Digital Music Representation
2) sound structure

Geoffroy Peeters
peeters@ircam.fr
CUIDADO I.S.T. - Ircam Analysis/Synthesis Team

ircam Centre Pompidou

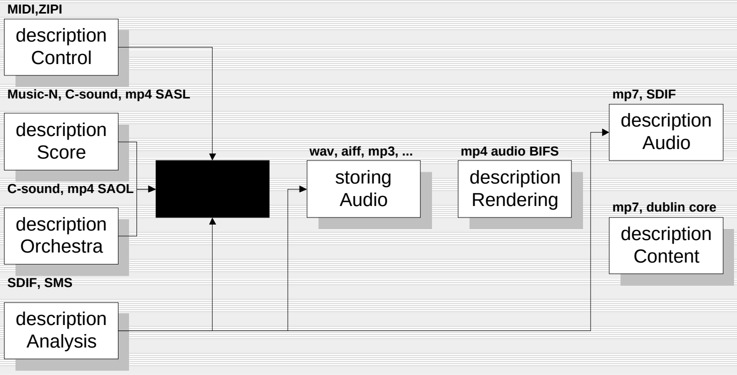
Introduction

- Four time-scales:
  - Eric Scheirer (1998) “The temporal structure of the pressure waveform has a significant effect on whether or not we find a sound pleasing or interesting.
    This structure happens on several different time scales.
    - Notes in a musical composition change on a time scale of hundreds of milliseconds.
    - The timbre of sounds changes on a time scale of tens of milliseconds.
    - The actual sound waveform changes on a time scale of tens of microseconds”
  - sound/music descriptions -> time-less
- Representations
  - to store audio: wav, aiff, mp3, …
  - to create sound: Music-N, C-sound, Mpeg-4 SA
  - to control sound: MIDI, ZIPI
  - to render sound: mp4 audio BIFS
  - of audio analysis : SDIF, SMS, …
  - of sound’s description: mp7, Dublin Core

ircam Centre Pompidou

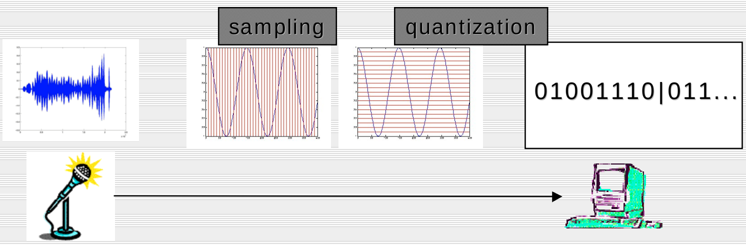
G. Peeters 2

Introduction



0. Signal representations basics
Time-domain

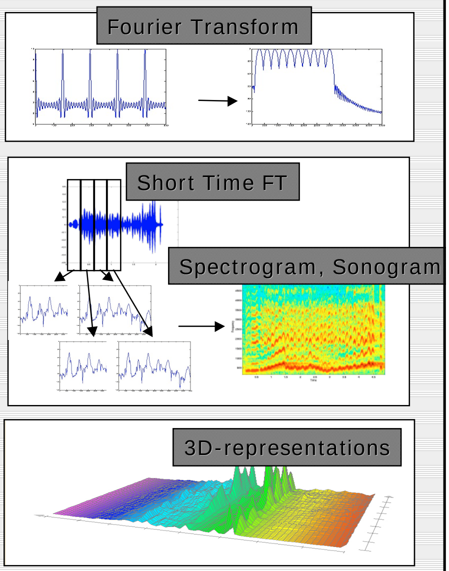
- sound pressure s(t)
  - discrete time (sampling rate: 11025, 22050, 44100, 48000, ...)
  - discrete values (quantization: 2^{8bits}=256 values, 2^{16bits}= 65536 values, ...)
  - signal s(n)



0. Signal representations basics
Time-frequency/Time-scale domain

è from the time-domain to the frequency domain
- è most-used transforms:
  - è Fourier Transform, Fast Fourier Transform (FFT) algorithm
    - Ô decompose the signal on a set of sinusoidal component
  - è Discrete Cosine Transform
  - è Wavelet Transform
    - Ô decompose the signal on a set of well-localized in time and frequency functions (time-scale transforms)

è time/frequency domain
- è perform a transform around a given time: Short-Time Fourier-Transform
  - è 2-D representations:
    Time | Frequency
    color represents amplitude
  - è 3-D representations:
    height represents amplitude

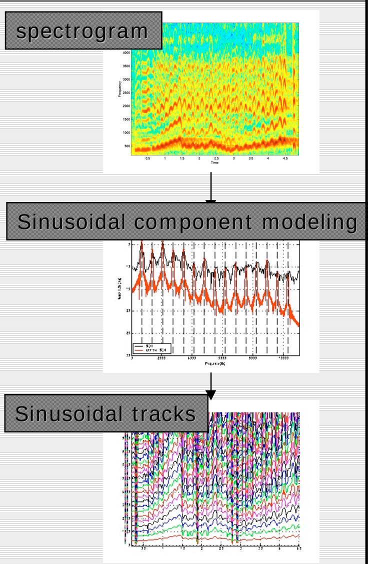


ircam Centre Pompidou

G. Peeters 5

0. Signal representations basics
Signal models domain

è Most common signal models:
- è Periodic/Harmonic model
- è Sinusoidal additive model + noise
- è Source / Filter decomposition
  - è LPC (linear predictive coding)
    - Ô estimates the filter coefficients or the resonance, (in the case of speech: formant/anti-formant)
  - è CELP: coded excitation LP (source signal decomposed on a codebook)
  - è CEPSTRUM: representation of the shape of the spectrum
  - è MEL-CEPSTRUM: CEPSTRE using a Mel scale for the spectrum (perceptual scale)
- è Granular, pitch-synchronous granular model
  - è represents the signal as a succession of grains possibly at a periodic rate

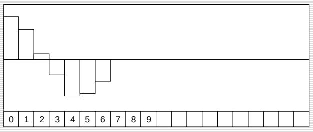


ircam Centre Pompidou

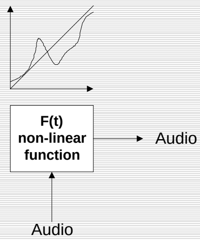
G. Peeters 6

0. Signal representations basics
Synthesis algorithms

- Wave-table synthesis



- Wave-shaping synthesis
  - apply a non-linear function to the amplitude of a sound (example: clipping that occurs when an audio amplifier is overdriven)
  - broad range of musically useful timbres
  - dynamic spectra

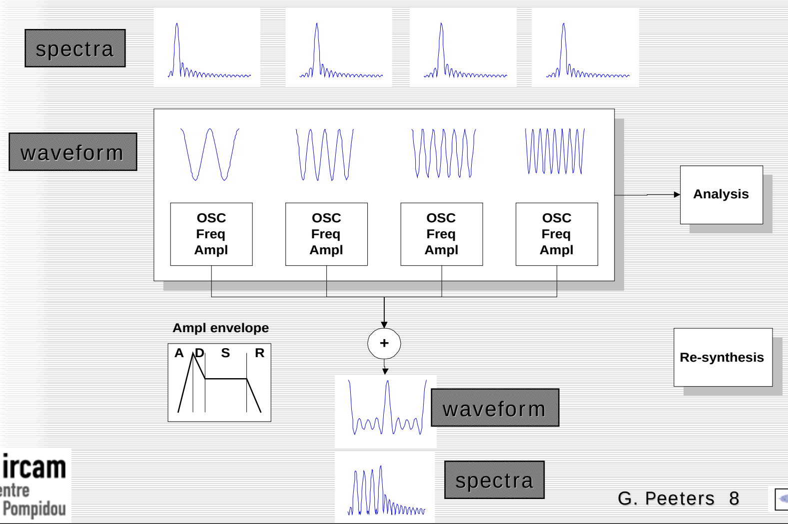


ircam Centre Pompidou

G. Peeters 7

0. Signal representations basics
Synthesis algorithms

- Additive synthesis:
  create a complex waveform by addition of basic waveforms (sinusoidal components)

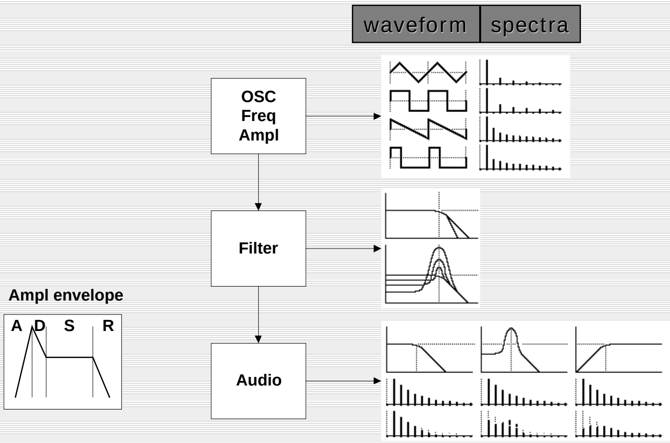


ircam Centre Pompidou

G. Peeters 8

0. Signal representations basics
Synthesis algorithms

è Subtractive synthesis:
remove components by filtering a spectrally rich waveform
(waveform=white noise, saw-tooth, triangular, square waveform)

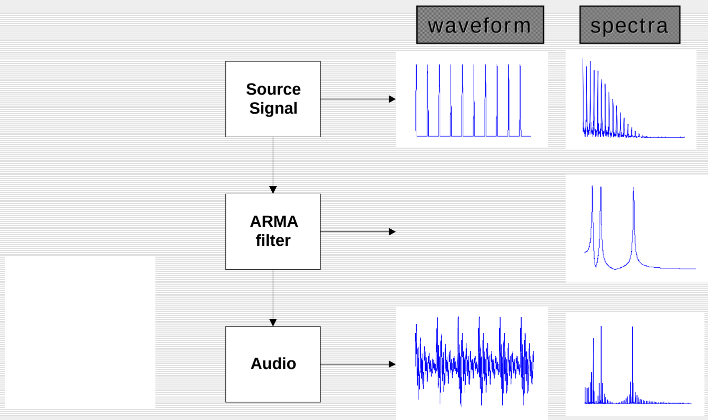


ircam Centre Pompidou

G. Peeters 9

0. Signal representations basics
Synthesis algorithms

è Source/Filter synthesis:
based on physics: a periodic “source” signal is filtered by a filter
(ARMA filter) (see vowel synthesis = formant synthesis)

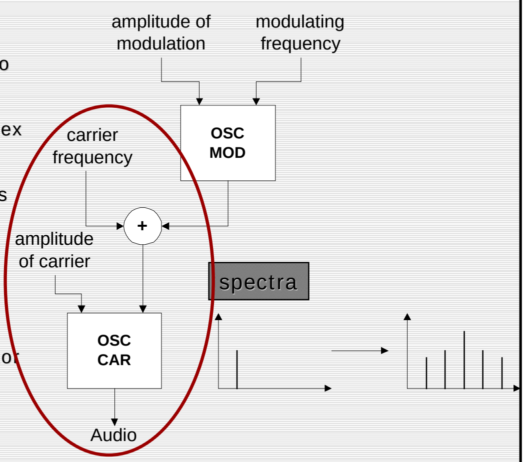


ircam Centre Pompidou

G. Peeters 10

0. Signal representations basics
Synthesis algorithms

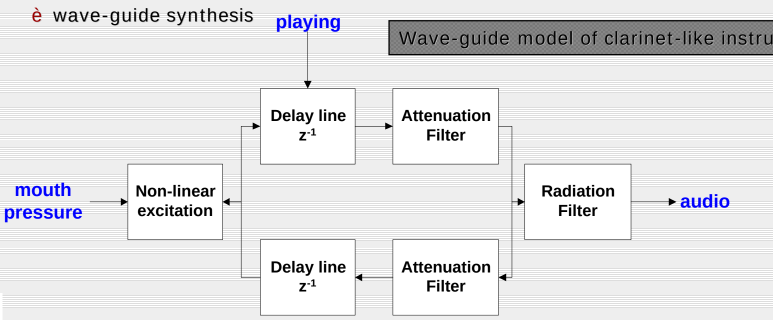
- FM synthesis:
  - Principle:
    - low frequency oscillator modulating a VCO -> vibrato effect
    - high frequency oscillator modulating a VCO -> complex spectrum (see Bessel functions)
    - Complex spectrum: depends on the magnitude of the modulation
  - a set of "operators"
  - Each operator consists of:
    - A digitally controlled oscillator (DCO)
    - An amplifier
    - An envelope generator
  - minimum: two operators
    - a modulator
    - a carrier



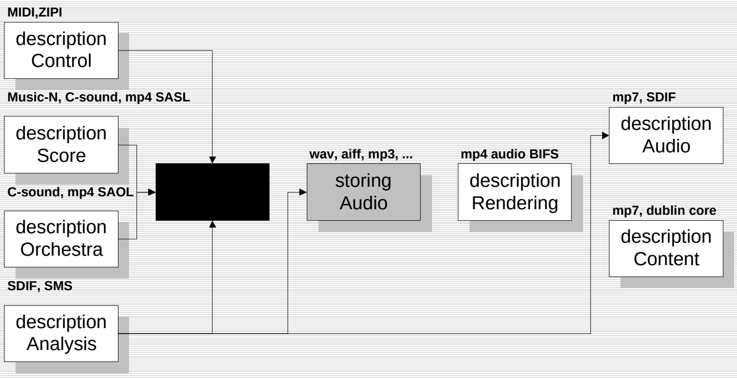
0. Signal representations basics
Synthesis algorithms

- Physical modeling synthesis:
  - represent the object producing the sound instead of the produced sound itself
  - Most well-known physical model: Karplus-strong (plucked string)
  - require a different model for each instrument
  - Various modeling techniques for physical models
    - mass-spring paradigm
    - modal synthesis
    - wave-guide synthesis

Wave-guide model of clarinet-like instrument



I. Representations to store audio

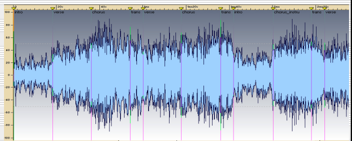


ircam Centre Pompidou

G. Peeters 13

I. Representations to store audio

- Non-compressed audio representations
  - Representation of the sound (sampled and quantified) pressure: s(n)
  - Using PCM (Pulse Code Modulation)
  - Store the audio signal and related information for reading it
  - Most format are based on "chunks"
    - data chunks
    - information chunks
      = information needed in order to read the file: big/little endian, sampling rate, quantization, nbchannels
    - extra chunks:
      - marker chunks
      - instrument chunks (play mode, loop points)
      - midi data chunk,
      - comments chunk (marker+text)
      - application specific chunks
      - meta-information: name, author, copyright, annotation chunks
      - wav: bitmap
  - Examples:
    - aiff (Audio Interchange File Format) (Mac)
    - wav PCM (RIFF wav) (Windows)
    - au, snd (sun/next station)



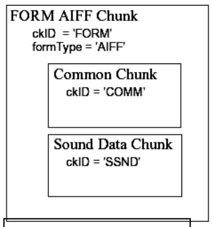
Media information
-Sampling rate
-Sample size
-number of channels
- ...

ircam Centre Pompidou

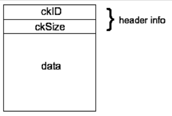
G. Peeters 14

I. Representations to store audio

AIFF structure



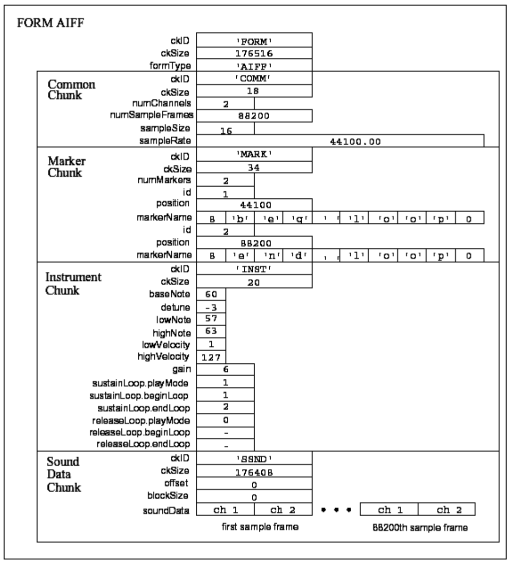
Chunk format



Common chunk

```
#define    CommonID  'COMM'    /* ckID for Common Chunk */

typedef struct {

    ID             ckID;
    long           ckSize;

    short          numChannels;
    unsigned long  numSampleFrames;
    short          sampleSize;
    extended       sampleRate;

}  CommonChunk;
```

Example

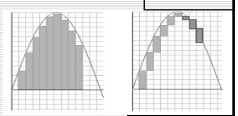


I. Representations to store audio

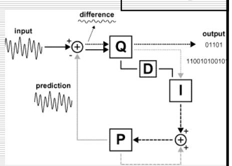
- Compressed audio representations
  - Not a direct representation of the sound pressure
  - Examples:
    - wav mu-law, a-law, DPCM, IMA ADPCM
    - aiff-c, aifc
    - MPEG-1 Layer III (subband-coding, masking)
    - MPEG-2 AAC

    - QT (QuickTime)
    - WMA (Windows Media Audio)
    - RA (Real Audio)
    - SWA (Shockwave Audio)
    - ATRAC3 (Adaptive Transform Acoustic Coding 3)

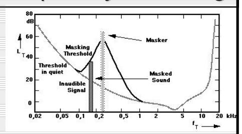
DPCM



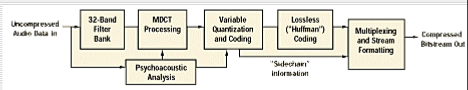
ADPCM



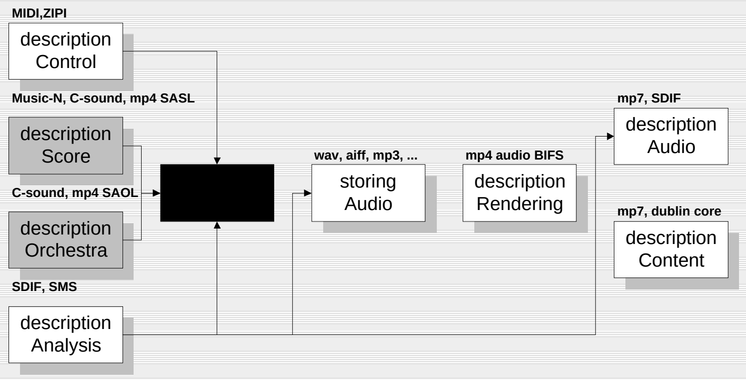
Frequency masking



Perceptual coders



II. Representations to create sound



G. Peeters 17

II. Representations to create sound
C-Sound

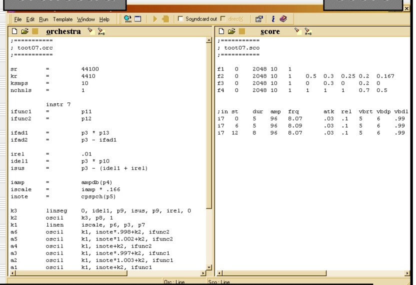
- History of computer sound synthesis
  - 1957: Max Mathews: Music I (IBM 704 computer)
  - ...
  - 1960: Music III: concept of modular instruments (unit generators)
  - 1962: Music IV: orchestra/score model
  - 1969: Music V
  - 1980: Richard Moore: C-music
  - 1985: Barry Vercoe: C-sound
- Concept:
  - Richard Boulanger
    - "C-sound is a sound renderer. It works by first translating a set of text-based instruments, found in the orchestra file, into a computer data-structure that is machine-resident. Then, it performs these user-defined instruments by interpreting a list of note events and parameter data that the program "reads" from: a text-based score file, a sequencer-generated MIDI file, a real-time MIDI controller, real-time audio, or a non-MIDI devices such as the ASCII keyboard and mouse."

G. Peeters 18

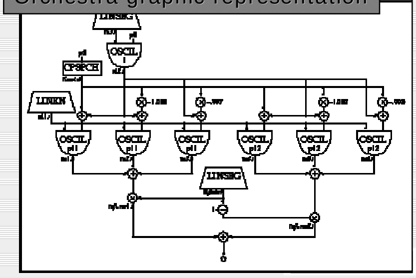
II. Representations to create sound
C-Sound

- C-sound concepts:
  - distinction between orchestra and score languages
  - orchestra: modular instruments, “unit generator”, macro language
  - distinction between sample rate and control rate
    - i-rate variables: changed at the note rate
    - k-rate variables: changed at the control signal rate
    - a-rate variables: changed at the audio signal rate
  - compilation creates algorithm which then creates sound
  - latest versions: MIDI compatible, real-time use
  - Syntax example:
  - Unit generator list:
    - sinusoidal generator
    - reverb generator, …
    - …

orchestra → score



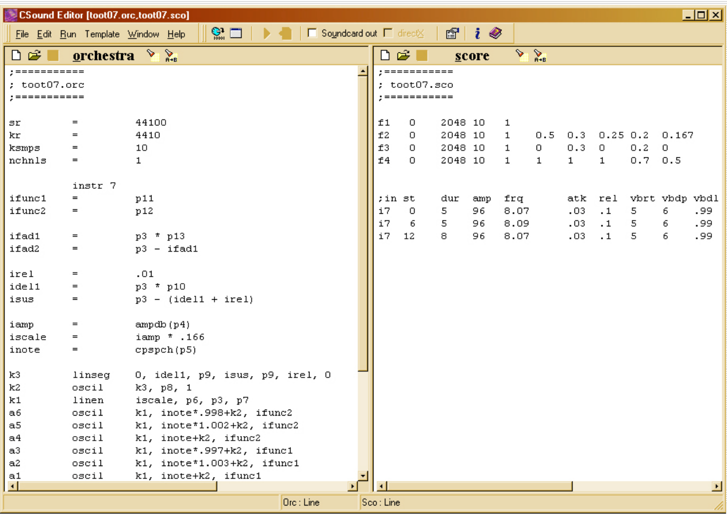
Orchestra graphic representation



ircam Centre Pompidou

II. Representations to create sound
C-Sound

orchestra → score



ircam Centre Pompidou

G. Peeters 20

II. Representations to create sound
MPEG-4 Structure Audio (Scheirer, 1998)

- History:
  - "Netsound" (Casey, Smaragdis 1996): using C-sound to transmit sounds over networks
  - C-sound intellectual properties issues -> Scheirer develops SA
- Structure Audio: "a Music Synthesis Language"
  - SAOL: Structure Audio Orchestra Language
  - SASL: Structure Audio Score Language

  - "highly readable", "highly modular", "highly expressive", "highly functional"
  - more understandable and concise than C-sound ("C-like" language)
  - new features
    - 100 built-in "unit generator", may be extended with new "unit-generator"
    - "cpuload" -> allows Dynamic voice-stealing algorithms in the orchestra
    - ...
  - SAOL may be controlled by
    - SASL
    - MIDI-file
    - real-time MIDI events

ircam Centre Pompidou

G. Peeters 21

II. Representations to create sound
MPEG-4 Structure Audio (Scheirer, 1998)

SAOL

```
instr beep3(pitch) {
  imports ksig amp, off;              // controllers
  ksig vol, init;
  table wave(harm,2048,1);
  asig sound;

  if (!init) {                        // first time we're called
    amp = 0.5; init = 1;
  }
  if (off) { turnoff; }               // we got the 'off' control
  vol = port(amp,0.2);                // smooth the volume signal
  sound = oscil(wave,pitch);
  output(sound * vol);
}
```

SASL

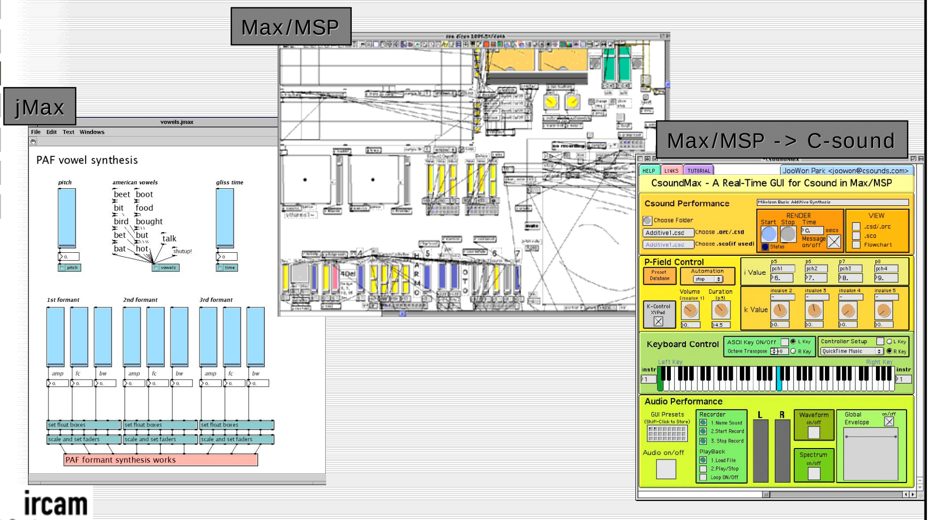
```
n1: 0.0 beep3 1 440                  // first note
    0.5 beep3 1 480                  // second note

n2: 1.0 beep3 -1 220                 // third note - unbounded initial duration
n2: 1.0 beep3 -1 440                 // fourth note
n3: 1.0 beep3 -1 660                 // fifth note

    2.0 control n2 amp               // applies to third and fourth notes
    2.5 control n2 amp 0.5
    3.0 control n3 amp 0.2           // applies to fifth note
    3.0 control n2 amp 0.2
    4.0 control n2 off 1
    4.0 control n3 off 1
```

ircam Centre Pompidou

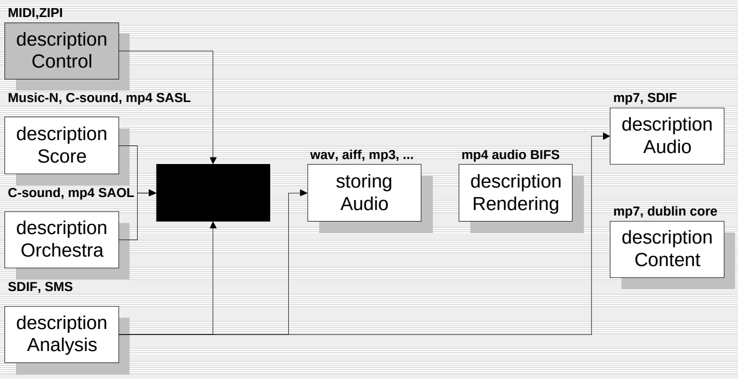
G. Peeters 22

II. Representations to create sound
"Unit generator", Modular graphical conception Synth

è Nyquist, CLM, SuperCollider, jMax, MAX-MSP, Pure Data, Reaktor, ...



III. Representations to control sound



III. Representations to control sound
MIDI

- MIDI (Musical Instrument Digital Interface) 1982/1983
  - Goal:
    allow musicians to connect synthesizers together
  - How:
    instruct a synthesizer which sounds to use, which notes to play, how loud to play it, ...
  - Characteristic:
    unidirectional asynchronous bit stream at 31.25 Kbits/sec
  - Today's usage:
    used in digitized audio in games and multimedia applications
  - Advantages
    - storage place:
      transmit only the playing information not the sound
      (1 minutes of audio: 10Mbytes in midi only 10Kbytes)
    - easiness of manipulation

G. Peeters 25

III. Representations to control sound
MIDI

- MIDI messages
  - Channel Voice Messages (16 channels)
    carry musical performance data
    - examples: Note On, Note Off, Polyphonic Key Pressure, Channel Pressure, Pitch Bend Change, Program Change, and the Control Change messages
  - Mode Messages
    affect the way a receiving instrument will respond to the Channel Voice messages
    - examples: reset all controller, Omni On/Off
  - System messages
    - System Common Messages:
      include MTC Quarter Frame, Song Select, Song Position Pointer, Tune Request, and End Of Exclusive (EOX) -> Midi Time Code
    - System Real Time Messages:
      used to synchronize all of the MIDI clock-based
      - Timing Clock, Start, Continue, Stop, Active Sensing, and the System Reset message
  - Midi limitations:
    - 1) limited bandwidth -> synchronization
      MIDI = 31.25 Kbit/s, serial transmission (possible that musical events which originally occurred at the same time may not actually be played at exactly the same time)
    - 2) dedicated to keyboard-like instruments

G. Peeters 26

III. Representations to control sound
Midi Files - GM - GS - XG

- Standard Midi Files:
  - why ? midi has been developed for real-time, need to add a "time-stamping" for the MIDI messages when stored
  - Format 0: single track, Format 1: several tracks,Format 2: several independent patterns
- General Midi (GM):
  - why ? problem with "pure" MIDI: no standard for the relationship between patch numbers and specific sounds for synthesizers
  - definition of a General MIDI Sound Set (a patch map): 128 basic patches, MIDI Channels 1-9 and 11-16
  - definition of a General MIDI Percussion map (mapping of percussion sounds to note numbers): MIDI Channel 10
- Roland GS
  - why ? extension of GM because of GM limitations
  - additional controllers (synthesizer parameters, effect parameters)
  - scheme of variation tones (banks of GM sounds)
  - scheme of tone 'fallback' (for non-existent banks)
- Yamaha XG
  - a more powerful extension of GM

ircam Centre Pompidou

G. Peeters 27

III. Representations to control sound
ZIPI

- ZIPI (Gibson/ Zeta Music, CNMAT)
- Motivation:
  - MIDI is based on keyboard controlling: a note cannot start without a pitch
  - not appropriate for controlling from guitar, violin, sustained instruments
- ZIPI characteristics
  - network oriented, dynamic assignment
  - larger bandwidth: lower bandwidth=10 times MIDI
- ZIPI
  - Music Parameter Description Language (MPDL)
  - MIDI
  - Data dump
  - Digital Audio

ircam Centre Pompidou

G. Peeters 28

III. Representations to control sound
ZIPI

- ZIPI Music Parameter Description Language (MPDL)
  - 1) note address
    - MIDI: address=channel, or address=pitch's note
    - ZIPI MPDL: address=note whatever the pitch
    - ZIPI MPDL: group of group of notes:
      "notes" (127) in "instruments" (127) in "families" (63) = 1.016.127 addresses
      - MIDI: no way to articulate an entire chord with a single message (note or channel based message)
    - each ZIPI device has its own address space
  - 2) note descriptors
    - syntax: descriptor ID | value
    - example of descriptor:
      - articulation, pitch, frequency in Hz, Amplitude, Loudness, Brightness, Even/odd harmonic balance, roughness, spatialization azimuth angle, timbre space X dimension, …
      - articulation: note sounds, note re-attacks
      - program change
      - Higher order-messages:
        - modulation, housekeeping (defines notes priority), querying a synthesizer, comments, time-tags (ensure synchronization when streaming ZIPI (minimum delay to respect))
  - 3) in ZIPI: distinction between Controller and Synthesizer Messages (in MIDI both are mixed)

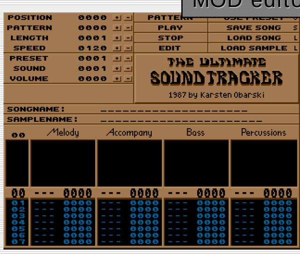
ircam Centre Pompidou

G. Peeters 29

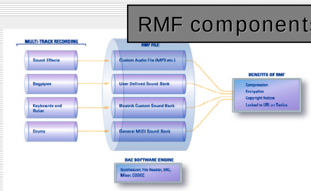
III. Hybrid sound/control representations
MOD/RMF

- MOD
  - tracker file format
  - contains both
    - musical score information
    - actual instrument sound samples that are used to play a tracker song.
- RMF (Rich Music Format)
  - three types of information:
    - highly compressed sound samples,
    - control data (playback instructions, like MIDI),
    - specifications for interactivity

MOD editor



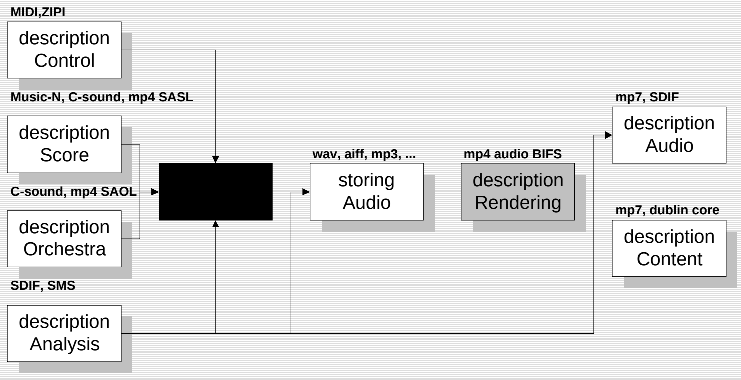
RMF components



ircam Centre Pompidou

G. Peeters 30

IV. Representations to render sound



ircam Centre Pompidou

G. Peeters 31

IV. Representations to render sound

- Multi-channel file formats
  - Dolby Digital (Dolby AC-3) 5.1
    - 5 discrete (independent) channels (center, left, right, surround left, surround right) (from 20Hz to 20,000 Hz)
    - 6th channel dedicated for low frequency effects (LFE) (3 Hz to 120 Hz).
  - DTS Digital Surround 5.1
  - ...
- Rendering description formats
  - MPEG-4 Audio BIFS
    - BIFS (Binary Format for Scenes) in SNHC (Synthetic/Natural Hybrid Coding) based on VRML
  - Sound attaches to objects: sound node used to put sound in the scene
  - MPEG-4 Version 2: Advanced Audio BIFS
    - Sound node
    - Sound BIFS
      - Acoustic/physical approach
      - Perceptual approach (SPAT)

Acousmonium

Espace de projection

ircam Centre Pompidou

G. Peeters 32

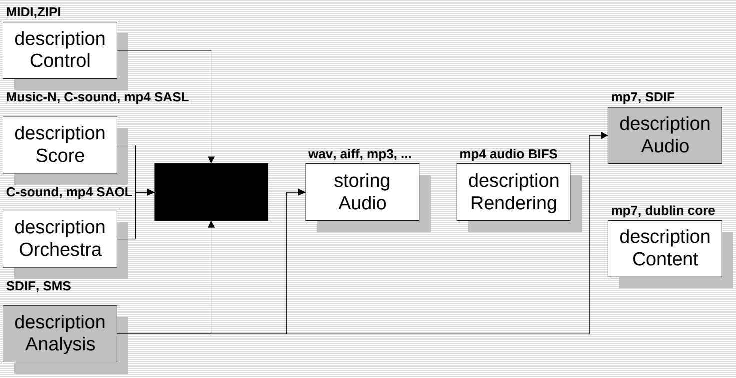
IV. Representations to render sound

- Two approaches:
- Acoustic/physical approach
  - Directive sound
    - source
    - location, direction, intensity, speedOfSound
    - angles, directivity filters
    - useAirabs, spatialize, roomEffects
  - Acoustic Scene
    - late reverberation, reverbTime, reverbLevel, reverbDelay
    - center, size of bounding box
  - AcousticMaterial
    - acoustic reflectivity and transmissions filters
    - ...
- Perceptual approach (SPAT)
  - Directive Sound
  - Perceptual Parameters
    - source presence,
    - warmth,
    - brilliance,
    - room presence,
    - envelopment,
    - running reverberance,
    - late reverberance,
    - heaviness, liveness

ircam Centre Pompidou

G. Peeters 33

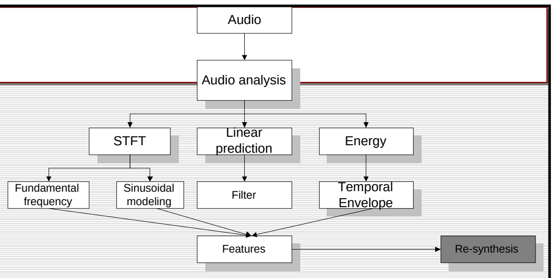
V. Representations of audio analysis



ircam Centre Pompidou

G. Peeters 34

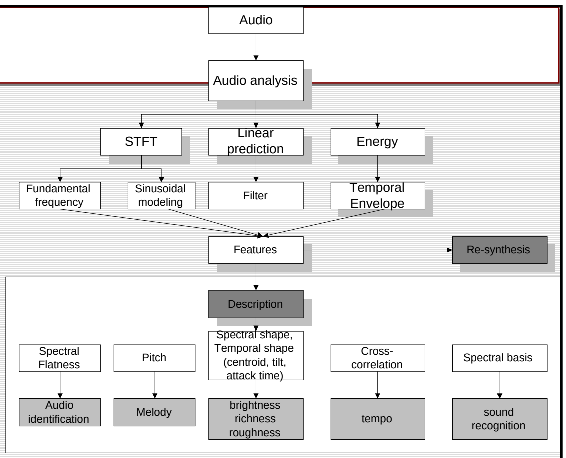
V. Representations of audio analysis



- What are the results of an audio analysis used for ?
  - Parameters for re-synthesis
    - STFT
    - Sinusoidal models
    - Source/filter
    - …

ircam Centre Pompidou

V. Representations of audio analysis

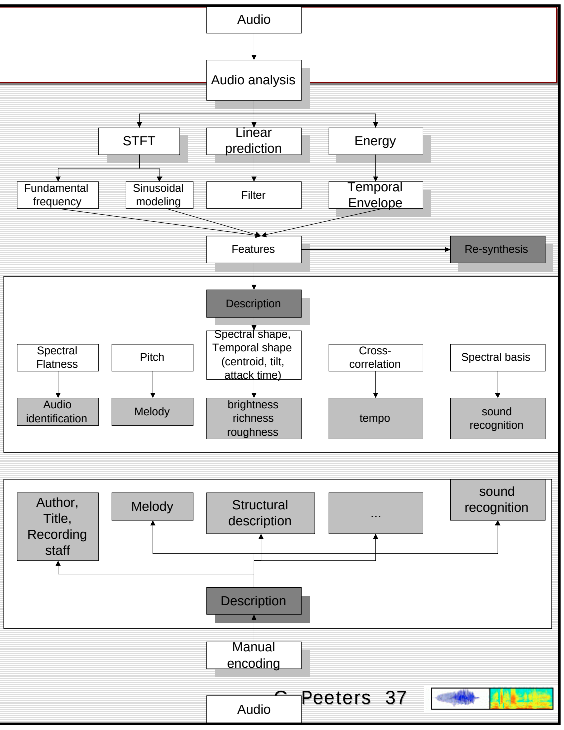


- What are the results of an audio analysis used for ?
  - Parameters for re-synthesis
    - STFT
    - Sinusoidal models
    - Source/filter
    - …
  - Parameters for description
    - Fundamental frequency
      - pitch
    - Spectral centroid
      - brightness,
    - Spectral Flatness
      - Audio ID / Fingerprint
    - classification, indexing
    - high-level synthesis
  - What about manual encoding ?

ircam Centre Pompidou

V. Representations of audio analysis

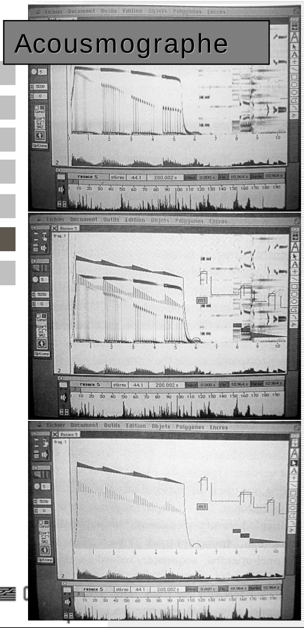
- What are the results of an audio analysis used for ?
  - Parameters for re-synthesis
    - STFT
    - Sinusoidal models
    - Source/filter
    - ...
  - Parameters for description
    - Fundamental frequency
      - pitch
    - Spectral centroid
      - brightness,
    - Spectral Flatness
      - Audio ID / Fingerprint
    - classification, indexing
    - high-level synthesis
  - What about manual encoding ?

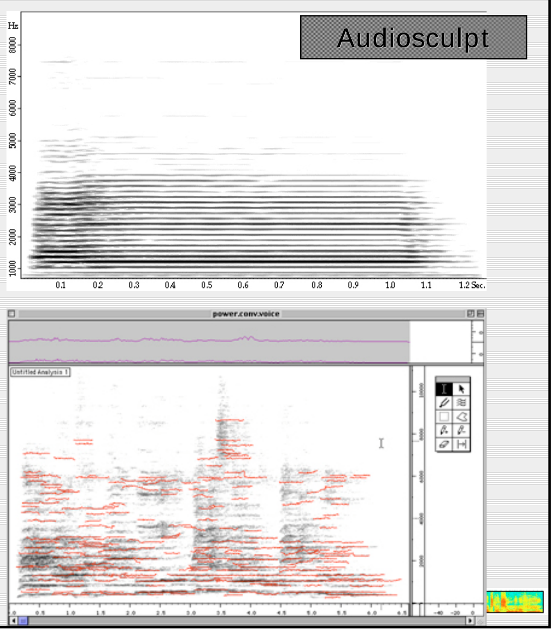


ircam Centre Pompidou

G. Peeters 37

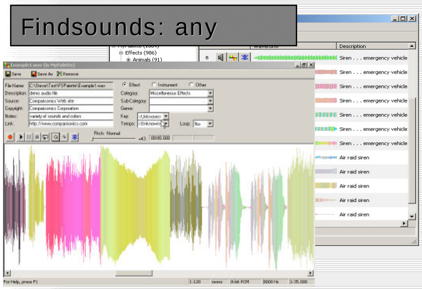
V. Representations of audio analysis
Graphical representations: spectral representations



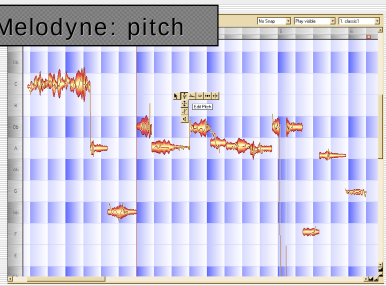


V. Representations of audio analysis
Graphical representations: descriptors representations

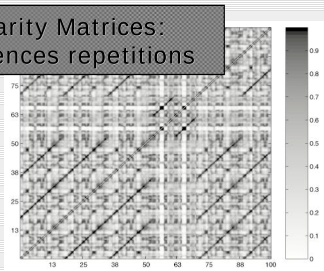
Findsounds: any



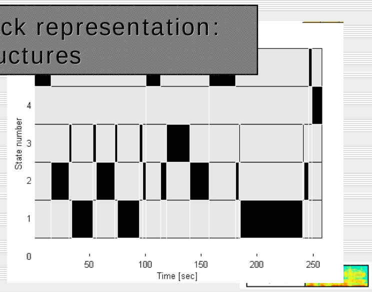
Melodyne: pitch



Similarity Matrices: sequences repetitions



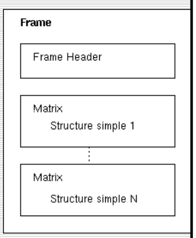
Track representation: structures



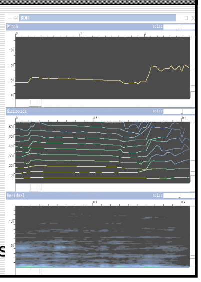
V. Representations of audio analysis
SDIF

- SDIF (Sound Description Interchange Format) 1998 (CNMAT, Ircam, IUA/UPF)
  - Meta-format (// XML)
  - Definition of a file formats and structures: chunk-based
  - Frame-based (self defined)
    - Stream ID
    - Frame type/matrix type examples:
      - 1FQ0 Fundamental Frequency Estimates,
      - 1STF Discrete Short-Term Fourier Transform,
      - 1TRC Sinusoidal Tracks,
      - 1RES Resonance / Exponentially Decaying Sinusoids
    - Extensible !
  - Tools:
    - compliant software for coding/ decoding, visualization
    - libraries for SDIF files manipulation
  - Conclusion:
    - allows any type of descriptions although only signal-based analysis/features are defined
  - SDIF extensions:
    - XML DTD for Frame/Matrix definition
    - SDIF / XML: for streams relationship
    - SDIF to MPEG-4 cross-coder source code

SDIF structure



SDIF viewer (IU/UPF)

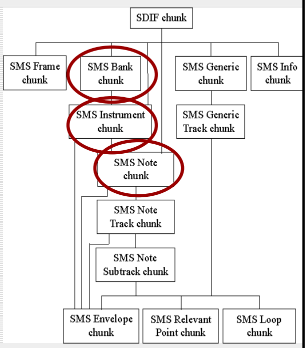


G. Peeters

V. Representations of audio analysis SMS

- SMS: File Format
  - based on SDIF
  - + a set of chunks specific to SMS
  - different levels of spectral abstraction useful in a variety of applications that go from high quality audio coding to synthesis applications that use banks of spectral data
    - SMS Generic chunk: mono, stereo, quadraphonic
    - SMS Generic Track chunk: Track_ID, Magnitude_Threshold, Spectral_Range
    - SMS Note chunk: Default_Magnitude_Threshold, Note_Type (pitched, unpitched), pitch
    - SMS Note Track chunk: Note_Track_Type (whole note, attack, steady state, release, articulation)

SMS file format structure

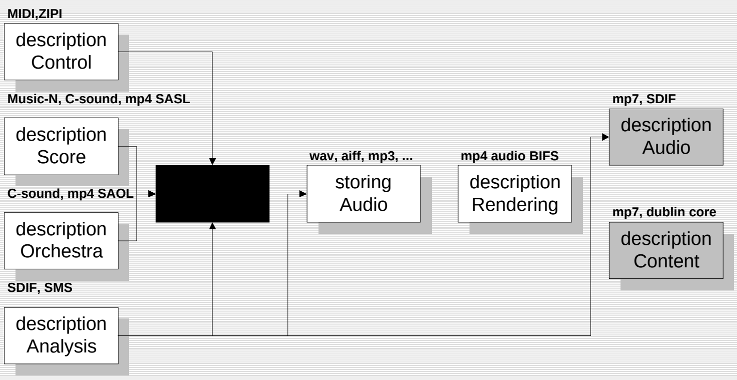


- SMS: Score Format
  - Time statements,
  - Event statements,
  - Synth parameters
  - Controller parameters:
    Key Velocity Wind pressure, Key Number Embouchure, Key Pressure Lip pressure, Pitch-bend wheel Lip, frequency, Mod wheel Wind keypads, Switch pedal, Continuous pedal Drum head striking X position, Drum, head striking Y position, ...

ircam Centre Pompidou

G. Peeters 41

VI. Representations of sound's description

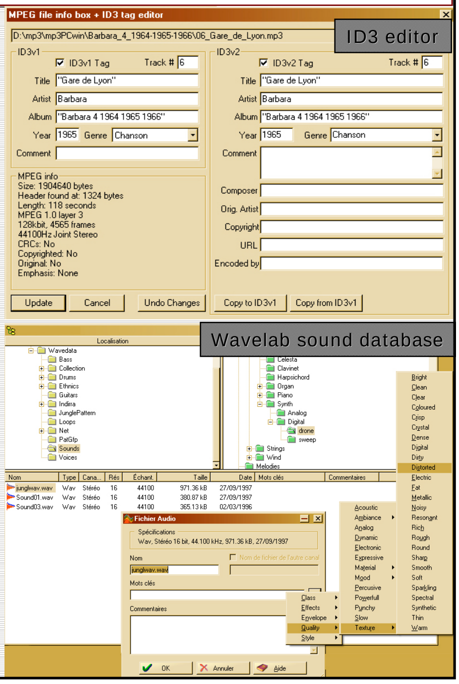


ircam Centre Pompidou

G. Peeters 42

VI. Representations of sound's description

- Meta-data description of a sound ?
  - Content description ?
  - Data related to the sound content
or
describing the sound content
  - Data contained in the sound content
or
not contained in
  - Examples
    - Information chunk in wav/aiff
    - ID3v1/v2
    - Dublin Core
    - MPEG-7 Multimedia Description Interface



ircam Centre Pompidou

VI. Representations of sound's description
Dublin Core

- Dublin Core (1995):
  - description of electronic media
  - 15 elements
  - web resources oriented
  - simpler than MARC

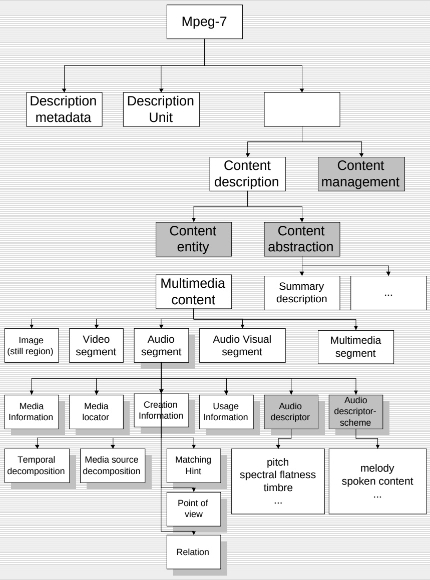
	DC Element	Definition
1	Title	A name given to the resource
2	Creator	An entity primarily responsible for making the content of the resource
3	Subject	The topic of the content of the resource
4	Description	An account of the content of the resource
5	Publisher	An entity responsible for making the resource available
6	Contributor	An entity responsible for making contributions to the content of the resource
7	Date	A date associated with an event in the life cycle of the resource
8	Type	The nature or genre of the content of the resource
9	Format	File format or mime type (MPEG-1, QuickTime, RealVideo…)
10	Identifier	An unambiguous reference to the resource within a given context
11	Source	A Reference to a resource from which the present resource is derived
12	Language	A language of the intellectual content of the resource
13	Relation	A reference to a related resource
14	Coverage	The extent or scope of the content of the resource
15	Rights	Information about rights held in and over the resource

ircam Centre Pompidou

G. Peeters 44

VI. Representations of sound's description MPEG-7

- MPEG-7 Multimedia Content Description Interface
  - new ISO standard ISO/IEC 15938
  - a meta-data standard, not a compression standard)
    - Question : " How to find ? " currently: efficient text-based search, "Audiovisual data should be just as searchable as text"
    - Descriptions: characterize, enable search/filtering, enable navigation
    - Description generated manually or automatically
      MPEG-7 : normalization of meta-data format (bit-stream), not of the extraction methods
    - Idea: description linked to the actual A/V data (not necessarily located on the same place)
    - Standardization: enables interoperability between metadata databases and applications

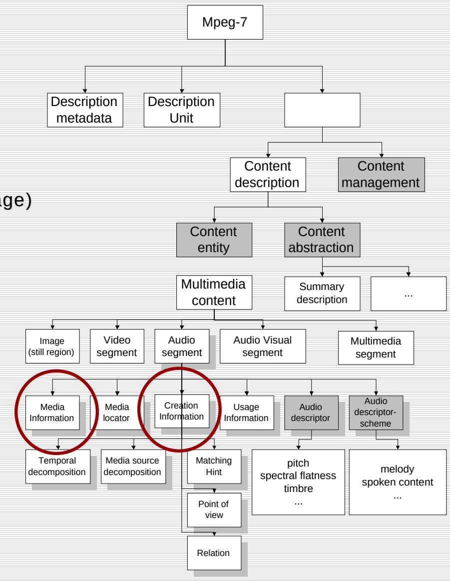


ircam Centre Pompidou

G. Peeters 45

VI. Representations of sound's description MPEG-7

- MPEG-7 Multimedia Content Description Interface
  - Bibliographic-like information
    - media
    - meta (creation / production / usage)

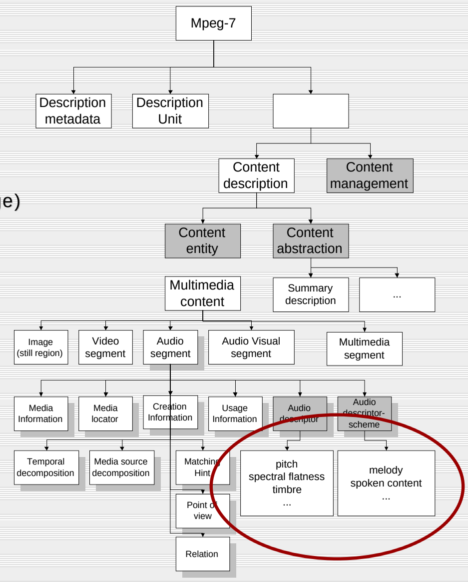


ircam Centre Pompidou

G. Peeters 46

VI. Representations of sound's description
MPEG-7

- MPEG-7 Multimedia Content Description Interface
  - Bibliographic-like information
    - media
    - meta (creation / production / usage)
  - Content description

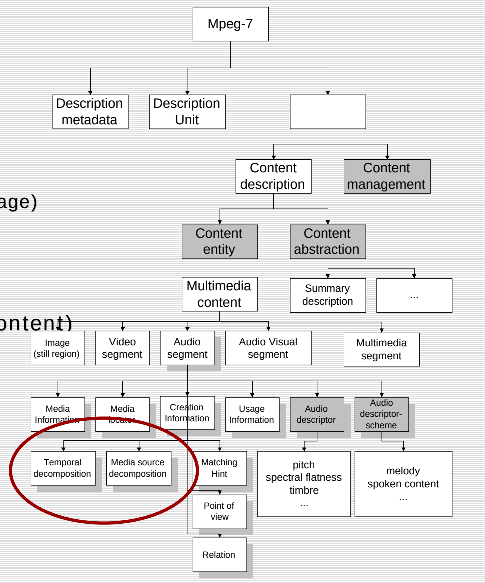


ircam Centre Pompidou

G. Peeters 47

VI. Representations of sound's description
MPEG-7

- MPEG-7 Multimedia Content Description Interface
  - Bibliographic-like information
    - media
    - meta (creation / production / usage)
  - Content description
  - Structure description (table of content)

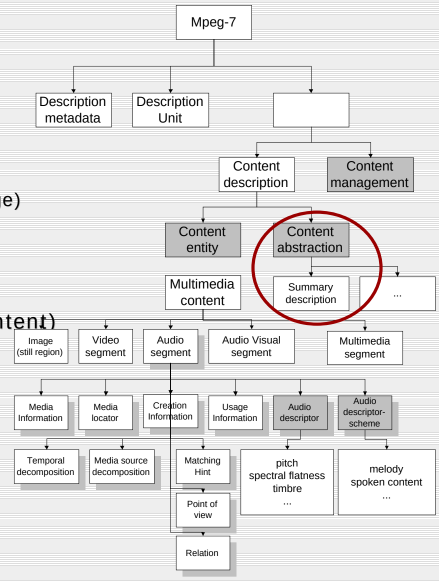


ircam Centre Pompidou

G. Peeters 48

VI. Representations of sound's description MPEG-7

- MPEG-7 Multimedia Content Description Interface
  - Bibliographic-like information
    - media
    - meta (creation / production / usage)
  - Content description
  - Structure description (table of content)
  - Semantic description (index)
  - Navigation and access control
    - summary

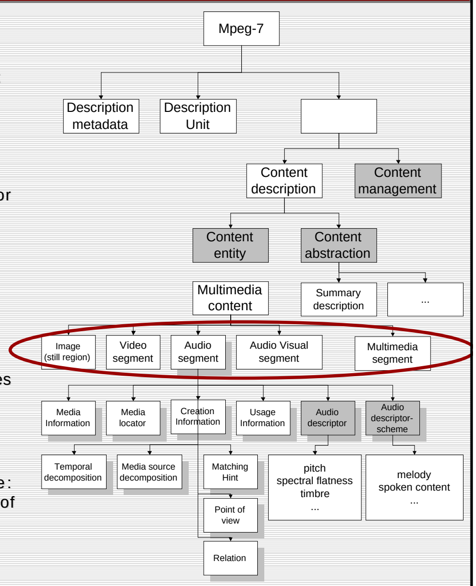


ircam Centre Pompidou

G. Peeters 49

VI. Representations of sound's description MPEG-7

- Part 1. MPEG-7 Systems: The tools that are needed to prepare MPEG-7 Descriptions for efficient transport and storage, and to allow synchronization between content en descriptions. Tools related to managing and protecting intellectual property
- Part 2. MPEG-7 Description Definition Language: The language for defining new Description Schemes and perhaps eventually also for new Descriptors.
- Part 3. MPEG-7 Visual: The Descriptors and Description Schemes dealing with (only) Visual descriptions
- Part 4. MPEG-7 Audio: The Descriptors and Description Schemes dealing with (only) Audio descriptions
- Part 5. MPEG-7 Multimedia Description Schemes: The Descriptors and Description Schemes dealing with generic features and multimedia descriptions
- Part 6. MPEG-7 Reference Software: A software implementation of relevant parts of the
- Part 7. MPEG-7 Standard MPEG-7 Conformance: Guidelines and procedures for testing conformance of MPEG-7 implementations.

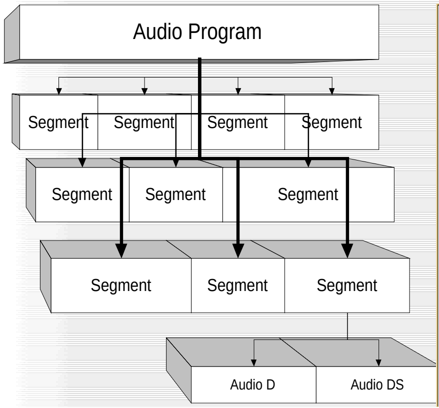


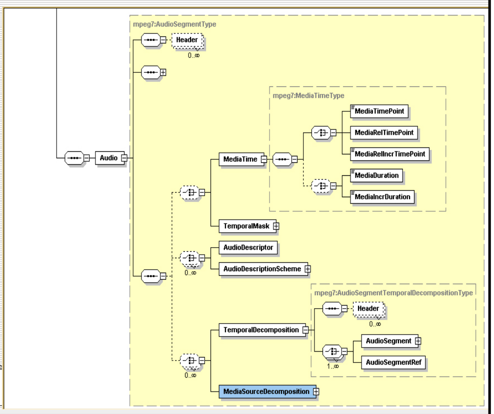
ircam Centre Pompidou

G. Peeters 50

VI. Representations of sound's description MPEG-7

è MPEG-7 Structure Description





ircam Centre Pompidou

G. Peeters 51

VI. Representations of sound's description MPEG-7

è MPEG-7 Audio Low-level audio descriptors

è Basic:
Instantaneous waveform and power values

è Basic Spectral:
A log-frequency power spectrum, and spectral features including spectral centroid, spectral spread, and spectral flatness

è Signal parameters:
Fundamental frequency of quasi-periodic signals, and harmonicity of signals

è Timbral Temporal:
Log attack time and temporal centroid

è Timbral Spectral:
Specialized spectral features in a linear-frequency space, including a spectral centroid, and spectral features specific to the harmonic portions of signals, including harmonic spectral centroid, spectral deviation, spectral spread, and spectral variation.

è Spectral Basis representations:
Features used primarily for sound recognition, but generally useful as projections into a low-dimensional space to aid compactness and recognition.

è Silence Descriptors .

ircam Centre Pompidou

G. Peeters 52

VI. Representations of sound's description
MPEG-7

è MPEG-7 Audio High-level audio description tools (Ds and DSs)

è **Musical Instrument Timbre description tools**
Timbre descriptors aim at describing perceptual features of instrument sounds. Timbre is currently defined in the literature as the perceptual features that make two sounds having the same pitch and loudness sound different. The aim of the Timbre description tools is to describe these perceptual features with a reduced set of descriptors. The descriptors relate to notions such as "attack", "brightness" or "richness" of a sound.

è **Sound recognition tools**
The sound recognition descriptors and description schemes are a collection of tools for indexing and categorization of general sounds, with immediate application to sound effects. Support for automatic sound identification and indexing is included as well as tools for specifying a taxonomy of sound classes and tools for specifying an ontology of sound recognizers. Such recognizers may be used to automatically index and segment sound tracks.

è **Spoken Content description tools**
The Spoken Content description tools allow detailed description of words spoken within an audio stream. In recognition of the fact that current Automatic Speech Recognition (ASR) technologies have their limits, and that one will always encounter out-of-vocabulary utterances, the Spoken Content description tools sacrifice some compactness for robustness of search. To accomplish this, the tools represent the output and what might normally be seen as intermediate results of Automatic Speech Recognition (ASR). The tools can be used for two broad classes of retrieval scenario: indexing into and retrieval of an audio stream, and indexing of multimedia objects annotated with speech.

è **Melody description tools (Melody, Melody contour)**
The Melody Contour DS is a compact representation for melodic information, which allows for efficient and robust melodic similarity matching, for example, in query-by-humming. The Melody Contour DS uses a 5-step contour (representing the interval difference between adjacent notes), in which intervals are quantized. The Melody Contour DS also represents basic rhythmic information by storing the number of the nearest whole beat of each note, which can dramatically increase the accuracy of matches to a query. For applications requiring greater descriptive precision or reconstruction of a given melody, the Melody DS supports an expanded descriptor set and high precision of interval encoding.

ircam Centre Pompidou

G. Peeters 53

VI. Representations of sound's description
MPEG-7

è MPEG-7 Audio Examples:
- è summary
- è melody
- è spectrum flatness

ircam Centre Pompidou

G. Peeters 54