Analysis of Polyphonic Audio Using Source-Filter Model and Non-Negative Matrix Factorization

Tuomas Virtanen and Anssi Klapuri
tuomas.virtanen@tut.fi, anssi.klapuri@tut.fi
Institute of Signal Processing, Tampere University of Technology, Tampere, Finland

Introduction

- Framework for (polyphonic) audio — linear signal model for magnitude spectrum $x_t(k)$:

$$\hat{x}_t(k) = \sum_{n=1}^{N} g_{n,t} b_n(k) \tag{1}$$

- $g_{n,t}$ is the gain of basis function n in frame t, and $b_n(k)$, $n = 1, \ldots, N$ are the basis functions
- Spectrogram modeled as a sum of components, each of which has a fixed magnitude spectrum and time-varying gain
- Applied e.g. in sound source separation: each sound represented with distinct basis functions
- Unsupervised learning estimation algorithms: ICA [1], NMF [3], sparse coding [4]
- Supervised learning: NMF, sparse coding, vector quantization
- Distinct bases required for each pitch/phoneme-combination — makes the estimation and clustering less reliable

Proposed Source-Filter Model

- Each basis $b_n(k)$ is described as a product of the magnitude spectra of an excitation (source) $e_i(k)$ and a filter $h_j(k)$.
- "Source" refers to a vibrating object such as a guitar string — varies with pitch
- "Filter" represents the resonance structure of the rest of the instrument which colors the produced sound — varies with timbre
- Model for the magnitude spectrum of mixture signal:

$$\hat{x}_t(k) = \sum_{i,j} g_{i,j,t} e_i(k) h_j(k) \tag{2}$$

- Smaller number of parameters — bases are restricted to $b_n(k) = e_i(k)h_j(k)$
- The model associates components with the same timbre (resp. pitch), leading to an automatic clustering of bases to sound sources (resp. musical notes).

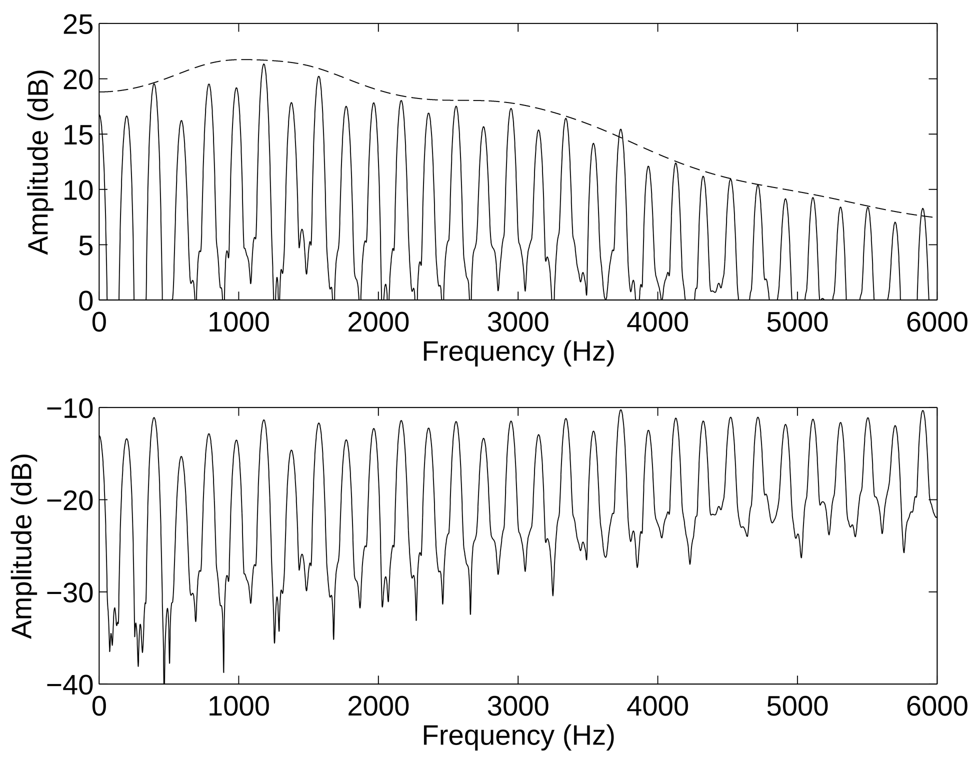

FIGURE 1: Upper plot: spectrum of a trumpet tone (solid line) and estimated filter (dashed line). Lower plot: excitation.

NMF Algorithm for Parameter Estimation

- On magnitude spectrograms non-negative matrix factorization produces representations where each basis corresponds to an individual sound source
- Estimation principle: minimize the reconstruction error between the magnitude spectrograms of the observed signal and model, while restricting the parameters to positive values
- We extend the algorithm to estimate the parameters of the source-filter model
- Initialize parameters with random positive values, apply multiplicative update rules sequentially. Each update decreases the reconstruction error

Representing Several Pitch Values with a Single Excitation

- Translating the bases in frequency allows representing several pitch values with a single basis (requires logarithmic frequency resolution)
- The amount of contribution of the n^{th} basis translated by τ frequency indices denoted by $g_{n,t,\tau}$
- Model can be written as a convolution between gain and basis:

$$\hat{x}_t(k) = \sum_{n,\tau} g_{n,t,\tau} b_n(k-\tau) \tag{3}$$

- Estimation algorithms extended from NMF proposed e.g. by FitzGerald [2] and Virtanen [5, pp. 57-65]

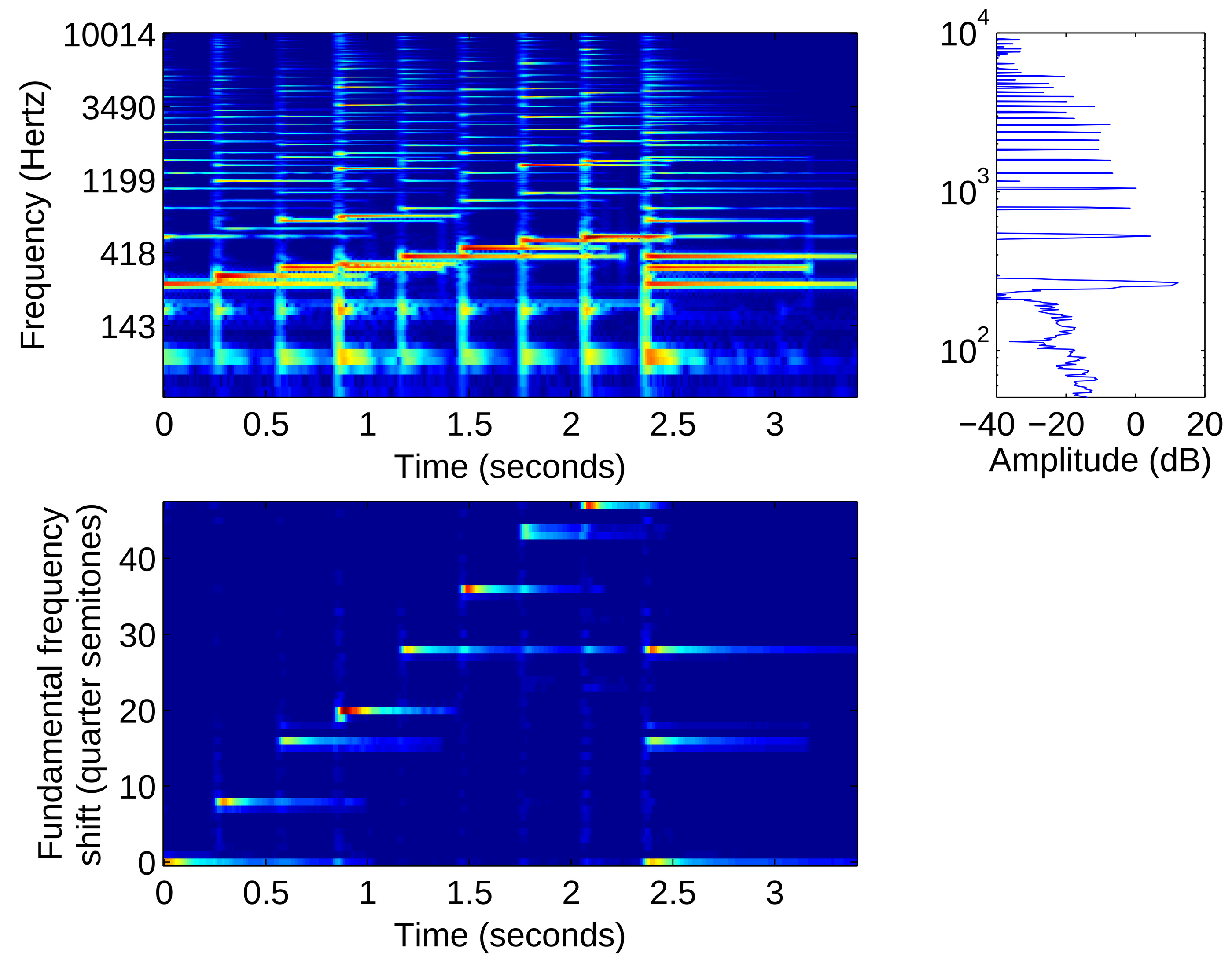

FIGURE 2: Representing several fundamental frequency values with a single basis.

- Drawback: the translation affects the entire basis function, and therefore the filter becomes translated
- Proposed model: different fundamental frequency values obtained by translating a single harmonic excitation while keeping the filter fixed

$$\hat{x}_t(k) = \sum_{i,j,\tau} g_{i,j,t,\tau} e_i(k-\tau) h_j(k). \tag{4}$$

- Estimation algorithm extended from NMF

References

[1] M. A. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. In *International Computer Music Conference*, Berlin, Germany, 2000.

[2] Derry FitzGerald, Matt Cranitch, and Eugene Coyle. Generalised prior subspace analysis for polyphonic pitch transcription. In *Proceedings of International Conference on Digital Audio Effects*, Madrid, Spain, 2005.

[3] Paris Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, 2003.

[4] Tuomas Virtanen. Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006. Accepted for publication.

[5] Tuomas Virtanen. *Sound Source Separation in Monaural Music Signals*. PhD thesis, Tampere University of Technology, 2006. available at http://www.cs.tut.fi/~tuomasv.