# From West to East: Who can understand the music of the others better?

Charilaos Papaioannou
NTUA, QMUL
cpapaioan@mail.ntua.gr

Emmanouil Benetos
QMUL
emmanouil.benetos@qmul.ac.uk

Alexandros Potamianos
NTUA
potam@central.ntua.gr

## Highlights

- **Existing music audio embedding models** can be used to transfer or learn representations to **"non-Western" cultures**
- The inverse transfer direction, utilizing learned **representations from non-Western** datasets, can be **beneficial for Western** target domains
- The aggregation of all the cross-domain knowledge transfers can **provide insights about the similarities between the domains**, seeking to answer the question of the paper

## 1. Datasets

### Western

- **MagnaTagATune**: 210 hours – top-50 tags
- **FMA-medium**: 208 hours – top-20 hierarchical genres

### Eastern Mediterranean

- **Lyra** (Greek folk): 80 hours – top-30 tags
- **Turkish-makam**: 215 (out of 359) hours – top-30 tags

### Indian

- **Hindustani**: 206 (out of 343) hours – top-20 tags
- **Carnatic**: 218 (out of 503) hours – top-20 tags

| MagnaTagATune | | FMA-medium | | Lyra | |
|---|---|---|---|---|---|
| Guitar | 18.76% | Rock | 28.41% | Voice | 76.21% |
| Classical | 16.52% | Electronic | 25.26% | Traditional | 76.05% |
| Slow | 13.71% | Punk | 13.28% | Violin | 57.34% |
| Techno | 11.42% | Experimental | 9.00% | Percussion | 53.71% |
| Strings | 10.55% | Hip-Hop | 8.80% | Laouto | 51.69% |
| Drums | 10.05% | Folk | 6.08% | Guitar | 37.34% |
| Electronic | 9.74% | Garage | 5.67% | Klarino | 31.05% |
| Rock | 9.17% | Instrumental | 5.40% | Nisiotiko | 26.85% |
| Fast | 8.92% | Indie-Rock | 5.17% | Place-None | 25.16% |
| Piano | 7.95% | Pop | 4.74% | Bass | 24.76% |

| Turkish-makam | | Hindustani | | Carnatic | |
|---|---|---|---|---|---|
| Voice | 63.33% | Voice | 83.90% | Voice | 82.35% |
| Kanun | 31.09% | Tabla | 53.03% | Violin | 78.45% |
| Tanbur | 27.93% | Khayal | 41.33% | Mridangam | 75.65% |
| Ney | 27.56% | Harmonium | 39.25% | Kriti | 70.87% |
| Orchestra | 26.38% | Teentaal | 35.35% | Adi | 51.88% |
| Oud | 24.36% | Tambura | 27.88% | Ghatam | 30.32% |
| Kemence | 22.79% | Ektaal | 21.58% | Khanjira | 17.65% |
| Cello | 17.83% | Pakhavaj | 7.88% | Rupaka | 11.98% |
| Violin | 17.62% | Sarangi | 7.30% | Mishra chapu | 7.27% |
| Hicaz | 10.63% | Dhrupad | 7.05% | Tana Varnam | 5.21% |

TABLE 1. Relative frequencies of the top-10 tags

## 2. Models

- **VGG-ish**: 7-layer CNN with 3×3 filters and 2×2 max-pooling, followed FC layers – mel-specs of 3.69sec length
- **Musicnn**: vertical and horizontal convolutional filters to capture timbral and temporal features followed by dense layers – mel-specs of 3sec length
- **Audio Spectrogram Transformer**: 16×16 patches of input, trainable positional embeddings, encoder part of the Transformer – mel-specs of 8sec length

| Model | VGG-ish | | Musicnn | | AST | |
|---|---|---|---|---|---|---|
| Metric / Dataset | ROC-AUC | PR-AUC | ROC-AUC | PR-AUC | ROC-AUC | PR-AUC |
| **MagnaTagATune** | 0.9123 | 0.4582 | 0.9019 | 0.4333 | **0.9172** | **0.4654** |
| **FMA-medium** | **0.8889** | 0.4949 | 0.8766 | 0.4473 | 0.8886 | **0.5024** |
| **Lyra** | 0.8097 | 0.4806 | 0.7391 | 0.4042 | **0.8476** | **0.5333** |
| **Turkish-makam** | **0.8696** | 0.5639 | 0.8505 | 0.5299 | 0.8643 | **0.5669** |
| **Hindustani** | **0.8477** | **0.6082** | 0.8471 | 0.6016 | 0.8307 | 0.5786 |
| **Carnatic** | 0.7392 | 0.4278 | 0.7496 | 0.4182 | **0.7706** | **0.4394** |

TABLE 2. Model performance on single domain auto-tagging tasks

## 3. Cross-cultural Music Transfer Learning

- Transfer of a trained model from to a target domain and **fine-tuning** of the **output layer** or of the **whole network**
- **Aggregation** of all knowledge transfer results to **specify which source is the best candidate for each target dataset** and derive insights about domain similarity

| Target domain | MagnaTagATune | | FMA-medium | | Lyra | | Turkish-makam | | Hindustani | | Carnatic | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| trainable layer(s) / Source domain | output | all | output | all | output | all | output | all | output | all | output | all |
| **VGG-ish** | | | | | | | | | | | | |
| **MagnaTagATune** | - | 91.23 | 88.11 | **92.39** | 74.69 | **85.40** | 76.79 | 86.84 | 76.09 | 85.04 | 67.19 | **74.71** |
| **FMA-medium** | 85.82 | **91.29** | - | 88.89 | 68.56 | 84.04 | 75.40 | **87.78** | 75.77 | 84.39 | 67.03 | 74.56 |
| **Lyra** | 84.34 | 90.93 | 82.84 | 92.10 | - | 80.97 | 76.98 | 87.21 | 77.41 | 84.24 | 67.30 | 73.52 |
| **Turkish-makam** | 85.19 | 90.90 | 84.41 | 91.74 | 70.93 | 82.38 | - | 86.96 | 77.54 | **85.32** | 67.16 | 73.50 |
| **Hindustani** | 84.24 | 91.02 | 83.83 | 91.91 | 66.27 | 79.71 | 77.25 | 87.63 | - | 84.77 | 66.72 | 74.63 |
| **Carnatic** | 84.18 | 91.00 | 82.62 | 91.73 | 61.59 | 76.72 | 77.07 | 87.40 | 78.19 | 84.81 | - | 73.92 |
| **Musicnn** | | | | | | | | | | | | |
| **MagnaTagATune** | - | 90.19 | 87.34 | **91.03** | 71.79 | 78.74 | 74.72 | **85.96** | 75.87 | 84.18 | 66.12 | 75.57 |
| **FMA-medium** | 85.52 | **90.35** | - | 87.66 | 65.94 | 77.59 | 75.51 | 85.13 | 73.16 | **85.49** | 66.38 | 75.77 |
| **Lyra** | 81.38 | 90.03 | 82.23 | 90.80 | - | 73.91 | 74.11 | 85.20 | 78.10 | 83.29 | 65.09 | 75.51 |
| **Turkish-makam** | 84.35 | 90.11 | 83.79 | 90.81 | 61.87 | **79.83** | - | 85.05 | 75.67 | 83.75 | 67.49 | 74.09 |
| **Hindustani** | 82.38 | 89.86 | 83.42 | 90.85 | 64.48 | 78.95 | 74.60 | 85.58 | - | 84.71 | 65.25 | **76.95** |
| **Carnatic** | 83.02 | 90.05 | 82.78 | 90.74 | 61.83 | 77.92 | 75.09 | 85.43 | 75.34 | 84.19 | - | 74.96 |
| **AST** | | | | | | | | | | | | |
| **MagnaTagATune** | - | **91.72** | 89.25 | 91.99 | 75.68 | 83.77 | 76.28 | 87.20 | 74.67 | **86.57** | 66.03 | 75.43 |
| **FMA-medium** | 88.63 | 91.62 | - | 88.86 | 65.72 | 82.17 | 76.37 | **87.43** | 74.51 | 85.76 | 67.33 | 75.98 |
| **Lyra** | 87.49 | 91.44 | 87.44 | **92.43** | - | **84.76** | 77.08 | 86.80 | 72.24 | 83.73 | 68.47 | 76.59 |
| **Turkish-makam** | 87.33 | 91.40 | 86.31 | 91.95 | 72.70 | 77.95 | - | 86.43 | 70.13 | 83.56 | 67.10 | 75.23 |
| **Hindustani** | 87.40 | 91.35 | 87.11 | 92.26 | 71.74 | 84.60 | 75.70 | 86.90 | - | 83.07 | 67.75 | 75.85 |
| **Carnatic** | 87.42 | 91.45 | 86.83 | 91.75 | 63.33 | 81.44 | 76.87 | 87.14 | 74.11 | 82.91 | - | **77.06** |

TABLE 3. ROC-AUC scores (%) when applying transfer learning across all models and domains

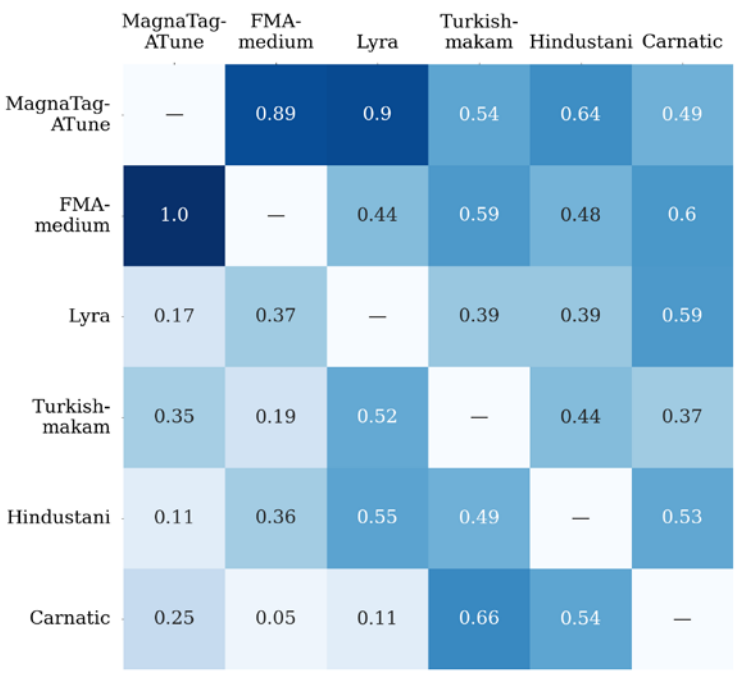


FIG 1. Cross-cultural music transfer learning results. **Rows** correspond to the **source** datasets and **columns** to the **target** dataset. The value of each cell (knowledge transfer) is **normalized and averaged** across all models and fine-tuning methods.

## 4. Conclusions

- State-of-the-art models can benefit from knowledge transfer not **only from Western to non-Western cultures but also the opposite** too
- **Transfer Learning** results can be interpreted to a degree **as a similarity metric between the music cultures**

National Technical University of Athens

Queen Mary University of London