# Towards Unsupervised Acoustic Guitar Transcription

**Andrew Wiggins, Youngmoo Kim**
Music and Entertainment Technology Laboratory, Drexel University, USA
{awiggins, ykim}@drexel.edu

## ABSTRACT

We introduce a deep neural network design for the unsupervised pitch estimation of acoustic guitar chords. The proposed system takes in a short audio clip containing a guitar chord or note and produces estimates for the pitches present and their amplitudes. It trains without requiring labeled data. In an analysis part of the network, a convolutional neural network produces pitch estimates from an input spectrogram. These pitch estimates are fed into a synthesis part that attempts to reconstruct the original input. The analyzer trains while the synthesizer remains fixed, and a reconstruction loss is minimized. As the network improves its reconstructions, it learns to produce accurate pitch estimates. We discuss two variants for the synthesis part: component note synthesis and Karplus-Strong synthesis. We hope that insights from this work can be integrated into a full network for unsupervised acoustic guitar transcription.

## 1. INTRODUCTION

Approaches to automatic music transcription often rely on access to labeled training data. Creating sufficiently large labeled datasets for such tasks is costly. However, for unsupervised systems, which require large datasets that need not be labeled, the required audio data is usually easily acquired. For the task of acoustic guitar transcription, the existing GuitarSet dataset [1] provides labeled data. We utilized GuitarSet to train TabCNN [2], our convolutional neural network approach to guitar tablature estimation. However, we found that the limited size of the dataset likely impedes the generalizability of the the network. We are interested in exploring an unsupervised network for acoustic guitar note transcription.

Recently, DrummerNet [3], a deep unsupervised approach to drum transcription was introduced. The network trains with unlabled drumstems, using an analysis network to create transcriptions, while a fixed synthesis network reconstructs the input. Inspired by this approach, we investigate the use of a similar paradigm for the area of acoustic guitar transcription. We introduce a network design for the task of predicting pitches present in an acoustic guitar chord. We briefly discuss two possible variants for the synthesizer part of our network.



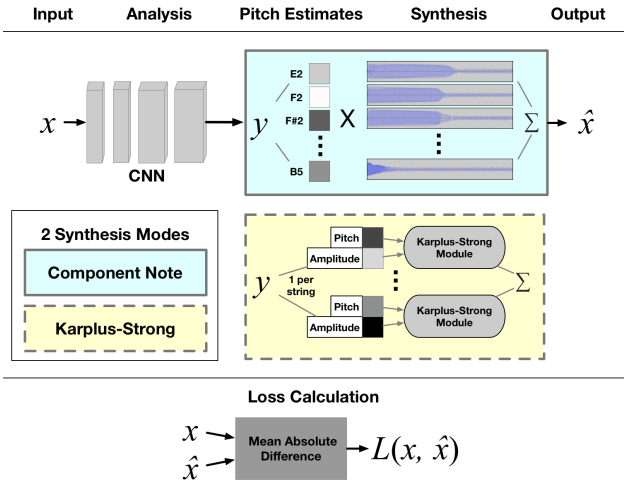


**Figure 1**. Overview of the proposed network design. An input spectrogram $x$ is analyzed by a convolutional neural network to produce a pitch estimation $y$. The vector $y$ is used to reconstruct the original spectrogram. (We consider two different synthesis modes.) A reconstruction loss is computed, measuring the difference between $x$ and its reconstruction $\hat{x}$.

## 2. NETWORK DESIGN

Our proposed network is outlined in Figure 1. It takes as input a spectrogram $x$ containing an acoustic guitar note or chord. A convolutional neural network (CNN) estimates the pitches and amplitudes of the notes present. This estimation $y$ is fed into a synthesizer which produces a reconstruction $\hat{x}$ of the original spectrogram. The pitch estimation CNN is trainable while the synthesizer is fixed. The network learns by minimizing a reconstruction loss, the mean absolute difference between $x$ and $\hat{x}$.

We introduce and discuss two variants for the synthesis module of the network: a component note synthesizer and a Karplus-Strong synthesizer.

### 2.1 Component Note Synthesis

The component synthesis technique is analogous to the approach used in DrummerNet [3]. The synthesizer has access to a recording of each single note playable on a stan-

dard acoustic guitar (E2–B5). It outputs a weighted sum of these notes depending on the pitch estimation vector $y$. We created 4 sets of note components using 2 acoustic guitars, a steel string and a nylon string, and 2 playing styles, fingered and picked. A random component set is selected for each training batch, so that the network does not overfit to a single string type or playing-style.

An advantage of this synthesis approach is the realistic acoustic guitar timbre, stemming from the use of actual guitar-note recordings. One downside for this style of synthesis is that a discrete set of pitches is required. The vector $y$ contains amplitude estimates for each of the 44 notes, but cannot, for instance, correctly estimate the pitch of a quarter-tone string bend.

### 2.2 Karplus-Strong Synthesis

The Karplus-Strong [4] variant uses a set of plucked-string physical-modelling synthesizers to produce the individual guitar notes. The synthesis technique uses a noise burst which is low-pass filtered in a feedback loop. This produces a pitched sound with higher harmonics that delay more quickly than lower ones, emulating a plucked string. The pitches of notes are determined by the length of the delay line, with an approximation of a continuous range of pitches producible via linear interpolation. The synthesizer outputs a weighted sum of the generated notes.

The advantage of using Karplus-Strong synthesis is that the pitch estimations are not as limited to a discrete set of pitches. In this variant, estimation $y$ contains pairs of continuous pitch and amplitude estimations. One pair for each voice is required, resulting in six estimation pairs for a standard acoustic guitar.

The disadvantage of this approach is the quality of the audio produced by the synthesizer. The resulting output does not sound like a realistic guitar, however, since the goal is accurate pitch estimation, the synthesis does not necessarily need to sound convincing. This technique is viable as long as the original spectrogram can be reconstructed closely enough to result in an accurate pitch transcription.

## 3. CONCLUSION AND FUTURE WORK

We introduced a network design for the unsupervised transcription of acoustic guitar chords. Our future work will focus on evaluating the proposed system, comparing the two synthesis approaches. We hope that this network design can be extended into a fully realized system for unsupervised acoustic guitar transcription, processing in-context guitar-playing rather than isolated chords.

## 4. REFERENCES


[1] Q. Xi, R. M. Bittner, J. Pauwels, X. Ye, and J. P. Bello, "Guitarset: A dataset for guitar transcription." in *ISMIR*, 2018, pp. 453–460.

[2] A. Wiggins and Y. Kim, "Guitar tablature estimation with a convolutional neural network." in *ISMIR*, 2019, pp. 284–291.

[3] K. Choi and K. Cho, "Deep unsupervised drum transcription," *arXiv preprint arXiv:1906.03697*, 2019.

[4] K. Karplus and A. Strong, "Digital synthesis of plucked-string and drum timbres," *Computer Music Journal*, vol. 7, no. 2, pp. 43–55, 1983.