abstract*—We present a simple algorithm which accelerates the sparsity-based audio inpainting. The algorithm optimally restricts the signal support around the missing data region. This way, increased computational efficiency is achieved by avoiding inclusion of unnecessary values in the optimization process. For testing purposes, we use the discrete Gabor transform as the sparsity promoting representation, but the method can be easily translated to other systems.**

## I. Introduction

In signal transmission or restoration applications, we are often tasked with the restoration of lost samples. If inference of this information from the rest of the signal is attempted, then the restoration process is referred to as *audio inpainting*. Although the distribution of missing samples could be general, in practice we mostly meet situations when a compact, connected part of the signal is missing (or corrupted so severely that it must be deleted, see for example [1]). The missing segment will be referred to as the *gap*.

Traditional approaches (see [2] and references therein) start from the observation that the signal can be usually locally modelled by an autoregressive (AR) process. Alternatively, the signal is decomposed into sinusoidal components which are individually interpolated into the interior of the gap. The behaviour of the sinusoidal components can be parametrized by an AR process, which allows inpainting sounds with tremolo or vibrato [3], [4]. This approach is successful for even very long gaps, but naturally holds only for gaps with a clear and stable harmonic structure around. Another class of methods, inspired by image inpainting, is exemplar-based, i.e. the signal is scanned for a segment resembling the part around the gap. The segment is then used to restore the missing information [5].

Yet another approach, which we investigate in this contribution, is based on sparsity [6]; it is assumed that an audio signal can be (approximately) represented by a relatively few coefficients in a properly chosen *synthesis* dictionary. Time-frequency (TF) representations like MDCT, STFT, Gabor transforms or constant-Q transforms are the dictionaries which have this property [7], [8], [9].

The idea behind the sparsity-based methods is to estimate the sparse coefficients of a signal from the reliable samples to recover the missing samples. Formally, we wish to solve the following optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \ \text{ subject to } \ \|\mathbf{y}^{\mathrm{r}} - \mathbf{D}^{\mathrm{r}}\mathbf{x}\|_2 \leq \delta. \qquad (1)$$

Finally, the solution $\hat{\mathbf{x}}$ is simply used as the synthesis coefficients for recovery. Here, $\|\mathbf{x}\|_0$ counts the nonzero elements of $\mathbf{x}$, $\mathbf{y}^{\mathrm{r}}$ represents the reliable samples of signal $\mathbf{y}$, $\mathbf{D}^{\mathrm{r}}$ is the corresponding part of the full dictionary $\mathbf{D}$, and $\delta$ is an allowed model error. See Fig. 1.

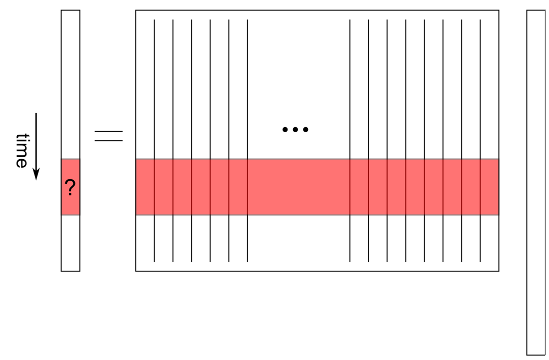


Fig. 1. Illustration of synthesis modelling of the signal as $\mathbf{y} = \mathbf{D}\mathbf{x}$, and its reliable samples, the subset $\mathbf{y}^{\mathrm{r}} = \mathbf{D}^{\mathrm{r}}\mathbf{x}$, respectively.

The problem above is NP-hard. Therefore, alternative algorithms have been proposed, providing approximate solutions. Examples include greedy algorithms [10], [11] or methods based on convex relaxation [12], [13], [14].

Sparse regression, although reasonably efficient, takes a substantial amount of time. The works referenced above introduce new concepts, and their software (if available at all) serves as a proof of this concept, which is why the length of signal around the gap is not considered an issue, or they use a simple overlap-add segmented approach. The manuscript at hand aims at accelerating the computation by a simple form of preprocessing.

### A. Basic idea

We argue that, in relation to (1), only a small time-frequency region in the neighborhood of the gap contains information relevant for the inpainting process, and we demonstrate how to select the minimal region extending around the gap such that the goal is achieved. This way, (1) alters to a new optimization problem

$$\hat{\mathbf{x}}_{\mathrm{R}} = \arg\min_{\mathbf{x}_{\mathrm{R}}} \|\mathbf{x}_{\mathrm{R}}\|_0 \ \text{ subject to } \ \|\mathbf{y}^{\mathrm{r}}_{\mathrm{R}} - \mathbf{D}^{\mathrm{r}}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}}\|_2 \leq \delta_{\mathrm{R}}. \qquad (2)$$

involving *restricted* versions of the signal $\mathbf{y}^{\mathrm{r}}_{\mathrm{R}}$, dictionary $\mathbf{D}^{\mathrm{r}}_{\mathrm{R}}$, and coefficients $\mathbf{x}_{\mathrm{R}}$.

Time-frequency transforms are constructed from one or more time- and frequency-localized windows and their translations. Our algorithm takes advantage of the fact that the optimization (1) searches for sparse coefficients for a "long" $\mathbf{y}^{\mathrm{r}}$, while only few coefficients contribute to the gap being recovered. Indeed, in the final synthesis $\hat{\mathbf{y}} = \mathbf{D}\hat{\mathbf{x}}$, only atoms with nonzero overlap with the gap are *relevant*.

Although our idea holds also for any TF system with compactly supported windows, we demonstrate our method using the discrete Gabor transform (DGT) [15], [16]. The DGT of finite-length signals is usually considered a periodic transform, i.e. there are atoms supported at both ends of the signal at the same time, see Fig. 2. Such windows, while necessary, carry useless information and they should be ignored.

In order to minimize computational effort, we propose to compute the DGT from a restricted signal with as few samples around the gap as possible, while at the same time obtaining identical TF coefficients in the *relevant* segment, illustrated in Fig. 2.

Notice that saying this virtually corresponds to the *analysis* sparse modelling [6], [17], [18], in contrast to the *synthesis* model (1). However, since we use tight frames, in particular tight Gabor frames (see below), the supports of the analysis and synthesis windows are identical, and this allows us to seamlessly use our restriction algorithm in both fomulations, and to switch between DGT analysis and synthesis during the following.

### *B. Setup and notation*

Vectors are indexed starting from one. For simplicity, we assume only a single gap in the signal, which is surrounded by a sufficient number of signal samples. We denote the length of $\mathbf{y}$, the original, corrupted signal, by $L$.

Various implementations of the DGT are available; we use the LTFAT Matlab/Octave toolbox [19], [20], [21], which utilizes the following definition. The DGT of a signal $\mathbf{y}$ of length $L$ is a set of TF coefficients computed as

$$x_{m,n} = \sum_{k=1}^{L} \mathbf{y}_k \, \mathrm{e}^{-\mathrm{j}2\pi \frac{m-1}{M}(k-1)} \, \mathbf{g}_{((k-1-(n-1)a) \bmod L)+1} \quad (3)$$

for $n = 1, \ldots, L/a$ and $m = 1, \ldots, M$. Here, $\mathbf{g}$ is the window, usually assumed to be real and symmetric with a good TF concentration [22] and $a, M \in \mathbb{N}$ are the time shift and number of uniformly distributed frequency channels, respectively. The set of vectors $\left\{ \mathrm{e}^{\mathrm{j}2\pi \frac{m-1}{M}(k-1)} \, \mathbf{g}_{((k-1-(n-1)a) \bmod L)+1} \right\}$ might form a frame for $\mathbb{C}^L$, when proper $\mathbf{g}, a, M$ are chosen [15], [16]. If this frame is even tight, which will be also our case, these vectors comprise the columns of matrix $\mathbf{D}$.

Notice that the DGT is computed such that effectively the signal is periodic. This is the reason why the signal length $L$ should be a multiple of shift $a$ — otherwise, the windows would not be uniformly spaced. Similarly, $L$ is usually required to be a multiple of $M$ — otherwise, a phase discontinuity would appear at the boundary. Therefore, the processed signal is usually padded with zeros to a new length, divisible by both $a$ and $M$. Notice that the possible phase jump is actually not an issue in our case, since the boundary windows are never used to recover the signal in the gap, and this is the reason why divisibility by $M$ could be ignored.

In the following, the symbol $w$ will be used to represent the *effective length* of the window, i.e. the length of its support. First of all, we assume $w \ll L$, whereby $w$ can be both even or odd. We regard $\left\lfloor \frac{w}{2} \right\rfloor + 1$ as the "central" index of the window, the support of which extends $\left\lfloor \frac{w}{2} \right\rfloor$ entries to the left from the central index, and $\left\lceil \frac{w}{2} \right\rceil - 1$ entries to the right of it. If $w$ is odd, the central entry lies indeed in the center of the window; for $w$ even this entry is right-biased by one sample. The window shift $a = 0$ producing the first set of DGT coefficients (in the following informally referred to as "the first window") is placed in a way that its central entry is at position 1. This window thus ends at signal index $\left\lceil \frac{w}{2} \right\rceil$. Generally, a window numbered $n+1$ has its central entry at index $1 + na$, its first entry at index $1 + na - \left\lfloor \frac{w}{2} \right\rfloor$, and its last entry at index $na + \left\lceil \frac{w}{2} \right\rceil$. Of course, all these values are considered in $L$-modular arithmetics due to the periodicity.

The symbols $s$ and $f$ denote the indices of start and the end of the gap within the *original* signal, respectively. The indexes $q, Q$ correspond to the first and the last index, respectively, of the desired shorter signal within the original signal. The indexes $p, P$ similarly denote the central indices of the first/last window having nonempty overlap with the gap. The numbers $S = (p-1)/a + 1$ and $F = (P-1)/a + 1$ reflect the order (or, translation number) of these two windows within the DGT of the original signal. The symbols $u, v, U, V$ have analogous meaning to $s, f, S, F$, but related to the *shortened* signal. Fig. 2 presents the symbols at hand.

## II. Derivation of algorithm

It is clear that the presented quantities are related to each other via the parameter $a$, such that for particular $k, l, m \in \mathbb{N}$ it holds $p = q + ka$, $P = q + la$, and $Q = q + ma$. We can derive the complete algorithm step-by-step thanks to this basic observation.

**Find the central index $p$ of the first window (from left) such that it overlaps with the gap.** Since the central index of the first window is placed at 1, we can treat this problem by virtually shifting the window by $a$ repeatedly, until its rightmost index overlaps with the gap. This means that we seek $k$ as low as possible, such that

$$p + \left\lceil \frac{w}{2} \right\rceil - 1 \geq s \quad \Leftrightarrow \quad 1 + ka + \left\lceil \frac{w}{2} \right\rceil - 1 \geq s,$$

which means $k \geq \left(s - \left\lceil \frac{w}{2} \right\rceil\right)/a$. Such a $k$ is related to the number of the window within the DGT, and it holds $S = \left\lceil \left(s - \left\lceil \frac{w}{2} \right\rceil\right)/a \right\rceil + 1$. Therefore, the desired $p \leftarrow 1 + (S-1)a$.

**Find the first index of the restricted signal, $q$.** In order to ensure that the first useful window is not periodized in the DGT (otherwise it would carry possibly wrong information), $q$ must not lie right of the leftmost index of the first useful window. Formally, we seek $k$ as low as possible, such that

$$q \leq p - \left\lfloor \frac{w}{2} \right\rfloor \quad \Leftrightarrow \quad p - ka \leq p - \left\lfloor \frac{w}{2} \right\rfloor.$$

Therefore $q \leftarrow p - \left\lceil \left\lfloor \frac{w}{2} \right\rfloor / a \right\rceil \cdot a$.

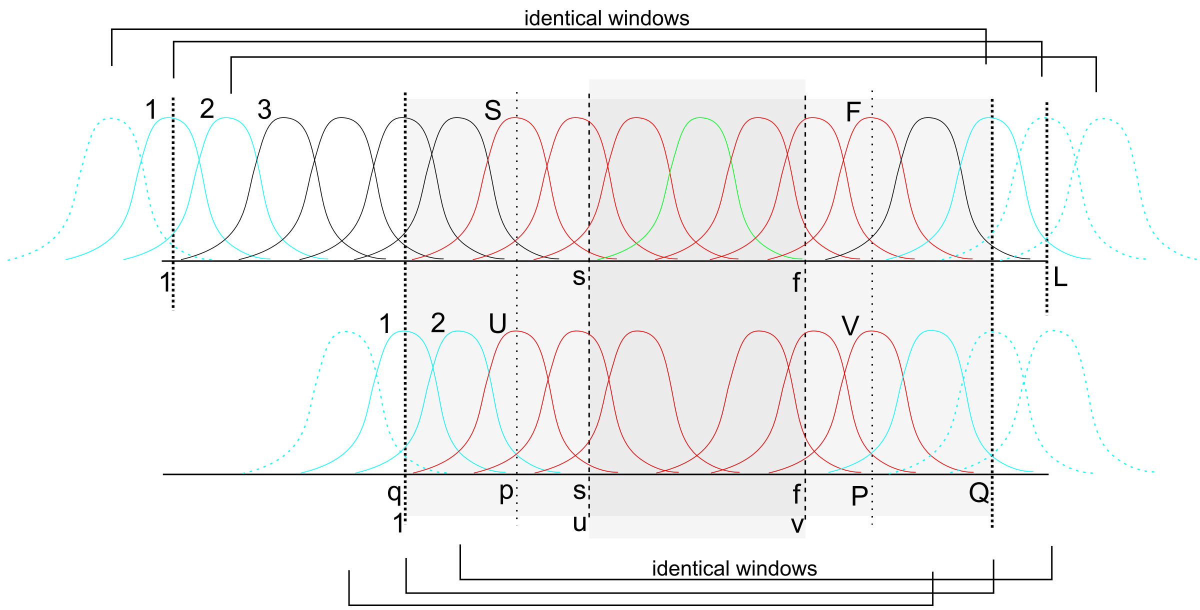


Fig. 2. Illustration of the restriction algorithm: Original longer piece of signal (top) with the gap at positions $s, f$. Individual window shifts of the related DGT are depicted schematically. Minimum-length excerpt of the signal (bottom). Only windows marked in red are relevant and will be used for data fitting and regularization. Cyan windows are corrupted with circular information, but must be present due to the periodicity of the DGT. They are irrelevant for the recovery, as are the black windows. The green window is completely contained in the gap and therefore does not carry any information. The plots above correspond to selecting $\mathit{offset} = 0$, see Sec. 2. The bottom graph contains two-fold axis: the first axis refers to the original setup, while the values at the second one are related to the new, shorter signal.

**Find the central index $P$ of the final window overlapping with the gap.** Starting from $p$, we look for the maximal shift by a multiple of $a$, such that the window has nonzero overlap with the gap. In other words, we seek the greatest $k$ for which

$$P - \left\lfloor \frac{w}{2} \right\rfloor \leq f \quad \Leftrightarrow \quad p + ka - \left\lfloor \frac{w}{2} \right\rfloor \leq f.$$

Such an optimal $k$ is again connected to the number $F$ of the window within the original DGT, and it holds $F = S + \left\lfloor (f + \left\lfloor \frac{w}{2} \right\rfloor - p)/a \right\rfloor$. Our $P$ is then $P \leftarrow p + (F - S)a$.

**Find the last index of the restricted signal, $Q$.** The index $Q$ is obtained in a similar fashion as $q$, i.e. the last useful window must not be periodized in the transform. Since the last window of the DGT coincides with the first one, it is not the case that the central window entry should be placed at index $Q$, rather the left neighbour of the central entry, i.e. we have $Q = P + ka - 1$ for some $k$. This virtual last window is marked by dots in Fig. 2. The rightmost entry of the last useful window lies at index $P + \left\lceil \frac{w}{2} \right\rceil - 1$, which leads to

$$Q \geq P + \left\lceil \frac{w}{2} \right\rceil - 1 \quad \Leftrightarrow \quad P + ka - 1 \geq P + \left\lceil \frac{w}{2} \right\rceil - 1,$$

i.e. $k \geq \left\lceil \frac{w}{2} \right\rceil / a$. Therefore, $Q \leftarrow P + \left\lceil \left\lceil \frac{w}{2} \right\rceil / a \right\rceil \cdot a$.

Optionally, one can check the divisibility of the new signal length, $Q - q + 1$, by $a, M$ (the reason has been explained above) and, in the negative case, one increases $Q$ to fulfill this requirement. If this new index $Q$ would be $Q > L$, the signal could be simply padded by zeros; note that this padding does not affect the useful coefficients.

Now, restricting the signal to the range from $q$ to $Q$ and computing the DGT results in the desired situation that all useful coefficients from the original DGT and from the restricted DGT coincide.

It could happen that $q < 1$ or $Q > L$, which means that there is not enough signal samples even in the original signal to compute the DGT. As long as at least $p - \left\lfloor \frac{w}{2} \right\rfloor \geq 1$ and $P - 1 + \left\lceil \frac{w}{2} \right\rceil \geq L$, zero values can be used at those positions without having any effect on the useful coefficients. There will be rare cases when $\mathit{offset}$ will help to satisfy the above inequalities, however in the remaining cases, the DGT performed of the signal around the gap will automatically introduce errors due to the transform's periodicity.

The position of the gap within the shortened signal can be easily computed as $u = s - q + 1$ and $v = f - q + 1$. Similarly, the order of the first and the last window (in the original signal they are $S$ and $F$) are $U = (p - q)/a + 1$ and $V = U + (F - S)$ for the shortened signal.

It may happen that the first or the last useful window of the restricted DGT has too small or too large overlap with the gap in relation to the signal recovery, see for example window $U$ in Fig. 2. We can shift all the transform to achieve a more beneficial layout of the windows. In our algorithm, this is done via choosing an offset parameter, which is by default set to $\mathit{offset} = 0$. A nonzero $\mathit{offset}$ clearly breaks the coefficients coincidence (which is no harm in practice) and the meaning of ordinals $S$ and $F$ gets lost. Notice that the overlaps of the gap with the windows can be simply deduced before the actual DGT is performed, and thus suitable $\mathit{offset}$ can be decided before the restricting algorithm is run.

The last item of the algorithm consists of computing vector $\mathit{overlap}$, gathering information about the lengths of overlaps, i.e. the number of samples which are shared by the gap and the windows. We can restrict ourselves to the useful windows

**Algorithm** (Finding the shortest restriction of the signal around the gap).
Input: $w, a, M, s, f$ (optionally *offset* $\in [0, w-1]$)
Output: $q, Q, p, P, S, F, u, v, U, V$

1) $S \leftarrow \left\lceil \left(s - \left\lceil \frac{w}{2} \right\rceil\right) / a \right\rceil + 1$
2) $p \leftarrow 1 + (S-1) \cdot a$
3) $p \leftarrow p + \mathit{offset}$
4) $q \leftarrow p - \left\lceil \left\lfloor \frac{w}{2} \right\rfloor / a \right\rceil \cdot a$
5) $F \leftarrow S + \left\lfloor (f + \left\lfloor \frac{w}{2} \right\rfloor - p)/a \right\rfloor$
6) $P \leftarrow p + (F - S) \cdot a$
7) $Q \leftarrow P + \left\lceil \left\lceil \frac{w}{2} \right\rceil / a \right\rceil \cdot a$
8) If $aM \nmid (Q - q + 1)$:
   $Q \leftarrow Q + a - [(Q - q + 1) \bmod \operatorname{lcm}(a, M)]$
9) $u \leftarrow s - q + 1, \quad v \leftarrow f - q + 1$
10) $U \leftarrow (p - q)/a + 1, \quad V \leftarrow U + (F - S)$
11) Determine the overlaps of the gap with individual windows, *overlap*.

Fig. 3. The resultant algorithm.

only, since for the others this value is zero. Therefore, *overlap* contains $F - S + 1$ values (for *offset* $\neq 0$, this number could be different!). A simple trick can be used for generating the vector: We construct an "indicator" signal of the restricted length $Q - q + 1$, i.e. the signal that takes value one at all indexes inside the gap (from $u$ to $v$) and zero otherwise. We then convolve it with the rectangular window ($w$ ones). Subsampling the convolution with factor $a$ leads to a vector carrying the desired values.

## III. Experiments and Software

The first simple experiment serves as a proof of concept of the algorithm. We take a short excerpt from the 'greasy' speech signal [20], [21]. This piece is 1100 samples long at 16 kHz sampling rate. We consider a virtual gap, starting at position 500, and 80 samples long. We compute the optimum restriction using the Algorithm in Fig. 3, using Matlab function `min_sig_supp.m`. The input values are $w = 256$ (Hann window), $a = 64$, $M = 64$, $s = 500$, $f = 580$, *offset* $= 0$. We receive output values $q = 257$, $Q = 833$, which means that the signal can be shortened by ca 50 % with no effect on the samples in the gap. We show the DGT spectrograms of the original and the shortened signal in Fig. 4, produced by `show_coefs_coincide.m`. From the total 18 vertical sets of Gabor coefficients, sets $S = 7$ to $F = 12$ in the original and $U = 3$ to $V = 8$ in the new coefficients coincide, which is marked by dashed boxes. The other coefficients do not have overlap with the intended gap. Actually, the overlaps with the gap can be computed by `overlap.m` and they are, from left to right: 13, 77, 81, 81, 68, 4 samples.

The Matlab files can be downloaded from URL http://www.utko.feec.vutbr.cz/~rajmic/software/accelerate_audio_inpaint.zip, and they require LTFAT [21] to be installed.

Next, we performed an experiment showing the ability of the algorithm to accelerate the inpainting process. First, the original corrupted signal of length 64 000 samples (4 s) has been inpainted. The gap size was 320 samples (20 ms). A tight Gabor system based on Hann window with $a = 683$ and $M = w = 2049$ was used. The sparse coefficients were computed by $\ell_1$-relaxation of (1) via Douglas-Rachford splitting

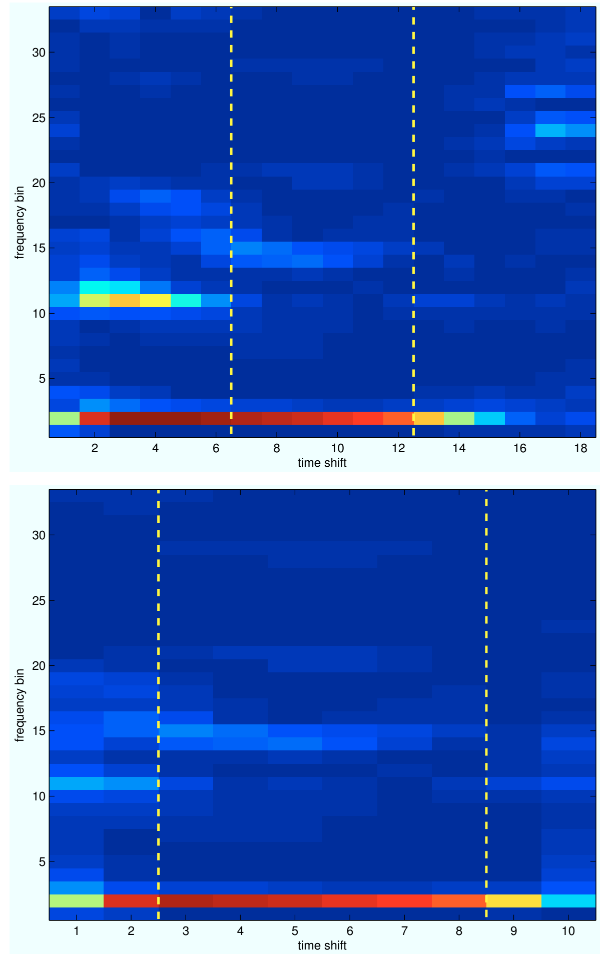


Fig. 4. Spectrograms of the original (above) and the shortened signals, respectively. They share the colormap and scale. Since the input signal is real, just the non-negative frequency bins are shown. Moduli of the DGT coefficients are displayed, nevertheless their phases are identical in the marked region as well.

[23]. This took 16 s on an ordinary PC with Intel Core2 Quad CPU @ 2.83 GHz. Then, the same gap was inpainted using the optimally restricted signal, which was only 4782 samples long. The optimization parameters has been tuned such that both experiments resulted in the same SNR. This way, the sparse regression took only 2 s. This experiment was performed using our audio-inpainting toolbox.

## IV. Discussion and Outlook

It is now clear how we turn to minimizing $\|\mathbf{x}_{\mathrm{R}}\|_0$ in (2) instead of solving the original problem (1). In fact, we note that sparsity of even the *subset* of $\mathbf{x}_{\mathrm{R}}$ should be minimized; it is the central subset, denoted $\mathbf{x}_{\mathrm{RR}}$, containing only the useful DGT window positions:

$$\hat{\mathbf{x}}_{\mathrm{R}} = \arg\min_{\mathbf{x}_{\mathrm{RR}}} \|\mathbf{x}_{\mathrm{RR}}\|_0 \text{ subject to } \|\mathbf{y}_{\mathrm{R}}^{\mathrm{r}} - \mathbf{D}_{\mathrm{R}}^{\mathrm{r}} \mathbf{x}_{\mathrm{R}}\|_2 \leq \delta_{\mathrm{R}}. \quad (4)$$

The remaining coefficients are not restricted since the periodic processing corrupts the coefficients and possibly destroys their sparse structure (compare coefficient strips adjacent to the rectangle borders in Fig. 4!). They do not influence the recovery in the gap. This is however not to say that the minimizers of

(1), (2) and (4) will be the same. The solutions are dependent on the content of $\mathbf{y}^{\mathrm{r}}$, on $\mathbf{D}^{\mathrm{r}}$, and on constants $\delta, \delta_{\mathrm{R}}$, and one can hardly hope for a simple relation between them, especially when $\ell_1$-relaxation is used.

Structured sparsity [17], [12], [14] has been evidenced to be beneficial for audio inpainting. Here, not the individual coefficients are sparse, but horizontal groups reflecting the harmonic structure of a sound are formed, and treated as sparse across *groups*. In such a case, our algorithm could be easily modified to accommodate this approach.

The algorithm could be also adopted to the case of multiple gaps or even more general layout of the missing samples. If two missing sample segments are well enough separated, each such gap can be treated individually. Actually, the proposed algorithm can be safely used on the gaps separately when the distance between the gaps' borders is greater or equal than the window size, $w$. In case when the distance is still at least greater or equal to $w - 1 - a$ the gaps could be separated if a proper $\mathit{offset}$ value is used. If pieces of missing samples are not sufficiently separated, then a longer signal segment can be cut using our algorithm, containing the full region of nonseparated gaps. However, if the treated signal misses entries whose distance is continuously below the limit, the only possibility is to use a blockwise overlap-add mechanism, like in [10].

A problem related to inpaiting is the so-called audio declipping [11], [14], [24], [25]. This is actually a better-conditioned problem than inpainting, since one has more prior information about the signal being restored. Results of this article can be seamlessly transferred to the sparsity-based declipping task, while keeping in mind that the spacing between clipped samples is usually low, which limits the use of the proposed algorithm for declipping.

## Acknowledgment

Research described in this paper was financed by the Czech National Sustainability Program under grant LO1401. For the research, infrastructure of the SIX Center was used. This work was also supported by the Austrian Science Fund (FWF) START-project FLAME (Frames and Linear Operators for Acoustical Modeling and Parameter Estimation; Y 551-N13).

## References

[1] V. Mach, "Digital restoration of recordings from the phonograph cylinders and their copies," in *As recorded by the phonograph: Slovak and Moravian songs recorded by Hynek Bím, Leoš Janáček and Františka Kyselková in 1909–1912*. Brno: The Institute of Ethnology of the Academy of Sciences of the Czech Republic, v.v.i., 2012, pp. 165–176.

[2] M. Fink, M. Holters, and U. Zölzer, "Comparison of Various Predictors for Audio Extrapolation," in *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13)*, Maynooth, 2013, pp. 1–7. [Online]. Available: http://dafx13.nuim.ie/papers/42.dafx2013_submission_27.pdf

[3] M. Lagrange, S. Marchand, and J.-b. Rault, "Long interpolation of audio signals using linear prediction in sinusoidal modeling," *J. Audio Eng. Soc*, vol. 53, no. 10, pp. 891–905, 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13390

[4] A. Lukin and J. Todd, "Parametric interpolation of gaps in audio signals," in *Audio Engineering Society Convention 125*, Oct 2008, pp. 3–6. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14664

[5] Y. Bahat, Y. Y. Schechner, and M. Elad, "Self-content-based audio inpainting," *Signal Processing*, vol. 111, no. 0, pp. 61–72, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168414005623

[6] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[7] 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 AES international conference on semantic audio*, London, UK, January 2014.

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

[9] G. A. Velasco, N. Holighaus, M. Dörfler, and T. Grill, "Constructing an invertible constant-Q transform with non-stationary Gabor frames," in *Proc. of the 14th Int. Conference on Digital Audio Effects DAFx11*, Paris, 2011, pp. 93–99. [Online]. Available: http://recherche.ircam.fr/pub/dafx11/Papers/47_e.pdf

[10] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. Plumbley, "Audio Inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, March 2012.

[11] ——, "A constrained matching pursuit approach to audio declipping," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 329–332.

[12] C. Kereliuk, P. Depalle, and P. Pasquier, "Audio interpolation and morphing via structured-sparse linear regression," in *Proceedings of the Sound and Music Computing Conference 2013*, Stockholm, 2013, pp. 546–552. [Online]. Available: http://tinyurl.com/kereliuk-structured

[13] D. Barchiesi, "Sparse approximation and dictionary learning with applications to audio signals," Ph.D. dissertation, Queen Mary University of London, 2013.

[14] K. Siedenburg, M. Kowalski, and M. Dorfler, "Audio declipping with social sparsity," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1577–1581.

[15] O. Christensen, *Frames and Bases, An Introductory Course*. Boston: Birkhäuser, 2008.

[16] K. Gröchenig, *Foundations of time-frequency analysis*. Birkhäuser, 2001.

[17] I. Bayram and D. Akykıldız, "Primal-dual algorithms for audio decomposition using mixed norms," *Signal, Image and Video Processing*, vol. 8, no. 1, pp. 95–110, 2014.

[18] S. Kitić, N. Bertin, and R. Gribonval, "Audio declipping by cosparse hard thresholding," in *2nd Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

[19] P. L. Søndergaard, B. Torrésani, and P. Balazs, "The Linear Time Frequency Analysis Toolbox," *International Journal of Wavelets, Multiresolution Analysis and Information Processing*, vol. 10, no. 4, 2012.

[20] Z. Průša, P. Søndergaard, P. Balazs, and N. Holighaus, "LTFAT: A Matlab/Octave toolbox for sound processing," in *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)*, Laboratoire de Mécanique et d'Acoustique. Marseille, France: Publications of L.M.A., October 2013, pp. 299–314.

[21] P. L. Søndergaard. (2013) LTFAT webpage. URL: http://ltfat.sourceforge.net.

[22] A. Nuttall, "Some windows with very good sidelobe behavior," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 29, no. 1, pp. 84–91, 1981.

[23] P. Combettes and J. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212, 2011.

[24] S. Kitic, L. Jacques, N. Madhu, M. Hopwood, A. Spriet, and C. De Vleeschouwer, "Consistent iterative hard thresholding for signal declipping," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 5939–5943.

[25] S. Kitić, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: A flexible non-convex approach," in *LVA/ICA 2015, LNCS 9237*, E. V. et al., Ed. Springer International Publishing Switzerland, 2015, pp. 243–250.