Expressive music performance

Chapt. 7
1. Why study expressiveness

- Understanding human communication strategies
  - Non-verbal communication
  - Expressiveness tells “how to take” the explicit message
  - Disambiguate language expressions (e.g. in a movie)

- To embody expressive knowledge in machines
- To adapt HCI to the basic forms of human behavior

- Artistic applications, games and entertainment
- Auditory Display, Music Retrieval, Education
What is expressiveness (in music performance)

- refers to the effect of auditory parameters of music performance covering acoustic, psychoacoustic, and/or musical factors
- refers to the variation of auditory parameters away from a prototypical performance, but within stylistic constraints
  - is not deviation relative the musical score,
  - but the sound world that the score attempts to represent
- is not the same as emotion
- is used in the intransitive sense of the verb
  - no specific feeling is necessarily being expressed;
  - rather the music performance sounds “expressive” to differing degrees
- depends on historical and cultural context

Fabian, Timmers, Schubert 2014
Expressiveness: content and layers

<table>
<thead>
<tr>
<th>Content</th>
<th>Layer</th>
<th>Sample Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Musical expressiveness</strong></td>
<td>Compositional</td>
<td>'objective' expressiveness</td>
</tr>
<tr>
<td></td>
<td>structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>expr. amount</td>
</tr>
<tr>
<td></td>
<td></td>
<td>appropriate expr.</td>
</tr>
<tr>
<td><strong>Emotional expressiveness</strong></td>
<td>Compositional</td>
<td>compositional features</td>
</tr>
<tr>
<td></td>
<td>structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>expressive deviations</td>
</tr>
<tr>
<td><strong>Sensory / motor expressiveness</strong></td>
<td>Compositional</td>
<td>sensory qualities</td>
</tr>
<tr>
<td></td>
<td>structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>KE space, action metaphor</td>
</tr>
</tbody>
</table>
1. Expressiveness in music performance

Expressiveness refers to:

- aspects of a musical performance that are under the control of the performer
- the means used by the performer to convey the composer’s message
- the performer’s contribution to enrich the musical message.

![Diagram showing key velocity and note number with bar numbers highlighted: bar 2, bar 4, bar 8, bar 10, bar 12, bar 16.]

![Musical staff with notes and arrows indicating dynamic profiles vs. musical structure.]

Dynamic profiles vs. musical structure.
Expressiveness related to the musical piece

- **Strategies**
  - to highlight the musical *structure*
  - to disambiguate passages where several structural interpretations are possible
- **Dramatic narrative** developed by the performer
- **Emotional** content or expressive intention
- **Physical** and motor constraints or problems (e.g. fingering)
- **Stylistic** expectation based on
  - cultural norm (e.g. jazz vs. classic music)
  - actual performance situation (e.g. audience engagement)
- **Musical instrument** (e.g. W. Carlos)

Palmer, 1997
Expressiveness added by the performer

Performers may play the same piece according diverse nuances:

- emotions, affects (perceived, induced)
- KANSEI (i.e. sensibility, feeling, sensitivity)
- expressive intention
  - include emotion, affects as well other sensorial and descriptive adjectives or actions
  - evidences the explicit intent of the performer in communicating expression

Artistic expression

Aggressive  Gloomy  Active
Music communication chain

Composer's mental representation $\xrightarrow{T_{CS}}$ Score $\xrightarrow{T_{SP}}$ Performance $\xrightarrow{T_{PL}}$ Listener's perception

accessible information

score

implicit information

expressive intentions

perform

performance expressive cues

Friberg 1995
Modelli computazionali per l’esecuzione espressiva della musica

Tradurre in suono i simboli in cui è codificata una partitura musicale

Optical Music Recognition → Modelli per l’esecuzione espressiva → Strumento o Modelli di sorgente/segnaletica

- partitura grafica
- partitura digitale (MIDI, MusiXml)
- segnali di controllo
- suono
Musical Turing Test

which audio clip was performed by a computer?

A

B

Prelude No. 3 by Nino Rota
Modelli computazionali per l’esecuzione espressiva della musica

Tradurre in suono i simboli in cui è codificata una partitura musicale

evento partitura

\[
\begin{array}{l|l}
P(n) & \text{pitch} \\
MP(n) & \text{metric position} \\
ML(n) & \text{metric length}
\end{array}
\]

evento sonoro

\[
\begin{array}{l|l}
FR(n) & \text{frequency} \\
O(n) & \text{onset time} \\
DR(n) & \text{duration} \\
IOI(n) & \text{inter onset interval} \\
L(n) & \text{legato} \\
I(n) & \text{intensity} \\
BR(n) & \text{brightness} \\
AD(n) & \text{attack duration} \\
EC(n) & \text{envelope centroid}
\end{array}
\]
Modelli computazionali per l’esecuzione espressiva della musica

Tradurre in suono i simboli in cui è codificata una partitura musicale

[Graph showing onset time vs. metric position]

Esecuzione nominale
Modelli computazionali per l’esecuzione espressiva della musica

Tradurre in suono i simboli in cui è codificata una partitura musicale

![Diagram](https://via.placeholder.com/150)

- Esecuzione nominale
- Esecuzione espressiva
Modelli computazionali per l’esecuzione espressiva della musica

- Le deviazioni espressive sono correlate a specifiche strutture ritmiche, melodiche e armoniche
- La sintassi del “linguaggio” musicale (suddivisione in motivi, frasi, periodi) ha un ruolo importante
- Modelli delle deviazioni espressive a partire dagli Anni ’80 (Sundberg et al. 1983, Todd 1992, Widmer 2003, …)
- Esistono anche aspetti connotativi (emozioni, sensazioni, …), che dipendono dalle intenzioni espressive dell’esecutore
Accessible information

**Symbolic information:**
- score:
  - note list
  - performance instructions: e.g. agogic marks
- features extracted from score
- musical structure is not evident
- annotations (e.g. structure, prosody, etc.)

**Physical information:**
- audio signal
  - acoustic and perceptual features extracted from audio
- performance actions: e.g. MIDI
- performer gestures
- listener’s physiological response
Accessible information

Expressive information

- basic emotions
  - Happiness, Anger, Sadness, Fear, Tenderness
- emotion space ➔ Valence / Arousal
- sensorial space ➔ Kinetics / Energy
- expressive intentions
- metaphorical description

Frequently, focus on expressive deviations
Expressive deviations (variations)

- Systematic presence of *deviations* from the musical notation
- Deviations are only the external surface

**Measurable:**
- from Midi performances: timing, key velocity
- from audio signal (partially)

**Reference** for computing deviations:
- *score*, both for theoretical (the score represents the music structure) and practical (it is easily available) reasons
- *global* (longer time spam) measurements as reference for local ones
- *mean* performance
- *neutral* performance

Mozart K622: *score*  
Mozart K622: *neutral*
Timing variations by 3 pianists

Interonset Interval deviations (%)

Horowitz 65

Schnabel

Brendel

(Träumerei first 8 bars)
Why do these deviations exist?

The score serves mainly:
- as a mean to *convey instructions* to the performer
- as an *aid for memorization* and preservation.

The score may serve as
- a *representation of cognitive elements* of the composition
- rather than the physical parameters.

The performer *interprets* these symbols, taking into account
- the implicit information
- his/her personal artistic feeling and aim.

The *liberty* of the musicians to exhibit their own *personal interpretation* has varied, but has rarely been completely denied.

see e.g. prosody in speech  ➔ *Musical prosody*
Information and music performance

Expressive performance parameters

► Physical information level
  ◆ timing of musical events and tempo,
  ◆ dynamics and articulation
  ◆ ...

► Symbolic information level
  ◆ Score
  ◆ Musical structure (?)
Example of symbolic information

Tree representing the hierarchical structure of the first 16 bars extracted from Mozart sonata K 545

- rectangles represent structural elements,
- circles represent notes.
Information representation

- Event information representation: $EV[n]$
  - score as a sequence of events: $EV[\cdot]$

- Time parameters
  - pitch value $FR[n]$, Onset time $O[n]$ and Duration $DR[n]$
    - $IOI[n] = O[n + 1] - O[n]$ \textit{Inter Onset Interval}
    - $L[n] = DR[n]/IOI[n]$ \textit{Legato}

- Intensity $I[n]$ or KeyVelocity $KV[n]$ for MIDI event

- Timbre-related parameters.
  - Brightness $BR[n]$
  - energy envelope
    - Attack Duration $AD[n]$
    - Envelope Centroid $EC[n]$
Time information representation

- **space metaphor:**
  - **score position:** $x[n]$
  - measured in
    - musical units (one musical unit = whole note value)
    - metrical units (beats or bars).
  - **length:** symbolic duration of an event
  - **distance:** interval between events
- Both can be measured in musical or metrical units
  - e.g. in a 3/4 rhythm
    - the length of 1 bar
      - = 3 beats (quarter note) = 0.75 musical units (whole notes)
    - the corresponding distance of the event at the beginning of a bar from the event at the beginning of the next bar
      - = 3 beats (quarter note)
      - = 0.75 musical units (whole notes) wide.
Time information representation

- **symbolic time** reference can be
  - absolute (as in the so-called piano roll notation)
  - or relative: the distance from a previous event is specified
- score as a sequence of events
- distance between two adjacent events: note or rest value NV[n]
- score position
  - \( x[n + 1] = x[n] + NV[n] \)
- symbolic time in seconds
- **nominal time** \( t_{nom} \) or score time
  - nominal onset time \( O_{nom}[n] \)
  - nominal duration \( DR_{nom}[n] \)
  - inter onset interval (nominal length)
    - \( IOI_{nom}[n] = O_{nom}[n+1] - O_{nom}[n] \)
Time information representation: Tempo

- **Tempo**: measured by metronome
- \( v = \frac{\text{score length}}{\text{performance duration}} \)
  - \( t_{\text{nom}} = \frac{x}{v} \)
  - \( O_{\text{nom}}[n] = \frac{x[n]}{v} \)

  e.g. with of \( MM = 120 \) quarters per minute,
  - \( v = 2 \) quarters (beat) per second
  - a beat will lasts 0.5 seconds.

  A piece of 16 bars with rhythm 3/4 (e.g. a waltz) will have total length of 12 musical units and 48 beats.
  - this piece played exactly at \( MM = 100 \) quarters per minute will lasts 28.8 seconds
    - \( v = \frac{12}{28.8} = 0.417 \) musical units per second
    - \( v = \frac{48}{28.8} = 1.67 \) beat per second.
Time information representation

- at physical level: performance time \( t \)
  - measured in seconds
  - Onset time: \( O[n] \)
  - Duration: \( DR[n] \)
  - Inter Onset Interval: \( IOI[n] = O[n + 1] - O[n] \)
    - IOI deviation: \( \Delta R = \frac{IOI[n]}{IOI_{nom}[n]} \% \)
  - Legato: \( L[n] = \frac{DR[n]}{IOI[n]} \)

![Graph showing time information parameters]
Time information representation

- the relation between performance time and score position
  - \( x(t) \) or \( t(x) \)

- Tempo
  - Mean tempo
  - Main tempo: prevailing tempo
  - Basic tempo: central tendency
  - Local tempo
  - Event tempo
  
\[
\begin{align*}
  v_{\text{mean}} &= \frac{x[N] - x[1]}{O[N] - O[1]} \\
  v_{\text{ev}[n]} &= \frac{x[n + 1] - x[n]}{O[n + 1] - O[n]} = \frac{NV[n]}{IOI[n]}
\end{align*}
\]

- Tempo as function of score position
  - Duration of a measure
  - Relative IOI

\[
IOI_{rel}[n] = \frac{IOI[n]}{NV[n]}
\]
Example of time map

- **Tempo** $v(x)$

- **Time shift**: timing measured as deviations from a regular pulse

- **Time map**: $t(x)$
Time shift and micro-pause

- **time shift**: deviation from expected

\[ TS[n] = O[n] - O_{exp}[n] \]

![Diagram showing time shift, micro-pause, and tempo](image-url)
Model representations

- **Time representation**
  - Discrete values ➔ events
    - e.g. articulation of timing of individual notes
  - Continuous values ➔ functions
    - e.g. vibrato

- **Granularity**
  - numerical (absolute, relative) values
    - e.g. time interval in ms, relative IOI
  - categorical
    - e.g. staccato vs. legato, shortening vs. lengthening

- **Time scale**
  - single note, e.g. attack time, vibrato
  - local: few notes, e.g. articulation of a melodic gesture
  - more global: e.g. phrase crescendo
2.1 Research on music performance: modeling

- ’30-’40: Seasore collected and analysed many objective measurements from performances
- ’40-’70: no much activity
- ’60: beginning of computer music
- ’70-...: computational approaches

Model development strategies:
- a. based on intuition
- b. based on measurements: statistical and mathematical models
- c. analysis by synthesis, e.g. rule based
- d. data driven
a. Models for understanding, based on intuition

- Human analyst
  - select musical parameters
  - devises a theory or a mathematical model
  - assesses empirical validity
    - by testing on real performance data
    - by expert evaluation

- Outputs:
  - expressive deviations
  - expressiveness recognition
Performance expressive parameters

Main expressive parameters:
- tempo and timing (global, local)
- dynamics
- articulation

for some instruments, voice
- vibrato, tremolo
- intonation

for contemporary music
- timbre
- virtual space position

How do parameters interact?
b. Modeling strategies for understanding

**Analysis-by-measurement**

- based on the analysis of deviations measured in recorded human performances
- aims at recognizing regularities in the deviation patterns and to describe them by means of a mathematical model, relating score to expressive values

**Performing controlled experiments:**

- by manipulating one parameter in a performance (e.g. the instruction to play at a different tempo)
- the measurements reveal something of the underlying mechanisms.
Analysis by measurements

Analysis-by-measurement

- Selection of performances.
- Measurement of the physical properties of every note.
- Reliability control and classification of performances.
- Selection and analysis of the most relevant variables.
- Statistical analysis and development of mathematical interpretation model of the data.
c. Analysis by synthesis

- **Analysis by synthesis** models
  - Development method
    - Model formalizes intuition, knowledge of expert musician
    - Performance generation by model simulation
    - Evaluation by experts
    - Refinements
  - The system acts as a student

- **KTH rule system**
Analysis by synthesis: method

- Collect data from analysis by measurements
- Further steps
  - Start with a hypothesized principle
  - repeat
    - realize it in terms of a synthetic performance
    - evaluate it by listening.
    - if needed,
      - modify the hypothesized principle or formulate new rule
  - until the performance is satisfactory
Analysis-by-synthesis: pros and cons

Advantages:
- perception of the music is directly used in evaluation
- general validity of the hypotheses can directly be tested by applying it to other music examples
- feedback loop is very short

Disadvantages
- conclusions are based on the expertise of just a few people.
- parameters can be chosen rather arbitrarily.

The success depends on:
- the formulation of hypotheses
- on competent listeners.
d. Data driven approach

Search for and discover dependencies on very large data sets
  via machine-learning and data-mining techniques;
  goal: ‘explaining’ the given data as well as possible.

No preliminary hypothesis: bottom up approach

Advantages:
  rooted in large amounts of empirical data
  the machine may discover truly novel and unexpected things
  good results in extracting local musically relevant patterns

Problems:
  difficult to incorporate musicological knowledge into the process
  challenging: human-interpretable high level models

Widmer et al.
Example

- Dynamics deviation learned from the training pieces applied to *Chopin Waltz Op.18, Op.64 no.2*
Example

Tempo deviation learned from the training pieces applied to Chopin Waltz Op.18, Op.64 no.2
Models

- Model based on measurements
  - Tod’s model,
  - Final retard

- Analysis by synthesis
  - KTH rule systems
  - CaRo system

- Data driven
  - SaxEx
  - YQX

- Interactive systems: Virtual Philarmony

- Expressive sound movements
d1. Case-based reasoning

**Case-based reasoning:**
- Appropriate for problems where:
  - many examples of solved problems can be obtained
  - a large part of the knowledge involved in the solution of problems is tacit, difficult to verbalize and generalize.
  - new solved problems can be revised by humans and memorized (= learn by experience)
- A case consists of a problem, its solution, and, typically, annotations about how the solution was derived.

**Saxex system (Arcos, 1997)**
- spectral processing
- example: Autumn lives
  - Inexpressive
  - Expressive
Example: KTH performance rule system

Reasoning: rule system
- if < condition >
  then < action >

- weighted addition of many rules
  \[ \text{dev}[n] = \sum k_i f_i (\text{note}[n]) \]

KTH performance rule system
- differentiation rules
- grouping rules
- synchronization rules
- meta-rules
**Types of rules**

- **Differentiation Rules** enhance the differences between scale tones, and between note values.

- **Grouping Rules** show which tones belong together and which do not.
  - In music the "belonging-togetherness" exists at several levels simultaneously.
  - Tones constituting melodic gestures, such as belong together, as do tones constituting phrases.
  - The rules mark, by means of micro-pauses and lengthening of tones, the boundaries between all these tone groups.

- **Ensemble Rules** keep the order in ensembles.
  - They achieve synchronization by lengthening and shortening the individual tones.

- **Meta-Rules** for emotional expressive performance
  - choice of rules and weight
KTH Rules examples

- Differentiation rules
  - Duration Contrast
  - High sharp
  - Melodic charge

- Grouping rules
  - Phrase arch
  - Punctuation
  - Final retard

- Ensemble rules
  - Melodic synchronization
  - Bar synchronization

- Macro-Rules
  - for emotional expressive performance
Differentiation rule: Duration Contrast

The contrast between long and short note values, such as half note and eighth note, is sometimes enhanced by musicians. They play the short notes shorter and the long notes longer than nominally written in the score.

No contrast:
Duration contrast

- Medium, $k = 2.2$
- Exaggerated, $k = 4.4$
Duration contrast

Inverted, $k = -2.2$
Differentiation of Pitch Categories (2)

- **MELODIC CHARGE:**
  - This rule accounts for the "remarkableness" of the tones in relation to the underlying harmony.
  - Sound level, duration and vibrato extent are increased in proportion to the melodic charge value

- Emphasis rule

<table>
<thead>
<tr>
<th>Tone</th>
<th>C</th>
<th>G</th>
<th>D</th>
<th>A</th>
<th>E</th>
<th>B</th>
<th>F#</th>
<th>D</th>
<th>Ab</th>
<th>Eb</th>
<th>Bb</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{mel}$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6.5</td>
<td>5.5</td>
<td>4.5</td>
<td>3.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 3.1: Melodic charge $C_{mel}$ for the various scale tones in a C major or minor scale.
Example: Melodic charge

- J S Bach, Fugue theme of the *Kyrie* movement from the *b-minor Mass*, BWV 232
- No rule
- $K = 1.5$
PUNCTUATION: The melody can be divided into small musical gestures normally consisting of a few notes. This rule tries to identify and perform these gestures. It consists of two parts:

- the gesture analysis
- application of these in the performance.

- The gesture analysis is a complex system of 14 subrules where concepts from LEAP ARTICULATION and ACCENTS are used.
- The identified gestures are performed by inserting micro-pauses at the boundaries
The main principles for the finder rules are:
- in melodic leap, with different weights for different contexts,
- after longest of five notes,
- after appoggiatura
- before a note surrounded by longer notes
- after a note followed by two or more shorter notes of equal duration

The eliminator rules remove marks or reduce weights in this cases
- after very short notes
- in a melodic step motion
- when several duration rules interact
- at two adjoining marks in a tone repetition
J S Bach: Fugue theme from *Fantasia und Fuga*, g minor, BWV 542.
In the example another rule is also applied

- No rule
- Medium
- Exaggerate
Macro-grouping rule: Phrase arch

**PHRASE ARCH**

- Tempo curves in form of arches with an initial accelerando and a final ritardando are applied to the phrase structure.
- Sound level is coupled with the tempo variation so as to create crescendi and diminuendi.

In this example two other rules are applied:

- **Durational contrast**
- **Punctuation**

Phrase arch

- **Medium, \( k = 1.5 \)**

- **Exaggerated, \( k = \)***
**Phrase arch**

- Inverted, $k = -1.5$
### Macro rules: for emotions

**Macro-Rules for Emotional Expressive Performance**

- **choice of rules and weight**

**Examples**

- **Fear**
- **Anger**

<table>
<thead>
<tr>
<th>Expressive Cue</th>
<th>Gabrielsson and Juslin</th>
<th>Macro-Rule in Director Musices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tempo</td>
<td>Irregular</td>
<td>Tone IOI is lengthened by 80%</td>
</tr>
<tr>
<td>Sound Level</td>
<td>Low Sound</td>
<td>Level is decreased by 6 dB</td>
</tr>
<tr>
<td>Articulation</td>
<td>Mostly staccato or non-legato</td>
<td>Duration Contrast Articulation rule</td>
</tr>
<tr>
<td>Time Deviations</td>
<td>Large</td>
<td>Duration Contrast rule</td>
</tr>
<tr>
<td></td>
<td>Structural reorganizations</td>
<td>Punctuation rule</td>
</tr>
<tr>
<td></td>
<td>Final acceleration (sometimes)</td>
<td>Phrase Arch rule applied on phrase level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phrase Arch rule applied on sub-phrase level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Ritardando</td>
</tr>
</tbody>
</table>
Emotional expression: SADNESS

IOI deviations

articulation

dB deviations
e. Modeling expressive intentions: CaRo

CaRo model:
- expressive deviations from an input neutral performance
- elementary transformations, that not destroy the relation between deviations and musical structure

Symbolic Description of Audio Perf. → Musical Expressiveness Model → Audio Performance (Sinusoidal Model) → Audio Processing → MIDI Rendering → Expressive Intentions

MIDI Performance → (Off-Line) Symbolic Description of Audio Perf.
Dynamics profiles show that there are many differences between the performances.

In spite of these differences, some similarities can be found amount the profiles.

Pianist A’s dynamics profiles measured in a single performance.
Normalized measured deviations

Pianist A’s dynamics profiles, normalized to zero mean and unitary variance
Average of measured deviations

Grand average of the pianist A’s dynamics profiles
Computing deviations: variables k and m

\[ P_e(n) = k_e \overline{P}_0 + m_e \left( P_0(n) - \overline{P}_0 \right) \]

\[ P(n) = k(x, y) \overline{P}_0 + m(x, y) \left( P_0(n) - \overline{P}_0 \right) \]
Examples

Input _______________________________ Output

Mozart K622 neutral

Mozart K622 nominal

Preludio

Corelli Op. V neutral

MIDI

bright

dark

hard

soft

light

heavy

Post processing

bright

heavy

light

hard
Interaction: the control space

- Interactive Interface
- Multimodal Control
- Multimedia Context
Human communicates his own intention in a multimodal way.

For the human-machine communication, we may distinguish two main classes of possible interfaces:

- **Graphic panel dedicated to the control.** The control variables are directly displayed on the panel and the user should learn how to use it.

- **Multimodal,** where the user interacts ‘freely’ through movements and non-verbal communication → Task of the interface is to analyze and to understand human intention.
Control spaces

The control space is the user interface, which controls, at an abstract level, the expressive content and the interaction between the user and the audio object of the multimedia product.

Perceptual expressive space, was derived by multidimensional analysis of various professionally performed pieces ranging from western classical to popular music.

Synthetic expressive space, allows the author to organize his own abstract space by defining expressive points and positioning them in the space.
Mapping strategies

- From position in control space \((x, y)\) to expressive variables \(k\) and \(m\)
  - Perceptual Expressive Space: linear transformation of position
    
    \[
    k(x, y) = a_{k,0} + a_{k,1}x + a_{k,2}y \\
    m(x, y) = a_{m,0} + a_{m,1}x + a_{m,2}y
    \]

  - Synthetic Expressive Space: quadratic interpolation

- From parameters \(P\) of neutral performance and score \(P\) to expressive parameters \(Pe\)
  
  \[
  P(n) = k(x, y)\overline{P}_0 + m(x, y)\left(P_0(n) - \overline{P}_0\right)
  \]
drawing a trajectory in the control space, the expressive content change accordingly
Expressive audio rendering

- Signal processing techniques based on sinusoidal model
  - Time stretching
  - Pitch Shifting
  - Amplitude Envelope Control
  - Brightness control

- Rendering of expressive deviations:
  - Local tempo:
    - time stretching (different for attack, sustain and release)
  - Legato:
    - time stretching and pitch shifting (with ad-hoc spectral processing)
  - Intensity:
    - amplitude envelope and brightness
  - Envelope Shape:
    - amplitude envelope (with triangular-shaped functions)
Example

- Expressive post-processing
- Time-frequency processing for legato
Real time expressive processing

Corelli - Violin Sonata in A dur op. V (excerpt)

Corelli Op. V neutral

Preludio
Once upon a time...

- The control space was also implemented as an Applet.
- It is used to change interactively the sound comment of the multimedia product.
- The author can associate different expressive intentions to the character.
Caro 2.0
YQX: a model for expressive music performance

Data
- Performances of Mozart Piano Sonatas by Roland Batik
- Performances of Chopin Piano Pieces by Nikita Magaloff

Method
- Model score-performance dependencies as a probabilistic network:
Probabilistic network

- Bayesian model
  - relations between *score context* and *target variables*

- Model $p(Y \mid Q, X)$ as a Gaussian distribution $N(\mu, \sigma^2)$ that varies linearly with $X$: when $Q = \bar{q}$ and $X = \bar{x}$

\[
\mu = d_{\bar{q}} + \bar{k}_{\bar{q}} \cdot \bar{x}
\]

- Estimate the parameters $d_{\bar{q}}$ and $\bar{k}_{\bar{q}}$ by linear regression (least squares)
  - $\sigma^2$ is the average residual error

- Trained by estimating, separately for each target variable,

- The dependency on the discrete variables $Q$ is modeled by computing a separate model for each possible combination of values of the discrete values.
Features/Targets

- **Targets (Y):** timing, loudness, articulation

- **Discrete Features (Q):**
  - pitch interval,
  - rhythmic context,
  - I-R label, I-R position

- **Continuous Features (X):**
  - I-R closure,
  - duration ratio

```
Features
IR-Label: Process
IR-Closure: -0.4332
Pitch-Interval: -1
Rhythmic context: s-s-1
Duration Ratio: 0.301
```

```
Targets
Timing: 0.036
Articulation: 1.22
Loudness: 0.2
```
Performance Rendering

1. Read Expressive hints from score (trill, cresc., rit., p, f, etc.)
   - Set up basic tempo curve
   - Set up basic loudness curve

2. Extract features from score

3. Predict expressive deviations from model;
   For every note and every target:
   - Enter features as evidence $\bar{q}, \bar{x}$ in network
   - Infer most likely timing/loudness/articulation value
     \[
     \mu = d_{\bar{q}} + k_{\bar{q}} \cdot \bar{x}
     \]

4. Combine basic curves and predicted deviations
   (for tempo/loudness respectively)
YQX system

Diagram:
- New Piece (MusicXML) connected to:
  - Training Data
  - Dynamic Annotations
  - Tempo Annotations
  - Feature Extraction
- Feature Extraction connected to:
  - Loudness Prediction
  - Local Tempo Prediction
  - Note Timing Prediction
  - Note Level Rules
  - Articulation Prediction
- Loudness Prediction connected to:
  - Tempo
  - Articulation
- Tempo Prediction connected to:
  - Tempo
  - Articulation
- Articulation Prediction connected to:
  - Expressive MIDI
YQX system at Rencon 2008
The composer asked for expressive sound movements

Experiment: stimuli
- speed (fast, slow)
- path (continuous circular, discontinuous random)
- articulation (legato, staccato)
- timbre (white noise, harmonic sound)

Results: valence - arousal dimensions
- similar space location for similar movements
- strong relation between speed and arousal axis
- timbre influence
- legato-staccato is a parameter stronger than path

De Gotzen, IEEE MM 0484

- Video opera: PalaFenice, Venezia, 2002
  - sound movement in space as a musical parameter
  - expressive matching between instrumental and sound movement gestures
  - metaphor: the public enters the physical movement

- trombone player’s gesture
- gesture analysis
- expressive spatialization
Movimento del suono delle trombe in funzione della sua dinamica

- *Guarnieri, Medea*
  - Viene estratto l’inviluppo di ampiezza e trasformato in movimento longitudinale di spazializzazione

Riduzione stereofonica

Movimento sinistra - destra
Medea, by Adriano Guarneri

Teatro La Fenice, Venezia, 2003
► 72-element choir, 62-element orchestra, 4 soloist singers
► Live electronics: reinforcement, spatialization, real time sound processing
Another test

Which is your favorite performance?

A

B

Allegro Burlesco by Kuhlau Op. 88, n. 3
Another test

Which is your favorite performance?

A

B
A dynamic model of phrasing (Todd)

Music and motion

- musical movement has two degree of freedom, tonal movement and rhythmic movement;
- this movement is similar to and imitates motion in physical space;
- the object of motion in physical space, to which musical movement alludes, is that of a body or limb.
Definitions

- **Definitions:**
  - \( x = \) score position
  - \( t = \) performance time
  - \( a = \) acceleration
  - \( u = \) initial velocity

- \( x(t) \) or \( t(x) \)

\[
\begin{align*}
a &= a(t) \\
v &= v(t) = \int a(t) \, dt \\
x &= x(t) = \int v(t) \, dt \\
a &= a(x) \\
v &= v(x) \\
t &= t(x) = \int \frac{1}{v(t)} \, dx
\end{align*}
\]
Linear tempo model

- tempo is supposed to vary linearly in time

\[
\begin{align*}
a(t) &= a \\
v(t) &= \int a(t) dt = u + at \\
x(t) &= \int v(t) dt = ut + \frac{at^2}{2}
\end{align*}
\]

\[
\begin{align*}
a(x) &= a \\
v(x) &= \sqrt{u^2 + 2ax} \\
x(t) &= \frac{\sqrt{u^2 + 2ax} - u}{a}
\end{align*}
\]
Energy, tempo and intensity

A piece can be decomposed in a hierarchical sequence of segments
Every segment is characterized by an accelerando-ritardando pattern
The analogy is the movement of a particle of mass $m$ in a V-shaped potential well of length $L$

Energy (E) = Kinetic energy + Potential energy (V)
  - Kinetic energy = $m v^2 / 2$

$$v(x) = \sqrt{2(E - V(x))/m}$$

$$v(x) = \sum_{j} \sqrt{2(E - V_j(x))/m_j}.$$  

The complete $x(t)$ mapping results shaped as piece-wise parabolas.
Example: simulation vs. performances

- Chopin Prelude
  - Simulation (bold line) vs Performance data (dotted line)
Example

Theme from the six variations composed by Beethoven over the duet *Nel cor più non mi sento*
Example

Nel cor più non mi sento
Music and Structure (Todd, 1985)

Phrase structure

Predicted timing (IOI)
Multilevel decomposition (Widmer Tobudic)
Multilevel decomposition of dynamics curve of performance of Mozart Sonata K.279:1:1, mm.31-38
Music and Motion (Friberg & Sundberg, 1999)

\[ v(x) = \left[ 1 + (v_{\text{end}}^q - 1)x \right]^{1/q}. \]
Final retard

- Final retard is a three phase quasi-mechanical movement:
  - given constant tempo
  - quadratic function (parabola) for tempo decrease from mechanical retard with constant force
  - linear final tempo curve
Mathematical models: Mazzola model

- Developed by Guerino Mazzola and colleagues in Zurich
- Mathematical music theory and performance model
  - Applies analysis and performance components
  - computer-aided analysis tools for musical structure
    - each aspect implemented in a Rubbette (plugin)
  - performance is generated with the Rubettes
  - uses “Stemma/Operator” theory for mapping
- MetroRUBETTE
  - (inner) metrical analysis
  - result is different than Lerdahl and Jackendoff metrical analysis
  - used linear mapping between metrical weight and tone intensity to generate a performance
- Empirical Evaluation
  - not evaluated against real performances
YQX: a model for expressive music performance

Data
- Performances of Mozart Piano Sonatas by Roland Batik
- Performances of Chopin Piano Pieces by Nikita Magaloff

Method
- Model score-performance dependencies as a probabilistic network:
Probabilistic network

- Bayesian model
  - relations between score context and target variables

Model $p(Y \mid Q,X)$ as a Gaussian distribution $N(\mu, \sigma^2)$ that varies linearly with $X$: when $Q = \bar{q}$ and $X = \bar{x}$

$\mu = \bar{d}_q + \bar{k}_q \cdot \bar{x}$

- Estimate the parameters $\bar{d}_q$ and $\bar{k}_q$ by linear regression (least squares)
  - $\sigma^2$ is the average residual error

- Trained by estimating, separately for each target variable
- The dependency on the discrete variables $Q$ is modeled by computing a separate model for each possible combination of values of the discrete values.
Features/Targets

- Targets (Y): timing, loudness, articulation
- Discrete Features (Q):
  - pitch interval,
  - rhythmic context,
  - I-R label, I-R position
- Continuous Features (X):
  - I-R closure,
  - duration ratio
Performance Rendering

1. Read Expressive hints from score (trill, cresc., rit., p, f, etc.)
   - Set up basic tempo curve
   - Set up basic loudness curve

2. Extract features from score

3. Predict expressive deviations from model;
   For every note and every target:
   - Enter features as evidence $\bar{q}$, $\bar{x}$ in network
   - Infer most likely timing/loudness/articulation value
   \[
   \mu = d_{\bar{q}} + \bar{k}_{\bar{q}} \cdot \bar{x}
   \]

4. Combine basic curves and predicted deviations
   (for tempo/loudness respectively)
YQX system
YQX playing "My Nocturne" at Rencon 2008

https://www.youtube.com/watch?v=IS3519TUkzg
Tempo and articulation prediction.

- Tempo and articulation are predicted by a Bayesian Network modelling dependencies between score and performance as conditional probability distributions.
- The score model comprises simple score descriptors (rhythmic, melodic and harmonic) and higher level features from the Implication-Realization (I-R) model of melodic expectation by E. Narmour (I-R-labels and a derivation of I-R-closure).
Tempo prediction

- **Local tempo**, a per-note prediction of long-term tempo changes;
  - take the surrounding performance context into account
  - the Bayesian network is unfolded in time and an adapted version of the Viterbi Algorithm for HMM is used to predict a series of tempo changes that is optimal with respect to the complete piece

- **Note timing**, note-to-note deviations from local tempo
  - using only the immediate score context and the value predicted for the previous note

- **Global tempo**, extracted from tempo annotations in the score (e.g. andante, a tempo).

- Combined to form the final tempo
- Articulations: only the immediate score context
Data driven rule extraction algorithm for musical expression.

Two of the rules are used to further enhance the aesthetic qualities of the rendered performances.

**staccato rule:**
- if two successive notes have the same pitch, and the second of the two is longer,
  - then the first note is played staccato.

**delay-next rule:**
- if two notes of the same length are followed by a longer note,
  - then the last note is played with a slight delay.
- Professional musicians tend to emphasize the melody by playing the melody notes slightly ahead of time.
A set of basis functions $\phi$ is combined in a function $f$, parametrized by a vector of weights $w$, to approximate the dynamics measured from a recorded performance of the score.

Dynamics as a function $f(\phi; w)$ of basis functions $(\phi_1, \ldots, \phi_K)$, representing dynamic annotations and metrical basis functions.
Interactive performance systems

- Real time control of the expressive performance systems
- VirtualPhilharmony (Baba, Hashida, Katayose)
  - conducting interface that enables users to perform expressive music with conducting action.
  - provide its player with the sensation of conducting a real orchestra.

![Diagram of Interactive performance systems]

- Conducting
  - Glove with accelerometer sensor
  - Wii Remote and infrared baton
  - MIDI theremin (capacitance sensor)

- Expressive performance template
  - CrestmuseXML
  - MusicXML (Score)
  - DeviationInstanceXML (Expression)

- Max/MSP
  - Detection of beat-point
  - Note with expression

- Scheduler
  - prediction of tempo of next beat

- Heuristics of experienced conductor

- MIDI

- Score
**Gesture sensors**

- **Glove interface using an accelerometer**
- **Baton using infrared sensor (Wii Remote)**
- **Capacitance sensor (MIDI Theremin)**

---

*Figure 2. Gesture Sensors*

*Figure 3. Real performance and V.P.*
Real performance and VP

- tempo and beat prediction
- control the performance model
- combine smoothly template and conductor

https://www.youtube.com/watch?v=t_ClcuIcn5k
VirtualPhilharmony

馬場 隆 橋田 光代 片寄 晴弘
(片寄グループ)
3. Models for music production

- Models for music performance production:
  - classic music
    - focus on pitched sounds
    - melodic and harmonic organization
    - linguistic paradigm
  - contemporary music
    - focus on sound, timbre
    - acoustics, psychoacoustic
    - performance as timbre control

- From structural models to sound control models
  - sound synthesis
  - live electronics
Performance models for sound synthesis

- mapping strategies
  - how to map input values and gestures to acoustic parameters
  - one to one, few to many, scaling

- processes, algorithmic control
  - low level parameters generated by algorithms
  - control of algorithms (e.g. masks)

- different abstraction levels
  - from abstract to concrete
  - time scales

- from note to sound event
Performance models for live electronics

- Live electronics performer processes instrumental sound
- combined effect (e.g. deviations of deviations)
Automatic performance: problems

The idea of automatic expressive music performance, especially when it is applied to the performance of classical music that was not written for this purpose, is questionable.

The possibility to completely model and render the artistic creativity implied in the performance is still to be demonstrated.

State of the art models

- rendering of some relevant aspects of a musically acceptable performance,
- but not sufficient for a full artistic appreciation.
Automatic performance can be acceptable:

- when a real artistic value is not necessary (even if useful)
- when the alternative is a mechanic performance of the score (as in many sequencers)
- in entertainment application
- when it not necessary to preserve the exact artistic environment of the composition, as in popular music

- when music is expressly created bearing in mind the use of technology

Toward model-listener interaction (conductor paradigm)?
In the era of information society, artists are always more frequently using technology in their artworks.

- sound generation and processing
- algorithmic composition
- interactive performance

How to control sound synthesis or processing engines (systems, algorithms, etc.)?

- typical performance topic
- need of establishing and computing the relation of musical and compositional aspects with sound parameters
- joint model development with composer

Performing environment
c. Performance models for education

Another important possible application of performance models, even of classical music, is in *education*.

The knowledge embodied in performance models can help the teacher to

- rationalize the performance strategies
- better convey his teaching effort.
Discussion

Need of a well-founded approach based on strong scientific knowledge
- from classical music performance studies to performance models of new music creation.
- from the practical knowledge of new music creators to new performance models.

Models for artistic creation:
- joint effort of scientists and musicians
- really new tools, not only inspired to problems and solutions of the past

Models for expressiveness extraction and processing
Toward models of/for performing arts