Refining mapping strategies to improve the sound quality of physically-controlled synthesis

V. Verfaille[♯], A. Rebillout[♭], Ph. Guillemain[♭], M. M. Wanderley[♯]

[♯]SPCL & IDMIL

CIRMMT — Centre for Interdisciplinary Research in Music Media and Technology

McGill

[♭]LMA–CNRS

LMA — CNRS — Centre National de la Recherche Scientifique

Agence Nationale de la Recherche — ANR

Fonds de recherche sur la nature et les technologies — Québec

Fonds de recherche sur la société et la culture — Québec

Développement économique, Innovation et Exportation — Québec

NSERC CRSNG

http://www.Consonnes.cnrs-mrs.fr/

♪ Acoustics'08 Paris — July 04, 2008

Physically-Controlled Synthesis

What is PCS?

- PCS = new synthesis technique
  combining a physical model with additive synthesis

  Guillemain, Verfaille, "Combining physical modeling and additive synthesis as a mapping strategy for realtime control", *Proc. Int. Conf. on Computer Music*, Vol I, pp. 442–9, 2007
- goal = build a synthesizer that
  - improves controllability (obeys to physical controls)
  - improves sound quality of digital sound synthesis
  - provides mapping/interpolation functionalities
- challenging problem: nonlinear functioning
- case study: clarinet model(s) + WX5 wind controller
- context: ANR project (ANR-05-BLAN-0097-01),
  Consonnes (control of natural and synthetic sounds),
  Axis II: digital models for realtime synthesis

Physically-Controlled Synthesis

A combination of synthesizers

PCS combines 2 clarinet sound synthesizers:

- a clarinet physical modelling (PM):
  - non-linear coupling of the bore with the reed
  - $\Longrightarrow$ odd/even balance: signature of the non-linearity strength

  Guillemain, Kergomard, Voinier, "Real-time synthesis of clarinet-like instruments using digital impedance models," JASA (118)1, pp. 483–494, 2005

  ▸ more
- a realtime additive synthesizer *Ssynth* (AS):
  - advanced and flexible control functionalities
  - additive database of natural instrumental sounds:
    - 3D mesh (pitch, dynamic, instrument) + time
    - navigation + interpolation, extrapolation, morphing
  - modular mapping (important feature of DMI design)

  Verfaille, Boissinot, Depalle, Wanderley, "*Ssynth*: a real time additive synthesizer with flexible control," *Proc. Int. Conf. Computer Music*, 2006

  ▸ more

Mapping and PCS

Interfacing with sound descriptors

Question: where/how to interface PM with AS?

1. spectrum analytical description in the physical model
   J. Kergomard and S. Ollivier and J. Gilbert; "Calculation of the spectrum of self-sustained oscillators using a variable truncation method: Application to cylindrical reed instruments", Acustica 86, pp. 685–703, 2000
   - pros: best for realtime
   - cons: too simple PM, not the best quality
2. audio-driven (proof of concept)
   - pros: modularity (any sound source, PM)
   - cons: latency (realtime) & biais (mapping)

Keypoint = sound descriptors

- used to interface the two synthesizers
- provide additive synthesis data by navigating the database and selecting neighbour additive frames to morph

Issues to consider:

- choice of perceptually relevant timbre descriptors
- quality: notorious critical task, can be affected by errors

Mapping and PCS

Sound descriptors set

Sound descriptors related to perception:

- pitch: fundamental frequency $F_0(k) \in \left[0, \frac{F_s}{2}\right]$,
- loudness: sound intensity
  $I(k) = \sqrt{E(k)} \in [0, 1]$ for normalized waveforms
- timbre:
  - non-linearity strength:
    - ratio of even/odd harmonics power: $r_{\mathrm{e/o}}(k) = \frac{E_{\mathrm{even}}(k)}{E_{\mathrm{odd}}(k)} \in [0, C]$, or
    - percentage of even harmonics power: $p_{\mathrm{e}}(k) = 100 \frac{E_{\mathrm{even}}(k)}{E_{\mathrm{even}}(k)+E_{\mathrm{odd}}(k)} \in [0, 100]$
  - brightness:
    - spectral centroid: $SC(k) = \frac{\sum_{k=0}^{N/2+1} |X(k)| \cdot k}{N \cdot \sum_{k=0}^{N/2+1} |X(k)|} \in \left[0, \frac{F_s}{2}\right]$, or
    - harmonic centroid: $HC(k) = \frac{\sum_{h=1}^{H(k)} a_h(k) \cdot f_h(k)}{\sum_{h=1}^{H(k)} a_h(k)} \in \left[0, \frac{F_s}{2}\right]$

Mapping and PCS

PCS as a question of mapping strategy

Mapping: combination + signal conditioning

Verfaille, Wanderley, Depalle; "Mapping strategies for gestural and adaptive control of digital audio effects", *Jour. New Music Research*, 35(1), pp.71–93, 2006

- combination/connection:
  - PM: 3-to-4; wind controller lip pressure, air pressure in the mouth, fingerings $\longrightarrow$ sound descriptors $(F_0, I, HC, r_{e/o})$
  - $4 \times$ 1-to-1: sound descriptors synthesizer interface
  - AS: 3-to-3N, database indexing: $(F_0, HC, r_{e/o}) \longrightarrow (a_i[n], f_i[n], \varphi_i[n])$
  - AS: 1-to-1, $I \longrightarrow$ sound level
- signal conditioning:
  - normalization: parameter range
  - warping: non-linear behavior to correct biais

Improving PCS = improving the mapping strategy between:

- PM and sound descriptors: low latency extraction ($\nearrow$ controlability)
- sound descriptors and AS: database search strategies

Issues and improvements

Issues raised by previous work

Some issues from [Guillemain, Verfaille, 2007]:

- latency = bottleneck on both sides:
  - $F_s = 44100$Hz
  - PM (∼23ms): 1024 samp. block size, 0% overlap
  - Ssynth (10ms): additive analysis sampled at 100Hz

  $\Longrightarrow$ to be reduced: higher analysis rate
- differences in sound descriptor behaviors

  $\Longrightarrow$ to be compensated
- limits of 2D indexing the database based on mean note values:
  - not enough information with 2D (either $(F_0(k), HC(k))$, or $(F_0(k), r_{e/o}(k))$
  - descriptor variations:
    - \+ ok for permanent regime
    - − transients handling not satisfactory

  $\Longrightarrow$ timbre to be better represented, transients to be better handled

Issues and improvements

Reducing latency due to sound descriptor computation

Block-by-block analysis driven by $\overline{F_0}$ (WX5):

1. higher analysis rate:
   - AS: $\nearrow$ database sampling:
     - 100 $\longrightarrow$ 200 Hz $\Longleftrightarrow$ τ_{AS} : 10ms $\longrightarrow$ 5ms
     - pros: no need to change the analysis block size
     - cons: bigger database cache
   - PM: $\searrow$ block size for STFT & ACF
     - cons: $\searrow$ quality of descriptors extracted

Issues and improvements

Reducing latency due to sound descriptor computation

Block-by-block analysis driven by $\overline{F_0}$ (WX5):

1. higher analysis rate:
2. PM: block-by-block $\longrightarrow$ sample-by-sample
   filters with $T_0/2$ latency: τ_{PM} : 23 $\longrightarrow$ [5, 1.25]ms for $F_0 = [100, 400]$Hz
   - computations:
     - F_0: from 0-crossing after band-pass gammatone filter (MIDI pitch centered)
     - $I_s(n)$: LP filtering $s(n)^2$ (0-freq. gammatone filter, $f_c < F_0(t)$)
     - $SC(n) = \frac{I_{s'}(n)}{I_s(n)}$
     - $r_{e/o}(n) = \left(\frac{I_{s,\text{even}}(n)}{I_{s,\text{odd}}(n)}\right)^2$ and $p_e(n) = 100\,\frac{I_{s,\text{even}}(n)^2}{I_{s,\text{even}}(n)^2 + I_{s,\text{odd}}(n)^2}$,
       $s_{\text{even}}(n)$ and $s_{\text{odd}}(n)$ obtained from modified comb-filters (wider peaks)
   - pros: $I_s, r_{e/o}, p_e$ are accurate
   - cons: $SC(n)$ biaised for polysinusoidal signals (exact if monosinusoidal)

Rebillout, "Interfacer deux modèles de synthèse sonore: un modèle de synthèse physique pour contrôler une synthèse additive", Sept. 2007

Issues and improvements

Compensation of sound descriptors: 1. range biais

Sound descriptors behave differently due to computation:

1. range biais:
   - mapping: normalization issue (signal conditioning)
   - PM vs AS: sound qualities & brightness range differ
   - solution: user control of range mapping

Issues and improvements

Compensation of sound descriptors: 2. content biais

Sound descriptors behave differently due to computation:

2. content biais:
   - computed from whole sound (PM) vs harmonics only (AS)
   - especially for transients
   - eg.: spectral vs harmonic centroid:
     - low-level signal: higher SNR
     - SC/HC tends to half Nyquist frequency; potentially to undefined values (0/0)
   - computation correction:
     - shifting: $HC_s(k, c_0) = \frac{\sum_{h=1}^{H(k)} a_h(k) \cdot f_h(k)}{c_0 + \sum_{h=1}^{H(k)} a_h(k)}$ $\Longrightarrow$ biaised for all values

       Beauchamp, "Synthesis by spectral amplitude and 'brightness' matching of analyzed musical instrument tones," JASA (30)6, pp. 396–406, 1982
     - truncating: $HC_t(k, c_0) = \frac{\sum_{h=1}^{H(k)} a_h(k) \cdot f_h(k)}{\max\left(c_0, \sum_{h=1}^{H(k)} a_h(k)\right)}$ $\Longrightarrow$ biaised only at low-level

       Guillemain, Verfaille, "Combining physical modeling and additive synthesis as a mapping strategy for realtime control", ICMC, Vol I, pp. 442–9, 2007

$\Longrightarrow$ mapping to reduce difference of descriptors range & behavior in particular conditions

Issues and improvements

Compensation of sound descriptors: 3. morphing/search biais

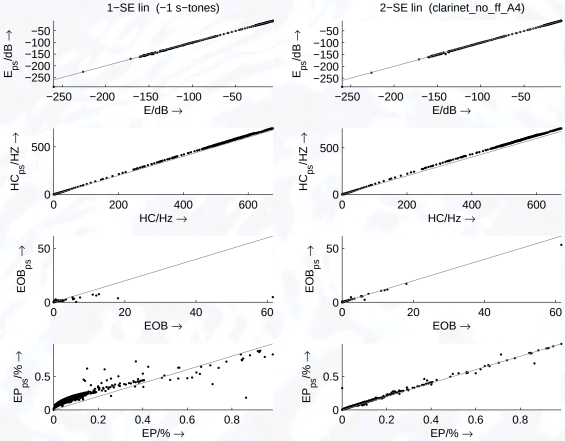
Sound descriptors behave differently due to computation:

3. morphing/search biais:
   - morphing =
     - pitch-shifting neighbor frames to target F_0
     - weighting according to distance
   - search according to descriptors before pitch-shifting
   - Are descriptor values similar after pitch-shifting with SE preservation?
   - investigation shifting ± 7 semi-tones using:
     - 1-SE: 1 linear SE $\Longrightarrow$ emphasis on artefacts due to a single & LQ SE
     - 2-SE: 1 linear SE for each harmonic comb (even & odd) more

Issues and improvements

Compensation of sound descriptors: 3. morphing/search biais

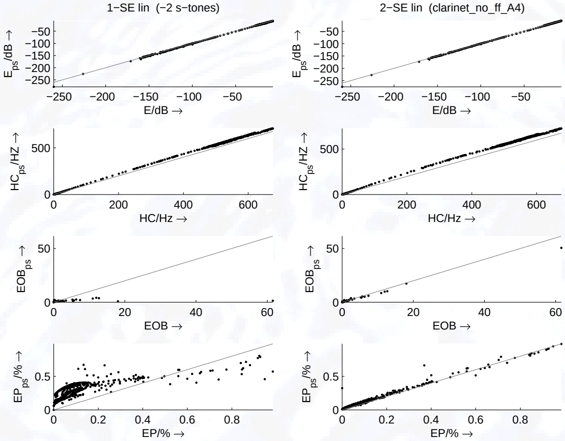
pitch-shifting: -1 semi-tone, $w \sim 0.85$



Issues and improvements

Compensation of sound descriptors: 3. morphing/search biais

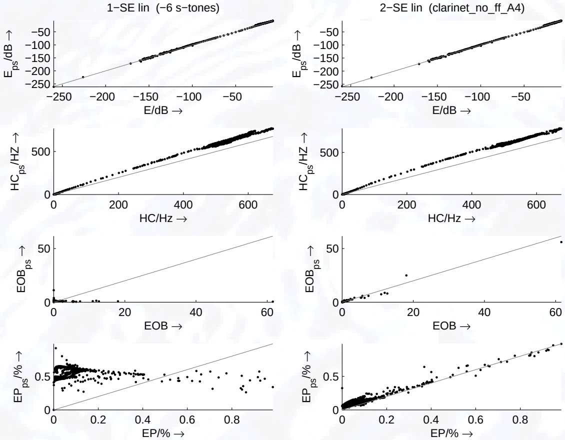
pitch-shifting: -2 semi-tones, $w \sim 0.71$



Issues and improvements

Compensation of sound descriptors: 3. morphing/search biais

pitch-shifting: -6 semi-tones, $w \sim 0.15$



▸ more

Issues and improvements

Compensation of sound descriptors: 3. morphing/search biais

Sound descriptors behave differently due to computation:

3. morphing/search biais:
   - observations:
     - more regular distribution of $p_e(k)$ in its range than $r_{e/o}(k)$ in both cases $\Longrightarrow p_e(k)$ more reliable!
     - 2-SE pitch-shifting better preserves timbre desc.: $SC(k)$ and $p_e(k) \Longrightarrow$ predictable (almost linear) transformation
   - we can compensate the error due to searching the database from descriptors before pitch-shifting

Issues and improvements

Better timbre representation and database indexing

Better timbre representation:

- 3D search: choose best neighbor frame with proper pitch $(F_0, HC, r_{e/o})$ instead of 2D search (F_0 & HC or $r_{e/o}$)
  - $+$ represents fine changes in timbre descriptors
  - $-$ faster changing trajectories in the database (potentially less smooth sounds)

Transient handling modified:

- $\nearrow$ database sampling rate: 100 Hz $\longrightarrow$ 200 Hz
- frame-by-frame neighbor search instead of mean value per note
- ignore the natural time unfolding in the database:
  - $+$ potentially looks for the 'nearest' frame (desc. value)
  - $+$ F_0 jitter is taken into account in the control data (instead of the database, via normal time unfolding)
  - $-$ even less smooth trajectories in the database
  - $-$ internal data handling modified for efficient realtime search (not finished)

Conclusions and Future Works

Contributions:

- better understanding of the interfacing via sound descriptors
- some solutions to compensate desc. behaviors

Further research:

- *Ssynth*: efficient realtime db search that ignores time-unfolding
- physical model: beating-reed phenomenon + network of toneholes
- add several versions of descr. computation to the database
  - use same computations (filter, block) on both sides
  - compare descriptor behavior depending on the mapping

Long-term:

- use truncation method in the PM to compute the additive descriptors
- perceptual evaluation of sound quality differences

Refining mapping strategies to improve the sound quality of physically-controlled synthesis

V. Verfaille[♯], A. Rebillout[♭], Ph. Guillemain[♭], M. M. Wanderley[♯]

[♯]SPCL & IDMIL

Centre for Interdisciplinary Research in Music Media and Technology

McGill

[♭]LMA–CNRS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Agence Nationale de la Recherche ANR

Fonds de recherche sur la nature et les technologies Québec

Fonds de recherche sur la société et la culture Québec

Développement économique, Innovation et Exportation Québec

NSERC CRSNG

http://www.Consonnes.cnrs-mrs.fr/

♪ Acoustics'08 Paris — July 04, 2008

Selected bibliography

J. W. Beauchamp, *"Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," J. Audio Eng. Soc., vol. 30, no. 6, pp. 396–406, 1982.*

P. Guillemain, J. Kergomard, and T. Voinier, "Real-time synthesis of clarinet-like instruments using digital impedance models," J. Acoust. Soc. Am, vol. 118, no. 1, pp. 483–494, 2005.

P. Guillemain, and V. Verfaille, Combining Physical Modeling and Additive Synthesis as a Mapping Strategy for Realtime Control, Proc. Int. Comp. Music Conf., Vol I, pp. 442–9, 2007.

R. T. Helland, "Synthesis models as a tool for timbre studies," Master's thesis, Norwegian University of Science and Technology, Department of Electronics and Communications, 2004.

J. Kergomard, S. Ollivier, and J. Gilbert, "Calculation of the spectrum of self-sustained oscillators using a variable truncation method: Application to cylindrical reed instruments," Acustica, vol. 86, pp. 685–703, 2000.

E. Tellman, L. Haken, and B. Holloway, " Timbre Morphing of Sounds with Unequal Numbers of Features," J. Audio Eng. Soc., vol. 43, no. 9, pp. 678–89, 1995.

*V. Verfaille, J. Boissinot, P. Depalle, and M. M. Wanderley, "*Ssynth*: a real time additive synthesizer with flexible control," Proc. Int. Computer Music Conf., 2006.*

M. M. Wanderley Ed., Organised Sound, Special issue on mapping. Francis & Taylor, 2002, vol. 7, no. 2.

T. A. Wilson and G. S. Beavers, "Operating modes of the clarinet," J. Acoust. Soc. Am, vol. 56, pp. 653–658, 1974.