Sparse Frequency Extrapolation of Spectrograms

Jabran Akhtar and Karl Erik Olsen
Norwegian Defence Research Establishment (FFI)
Box 25, 2027 Kjeller, Norway
Email: jabran.akhtar@ffi.no, karl-erik.olsen@ffi.no

***Abstract*—The short-time Fourier transform is a prevalent method used to analyze the frequency composition of a signal as a function of time. In order to achieve high resolution in frequency a large sliding window needs to be applied which degrades the time resolution. This paper proposes the adoption of sparse reconstruction as a mean to extrapolate supplementary values in time domain for each segment. Over short durations a signal's frequency content is likely to contain a limited number of effective frequencies and a sparse regeneration approach can be advantageous as an extrapolating mechanism. An enlarged number of samples can thus yield spectrograms with high frequency resolution. The capabilities of the proposed techniques are demonstrated on several synthetic and real data signals.**

***Index Terms*—spectrogram, short-time Fourier transform, time-frequency analysis, super-resolution, signal extrapolation**

I. Introduction

The short-time Fourier transform (STFT) is one of the classical and widely used techniques to determine the frequency content of a signal over shorter time intervals. A limited section of a signal is extracted through the means of a sliding window, multiplied by a tapering function and transformed to the Fourier domain. Plotting the magnitude of the transform as a function of time provides a spectrogram on how the frequencies of the signal vary as a function of time [1]. An important aspect in this regard is the frequency resolution of the spectrogram which in essence is determined by the number of data samples available for the transform. Employing a larger window with a larger set of samples increases the frequency resolution at the expense of poorer time localization. Obtaining high accuracy in time as well as high frequency resolution is therefore generally contradictory. Nevertheless, there have been several approaches presented in the literature on how to improve upon the traditional STFT by using various alternative time-frequency representations and transformations [2]–[6].

The last couple of years have also seen an increased focus on compressed sensing and sparse reconstruction techniques based on the L_1-norm optimization [7], [8]. These methods have typically been developed for use in various settings to reconstruct a signal or image where the data acquisition may have been carried out in a compressed or irregular manner. Under certain conditions a sparse reconstruction approach can guarantee perfect recovery even when parts of data may not be available. In many applications, such as audio recordings, sampling and data collection in itself is not really a bottleneck issue rather detailed and accurate signal analysis is the more prominent aspect.

In this paper we propose an alternative utilization of sparse reconstruction techniques for increasing the frequency resolution of the STFT. The Fourier transform with a sliding window is retained as the main transformation component and the frequency resolution is increased by a sparse extrapolation process in time domain. This provides an alternative resolution enhancement strategy compared to previously proposed approaches.

The motivation for selecting sparse reconstruction in this context comes from the fact that a frequency transform over a restricted number of samples is likely to contain a limited number of preeminent frequency elements and can thus be considered to be sufficiently sparse. Sparse reconstruction procedures may therefore be employed to extrapolate the dominant signal oscillations appropriately and narrow them down more precisely in Fourier domain. Another incentive for extrapolation comes from the fact that applying a windowing function weights down data entries at the beginning and end of the sequence. In an enlarged extrapolated data set the extrapolated values are the ones who are heavily scaled down while the potential of the original data can be utilized fully. The idea of extrapolating each segmented signal to increase its time-frequency resolution has also been proposed earlier in e.g. [2], [9] using more standard signal modeling and weighted norm approaches. Some recent applications of sparse reconstruction techniques with inter- and extrapolations have been applied in phase array antennas [10] and in multifunction radar [11] settings to compensate for data gaps and improve overall system performance.

A particular feature often linked with sparse reconstruction is that the obtained results are indeed sparse, i.e. contain significant number of exact zero values. In order to regenerate and retain the original properties of the spectrogram we additionally propose a merger of the extrapolated data with real data. This has the benefit that noise, less prevalent signal frequencies and other inaccuracies are fully preserved. Several simulated and real world examples are examined to demonstrate the principles introduced in this paper.

II. Spectrogram model

We consider a discrete-time signal $p(t) \in \mathbb{C}^{T\times 1},\ t = 1, ..., T$ sampled at regular intervals for which a time-frequency representation through STFT is desired. For this, a shorter segment $x(\hat{t})_k \in \mathbb{C}^{N\times 1},\ \hat{t} = 1, ..., N$ with N samples is extracted from within $p(t)$. It is assumed that $k = 1, ..., K$ and K denotes the total number of segmented intervals; the

ISBN 978-0-9928626-7-1 © EURASIP 2017

exact value of K depending upon the lengths of the signal and the segmented window alongside the amount of overlap between two consecutive sections. In the reminder of the text we use $\mathbf{x}_k \in \mathbb{C}^{N\times 1}$ to denote $x(\hat{t})_k$ for any particular value of k.

For further processing $\mathbf{x}_k$ is typically multiplied element-wise, designated by $\odot$, with a windowing function $\mathbf{w} \in \mathbb{C}^{N\times 1}$ and afterwards it may potentially also be zero-padded. Performing a Fourier transform results in $\mathbf{s}(\omega)_k$ expressed as:

$$\mathbf{s}_k = \mathbf{F}(\mathbf{w} \odot \mathbf{x}_k) \in \mathbb{C}^{N\times 1}. \tag{1}$$

$\mathbf{F}$ is the discrete Fourier matrix of size $N \times N$, $\mathbf{F}_{m,n} = \exp(-2\pi jmn/N)$. The above process is independent across each time segment and may be executed through an FFT to make it computationally more efficient. Stacking together all segments we arrive to the STFT matrix:

$$\mathbf{S}(k,\omega) = [\mathbf{s}_1 \ ... \ \mathbf{s}_K] \in \mathbb{C}^{N\times K}. \tag{2}$$

A. Sparse Extrapolation

Assuming each segmented spectrum $\mathbf{s}_k$ contains relative limited number of active frequencies, or is sufficiently sparse within a tolerance level, one can attempt to extrapolate it in time domain. An extrapolation process constructs additional samples from available data and in this proposition the procedure needs to materialize with respect to the main dominating frequencies as only the extrapolation of these frequencies can force the spectrum to remain sparse. The optimal solution will thus be the one that maximizes sparsity in frequency while still preserving, to a certain extent, the original signal's integrity, as specified later.

The new regenerated profile for a given segment is specified by $\hat{\mathbf{x}}_k \in \mathbb{C}^{L\times 1}$ and the relationship between time domain and frequency domain is as previously governed by

$$\hat{\mathbf{s}}_k = \hat{\mathbf{F}}(\hat{\mathbf{w}} \odot \hat{\mathbf{x}}_k) \in \mathbb{C}^{L\times 1} \tag{3}$$

where $\hat{\mathbf{F}}$ is now an $L \times L$ Fourier matrix and $\hat{\mathbf{w}} \in \mathbb{C}^{L\times 1}$ is the windowing function. L indicates the length of the entire extrapolated segment, where $L > N$ and is chosen freely. For simplicity we presume

$$Q = L - N \tag{4}$$

is an even number and expresses the total number of extrapolated samples. From the original segment, $Q/2$ number of samples are therefore extrapolated on both ends in time domain.

For the sparse reconstruction process we furthermore define a binary selection matrix $\mathbf{M} \in \mathbb{B}^{N\times L}$ by taking an $L \times L$ identity matrix $\mathbf{I}_{L\times L}$ and removing the first $Q/2$ and the last $Q/2$ rows. This eliminates the respective rows for which no samples are available. The purpose of the selection matrix is to extract values from positions that contain data.

The extrapolated and regenerated solution should have comparable values to those measured at their respective placements which can be expressed as

$$\mathbf{M}\hat{\mathbf{x}}_k = \mathbf{x}_k. \tag{5}$$

With windowing functions incorporated the requirement becomes

$$\mathbf{M}(\hat{\mathbf{w}} \odot \hat{\mathbf{x}}_k) = (\mathbf{M}\hat{\mathbf{w}}) \odot \mathbf{x}_k, \tag{6}$$

or equivalently

$$\begin{aligned} \mathbf{M}\,\hat{\mathbf{F}}^*\hat{\mathbf{F}}(\hat{\mathbf{w}} \odot \hat{\mathbf{x}}_k) &= (\mathbf{M}\hat{\mathbf{w}}) \odot \mathbf{x}_k \\ \mathbf{G}\,\hat{\mathbf{s}}_k &= \bar{\mathbf{w}} \odot \mathbf{x}_k, \end{aligned} \tag{7}$$

where $\bar{\mathbf{w}} = \mathbf{M}\hat{\mathbf{w}} \in \mathbb{C}^{N\times 1}$ is the truncated tapering function and $\mathbf{G}$ is the partial inverse Fourier matrix $\mathbf{G} = \mathbf{M}\hat{\mathbf{F}}^* \in \mathbb{C}^{N\times L}$ with $\hat{\mathbf{F}}^* \in \mathbb{C}^{L\times L}$ being the inverse Fourier matrix.

As the STFT can be presumed to be reasonably sparse for each segment, the optimal sparse solution $\hat{\mathbf{s}}_k$ must be found with respect to frequency domain. The extrapolating regenerating process can under convex relaxation therefore be formulated as

$$\hat{\mathbf{s}}_k = \arg\min ||\dot{\mathbf{s}}_k||_1 \tag{8}$$

$$s.t. \quad || \mathbf{G}\,\dot{\mathbf{s}}_k - \bar{\mathbf{w}} \odot \mathbf{x}_k \,||_2 \leq \varepsilon \tag{9}$$

where ε is acceptable error and $||\ ||_1$ indicates the L_1 norm. The constrain (9) is a relaxed version of (7) in order to accommodate for the presence of noise and other inaccuracies. (8) and (9) together form a standard sparse reconstruction problem where the selection of ε determines the nature of the solution. Generally, the tolerance level may be set relative to the average noise floor. A larger value can provide more flexibility in determining a sparse solution though it may then also deviate somewhat from the measured data set. On the other hand, a tighter constrain on ε forces the solution to be more closer to the original data which may retain peculiar properties including for example noise.

Stacking together the optimal solutions from each segment provides the regenerated STFT matrix with a greater number of bins in frequency:

$$\hat{\mathbf{S}}(k,\omega) = [\hat{\mathbf{s}}_1 \ ... \ \hat{\mathbf{s}}_K] \in \mathbb{C}^{L\times K}. \tag{10}$$

We remark that partial Fourier matrices have been well studied in the literature and have been shown to provide capable outcomes in sparse reconstruction applications where also several efficient algorithms have also been proposed on solving these types of problems [7], [12], [13]

B. Merged Extrapolation

For general purpose signal analysis a possible drawback with the extrapolated solution (10) is the inherit sparsity. By selecting an appropriate value for ε noise and other inaccuracies can be eliminated from the sparse solution which is also important as an extrapolation of e.g. noise is typically not desired. Nevertheless, for many algorithms and detailed spectrogram analysis the more subtle fluctuations and alterations within the original noisy signal may still remain of interest. A strategy to alleviate these issues is to first determine an extrapolated solution and then utilize the obtained extrapolated samples only for extensional purposes in time domain where the original signal data remains preserved unaltered in the

ISBN 978-0-9928626-7-1 © EURASIP 2017

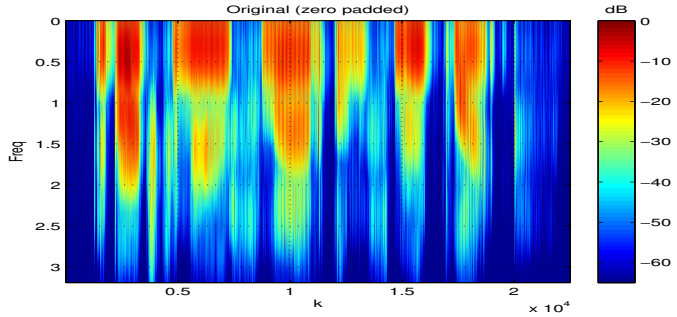


Fig. 1: Audio signal: standard spectrogram, $N = 16$

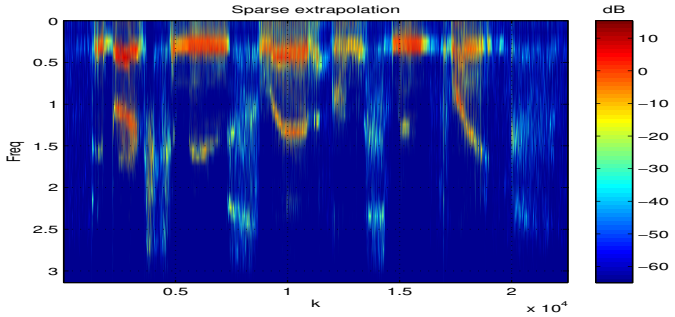


Fig. 2: Sparse extrapolated spectrogram, $N = 16, L = 64$

middle. This can be accomplished by transforming $\hat{\mathbf{s}}_k$ back to time domain

$$(\hat{\mathbf{w}} \odot \hat{\mathbf{x}}(t)_k) = \hat{\mathbf{F}}^* \, \hat{\mathbf{s}}_k \tag{11}$$

which is then merged with the original data incorporating the tapering function

$$(\hat{\mathbf{w}} \odot \tilde{\mathbf{x}}(t)_k) = \begin{cases} \bar{\mathbf{w}} \odot \mathbf{x}(t - Q/2)_k, & Q/2 < t < L - Q/2 \\ \hat{\mathbf{w}} \odot \hat{\mathbf{x}}(t)_k, & otherwise \end{cases} \tag{12}$$

where t now runs through $t = 1, ..., L$. The time domain solution accordingly contains the original signal, windowed correspondingly, in the center. A Fourier transform

$$\tilde{\mathbf{s}}_k = \hat{\mathbf{F}} \, (\hat{\mathbf{w}} \odot \tilde{\mathbf{x}}(t)_k) \tag{13}$$

across all segments yields the final merged spectrogram:

$$\tilde{\mathbf{S}}(k, \omega) = [\tilde{\mathbf{s}}_1 \, ... \, \tilde{\mathbf{s}}_K] \; \in \mathbb{C}^{L \times K}. \tag{14}$$

The merged spectrogram permits usage of standard filtering, detection and classification algorithms who may otherwise require modifications for sparse spectrograms.

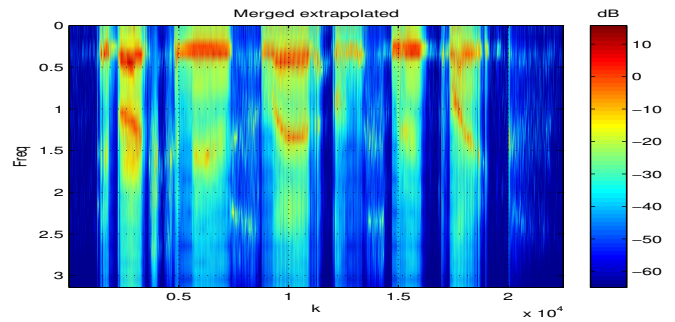


Fig. 3: Extrapolated merged spectrogram $N = 16, L = 64$

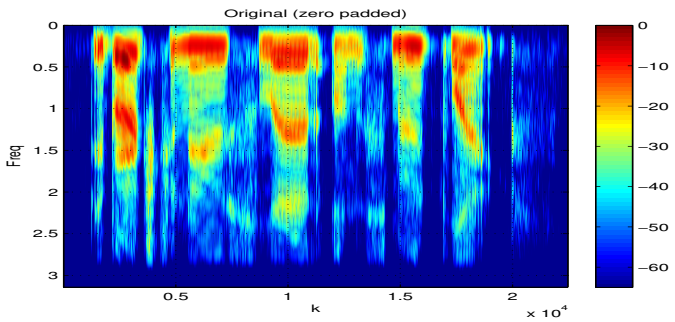


Fig. 4: Audio signal: standard spectrogram, $N = 64$

III. Results and Discussion

A. Audio test signal

To demonstrate the principles of the proposed sparse extrapolation approach, a clean audio recording of a male voice at 8 kHz was taken advantage from the freely available NOIZEUS database [14]. The samples were first run through the standard STFT with a window length of 16 samples and a hop size of only 1 sample between segments. This provides a versatile and large trial set of over 22500 STFT time bins. Each segment was tapered with the Hanning window.

The standard spectrogram of the original signal can be seen in figure 1 which was also zero-padded by 48 to bring the number of bins in frequency to 64. The limitations of a small window size are nevertheless very noticeable as the resolution in frequency is not sufficient to clearly separate the various components of the audio signal without a lot of smearing.

The spectrogram obtained through the sparse solutions, as per the procedure described in the previous section, with an extrapolation of 24 samples on each side is depicted in figure 2. The sparse reconstruction process was carried out using the SPGL1 [13] algorithm and with $\varepsilon = 0.05||\hat{\mathbf{w}}\mathbf{x}_k||$ to sanction a five percent norm deviation from the original extracted signal for each segment. All other parameters, including the length of the sliding window, are kept identical to those of the standard spectrogram. As one can observe in the figure the major features of the voice sample stand out and are now much more clearly located at specific frequency bands. The solution is also sufficiently sparse for association and audio analysis purposes. Note that extrapolation process contributes with additional integration gain and the power levels are given relative to the standard spectrogram. The merged solution combining real and extrapolated data is given in figure 3 which is now no longer sparse but due to extra samples offers a significant improvement over the original spectrogram in terms of frequency resolution. The convenience of having augmented extrapolated samples is substantial with more than 10dB.

Audio signals are commonly analyzed with various windowing lengths. The second case therefore inspects a window length of 64 pulses, still with a hop of 1 sample. 64 samples were extrapolated on each side for each segment by the sparse reconstructing process. The original plot, zero-padded

ISBN 978-0-9928626-7-1 © EURASIP 2017

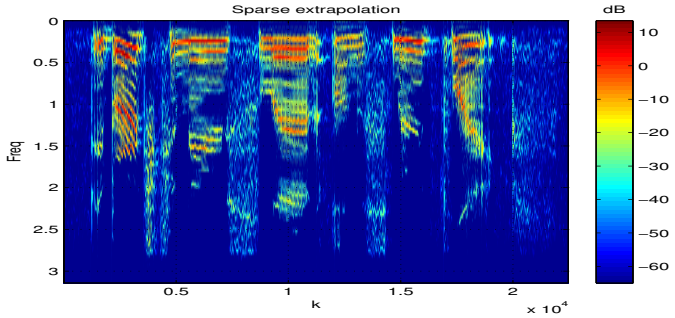


Fig. 5: Sparse extrapolated spectrogram, $N = 64, L = 192$

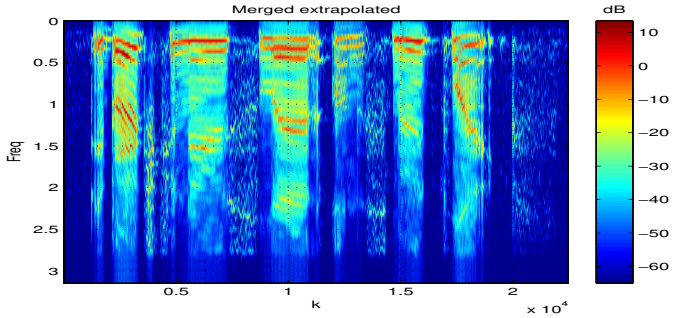


Fig. 6: Extrapolated merged spectrogram, $N = 64, L = 192$

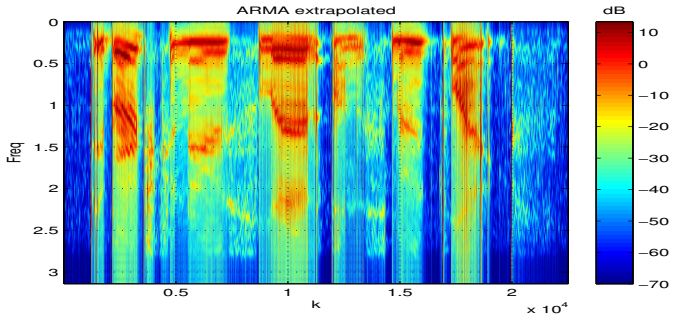


Fig. 7: ARMA extrapolated spectrogram, $N = 64, L = 192$

to length of 192 samples, can be seen in figure 4. This resembles the previously extrapolated figure 3 where many of the same frequencies stand out. The sparse extrapolation process (figure 5) correctly splits the various bands and a full discrimination is now possible. This remains the case even for the merged solution of figure 6. Subjectively, listening to the sparse or the merged sample sounds very similar to the original recording.

To compare the outcome against more traditional methods, each segment of the signal was extrapolated, on both ends, through two independent ARMA(10,40) processes using Prony's method [1]. A total of 64 additional samples were generated on each side, and the final spectrogram can be seen in figure 7. It can be observed that the model has not been able to divide the major frequency bands as successfully and there is marked leakage. For comparison, the standard spectrogram with 192 sliding window samples can be seen in figure 8.

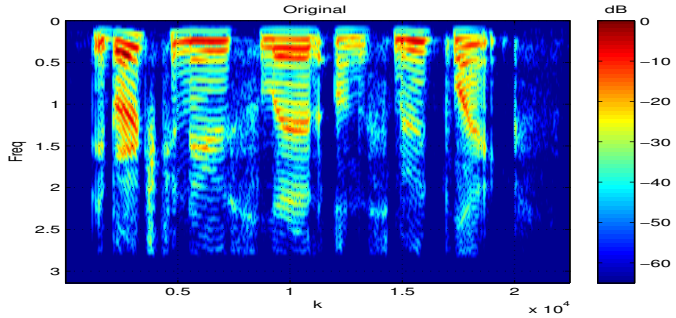


Fig. 8: Audio signal: standard spectrogram, $N = 192$

B. Noisy phonocardiogram signal

In order to investigate the performance under more demanding circumstances, a highly noisy simulation of a fetal phonocardiogram (PCG) recording sampled at 1 kHz with an SNR of -15.1dB was put to use [15]. The main objective to demonstrate that sparse reconstruction can also be highly useful in challenging noisy conditions where traditional extrapolation breaks down rapidly.

The original spectrogram with sliding Hanning window of 24 samples with a hop of 8 samples can be depicted in figure 9. The more abnormal properties as well as the low frequency heartbeats appear in the spectrogram though the latter is more easily observable in the magnified portion of the plot on the right side. The noise is otherwise quite dominating and the sparse reconstruction process must therefore take that into account as a highly sparse solution on it's own may not be able to capture all activity. A possible choice for ε can therefore be $\varepsilon = 0.5||\hat{\mathbf{w}}\mathbf{x}_k||$ to allow for an up to fifty percent norm disparity from each of the original segmented signals. The result of sparse extrapolation with this selection can be seen in figure 10 and the merged solution in figure 11. The outcome is not sparse as some noise is retained, except at positions enclosing high frequency anomalies. Otherwise, the main heart beating frequencies have clearly been enhanced and narrowed down in frequency. This is further evident in the hybrid spectrogram where the frequency spread is much cramped.

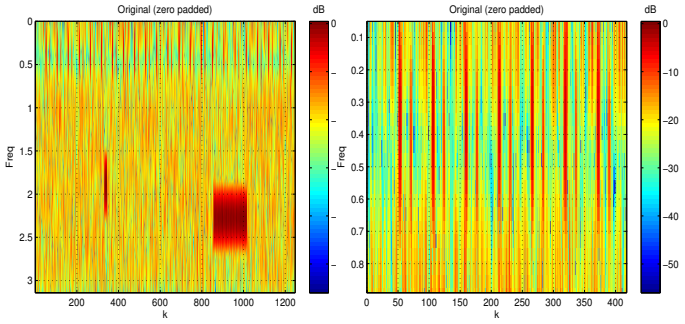
To realize more sparser images the acceptable error can be raised, albeit that may come at the expense of somewhat reduced sensitivity to the subordinate frequencies due to the very low SNR. Figures 12 and 13 illustrate this in practice where $\varepsilon = 0.95||\hat{\mathbf{w}}\mathbf{x}_k||$ is applied. The main features of the heart beats and the abnormalities are nevertheless preserved and stand out easily distinguishable even if the intensity levels are reduced.

Overall, the proposed sparse extrapolation and merger techniques have successfully managed to generate spectrograms with high frequency resolution using sliding window of the same length.

IV. Conclusion

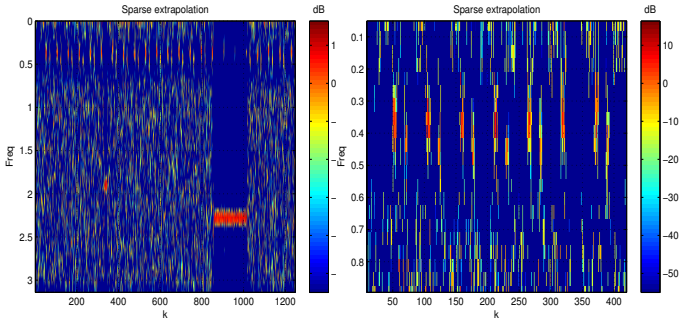
The short-time Fourier transform is an important tool in many signal processing applications and improving the time and frequency localization is of great interest. In this paper

ISBN 978-0-9928626-7-1 © EURASIP 2017



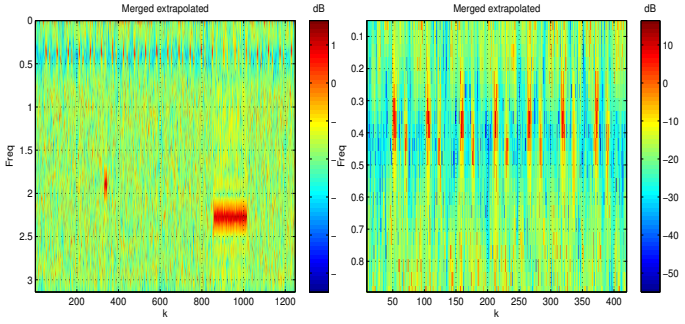
(a) Full (b) Magnified

Fig. 9: PCG: standard spectrogram, $N = 24$



(a) Full (b) Magnified

Fig. 10: Sparse extrapolated, $N = 24$, $\varepsilon = 0.5||\hat{\mathbf{w}}\mathbf{x}_k||$



(a) Full (b) Magnified

Fig. 11: Extrapolated merged, $N = 24$, $\varepsilon = 0.5||\hat{\mathbf{w}}\mathbf{x}_k||$

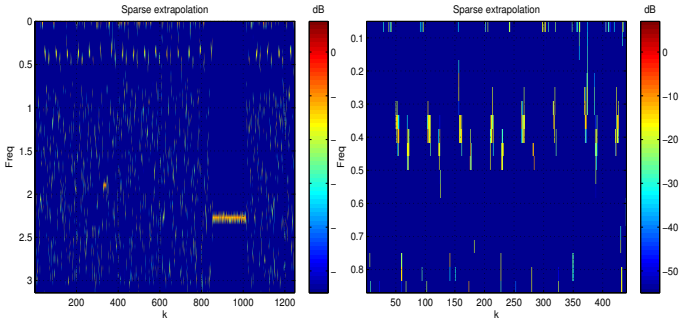
a technique based on sparse extrapolation of signals was proposed for this objective. Each segment of the signal is extrapolated in time in order to retrieve a sparse representation in frequency. This allows for a simple yet effective strategy to attain super-resolution sparse spectrograms. The extrapolated data may further be merged with the original statistics to obtain non-sparse high-resolution spectrograms. The practicability of this was demonstrated through several examples.

REFERENCES

[1] M. H. Hayes, *"Statistical Digital Signal Processing and Modeling"*. New York: Wiley, 1996.

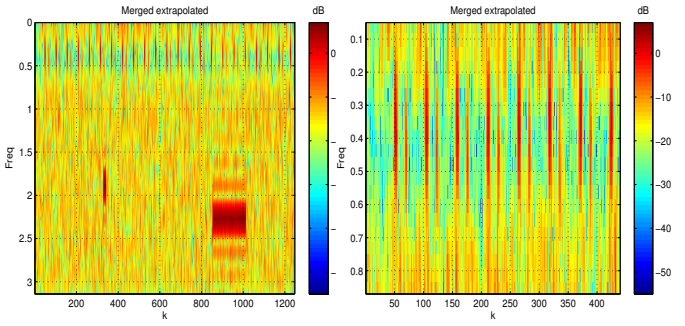
[2] G. Thomas and S. D. Cabrera, "Resolution enhancement in time-frequency distributions based on adaptive time extrapolations," in *Proc. of International Symposium on Time- Frequency and Time-Scale Analysis*, 1994, pp. 104–107.

[3] J. Nam, G. Mysore, J. Ganseman, K. Lee, and J. S. Abel, "A super-resolution spectrogram using coupled PLCA," in *Proc. of the 11th Conference of the International Speech Communication Association*, 2010.



(a) Full (b) Magnified

Fig. 12: Sparse extrapolated, $N = 24$, $\varepsilon = 0.95||\hat{\mathbf{w}}\mathbf{x}_k||$



(a) Full (b) Magnified

Fig. 13: Extrapolated merged, $N = 24$, $\varepsilon = 0.95||\hat{\mathbf{w}}\mathbf{x}_k||$

[4] R. Maleh and F. A. Boyle, "Exploiting spectral leakage for spectrogram frequency super-resolution," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, 2013.

[5] M. I. Mandel and Y. S. Cho, "Audio super-resolution using concatenative resynthesis," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2015, pp. 1–5.

[6] G. Schamberg, D. Ba, M. Wagner, and T. Coleman, "Efficient low-rank spectrotemporal decomposition using ADMM," in *IEEE Statistical Signal Processing Workshop*, 2016.

[7] E. Candès, J. Romberg, and T. Tao, "Stable signal receovery from incomplete and inaccurate measurments," *Communication in Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[8] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[9] X. Dong and Z. Zhu, "Digital extrapolation spectral analysis based on ARMA model," in *Proc. of the National Aerospace and Electronics Conference*, 1992.

[10] L. Anitori, W. van Rossum, and A. Huizing, "Array aperture extrapolation using sparse reconstruction," in *IEEE Radar Conference*, 2015, pp. 237–242.

[11] J.Akhtar and K. E. Olsen, "Formation of range-doppler maps based on sparse reconstruction," *IEEE Sensors Journal*, vol. 16, no. 15, pp. 5921–5926, Aug. 2016.

[12] N. Y. Yu and Y. Li, "Deterministic construction of Fourier-based compressed sensing matrices using an almost difference set," *EURASIP Journal on Advances in Signal Processing*, pp. 805–821, Oct. 2013.

[13] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[14] Y. Hu and P. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," in *Speech Communication*, 2007, pp. 588–601.

[15] M. Cesarelli, M. Ruffo, M. Romano, and P. Bifulco, "Simulation of foetal phonocardiographic recordings for testing of FHR extraction algorithms," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 513–523, Sept. 2012.

ISBN 978-0-9928626-7-1