Singing Voice Resynthesis

Nuno Fonseca

Introduction

The concept of capturing the sound characteristics of “something” is not new

Examples

- A sampler that replicates a Stradivarius violin
- A reverberation unit that replicates the reverb of Sydney Opera House
- An effects unit that replicate the “valve” characteristics of legendary audio devices

Typical Applications

	Capturing	
	Original Sound	Sound Transformation
Synthesis	Concatenative Synthesis	-
Audio Effect	-	Convolution

Typical Applications

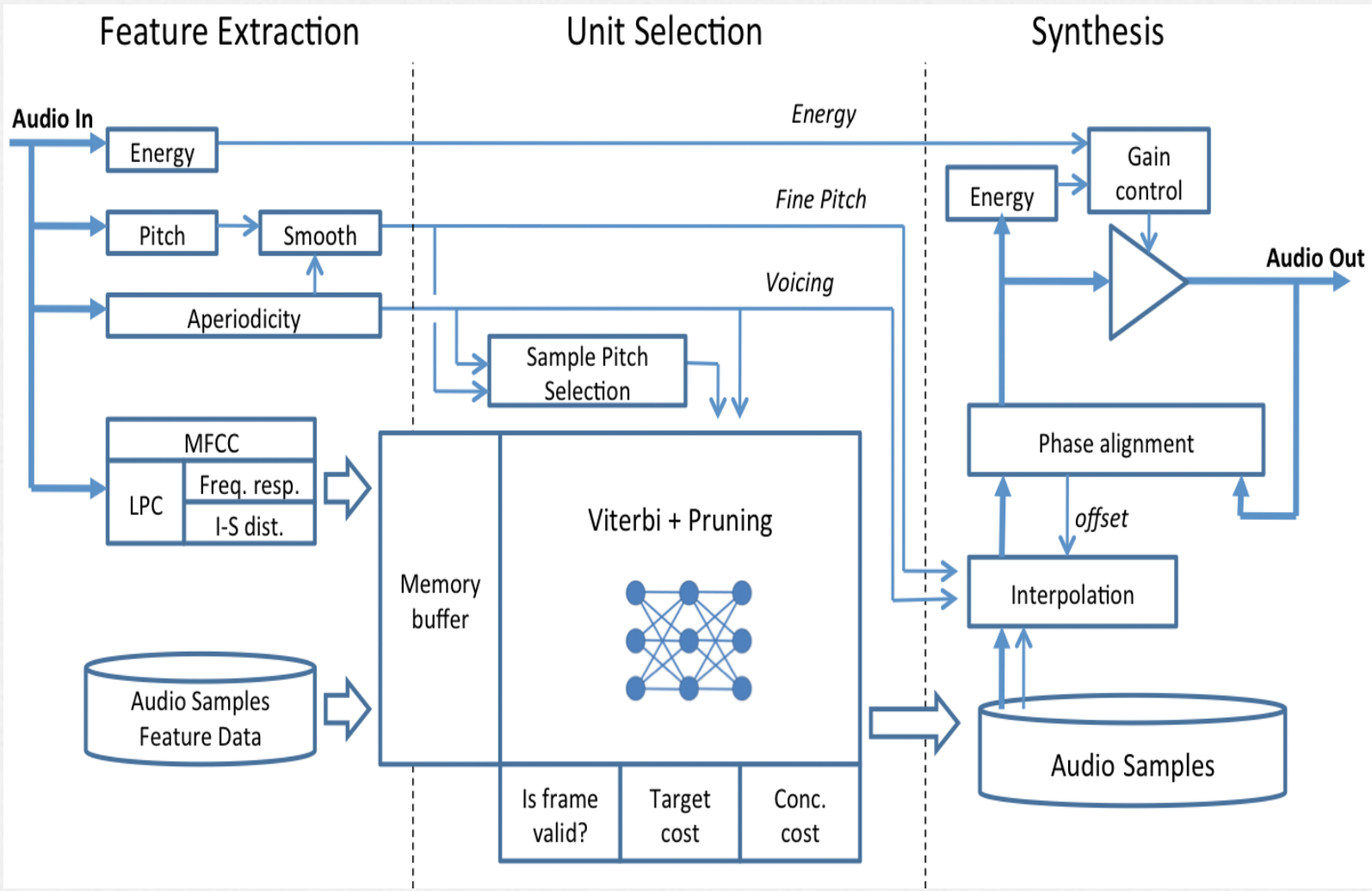
	Capturing	
	Original Sound	Sound Transformation
Synthesis	Concatenative Synthesis	-
Audio Effect	*Concatenative Resynthesis*	Convolution

Concatenative Resynthesis

- Concept is simple
  - e.g. feature extraction -> synthesis
  - using pre-recorded material
- Implementation is very complex…

Concatenative Resynthesis

- For singing voice
- Receives an audio stream
  - monophonic singing voice
- Outputs an audio stream
  - Same semantic information
    - Pitch, Phonemes, Dynamics
  - Uses an internal sound library


Feature Extraction
Unit Selection
Synthesis
Audio In
Energy
Pitch
Smooth
Aperiodicity
MFCC
LPC
Freq. resp.
I-S dist.
Audio Samples
Feature Data
Energy
Fine Pitch
Voicing
Sample Pitch
Selection
Memory
buffer
Viterbi + Pruning
Is frame
valid?
Target
cost
Conc.
cost
Energy
Gain
control
Audio Out
Phase alignment
offset
Interpolation
Audio Samples

Feature Extraction

- Dynamics (sound intensity)
  - Energy
- Pitch
  - YIN + Aperiodicity + smoothing
- Phonetic related info
  - MFCC
  - LPC
  - Aperiodicity (Voicing)

  [Not a phoneme recognizer]

Unit Selection

- Phonetic similarity
- Target Cost + Concatenation Cost

$$D_{i,j} = \frac{\sqrt{D_{MFCC}(i,j)^2 + D_{LPC\ resp}(i,j)^2 + D_{LPC\ dist}(i,j)^2 + D_{Ap}(i,j)^2}}{2}$$

- Viterbi + Pruning

Synthesis

- Pitch adjustments
  - Interpolation
- Phase alignment
  - offset tests
- Gain control

Some tests

- Amazing Grace (LeAnn Rimes)
- Frozen (Madonna)
- Tom's Diner (Susanne Vega)
- Whenever (Shakira)



Conclusions

- High level/semantic audio effect
  - Music/Speech
  - Analysis/Synthesis
- Much more work required
  - Audio artifacts
- Potential for future applications

Questions

Nuno Fonseca
(nuno.fonseca@ipleiria.pt)