Guitar Music Transcription: A New Design Concept

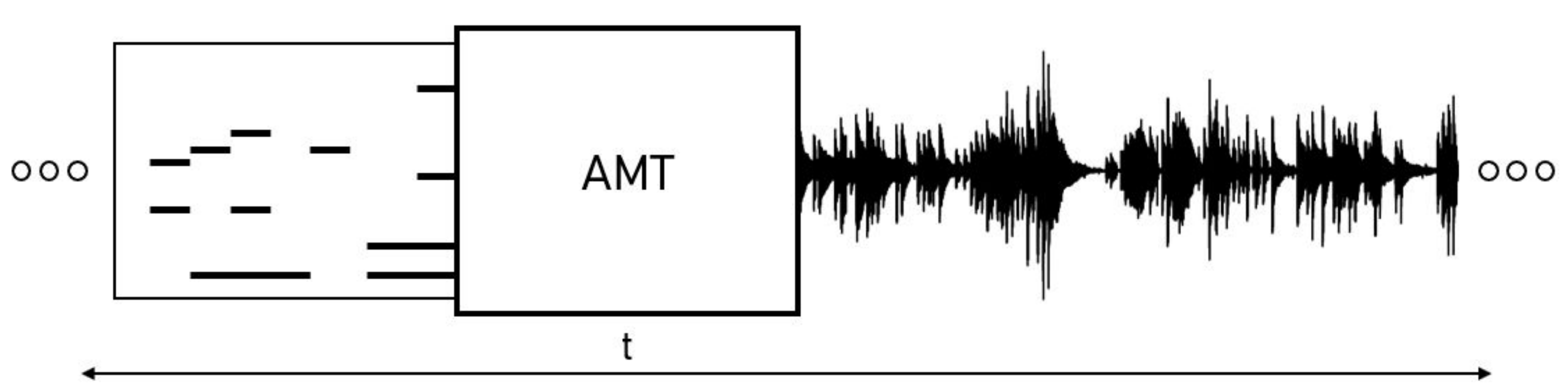
Frank Cwitkowitz

Department of Electrical and Computer Engineering, University of Rochester

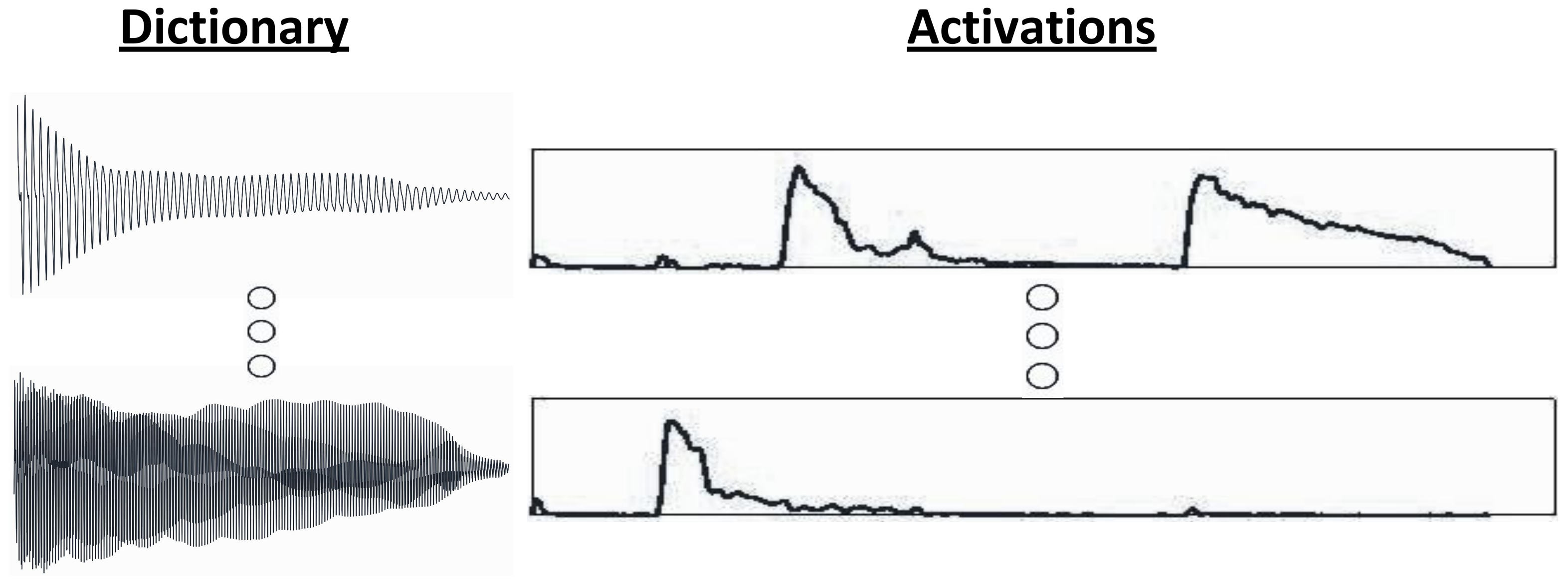
Abstract

- **Automatic Music Transcription (AMT)** approaches are often tailored towards solo piano recordings
  - Guitar has additional **expressive dimensions**
  - There is less data for solo guitar recordings
- Recent dataset (GuitarSet [2]) has created new opportunities for data-driven approaches
- Model-based approach to guitar-specific AMT using **Convolutional Sparse Coding (CSC)**
  - Proof of concept for solo piano recordings [1]
  - Dictionary generated from **real occurences** of each note (model)
  - Activations provide **latent representation** for inference

Problem

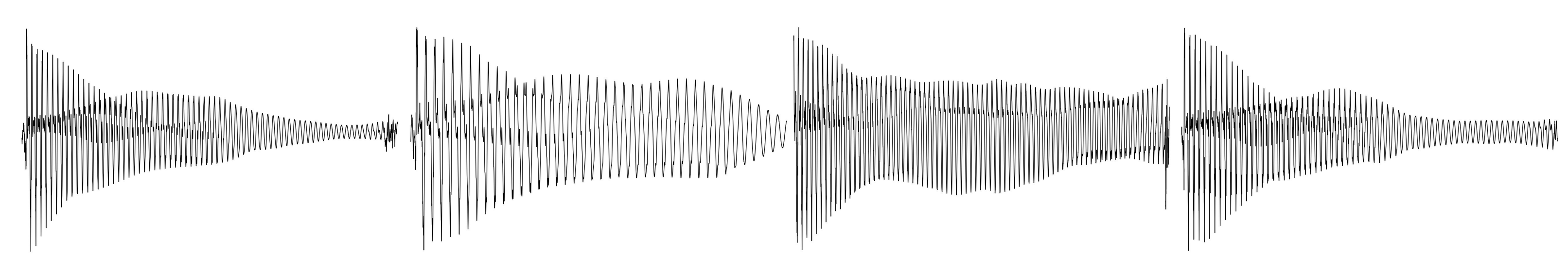


- Algorithmically recover the information sufficient to form a **symbolic representation of the music** inherent in an audio signal
  - Several types of symbolic representation, *e.g.* **tablature or piano-roll**, sheet music, MIDI, *etc.*
- Plenty of applications
  - Real-time instructional scenarios which listen and provide **feedback** for a human player
  - Mid-level music representations for **database querying**
  - Improvement of methods for **other music analysis problems**
- Represent time-domain signal as the sum of **dictionary elements activated across time**



Dictionary Generation

1. Isolate a training split, *i.e.* 5 out of 6 guitarists
2. Acquire **waveform of all notes**, grouping by string/fret
3. Enforce **minimum number of elements** per string/fret
   a. Pitch-shift elements from previous if necessary
4. **Reduce** the number of elements per string/fret to a fixed amount
   a. Iteratively find **element most dissimilar** to all elements in current group for inclusion
      i. Currently using summed dot-products



Estimating Activations

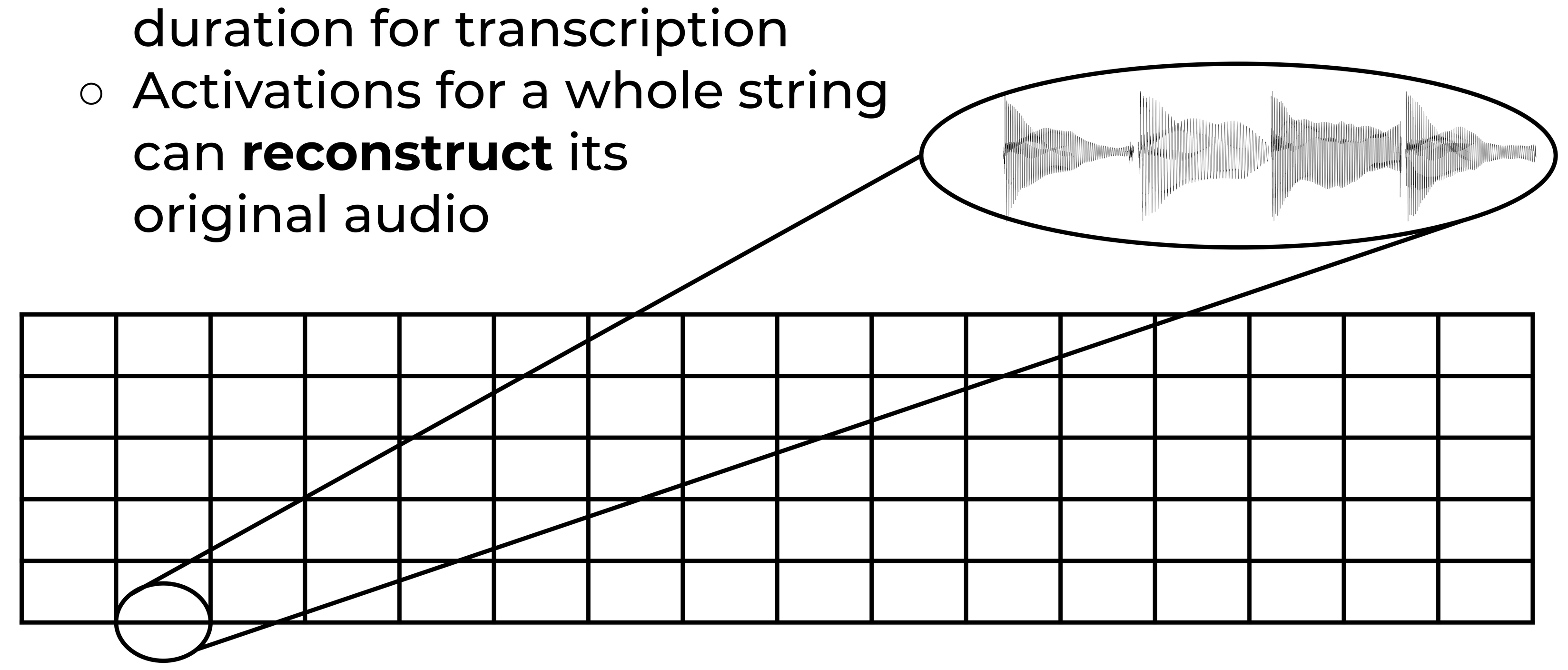
- Alternating Direction Method of Multipliers **(ADMM)** algorithm for Convolutional Basis Pursuit Denoising (**CBPDN)** problem
- Squared reconstruction error, Sparsity, and Lateral inhibition terms

$$\arg\min_{\{\boldsymbol{x}_m\}} \frac{1}{2}\left\|\sum_m \boldsymbol{d}_m * \boldsymbol{x}_m - \boldsymbol{s}\right\|_2^2 + \lambda\sum_m \|\boldsymbol{x}_m\|_1 + \mu\sum_m \Gamma(\{\mathbf{x}_m\})$$

$$\Gamma(\{\mathbf{x}_m\}) = |\boldsymbol{x}_m(t)|\left[\left(\sum_{\substack{n\in G(m)\\ |t-\tau|<T}} |\boldsymbol{x}_n(\tau)|\right) - |\boldsymbol{x}_m(t)|\right]$$

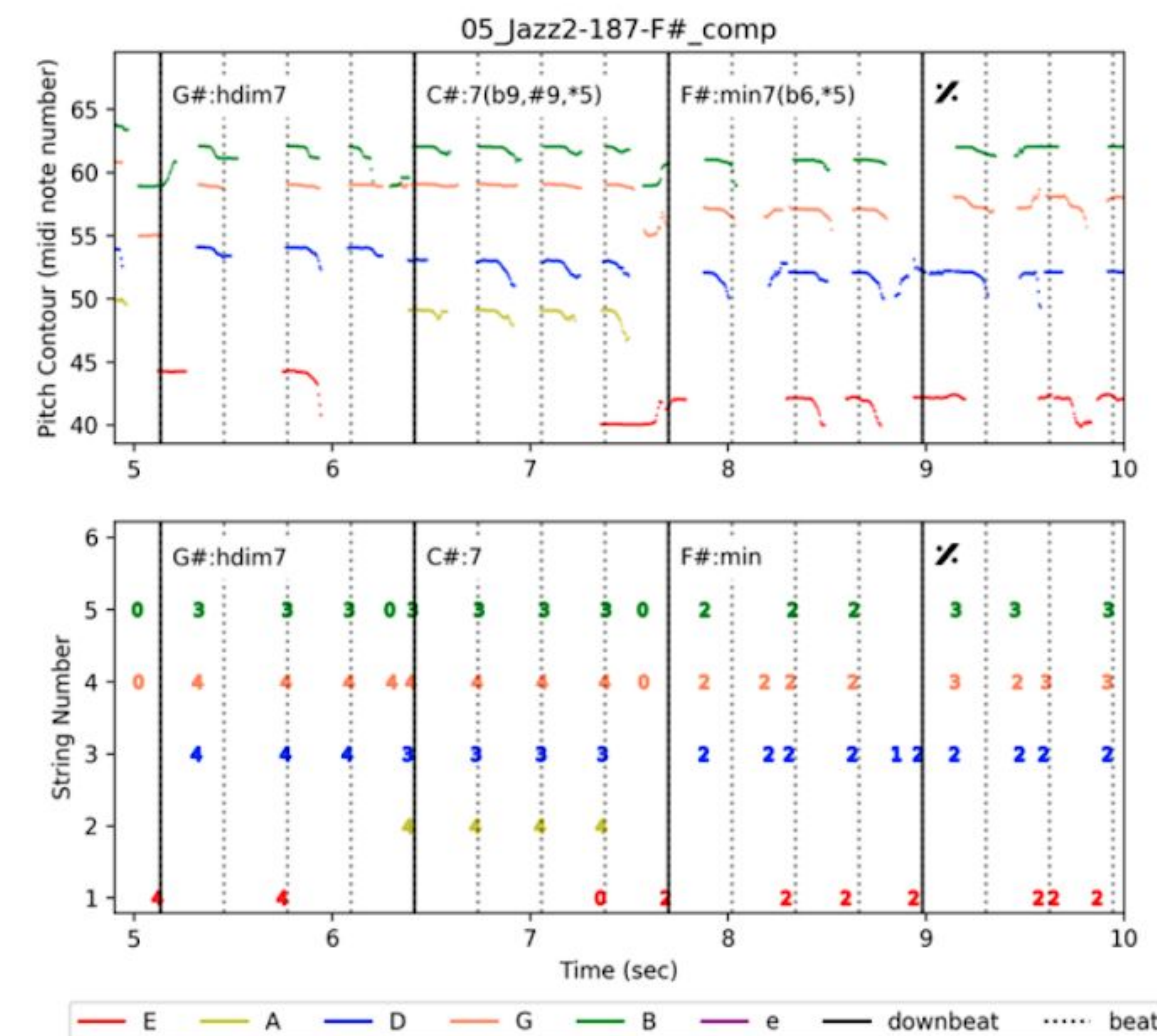
Inference

- **Grouping is maintained** at the activation level
  - Strongest activation by string gives pitch, onset, and duration for transcription
  - Activations for a whole string can **reconstruct** its original audio



GuitarSet [2]

- Recordings of **solo guitar pieces** (~30 seconds)
  - Mono-channel microphone
  - Hexaphonic pickup
- About **three hours of audio** from six guitarists
- Different styles, keys, genres, comp/solo, *etc.*



Evaluation - Results - Future Work

- Precision, recall, F1 score
  - Frame-wise predictions (Multiple F0 estimation)
  - Note-wise predictions (AMT)
    - Correct onset and pitch
    - Above, and correct offset
- Signal-to-Distortion Ratio (SDR)
- Need to implement in a **memory efficient** way
- Take closer look at the implementation specifics of [2]
- **Generate dictionary iteratively** from estimated guitar model parameters

References

1. Cogliati, Andrea, Zhiyao Duan, and Brendt Wohlberg. "Piano transcription with convolutional sparse lateral inhibition." IEEE Signal Processing Letters 24.4 (2017): 392-396.
2. Xi, Qingyang, et al. "GuitarSet: A Dataset for Guitar Transcription." ISMIR. 2018.