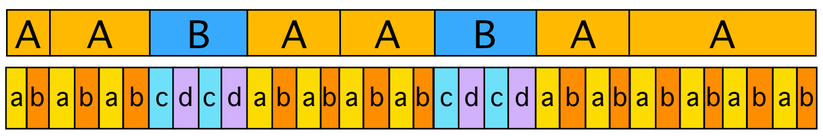
Multi-part pattern analysis:

Combining structure analysis and source separation to discover intra-part repeated sequences

Jordan B. L. Smith
Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST), Japan

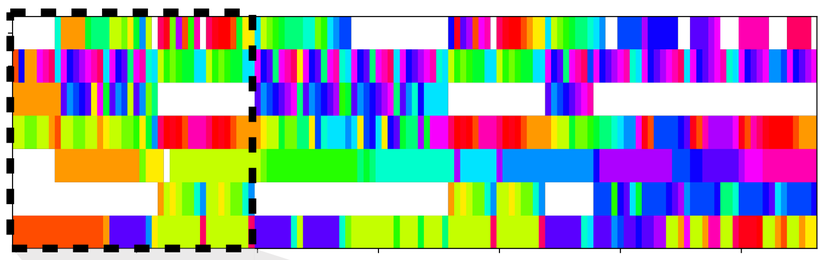
1. Motivation

Structural descriptions are usually single-dimensional, or perhaps hierarchical:

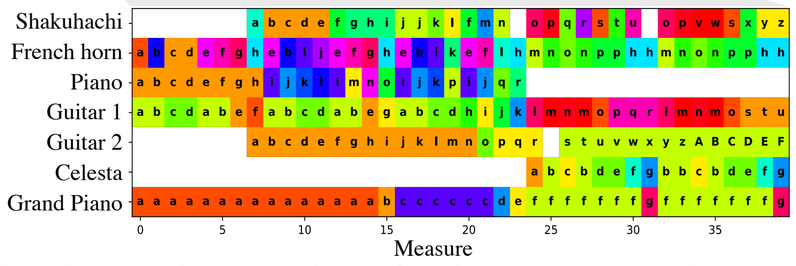


"While My Guitar Gently Weeps", by The Beatles (running example throughout)

But pieces of music are multi-layered, with each instrument playing its part according to its own patterns:



Repeating sequences and steady states recur independently among the parts:



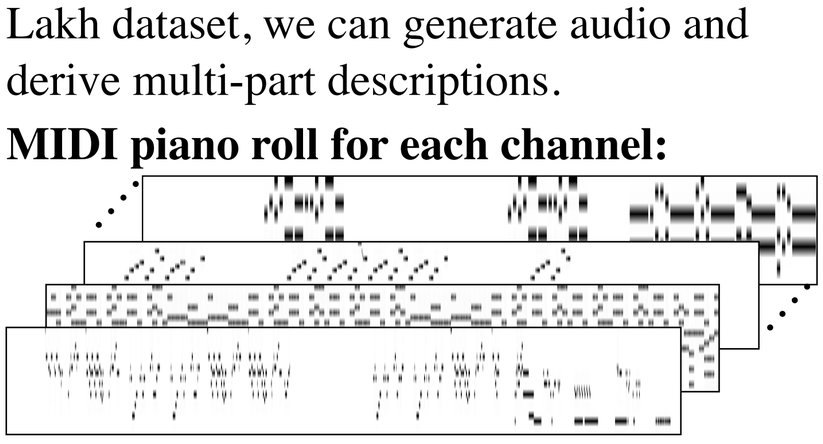
(In these plots, colour indicates a steady state *or* a consistent sequence.)

We propose a new task, **multi-part pattern analysis**, in which we aim to estimate such descriptions, directly from audio.

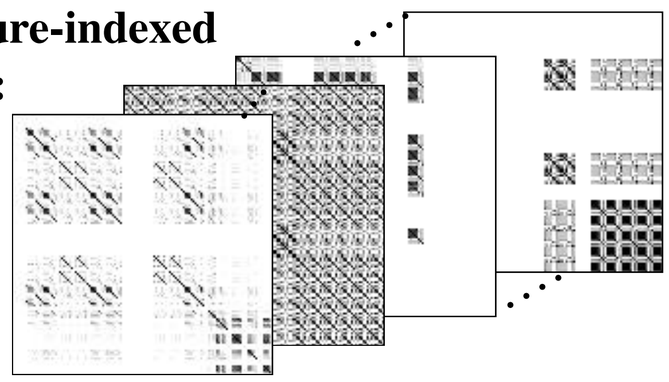
2. Data

No annotation corpora exist yet, and they would be costly to collect. Instead, using multi-channel MIDI files from Raffel's Lakh dataset, we can generate audio and derive multi-part descriptions.

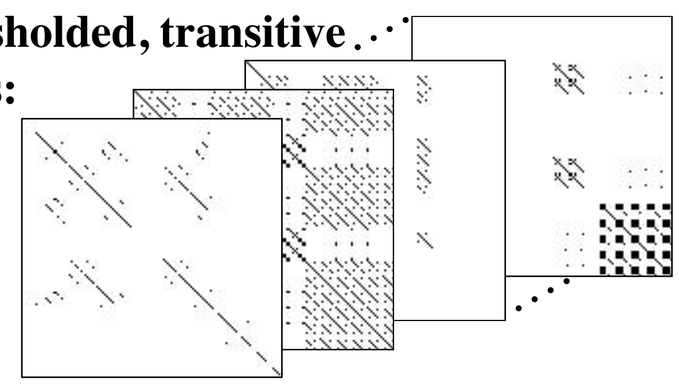
MIDI piano roll for each channel:



Measure-indexed SSMs:



Thresholded, transitive SSMs:

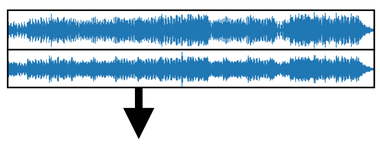


Transitive SSMs are equivalent to measure-wise labellings, which we visualize in the colourful plots above.

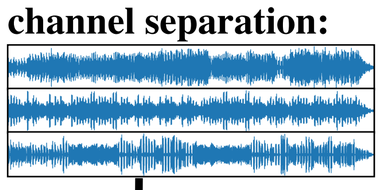
3. Algorithm

Briefly put: we perform source separation, then perform structural analysis on each estimated source.

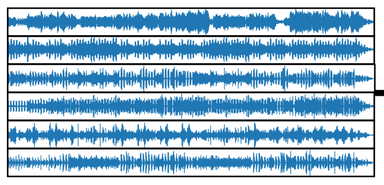
1. Stereo audio:



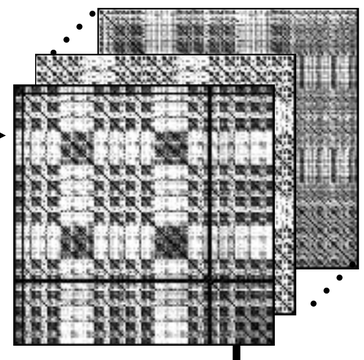
2. Left / right / centre channel separation:



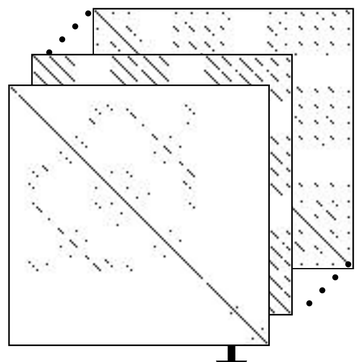
3. Harmonic / percussive separation:



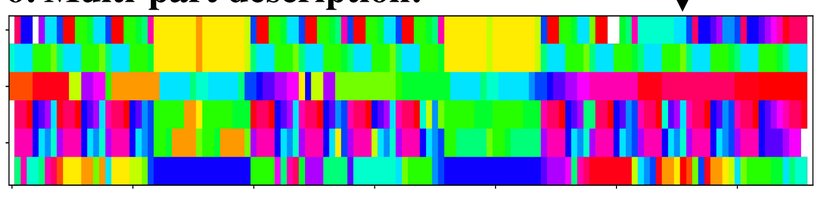
4. Chroma-based SSMs:



5. Thresholded SSMs:

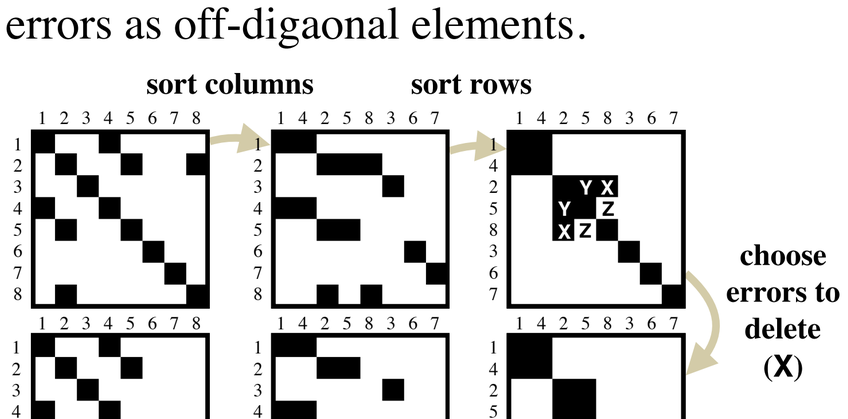


6. Multi-part description:



We introduce a novel method of enforcing transitivity in the structural analysis.

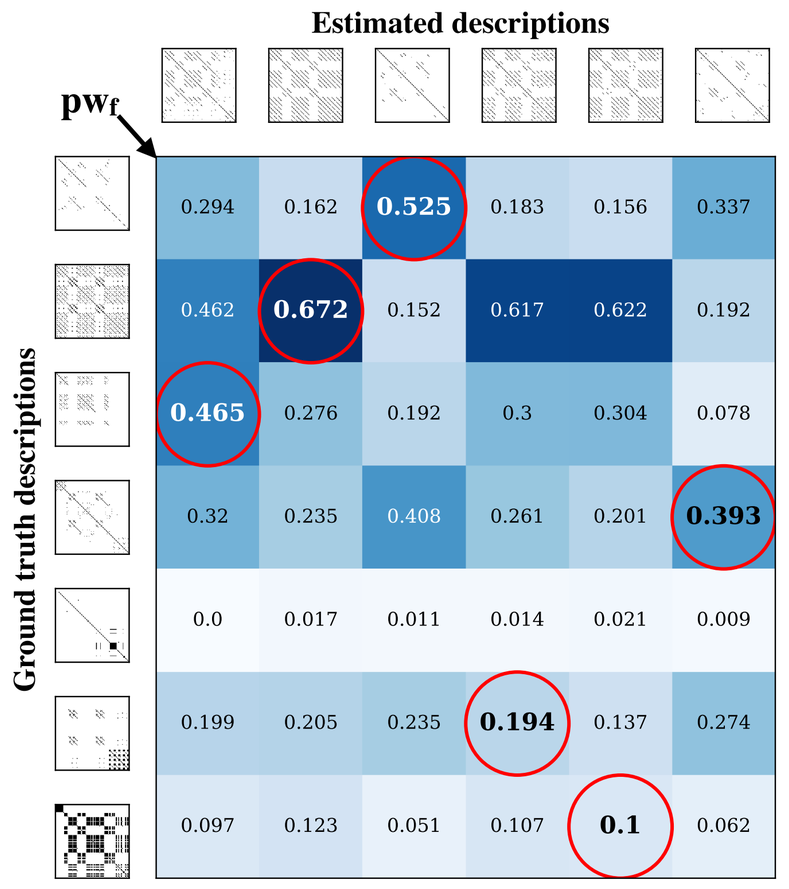
Lexical sorting transforms related sequences into blocks, exposing transitivity errors as off-digaonal elements.



We process the SSM with many thresholds, and pick the one that describes as much of the piece as possible (*coverage*) while minimizing the number of transitivity errors (*strain*).

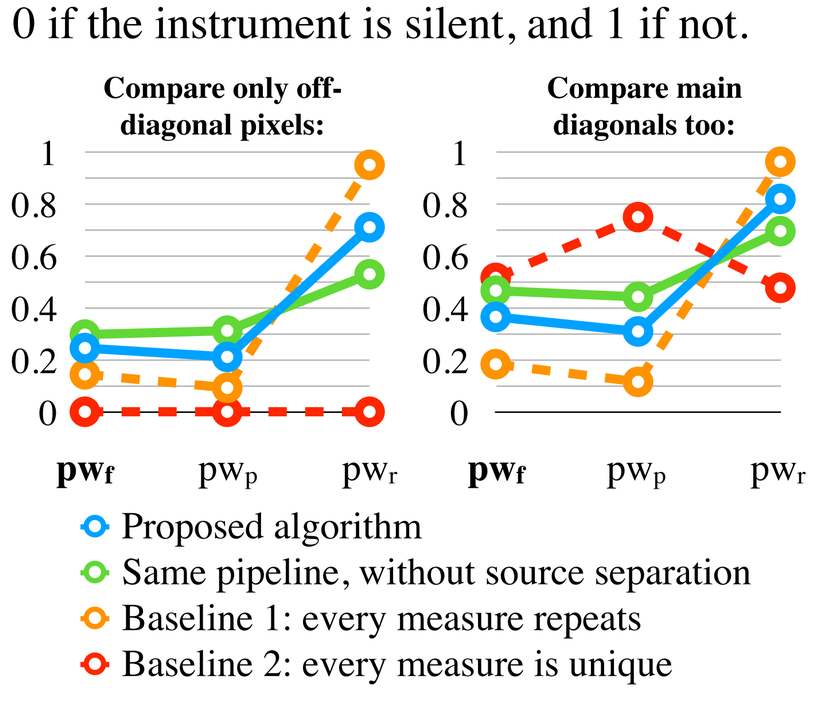
4. Evaluation

We compare each of the estimated descriptions to each of the ground truth parts, and report the best permutation.



We use conventional structure metrics: pairwise precision, recall and *f*-measure — but, should we count pixels on the main diagonal, or not?

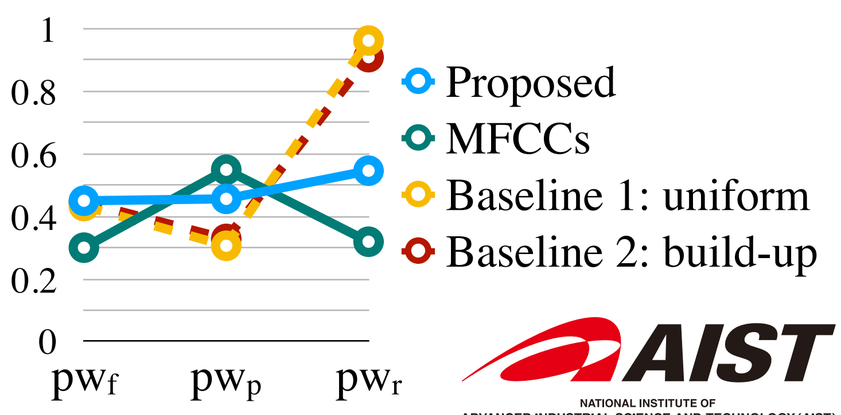
Main diagonal pixels are usually trivial and ignored, but here, main diagonal values are 0 if the instrument is silent, and 1 if not.



Comparing off-diagonal pixels, we evaluate "repetition detection" quality: our algorithm does better if we skip the source separation!

But, including diagonals, Baseline 2 does best of all — and its recall shows that 50% of the data lies on the main diagonal!

But there's one bright spot: our algorithm was best (narrowly) at labelling instrument mixtures consistently:



This work was supported in part by JST ACCEL Grant Number JPMJAC1602, Japan.