2023 IEEE International Conference on Acoustics, Speech and Signal Processing

jazznet: A Dataset of Fundamental Piano Patterns for Music Audio Machine Learning Research

Tosiron Adegbija
Department of Electrical and Computer Engineering
University of Arizona, USA
(tosiron@arizona.edu)

THE UNIVERSITY OF ARIZONA

Highlights

A dataset of over 162k labeled piano patterns–*chords, arpeggios, scales*, and *chord progressions*–and an open-source tool to automatically generate even more patterns!

Motivation

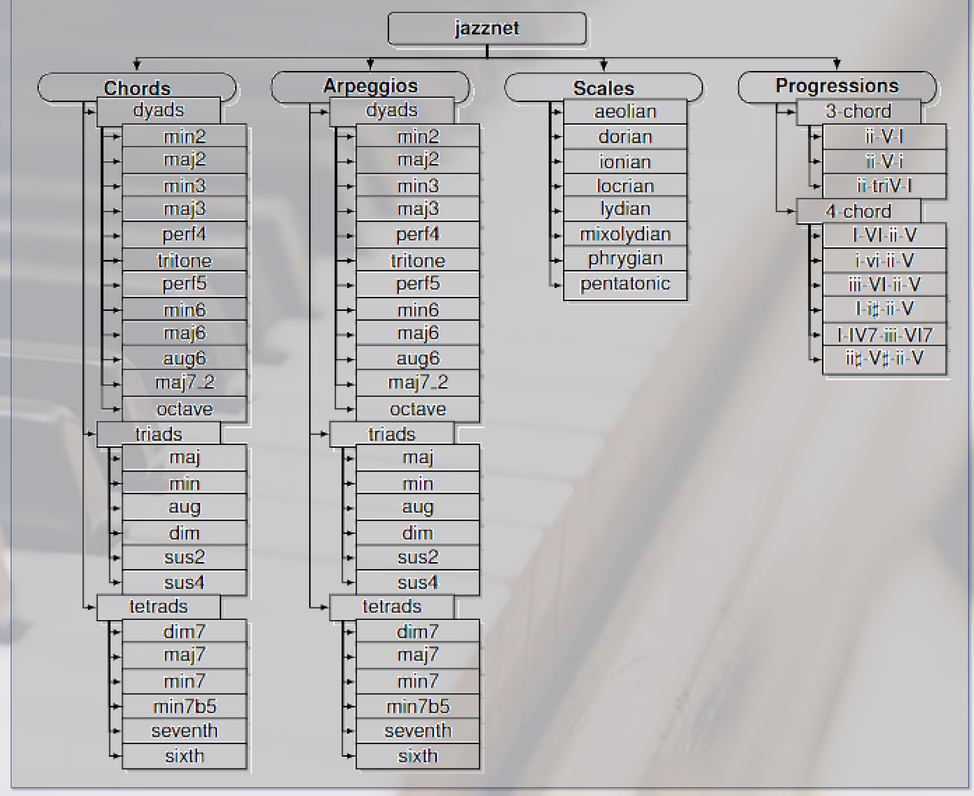
- Music lags other fields (e.g., image processing) in availability of high-quality datasets
- Most music datasets feature music pieces
- Humans learn better when they understand patterns before learning pieces
- Need music data for machine learning to mimic data for humans—patterns rather than pieces
- Need an easy way to generate music patterns

Distance-based pattern structures (DBPS)

- An algorithm to automatically generate patterns in all keys and inversions of an 88-key piano based on the distances between the pitches, represented using MIDI pitches
- E.g., Major 7th = [4, 3, 4]

Input: distance = $[d_0, d_1, d_2, ..., d_n]$; type="chord" or "arpeggio" or "scale"
Output: Pattern in all keys

```
foreach MIDI_pitch in range(24, 109) do
    note_0 ← MIDI_pitch
    time ← time #current time
    for i in range(0, len(distance)) do
        note_{i+1} ← note_i + distance[i]
        if type == "arpeggio" or "scale" then
            time ← time + 1
        end
    end
end
```



jazznet dataset

- Focuses on jazz music to make the dataset tractable; jazz music encompasses several other genres
- Piano patterns
  - Chords and arpeggios: *dyads* (2-note), *triads* (3-note), *tetrads* (4-note)
  - Scales: *diatonics* (7-note), *pentatonic* (5-note)
  - Progressions: *3-chord, 4-chord*
- Formats: MIDI and WAV
- Suggested subsets: *small* (15640 clips), *medium* (14688 clips), *large* (36720 clips)

Experiments and conclusion

- Two models using medium subset:
  - Convolutional recurrent neural network (CRNN) and deep CNN (M5)
    - 4 convolutional layers, ReLU activation, Adam optimizer, categorical cross-entropy loss function
    - CRNN input is Mel-spectrograms and M5 input is time series waveform
    - Mean average precision (mAP) for mode prediction (e.g., aug vs. dim) very low for both CRNN and M5: 0.63 and 0.30, respectively; AP as low as 0.06 for M5 on scales
    - Opportunities for improvement given the complexity of the dataset
- Dataset can be used for several tasks, e.g., music recommendation systems, music generation, music transcription, etc.
- We hope the dataset helps researchers to benchmark machine learning models for complex music information retrieval (MIR) tasks

Dataset statistics

Type	#modes	#total	time (s)	# hours	size (GB)
Chords	24	5,525	3	259	1
Arpeggios	24	5,525	4; 5; 6	433	1.7
Scales	8	4,590	9; 7	674	6.3
Progressions	9	146,880	7; 10	25,568	85.6
Total	65	162,520		26,934	95

Github page + paper + Zenodo repo