

Automatic Chord Recognition by Music Source Separation

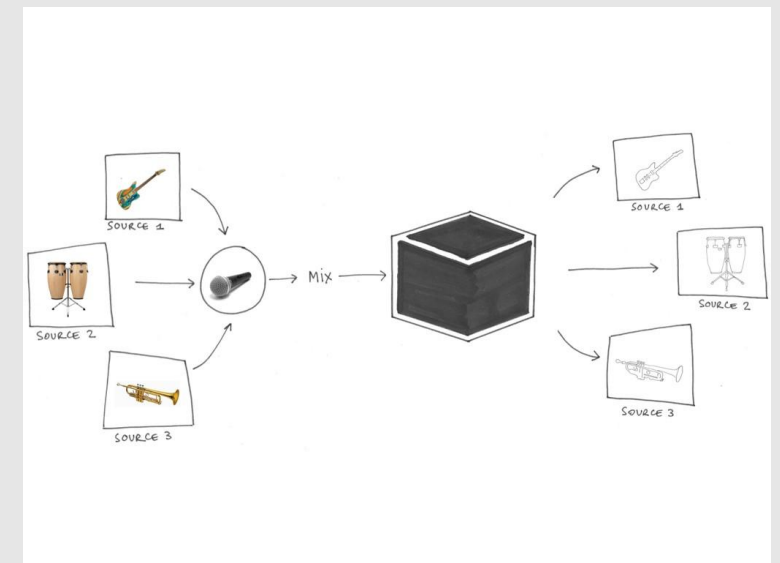
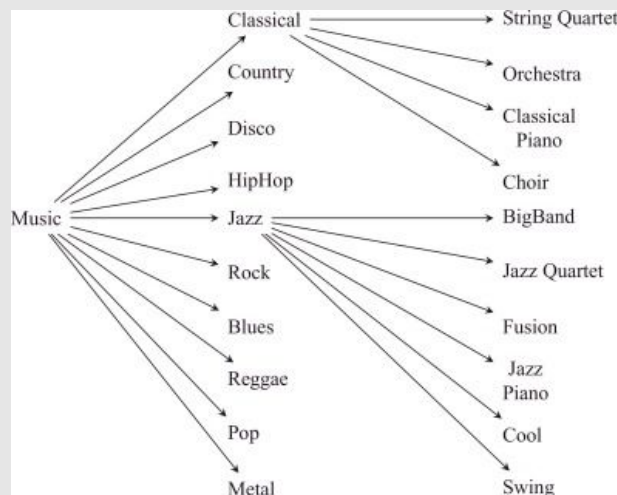
Daniel Ko
Fall 2020

CS 639: Undergraduate Computer Vision



Intro: What is ACR?

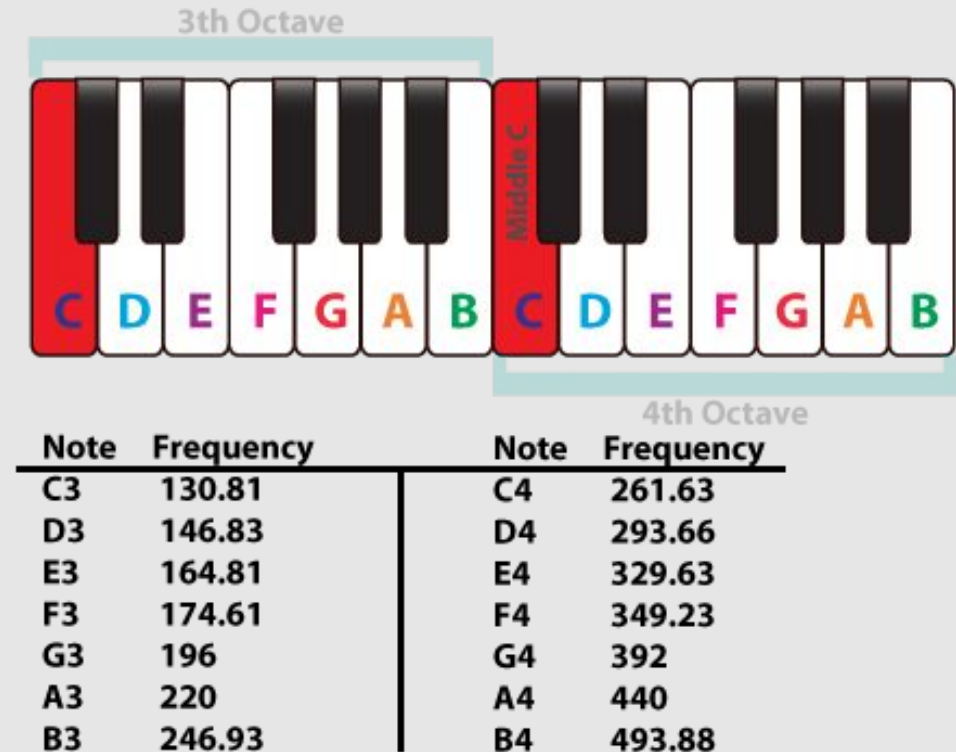
- Subfield of Music Information Retrieval (MIR)
 - Musical genre categorization
 - Music recommendation systems (i.e. personalized playlist)
 - Music source separation
- Automatic Chord Recognition (ACR) is the task of transcribing chords from music



Intro: Music Theory



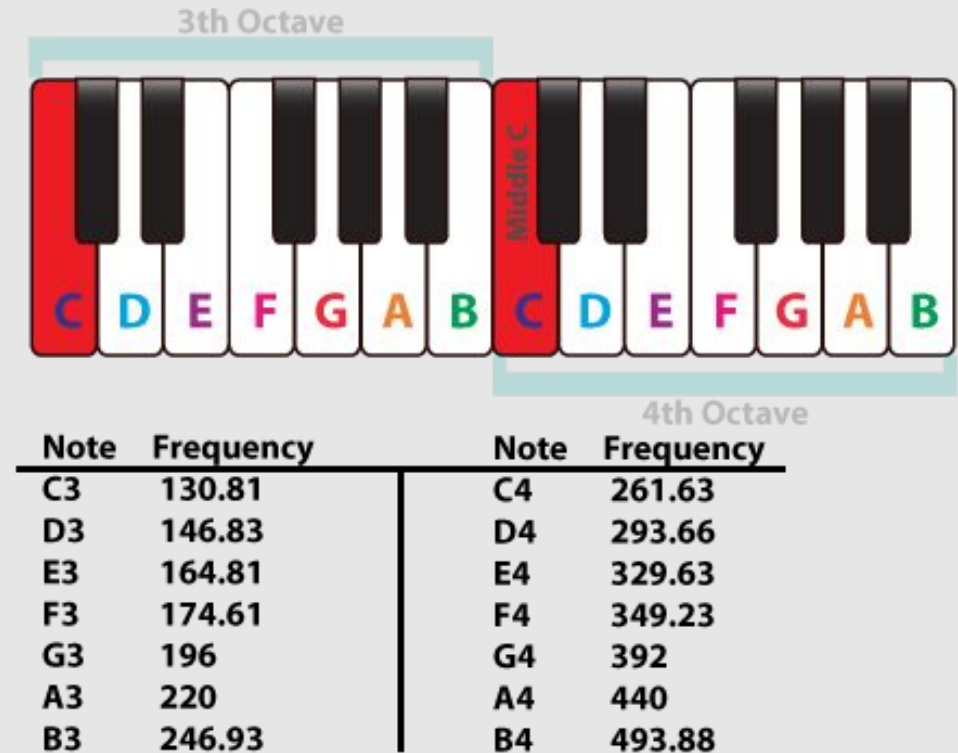
- A note represents a musical sound
 - Vibrations in the air
 - Measured by Hz
 - Perceived as pitch
 - 12 notes per octave
 - Octave is the interval of a note and the note with double its Hz
 - Logarithmic scale
 - Adjacent notes have same ratio: $2^{1/12}$



Intro: Music Theory



- Chords are a collection of notes played in unison (i.e. guitar or piano)
- Named based on the distance of notes played between notes
 - Major Triad
 - {0,4,7}
 - C major: {C,E,G}
 - Minor Triad
 - {0,3,7}
 - D minor: {D,F,A}
- Example: Let It Be - The Beatles
 - C major, G major, A minor, F major



Intro: Motivation



- Non trivial task for humans
 - Typical song takes 8 to 18 minutes by professional
 - Infeasible to transcribe large amounts at a time
- Useful to musicians
 - Medium to exchange and store musical ideas
 - Huge market for chords for songs
 - Real Book, Ultimate Guitar, etc

346

www.virtualsheetmusic.com

WHAT A WONDERFUL WORLD

- GEORGE DAVID WEISS / BOB THIELE

I see trees of green, red ros-es too, I see them bloom
for me and you, and I think to my-self what a won-der-ful
world. I see skies of blue and clouds of white, the
bright_ bless-ed day, the dark_ sa-cred night, and I think to my-self
what a won-der-ful world. The
colors of the rain-bow, so pret-ty in the sky are al-so on the fac-es of

Copyright © 1967 by Range Road Music Inc., Quartet Music and Abilene Music, Inc.
Copyright Renewed

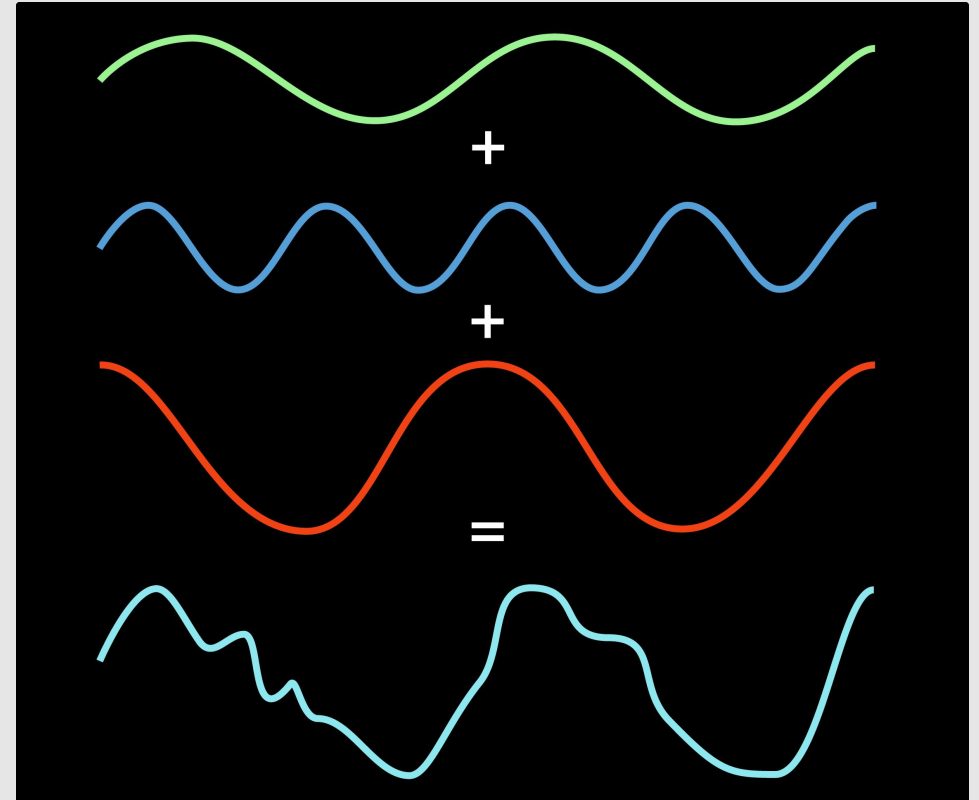
Low resolution sample

© 1999-2019 Virtual Sheet Music, Inc. & Hal Leonard Co.

Intro: Music Theory and Computer Vision



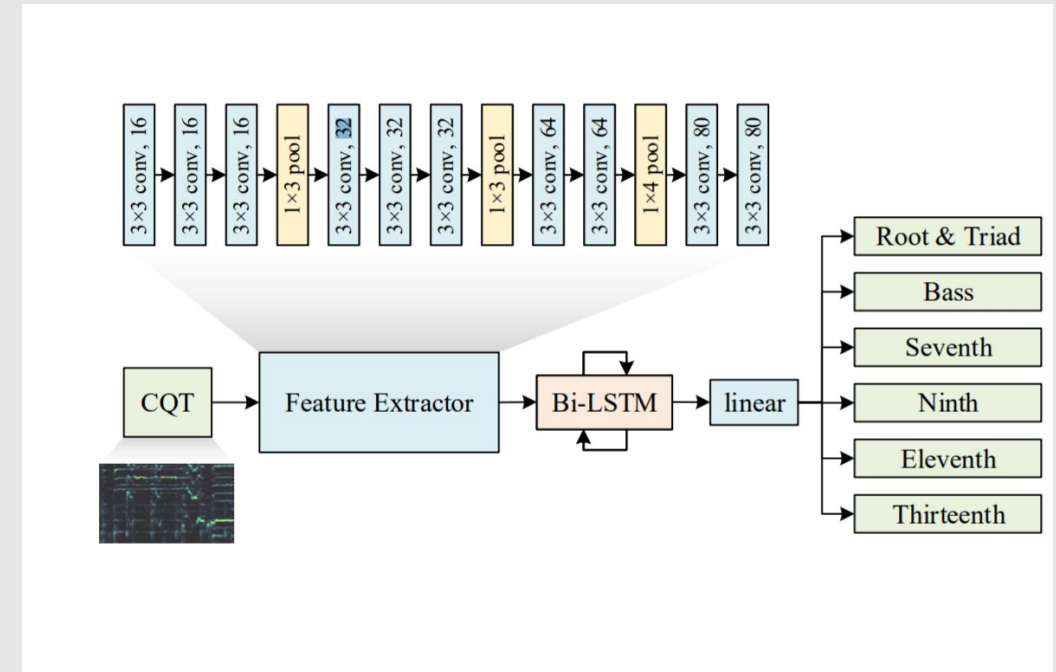
- Isn't this a computer vision class?
 - Both dealing with signals
- Decompose chord using Short-time Fourier transform
 - Distribution of frequencies per segment in song
- Convolutions to detect features



State of the Art: Junyan Jiang, et. al



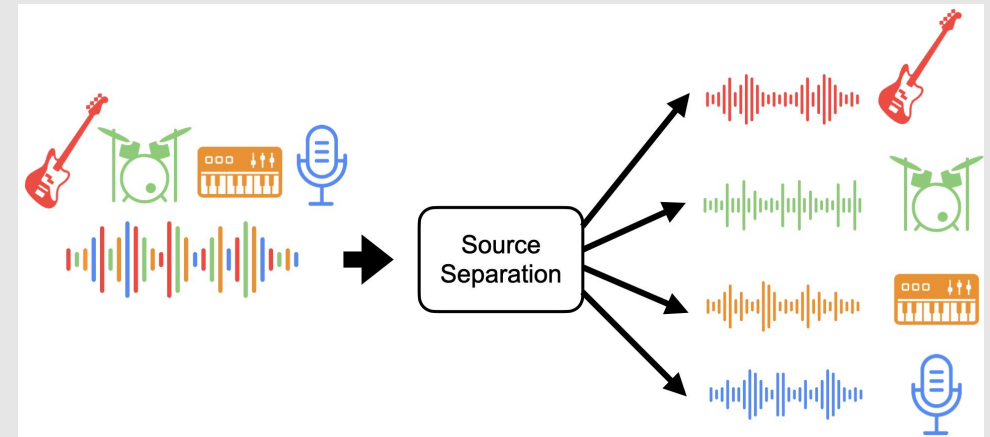
- Constant-Q transform (CQT)
 - Frequency of notes are logarithmic
⇒ using a logarithmic window function
- Convolutional Recurrent Neural Network (CRNN)
 - 9 layers of CNN with 3x3 kernel size
 - Batch normalization via max pooling
 - Bi-directional Long Short-Term Memory layer
 - Prevent frequent chord switch
- Linear output layer



Novel Approach: Music Source Separation



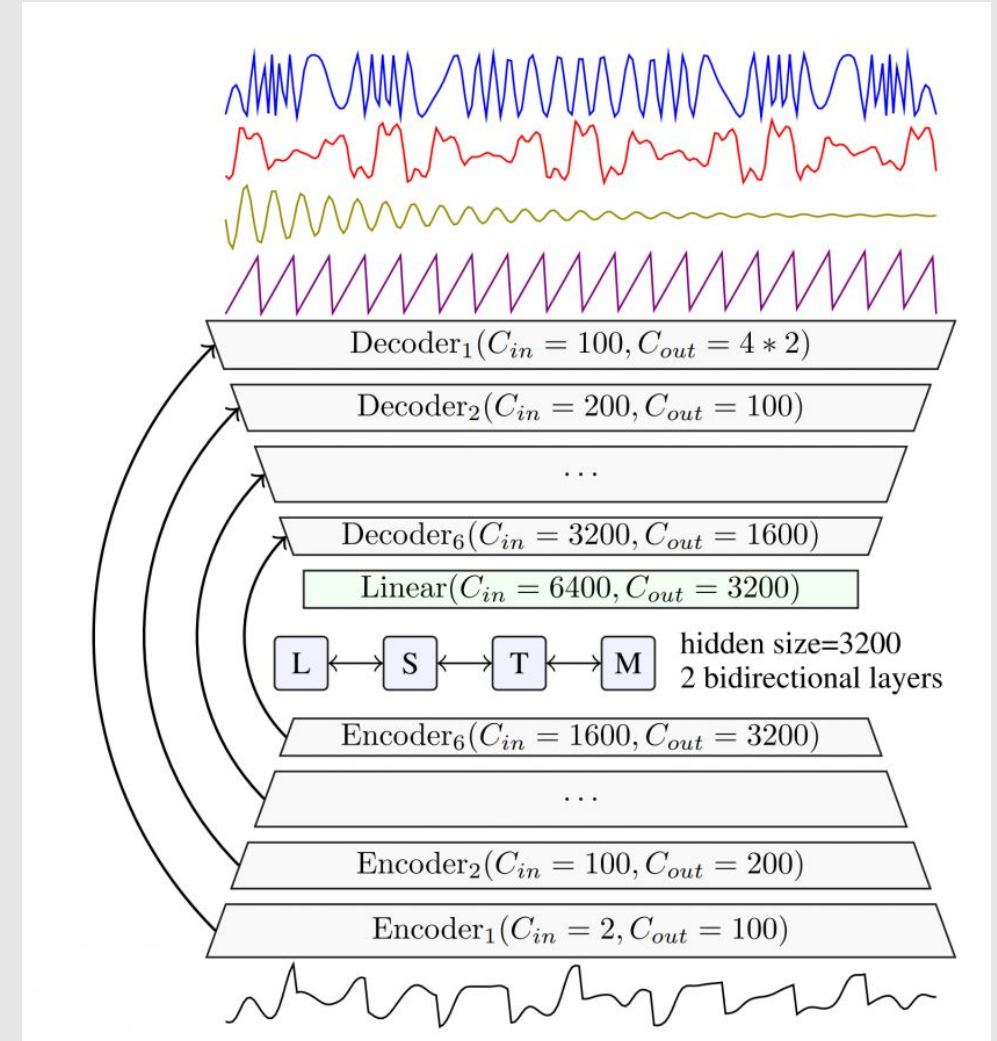
- Input before CQT/STFT is not processed
- Use Music Source Separation model to remove non chord elements
 - Song contains non chord elements i.e. drums, vocals
- Intuitively analogous to image segmentation and noise reduction



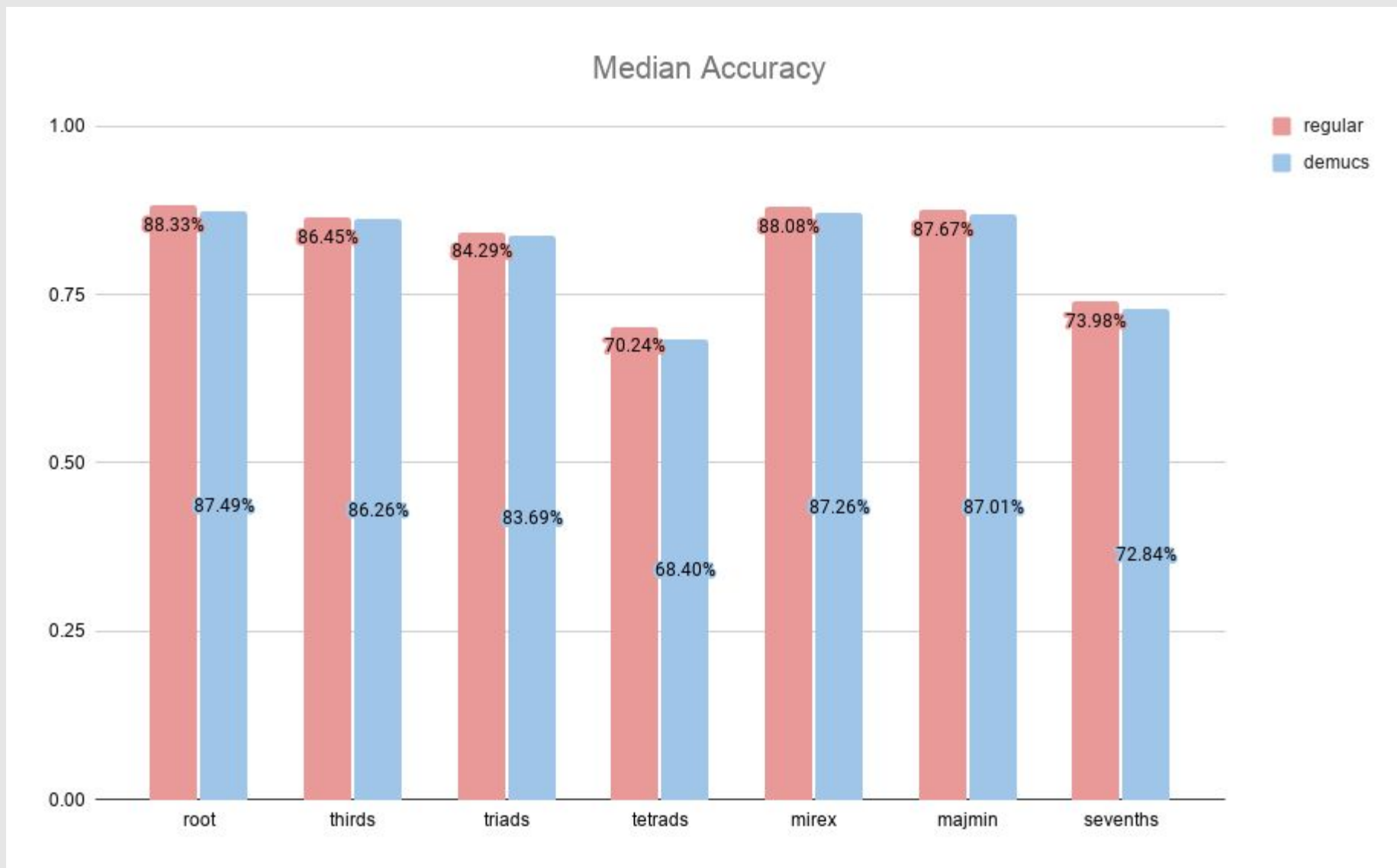
Novel Approach: Method



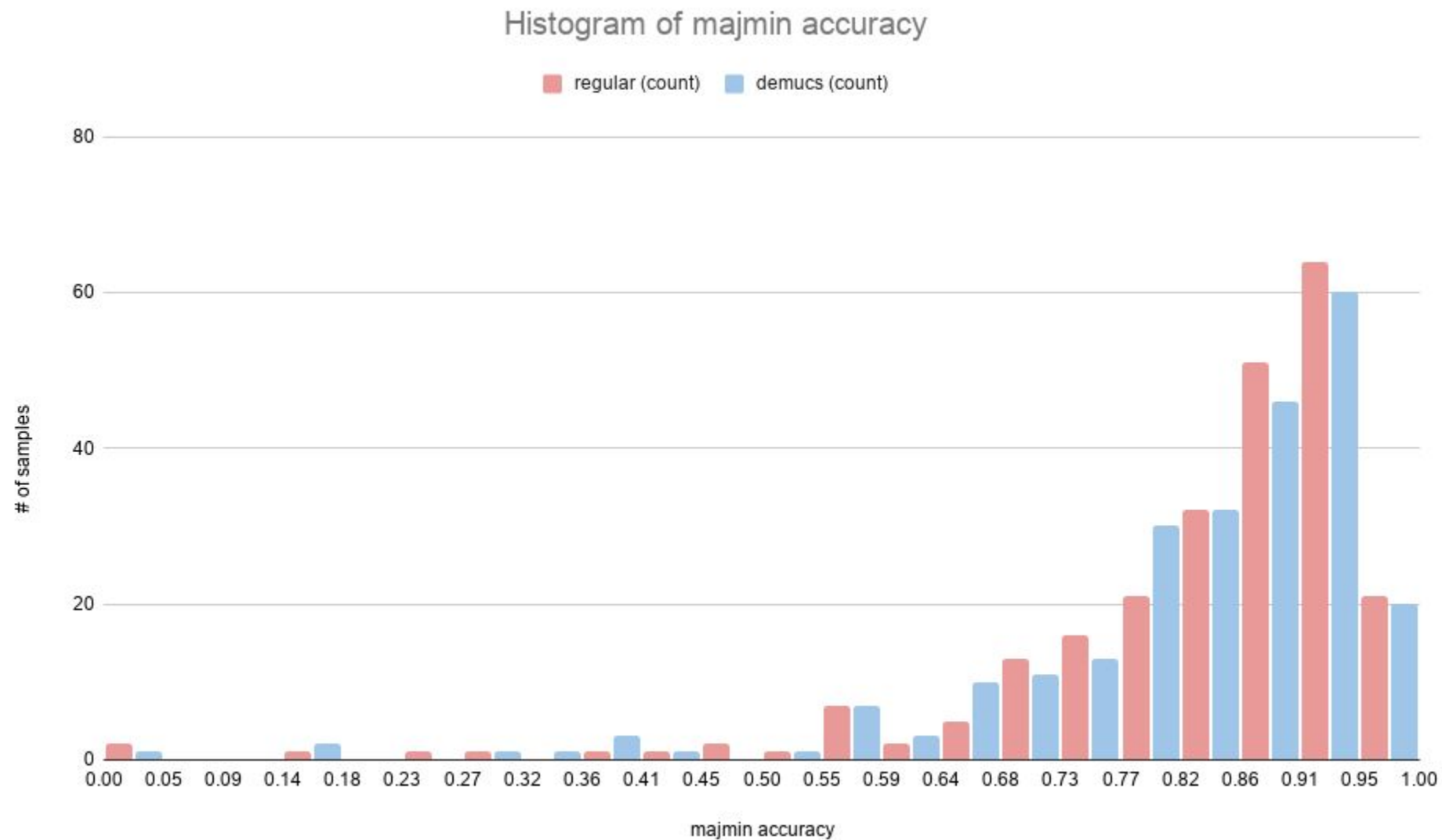
- Demucs
 - State of the art by Facebook Research
 - Bass, Drums, Other, Vocals
- Two models
 - Demucs model with Bass + Other
 - Original audio file
- Labeled dataset of 1217 songs
 - 60%, 20%, 20% for training, validation, testing
- How much do non chord elements impact current state of the art ACR systems?



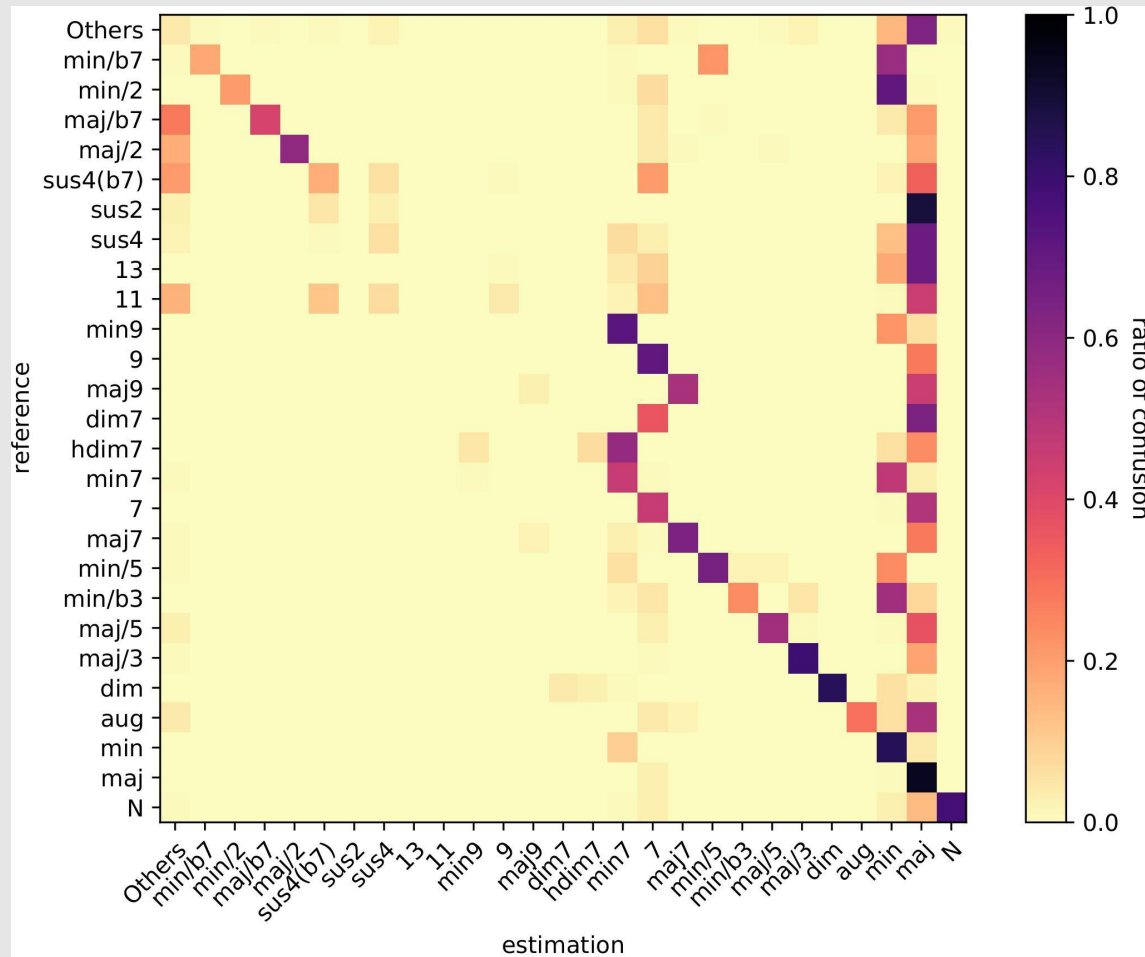
Results



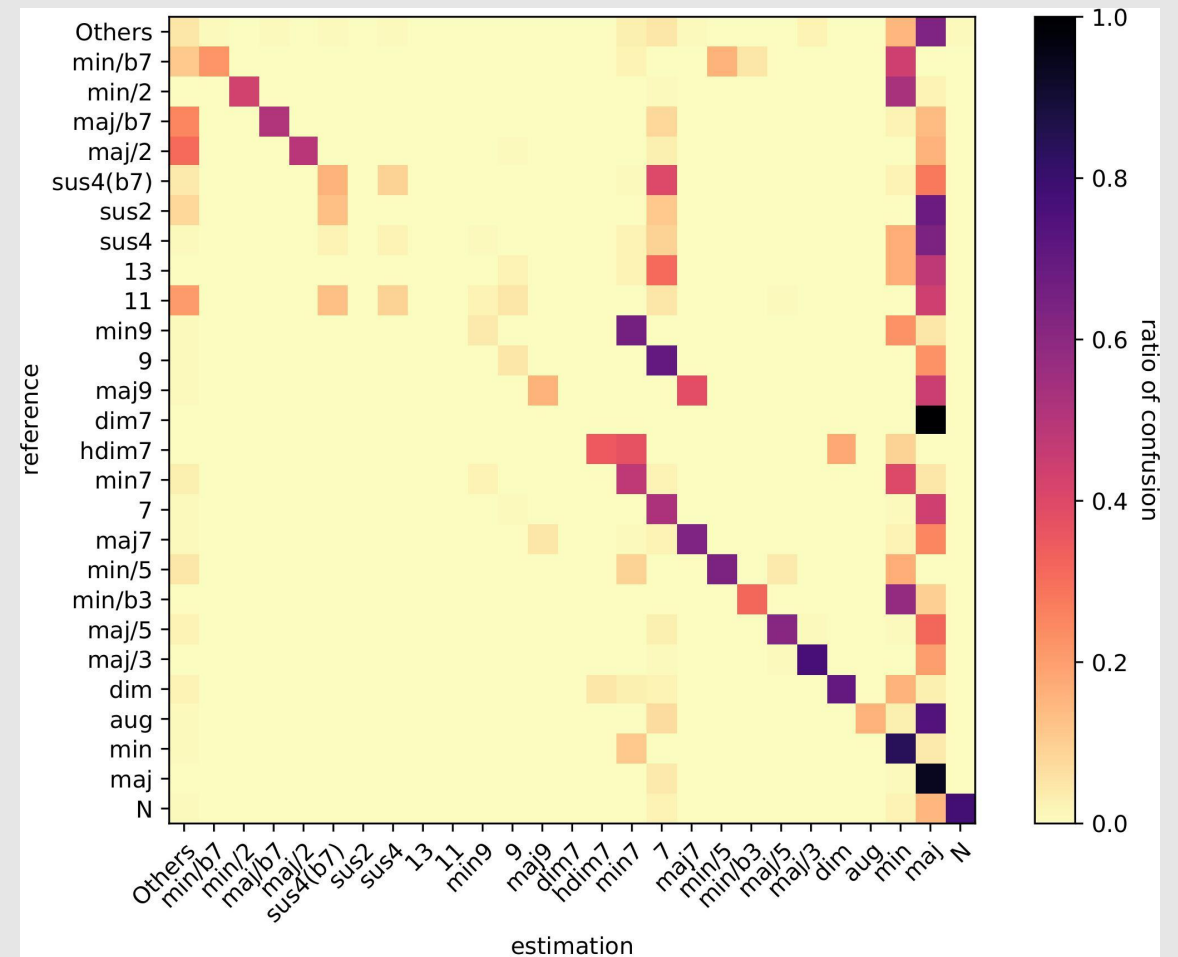
Results



Results



Demucs Model



Regular Model

Discussion



- Noise from other instruments has little impact
 - Model is deep enough
 - Poor separation
- Future direction
 - Training with music separation dataset
 - Music source separation with other ACR systems
 - Other state of the art models
 - Shallower models

Acknowledgments



- UW-Madison Center For High Throughput Computing for providing compute resources
- Junyan Jiang for providing reference model
- NYU Music and Audio Research Laboratory for dataset



References



John Ashley Burgoyne, et al. "An expert ground truth set for audio chord recognition and music analysis", 12th International Society for Music Information Retrieval Conference, 2011.

Junyan Jiang, et al. "Large-Vocabulary Chord Transcription via Chord Structure Decomposition", 20th International Society for Music Information Retrieval Conference, Delft, The Netherlands, 2019.

Alexandre Défossez, Nicolas Usunier, Léon Bottou, Francis Bach, "Music Source Separation in the Waveform Domain", fhal-02379796f, 2019

J. Pauwels and G. Peeters, "Evaluating automatically estimated chord sequences", 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013