CON ESPRESSIONE!
AI, Machine Learning, and Musical Expressivity

Gerhard Widmer
Institute of Computational Perception, Johannes Kepler University Linz (JKU)
AI Lab, Linz Institute of Technology (LIT)
ELLIS Unit Linz

AI & MUSIC?
A LITTLE EXPERIMENT

IDENTIFY THE HUMAN PIANIST


(1) (2) (3)
**COMPUTATIONAL MUSIC PERCEPTION**

**MUSIC DETECTION**


<table>
<thead>
<tr>
<th>Test week</th>
<th>GT mode</th>
<th>True ratio (%)</th>
<th>Est. Ratio (%)</th>
<th>Accuracy (%)</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F-Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>min</td>
<td>77.63</td>
<td>77.39</td>
<td>96.66</td>
<td>96.69</td>
<td>97.84</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>max</td>
<td>78.95</td>
<td>77.39</td>
<td>96.59</td>
<td>97.80</td>
<td>97.82</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ave</td>
<td>78.29</td>
<td>77.39</td>
<td>96.62</td>
<td>96.40</td>
<td>97.27</td>
<td>97.83</td>
</tr>
<tr>
<td>1</td>
<td>same</td>
<td>78.67</td>
<td>77.39</td>
<td>97.25</td>
<td>97.69</td>
<td>98.24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>min</td>
<td>80.09</td>
<td>79.43</td>
<td>96.73</td>
<td>96.36</td>
<td>97.95</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>max</td>
<td>82.07</td>
<td>79.43</td>
<td>96.21</td>
<td>95.27</td>
<td>96.09</td>
<td>97.65</td>
</tr>
<tr>
<td>2</td>
<td>ave</td>
<td>81.08</td>
<td>79.43</td>
<td>96.47</td>
<td>98.81</td>
<td>96.81</td>
<td>97.80</td>
</tr>
<tr>
<td>2</td>
<td>same</td>
<td>81.70</td>
<td>80.29</td>
<td>97.40</td>
<td>99.27</td>
<td>97.54</td>
<td>98.40</td>
</tr>
<tr>
<td>3</td>
<td>min</td>
<td>74.42</td>
<td>73.89</td>
<td>94.13</td>
<td>96.38</td>
<td>95.70</td>
<td>96.04</td>
</tr>
<tr>
<td>3</td>
<td>max</td>
<td>76.04</td>
<td>73.89</td>
<td>94.30</td>
<td>98.27</td>
<td>94.26</td>
<td>98.22</td>
</tr>
<tr>
<td>3</td>
<td>ave</td>
<td>75.73</td>
<td>73.89</td>
<td>94.21</td>
<td>97.33</td>
<td>94.97</td>
<td>96.13</td>
</tr>
<tr>
<td>3</td>
<td>same</td>
<td>76.43</td>
<td>74.45</td>
<td>95.40</td>
<td>98.24</td>
<td>95.70</td>
<td>96.95</td>
</tr>
</tbody>
</table>
COMPUTATIONAL MUSIC PERCEPTION

RHYTHMIC STRUCTURE: Beat, Tempo, Measures/Bars, Rhythm

MUSIC DETECTION


Florian Krebs

IEEE Signal Processing Cup 2017
ICASSP 2017, New Orleans
COMPUTATIONAL MUSIC PERCEPTION

HARMONIC STRUCTURE: Key, Chords, Chord Sequences
RHYTHMIC STRUCTURE: Beat, Tempo, Measures/Bars, Rhythm

MUSIC DETECTION

Key: C major

E7/b5  A7  D7/b9  G7  Db7  Cmaj7


COMPUTATIONAL MUSIC PERCEPTION

SEGMENT STRUCTURE: Chorus, Verse, Song Boundaries, ...
HARMONIC STRUCTURE: Key, Chords, Chord Sequences
RHYTHMIC STRUCTURE: Beat, Tempo, Measures/Bars, Rhythm

MUSIC DETECTION

COMPUTATIONAL MUSIC PERCEPTION

GENRE, STYLE: Jazz / Rock / Pop / Folk / HipHop / ...
SEGMENT STRUCTURE: Chorus, Verse, Song Boundaries, ....
HARMONIC STRUCTURE: Key, Chords, Chord Sequences
RHYTHMIC STRUCTURE: Beat, Tempo, Measures/Bars, Rhythm
MUSIC DETECTION

GENERAL MUSICAL SIMILARITY (e.g., for recommendation)
GENRE, STYLE: Jazz / Rock / Pop / Folk / HipHop / ...
SEGMENT STRUCTURE: Chorus, Verse, Song Boundaries, ....
HARMONIC STRUCTURE: Key, Chords, Chord Sequences
RHYTHMIC STRUCTURE: Beat, Tempo, Measures/Bars, Rhythm
MUSIC DETECTION
REAL-TIME MUSIC TRACKING

[Waveform image]

Video

REAL-TIME PIECE IDENTIFICATION AND TRACKING

Pianist: Cynthia Liem

REAL-TIME PIECE IDENTIFICATION

- Live audio
- Note transcription (Deep NN)
- Fingerprint generation
- Database lookup
- Tempo correction & Candidate evaluation
- Multi-agent tracking
- Piece ID & Score Position

CONCERTGEBOUW AMSTERDAM, Dec. 20, 2014

Royal Concertgebouw Orchestra, Amsterdam
Mariss Jansons, conductor

COMPUTATIONAL MUSIC PERCEPTION

REAL-TIME IDENTIFICATION AND TRACKING
GENERAL MUSICAL SIMILARITY (e.g., for recommendation)
GENRE, STYLE: Jazz / Rock / Pop / Folk / HipHop / ...
SEGMENT STRUCTURE: Chorus, Verse, Repeats, ....
HARMONIC STRUCTURE: Key, Chords, Chord Sequences
RHYTHMIC STRUCTURE: Beat, Tempo, Measures/Bars, Rhythm
MUSIC DETECTION

THEN WHAT IS HARD?
COMPOSING NEW MUSIC

MuseNet
We've created MuseNet, a deep neural network that can generate 4-minute musical compositions with 10 different instruments, and can combine styles from country to Mozart to the Beatles. MuseNet was not explicitly programmed with our understanding of music, but instead discovered patterns of harmony, rhythm, and style by learning to predict the next token in hundreds of thousands of MIDI files. MuseNet uses the same general-purpose unsupervised technology as GPT-2, a large-scale transformer model trained to predict the next token in a sequence, whether audio or text.

PLAYING MUSIC EXPRESSIVELY

Frederic Chopin, Nocturne Op. 9 No. 1 Bb minor

Arthur Rubinstein (1967):

Nikita Magaloff (1989):
EXPRESSIVE MUSIC PERFORMANCE

Chopin: Etude Op.10 no.3, E major, mm. 1-21

EXPRESSIVE MUSIC PERFORMANCE

Chopin: Etude Op.10 no.3, E major, mm. 1-21
EXPRESSIVE MUSIC PERFORMANCE

- Communicate musical structure
- Communicate artistic individuality and intentions
- Communicate expressive qualities
- Communicate / evoke emotions

→ need to understand how music is constructed, and how it is perceived

THE CON ESPRESSIONE PROJECT

European Research Council
Established by the European Commission

THE CON ESPRESSIONE PROJECT

- What kinds of expressive qualities do listeners perceive / differentiate?
- What is it in a performance that communicates an expressive quality?
- Can machines learn to recognise expressive qualities?
- Can machines learn to play music “expressively”?
- Can machines become truly ‘musical’ companions?
- … and what can we learn from all this?
THE CON ESPRESSIONE GAME

1.515 individual descriptions for a total of 3.166 terms, of which 1.415 are unique (approx. 45%)
THE CON ESPRESSIONE GAME

THE CON ESPRESSIONE PROJECT

- What kinds of expressive qualities do listeners perceive / differentiate?
- What is it in a performance that communicates an expressive quality?
- Can machines learn to recognise expressive qualities?
- Can machines learn to play music "expressively"?
- Can machines become truly 'musical' companions?
- … and what can we learn from all this?
PREDICTING AND EXPLAINING EXPRESSIVE QUALITIES?

Intuitive musical percepts ("mid-level features")

Perceived expressive qualities

PREDICTING AND EXPLAINING EXPRESSIVE QUALITIES?

Perceived expressive qualities

<table>
<thead>
<tr>
<th>Perceptual Feature</th>
<th>Question asked to human raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melodiousness</td>
<td>To which excerpt do you feel like singing along?</td>
</tr>
<tr>
<td>Articulation</td>
<td>Which has more sounds with unclear articulation?</td>
</tr>
<tr>
<td>Rhythmic Stability</td>
<td>Imagine marching along with the music. Which is easier to march along with?</td>
</tr>
<tr>
<td>Rhythmic Complexity</td>
<td>Is it difficult to find the meter? Does the rhythm have many faults?</td>
</tr>
<tr>
<td>Dissonance</td>
<td>Which excerpt has more dissonance?</td>
</tr>
<tr>
<td>Temporal Stability</td>
<td>Was it easier to determine the tonic and key? In which excerpt are there more modulations?</td>
</tr>
<tr>
<td>Modality (&quot;Minlessness&quot;)</td>
<td>Imagine accompanying this song with chords. Which song would have more minor chords?</td>
</tr>
</tbody>
</table>

PREDICTING AND EXPLAINING EXPRESSIVE QUALITIES?

THE CON ESPRESSIONE PROJECT

- What kinds of expressive qualities do listeners perceive / differentiate?
- What is it in a performance that communicates an expressive quality?
- Can machines learn to recognise expressive qualities?
- Can machines learn to **play music** “expressively”?
- Can machines become truly ‘musical’ companions?
- … and what can we learn from all this?
COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE

Predictive Model

Predictive Model

<table>
<thead>
<tr>
<th>note</th>
<th>( \phi_{n=0}(\cdot) )</th>
<th>( \phi_{n=1}(\cdot) )</th>
<th>( \phi_{n=2}(\cdot) )</th>
<th>( \phi_{n=3}(\cdot) )</th>
<th>( \phi_{n=4}(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>1 0 ( \frac{23}{52} ) ( \frac{31}{50} )</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n2</td>
<td>0 0 ( \frac{12}{50} ) ( \frac{11}{25} )</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n3</td>
<td>0 0 ( \frac{23}{52} ) ( \frac{31}{50} )</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n4</td>
<td>0 0 ( \frac{12}{50} ) ( \frac{11}{25} )</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n5</td>
<td>0 1 ( \frac{12}{50} ) ( \frac{11}{25} )</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n6</td>
<td>0 0 ( \frac{23}{52} ) ( \frac{31}{50} )</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

J\textsuperscript{LYU}

J\textsuperscript{LYU}
COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE

Linear mapping
\[ \hat{y} = f(x; w) = w^T \varphi(x) \]

Non-linear mapping
\[
\begin{align*}
\hat{y}_i &= f^{(L)} \left( w^{(L)}^T h_i^{(L-1)} + w_0^{(L)} \right) \\
h_i^{(l)} &= f^{(l)} \left( w^{(l)}^T h_i^{(l-1)} + w_0^{(l)} \right) \\
h_i^{(0)} &= \varphi(x_i)
\end{align*}
\]

*Cancino Chacón, C., Gadermaier, T., Widmer, G. and Grachten, M. (2017)*
An Evaluation of Linear and Non-linear Models of Expressive Dynamics in Classical Piano and Symphonic Music.
Machine Learning 106(6), 887-908.
**COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE**


<table>
<thead>
<tr>
<th>node</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Non-linear mapping


**COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE**

<table>
<thead>
<tr>
<th>node</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Non-linear mapping

COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE

Formel für Nested multi-layer non-linear Model (DNN)


COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE

Formel für Nested multi-layer non-linear Model (DNN)


COMPUTATIONAL MODELS OF EXPRESSIVE PERFORMANCE

Non-linear mapping

\[ IC(v_n) = - \log p(v_n | v_{n-1}, v_{n-2}, \ldots) \]

\[ H(v_n) = E[ - \log p(v_n | v_{n-1}, v_{n-2}, \ldots) ] \]

QUANTITATIVE EVALUATION

MUSICAL INSIGHTS

Figure 5. Distribution of rubato model parameter values over pianists; the size of the symbols is proportional to goodness of fit of the parameters to the data.

Non-expressive literal performance:

The bi-directional recurrent non-linear model (2017):
A MUSICAL “TURING TEST” ...

Algorithms can Mimic Human Piano Performance: The Deep Blues of Music

Emory Schubert¹, Sergio Canazza², Giovanni De Poli² and Antonio Rodà²

¹UNSW Australia, Australia; ²University of Padova, Italy

(Received 14 July 2016; accepted 13 November 2016)

Abstract

Can a computer play a music score, e.g. via a Disklavier, in a way that cannot be distinguished from a human performance of the same music? One hundred and seventy-two participants with a wide range of music playing backgrounds rated sound recordings of 7 performances of piano music by Kohler, one played by a human, and six generated by algorithms, including a ‘mechanical’ and an ‘unnatural’ version. Participants rated the extent to which each performance was by a human and explained their answers. The human performance had the lowest mean rating, but the human performance was rated as statistically identical to the other stimuli. There were no differences between ratings made by classical piano experts and lay listeners, but despite this, the musicians were more confident with their ratings. Qualitative analysis revealed five broad themes that contribute to judging whether a piece appears to be human. The themes were labelled (in descending order of frequency) intuitive, aesthetic, structural, thematic and temporal.

1. Introduction

Computer software has been able to send messages to acoustic pianos since 1986 by allowing the development of automated performances of piano music on the piano (De Poli, 2004). With recent improvements in algorithmic generation of standard (rocky piano) classical/romantic repertoire, a new research question has been emerging: Will there ever be a time when a listener cannot distinguish between an algorithm performing a piece (for example, via a Yamaha Disklavier the Bösendorfer SE reproducing piano, Gould & Bresee, 2003) versus a recording of an expert human performer (playing on the same device)? The ability of an algorithmic robot to be human-like has been a matter of fascination as humans are passionate about the possibilities of automation and robotics among the human species, the demystification of music, see Kaptel (2010). A famous example is a robot that could beat a world champion chess player. The ‘Deep Blue’ was able to achieve this milestone by the world champion, Gary Kasparov, in an official event in 1997, following several years of failures.
2017: A MUSICAL “TURING TEST” …

The Piece:

The Contestants:
- 4 algorithms (CaRo, Director Musices, VirtualPhilharmony, Basis Function Model)
- 1 human “internationally renowned pianist” [Schubert et al., 2017]
- 1 mechanical performance (deadpan)
- 1 “unmusical” performance (CaRo with inverted parameters)

The Evaluators:
- 172 listeners, different musical backgrounds, including pianists

“Humanplayer” rating by stimuli and expertise (mean and 1SE).
From (Schubert et al., JNMR 2017)
2017: A MUSICAL “TURING TEST” …

“This paper presents new evidence systematically demonstrating that algorithm-generated performances of piano music can be indistinguishable from human performances, suggesting some parallels with the 1990s victory of the Deep Blue computer over the world chess champion (human) chess player.”


Friedrich Kuhlau, Allegro Burlesco, Op.88 No.30

Computer (mechanical)
2017: A MUSICAL “TURING TEST” …

Friedrich Kuhlau, Allegro Burlesco, Op.88 No.30

(1) Computer (mechanical)  (2) Computer (Basis Function Model)  (3) Human Pianist

THE CON ESPRESSIONE PROJECT

- What kinds of expressive qualities do listeners perceive / differentiate?
- What is it in a performance that communicates an expressive quality?
- Can machines learn to recognise expressive qualities?
- Can machines learn to play music “expressively”?
- Can machines become truly ‘musical’ companions?
- … and what can we learn from all this?
EXPRESSIVE REACTIVE ACCOMPANIMENT: THE ACCompanion

The CON ESPRESSIONE Project

- What kinds of expressive qualities do listeners perceive / differentiate?
- What is it in a performance that communicates an expressive quality?
- Can machines learn to recognise expressive qualities?
- Can machines learn to play music “expressively”?
- Can machines become truly ‘musical’ companions?
- ... and what can we learn from all this?
THANK YOU!

This research receives funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 670035 ("Con Espressione").

JOHANNES KEPLER UNIVERSITY LINZ