Single Sensor Source Separation for Acoustical Machine Diagnostics

Martin Spiertz[1], Volker Gnann[1], Pascal Dietrich[2], Jens-Rainer Ohm[1], Michael Vorländer[2]

[1] *Institut für Nachrichtentechnik, RWTH Aachen University, D-52056 Aachen, Deutschland*

[2] *Institut für Technische Akustik, RWTH Aachen University, D-52056 Aachen, Deutschland*

Email: spiertz@ient.rwth-aachen.de

Introduction

Acoustical recordings in the context of machine diagnosis usually consist of a multitude of signal components. In order to analyse the individual components of the machine regarding their condition or even to detect faults it is advantageous to handle the signals separately according to their origin of generation. To become independent from recording situations, we propose single sensor source separation algorithms as preprocessing step for automatic machine diagnostic algorithms. Because of the underlying purely additive model, one promising method for source separation is the non-negative spectrogram analysis, e.g. the non-negative matrix factorization (NMF). The basic NMF can be extended e.g. by separating steady state signals, transient signals, and even harmonic components with time-varying pitch and partial amplitudes. With such extensions, the non-negative spectrogram analysis can be easily adopted to a wide range of separation tasks. In this contribution an algorithm is proposed to separate a measured signal of an electrical machine, containing the superposition of a steady state motor noise signal and the transient signal of a defect bearing.

Signal Analysis

We assume mixtures of three different types of input signals: steady state (s_{st}) and transient (s_{tr}) components with arbitrary, constant spectra and harmonic signals (s_{ha}) with time-varying pitch. x is the monaural time-discrete mixture:

$$x = s_{\mathrm{st}} + s_{\mathrm{tr}} + s_{\mathrm{ha}} \ . \tag{1}$$

For analysis, the mixture is transformed by the short-time Fourier transform (STFT) into a spectrogram $\underline{\mathbf{X}}(k,t)$ with frequency bins $0 \leq k < K$ and temporal indices $0 \leq t < T$. We use the square root of *Hann* window as analysis and synthesis window with an overlap of 50 %.

Non-Negative Spectrogram Separation

The components s_{st} and s_{tr} can be separated by non-negative matrix factorization, explained e.g. in [1]. In the same thesis, separation of components with time-varying pitch is done by non-negative matrix deconvolution (NMD). The main restrictions for NMD are first, the necessity of logarithmic frequency analysis making the signal synthesis more complex and second, the assumption of constant relative amplitudes for the single partials. This assumption is generally not true, e.g. in the case of resonances in the transfer path between sound source and sound receiver. Therefore, we propose a non-negative analysis with different algorithms for harmonic factorization and noise factorization.

In the following, the separation of $\underline{\mathbf{X}}(k,t)$ is done by minimizing the β-divergence between the approximating model and the magnitude-spectrogram in the case of $\beta = 2$ (Euclidean distance) or the power-spectrogram in the case of $\beta = 0$ (Itakura-Saito distance), see also [2]. Therefore, we define the matrix $\mathbf{V}$ to approximate by

$$\mathbf{V}(k,t) = |\underline{\mathbf{X}}(k,t)|^{2-\frac{\beta}{2}} \ . \tag{2}$$

For simpler description in frequency domain, we approximate the main-lobe of the analysis window by

$$\widetilde{W}(k,k_0) = e^{-\frac{(k-k_0)^2}{\sigma^2}} \ . \tag{3}$$

Because the main-lobe of the analysis window depends on the definition of $\mathbf{V}$, we set $\sigma^2 = \frac{3}{2\log(4-\beta)}$ as a good heuristic. The overall model of $\mathbf{V}(k,t)$ is described by

$$\mathbf{V}_{\mathrm{ha}}(k,t) = \sum_{n=1}^{N} a_n(t)\widetilde{W}\left(k, k_0(t,n)\right) \tag{4}$$

$$\mathbf{V}_{\mathrm{st}}(k,t) = \mathbf{B}(k,1) \cdot \mathbf{G}(1,t) \tag{5}$$

$$\mathbf{V}_{\mathrm{tr}}(k,t) = \mathbf{B}(k,2) \cdot \mathbf{G}(2,t) \tag{6}$$

$$\widetilde{\mathbf{V}}(k,t) = \mathbf{V}_{\mathrm{ha}}(k,t) + \mathbf{V}_{\mathrm{st}}(k,t) + \mathbf{V}_{\mathrm{tr}}(k,t) \tag{7}$$

$$\Rightarrow \mathbf{V}(k,t) \approx \widetilde{\mathbf{V}}(k,t) \tag{8}$$

with $a_n(t) \geq 0$ being the non-negative amplitude, and $k_0(t,n)$ being the pitch in discrete frequency-bins of the n-th partial in time frame t. The noise is modelled by NMF with two constant spectra stored in the columns of matrix $\mathbf{B}$ and two envelopes stored in the rows of matrix $\mathbf{G}$. The non-negativity of all components ensures herein a purely additive model.

The multiplicative update rules of NMF

$$\mathbf{B}(k,i) = \mathbf{B}(k,i) \cdot \frac{\sum_t \frac{\mathbf{V}(k,t)}{\widetilde{\mathbf{V}}^{\beta-2}(k,t)} \cdot \mathbf{G}(i,t)}{\sum_t \widetilde{\mathbf{V}}^{\beta-1}(k,t) \cdot \mathbf{G}(i,t)}, \text{ and} \tag{9}$$

$$\mathbf{G}(i,t) = \mathbf{G}(i,t) \cdot \frac{\sum_k \frac{\mathbf{V}(k,t)}{\widetilde{\mathbf{V}}^{\beta-2}(k,t)} \cdot \mathbf{B}(k,i)}{\sum_k \widetilde{\mathbf{V}}^{\beta-1}(k,t) \cdot \mathbf{B}(k,i)}, \tag{10}$$

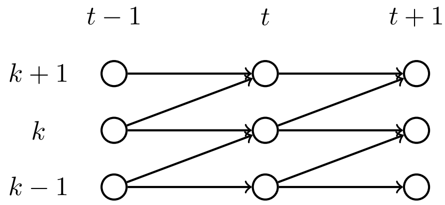


Figure 1: Pitch detection with Trellis diagram ($\Delta k_{\max} = 1$).

used e.g. in [2] lead to approximations of $\mathbf{V}_{\mathrm{st}}(k,t)$ and $\mathbf{V}_{\mathrm{tr}}(k,t)$.
For each of the N partials of $\mathbf{V}_{\mathrm{ha}}$ we initialize

$$\mathbf{Y}(k,t) = \max\Big(0, \mathbf{V}(k,t) - \sum_{l=1}^{n-1} a_l(t)\widetilde{W}(k, k_0(t,l)) - \sum_{i=1}^{2} \mathbf{B}(k,i)\cdot\mathbf{G}(i,t)\Big) \; . \quad (11)$$

For a given pitch $k_0(t,n)$ and non-overlapping functions $\widetilde{W}(k, k_0(t,n))$ the amplitudes $a_n(t)$ are evaluated by

$$\beta = 2\colon\; a_n(t) = \frac{\sum_k \mathbf{Y}(k,t)\cdot\widetilde{W}(k,k_0(t,n))}{\sum_k \widetilde{W}^2(k,k_0(t,n))}, \text{ or} \quad (12)$$

$$\beta = 0\colon\; a_n(t) = \sum_k \frac{\mathbf{Y}(k,t)}{\widetilde{W}(k,k_0(t,n))}. \quad (13)$$

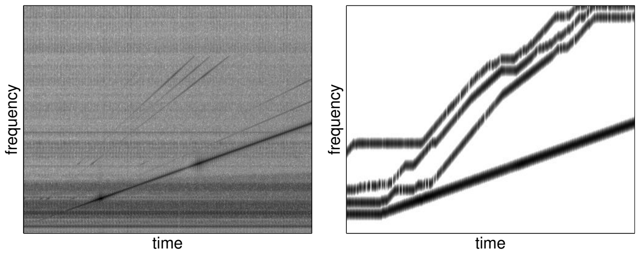
To find the optimal pitch, we use the Viterbi algorithm as proposed e.g. in [3]. We limit k_0 to $k_0(t,n-1)+c < k_0(t,n) < k_{\max}$, with $k_{\max}$ and c being user defined constants. The possible values for k_0 are the possible states in the Trellis diagram. For each frame t and each possible $k_0(t,n)$ the amplitudes $a_n(t)$ are evaluated according to Equation (12) or (13). With these amplitudes, the local cost d_{local} of each node is defined by the β-divergence d_β:

$$d_{\mathrm{local}} = \sum_k d_\beta\left(\mathbf{Y}(k,t), a_n(t)\widetilde{W}(k,k_0(t,n))\right) \; . \quad (14)$$

The final cost for each node is defined as the cost in Equation (14) and the minimum cost of all its predecessor-states. The predecessor-states are defined by the simple rule, that $k_0(t,n)$ is assumed to be non-decreasing. The maximum slope of $k_0(t,n)$ is limited by the user-defined parameter $\Delta k_{\max}$, as shown in Figure 1. The final values of $k_0(t,n)$ are defined by the path through the Trellis diagram with smallest cost.
One iteration of the final non-negative spectrogram analysis can be summarised as follows:

- Set $G(1,t) = 1$ to distinguish between steady state and transient parts of noise.
- Update matrices $\mathbf{B}$ and $\mathbf{G}$ (Equations (9-10)).
- Update $a_n(t)$ and $k_0(t,n)$.

The distinction between transient and steady state components can be also reached by a temporal continuity criterion, as proposed in [1], but, in our case, steady state can be modeled by a constant noise during the whole analysis. Because the update of the harmonic model needs much more evaluation time and the update of $\mathbf{B}$ and $\mathbf{G}$ needs more iterations than the harmonic model, the first two steps are repeated ten times for each iteration, to reduce the overall number of updates of the harmonic model



(a) $\mathbf{V}$ (b) $\mathbf{V}_{\mathrm{ha}}$

Figure 2: Mixture of a run up of an machine and an additive transient noise (a defect bearing). The right Figure shows the estimated harmonic approximation.

Signal Synthesis

The estimated spectrograms of the harmonic component is evaluated by a filtering operation according to [2]:

$$\underline{\widetilde{S}}_{\mathrm{ha}}(k,t) = \underline{\mathbf{X}}(k,t)\cdot\frac{\mathbf{V}_{\mathrm{ha}}(k,t)}{\widetilde{\mathbf{V}}(k,t)} \; . \quad (15)$$

The corresponding separated time-domain signals are evaluated by the inverse STFT. The other components are evaluated in an analogue way.

Experimental Results

In Figure 2, a run up of an electrical machine is shown. The left figure shows the increasng frequency and the time-varying amplitudes of the N different partials due to resonances. In the right figure, the approximation $\mathbf{V}_{\mathrm{ha}}$ is shown. It can be seen, that the pitch can be detected without any errors. The data of the partials (pitch and amplitudes) is only estimated correctly in the regions of high amplitudes (resonance frequencies). For such resonances, the proposed pitch detection leads to very robust results: Even in the case of wrong classification of partials the algorithm is able to detect correct partials in later time frames.

References

[1] T. Virtanen, *Sound Source Separation in Monaural Music Signals*, Ph.D. thesis, Tampere University of Technology, 2006.

[2] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[3] H. Quast, O. Schreiner, and M.R. Schroeder, "Robust pitch tracking in the car environment," in *Proc. IEEE Int. Conference on Acoustic Speech and Signal Processing ICASSP '02*, 2002.