Super-resolution of Musical Signals Using Approximate Matching Pursuit

Brennan P. Keegan[1], **Steven K. Tjoa**[2], **and K. J. Ray Liu**[1]

[1]Signals and Information Group, Dept. of Electrical and Computer Eng., Univ. of Maryland – College Park, MD, USA

[2]Imagine Research, San Francisco, CA, USA

Introduction

Super-resolution is well-studied for video and images but not for musical signals.

One idea [Smaragdis, et al., 2009]: For each low-resolution (LR) input spectrum, find its coefficients with respect to a low-resolution basis. Using the same coefficients, reconstruct a high-resolution (HR) output from a high-resolution basis.

But what if you use a **very large, overcomplete dictionary of real-world musical atoms**?

We propose a **super-resolution method using Approximate Matching Pursuit**.

Approximate Matching Pursuit

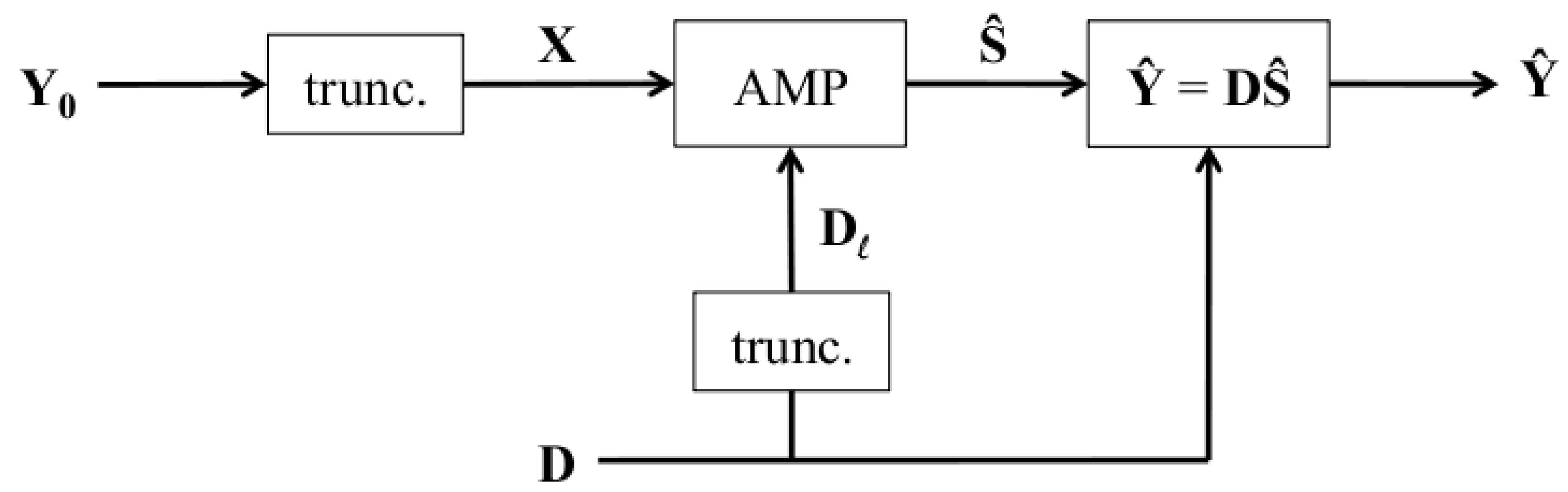
See [Tjoa and Liu, ISMIR 2011].

Basic Idea: Inside matching pursuit, match using an approximate nearest neighbor (ANN) method, not exhaustive linear search.

1. Input: spectrum $\mathbf{x} \in \mathbb{R}^M$; dictionary $\mathbf{D} \in \mathbb{R}^{M \times K}$.
2. Output: coefficients $\mathbf{s} \in \mathbb{R}^K$.
3. Initialize: $\mathbf{s} \leftarrow \mathbf{0}$; active set: $\mathcal{S} \leftarrow \emptyset$; residual $\mathbf{r} \leftarrow \mathbf{x}$; threshold $\epsilon > 0$.
4. While $||\mathbf{r}|| > \epsilon$:
   - **Find any k such that dictionary atom $\mathbf{a}_k$ and residual $\mathbf{r}$ are "close enough".**
   - Add atom to active set: $\mathcal{S} \leftarrow \mathcal{S} \cup k$
   - Solve for $\{s_j | j \in \mathcal{S}\}$: $\min_{s_j | j \in \mathcal{S}} ||\mathbf{x} - \sum_{j \in \mathcal{S}} \mathbf{a}_j s_j||_2$
   - Update residual: $\mathbf{r} \leftarrow \mathbf{x} - \mathbf{D}\mathbf{s}$
5. Return $\mathbf{s}$.

Proposed Super-resolution Method

1. Given a HR dictionary $\mathbf{D}$, truncate $\mathbf{D}$ in frequency to obtain the LR dictionary, $\mathbf{D}_\ell$.
2. Given LR input spectrogram $\mathbf{X}$ and LR dictionary $\mathbf{D}_\ell$, use AMP to obtain the coefficient matrix $\hat{\mathbf{S}}$.
3. Construct the HR output spectrogram, $\hat{\mathbf{Y}} = \mathbf{D}\hat{\mathbf{S}}$.

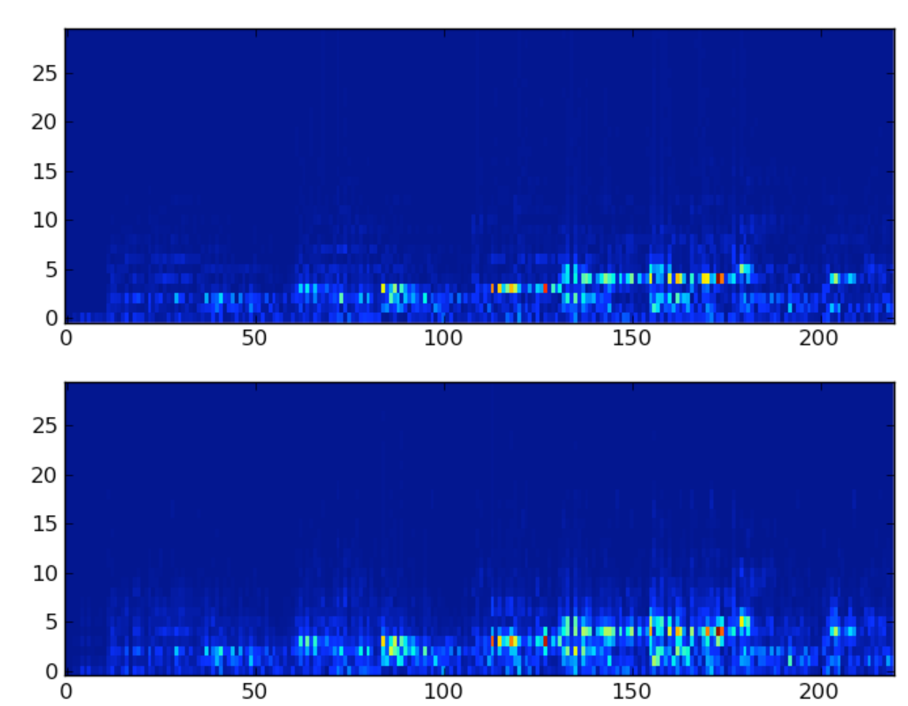


Evaluation: Start with ground-truth HR spectrogram, $\mathbf{Y}_0$. Evaluate the reconstruction error, $F = ||\mathbf{Y}_0 - \hat{\mathbf{Y}}||_F$.

Dictionary: Piano spectra from the U. Iowa Dataset.

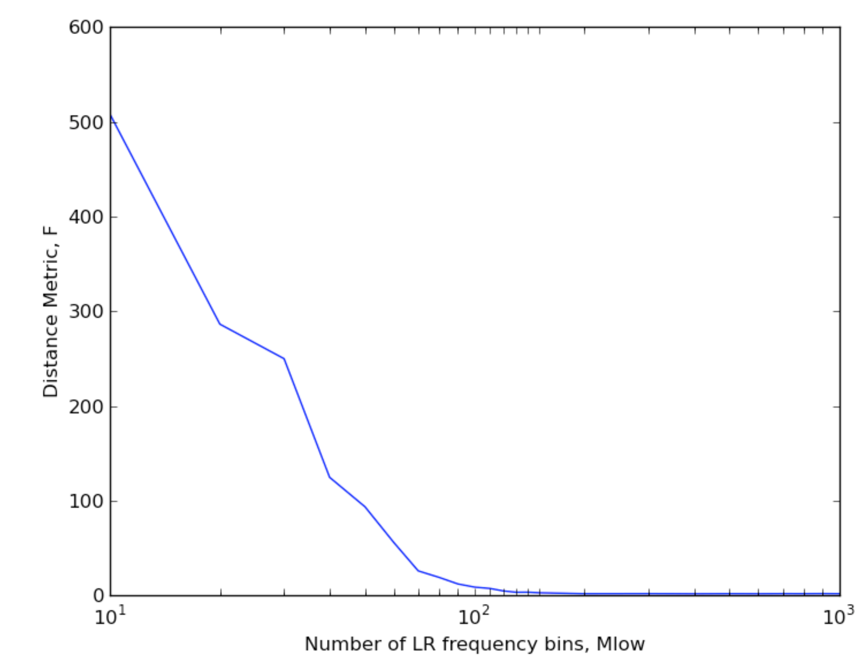
Example

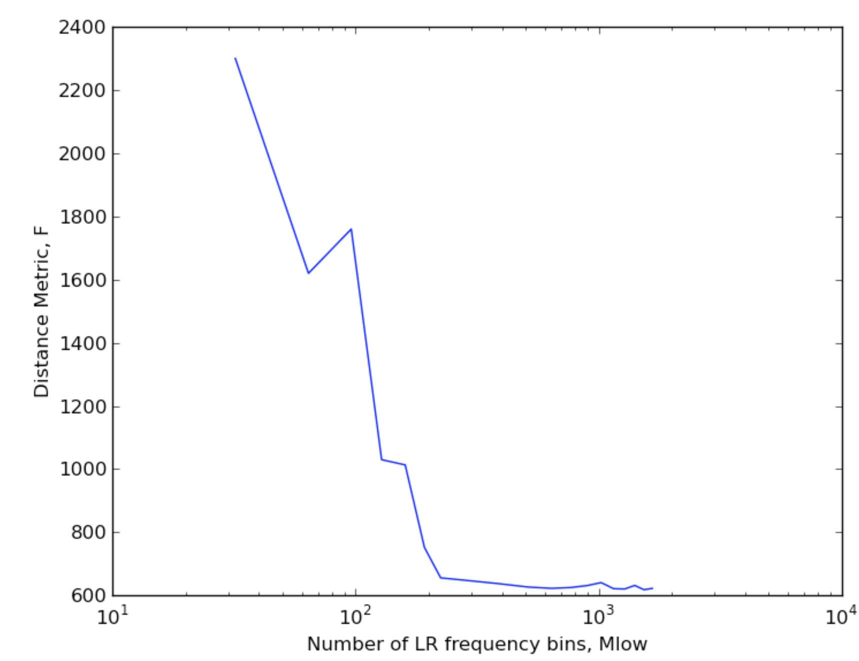
First five measures of Mozart's "Rondo Alla Turca." Spectrograms $\mathbf{X}$ (top) and $\mathbf{D}\hat{\mathbf{S}}$ (bottom).

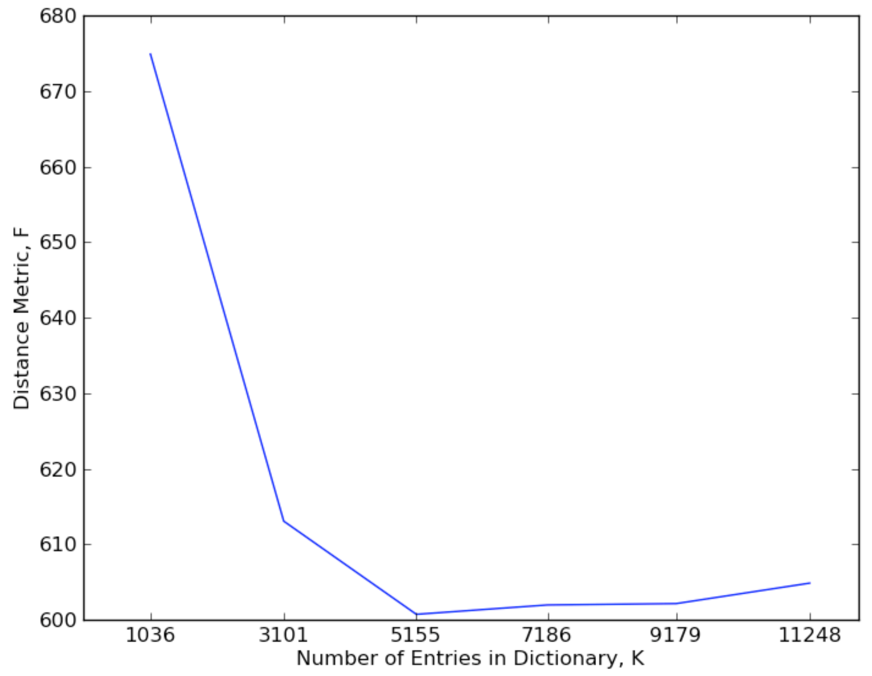


Experiments

1. F vs. number of LR frequency bins, M_{low} (synthetic input)
2. F vs. number of LR frequency bins, M_{low} (real musical input)
3. F vs. dictionary size, K (real musical input)







Conclusions

- Super-resolution is achieved with a dictionary of musical spectra.
- Proposed method works well for very LR inputs.
- Larger dictionaries generally result in higher accuracy.

Future Work

- Test on heterogeneous musical inputs with acoustic overlap.
- Larger, more representative dictionaries.
- Use AMP for time-domain decomposition.
- Image and video processing.