Can the Computer Learn to Play Music Expressively?

Christopher Raphael
University of Massachusetts at Amherst
2001

presented by
Vincent Chapdelaine-Couture
Music++ (Music Plus One): REAL-TIME musical accompaniment for a LIVE soloist playing a piece of NON-IMPROVISED music.

Required Learning Input: score, previous (repeatable) soloist and accompaniment performances.

Thus, it is usually NOT used in the following cases:
**Main components**

**Goal:** synchronize accompaniment with soloist based on predictions:
- learn from past soloist performances in order to make predictions;
- respond in real-time to the soloist’s tempo changes and expressiveness.

**Music++ has 2 main components:**
- **Listen:** estimate note onsets from acoustic signal (soloist or accompaniment).
  

- **Play:** Bayesian Belief network that can be used to schedule an optimal accompaniment.
  
Listen is trainable for new solo instrument, room acoustics, microphone placement:

- Divide signal in blocks of, say, 32ms.
- Assign each block a label from the set \( \{ \text{rest, pitch}_i, \text{artic}_i \} \),
  where \( 1 \leq i \leq n \), \( n \) being the number of notes in the score.
- Use HMM to produce (real-time but delayed) \( n \) note onsets.

Robust to inaccuracies and embellishments by soloist?
Play is a Bayesian Belief Network with:

- Observable variables: note onsets, known note times from score.
- Unobservable variables: local tempo, change in tempo, and rhythmic stress.

Offline: training from soloist and accompaniment performances, separately first and then joined.

Online: compute in real-time the optimal accompaniment.
Musical score:

- gives the pitches, positions and relative durations of the various notes
- points of synchronization between the soloist and accompaniment
Previous soloist and accompaniment performances (not playing at the same time).

- The solo data is used to learn how to predict the future evolution of the solo part and to know what can and cannot be predicted reliably.
- The accompaniment data is used to learn musicality (onset sequence, volumes).
Soloist’s live performance:

- Note onsets;
- Delayed output to resolve local ambiguities (accompaniment will start late; no visual cues from conductor).
The solo model

\[
\begin{pmatrix}
  t_{n+1} \\
  s_{n+1}
\end{pmatrix} =
\begin{pmatrix}
  1 & l_n \\
  0 & 1
\end{pmatrix}
\begin{pmatrix}
  t_n \\
  s_n
\end{pmatrix} +
\begin{pmatrix}
  \tau_n \\
  \sigma_n
\end{pmatrix}
\]

where \( l_n \) is the musical length of the \( n^{th} \) note in beats and the \( (\tau_n, \sigma_n)^t \) and \( (t_0, s_0)^t \) are mutually independent Gaussian random vectors.

Mean of \( \sigma_n \) is:
- negative: soloist is usually speeding-up;
- positive: soloist is usually slowing-down.

Mean of \( \tau_n \) is:
- negative: compressions in time (without any change in tempo)
- positive: stretches in time (without any change in tempo)

Changes are nearly deterministic (low variance) or quite variable (high variance).
The solo model (cont.)

\[
\begin{pmatrix}
    t_{n+1} \\
    s_{n+1}
\end{pmatrix}
= \begin{pmatrix}
    1 & l_n \\
    0 & 1
\end{pmatrix}
\begin{pmatrix}
    t_n \\
    s_n
\end{pmatrix}
+ \begin{pmatrix}
    \tau_n \\
    \sigma_n
\end{pmatrix}
\]

is rewritten as:

\[
x_{n+1}^{\text{solo}} = A_n x_n^{\text{solo}} + \xi_n
\]

Unknown distributions are \( x_0^{\text{solo}} \) and \( \{ \xi_n^{\text{solo}} \} \).
Training the solo model

Noisy observations: quantization, misfires from Listen process. Imprecisions from soloist,....

\[ x_{n}^{\text{obs}} = B x_{n}^{\text{solo}} + \xi_{n}^{\text{obs}} \]  \hspace{1cm} (3)

B is (1,0). The \{\xi_{n}^{\text{obs}}\} have 0-means and fixed variances.

Use EM to learn a posteriori distribution of the unobserved \{x_{n}^{\text{solo}}\}. 

Music and Machine Learning 2008  
presented by Vincent Chapdelaine-Couture
We now have estimates of the black and white nodes below...

Use EM to learn a posteriori distribution of the unobserved \( \{ \xi_{n}^{\text{solo}} \} \) vectors.
Adding the accompaniment

MIDI variables for the accompaniment:

- Damping times is deterministic (legato, staccato).
- Velocities (volumes) are learned accompaniment performance (e.g., volume w.r.t score position, phrase), and applied in deterministic manner.
- Note onsets determines synchronisation: this is the more important and difficult part!

Note: no dependency between velocities and onset times.
The accompaniment model

The accompaniment is first trained as with the soloist to get Gaussian distributions for each note.

The Tricky Part: accompaniment must follow the soloist!

Solution: the conditional distribution of the accompaniment part GIVEN the solo part.
Combining the solo and accompaniment models

Consider a section of the accompaniment part "sandwiched" between two solo notes:
The big picture...

- \( \{ X_n^{\text{obs}} \} \)
- \( \{ X_n^{\text{solo}} \} \)
- Phantom nodes
- \( \{ X_{m_i}^{\text{cond}} \}, \{ X_{m_r}^{\text{cond}} \} \)
- Sandwiches nodes \( \{ X_{m_i+1}^{\text{cond}} \}, \ldots, \{ X_{m_r-1}^{\text{cond}} \} \)
- Actual accompaniment times \( \{ X_i^{\text{accom}} \} \)
Predicting the future

Predictive model: scheduling the pending accompaniment note every time new information becomes available (see red lines).

Note: onsets are detected with a small amount of latency (see green lines).
What if:

- a wrong note is played
- thrills are added, or other embellishment
- a measure is repeated

Mozart Rondeau example:: Noise (eg./ croud), huge tempo changes.
Questions?