UNIVERSITÀ DEGLI STUDI DI FIRENZE

DISIT - Distributed Systems and Internet Technology Lab

http://www.disit.dinfo.unifi.it/

Relazione Annuale 2011

Presentazione della Tesi di Dottorato:

“Signal Processing Techniques Applied to Automatic Music Transcription”

Dottorato in Telematica e Società dell’Informazione

XXIV ciclo - III anno

16 Dicembre 2011

Coordinatore:	Prof. Dino Giuli
Tutor:	Prof. Paolo Nesi
Co-Tutor:	Prof. Alessandro Fantechi
Candidato:	Dr. Gianni Pantaleo

http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

Indice

1. Introduzione

http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

- **Trascrizione Automatica di brani musicali**:

Processo di conversione di un brano audio (file PCM Wave) in rappresentazione notazionale (MIDI file, spartito o formati equivalenti)

http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

AUTOMATIC MUSIC TRANSCRIPTION – 1. *Introduzione*

- **Trascrizione Automatica di brani musicali**:

  Processo di conversione di un brano audio (file PCM Wave) in rappresentazione notazionale (MIDI file, spartito o formati equivalenti)

- **Ricavare Informazioni** sulle note suonate in un brano audio:
  - Altezza (*pitch*)
  - Istante di attacco (*onset*)
  - Durata (*Duration*)
  - Intensità (*Loudness*)

http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

AUTOMATIC MUSIC TRANSCRIPTION – 1. *Introduzione*

- **Trascrizione Automatica di brani musicali**:

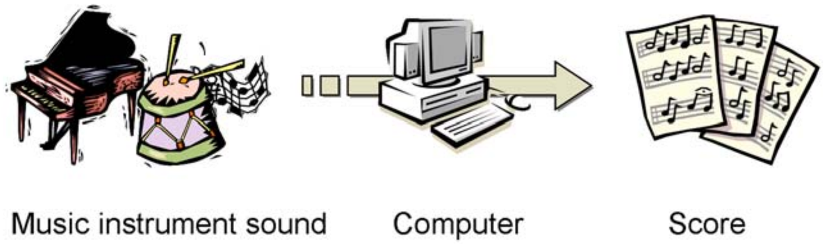
Processo di conversione di un brano audio (file PCM Wave) in rappresentazione notazionale (MIDI file, spartito o formati equivalenti)

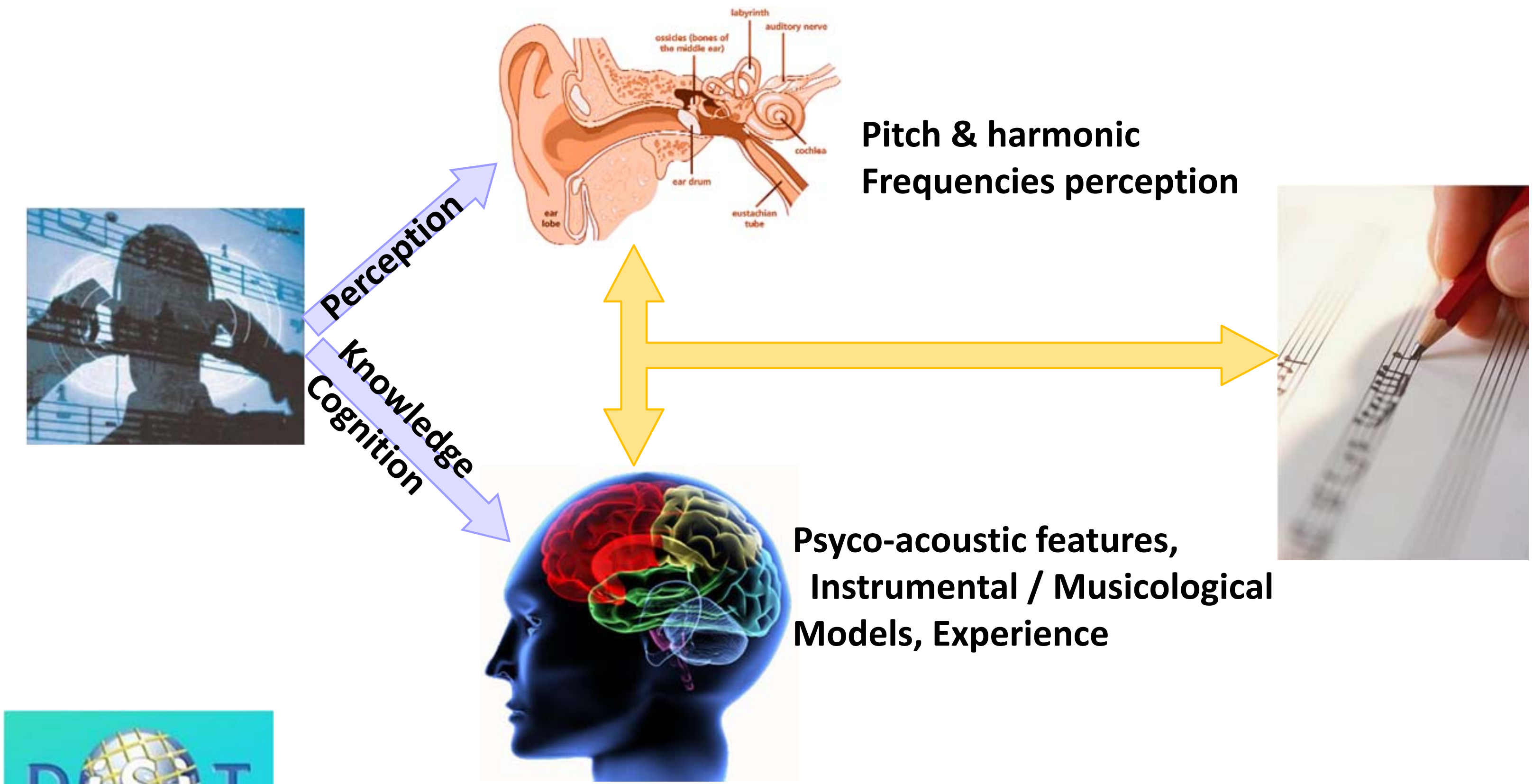
- **Ricavare Informazioni** sulle note suonate in un brano audio:
  - Altezza (*pitch*)
  - Istante di attacco (*onset*)
  - Durata (*Duration*)
  - Intensità (*Loudness*)

- **Applicazioni**:
  - Music track / performance recognition;
  - Music education / collaborative tools / e-learning;
  - Interactive performance / audio resynthesis;

Automatic Transcription VS Human Ear-Trained Transcription

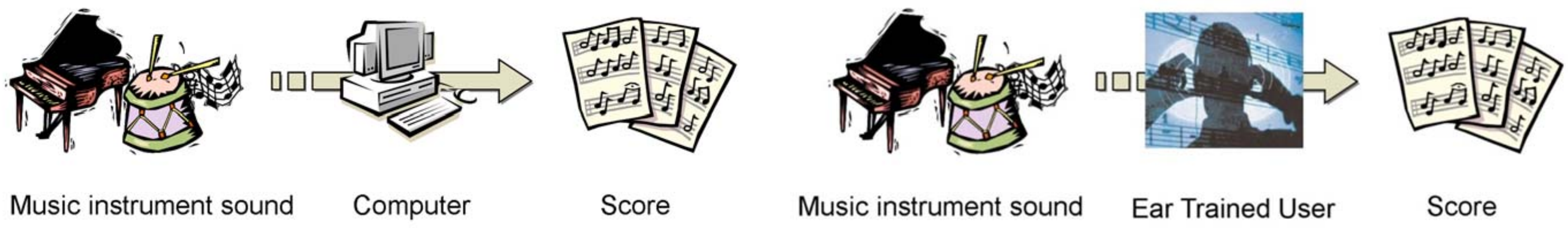
Automatic Transcription VS Human Ear-Trained Transcription

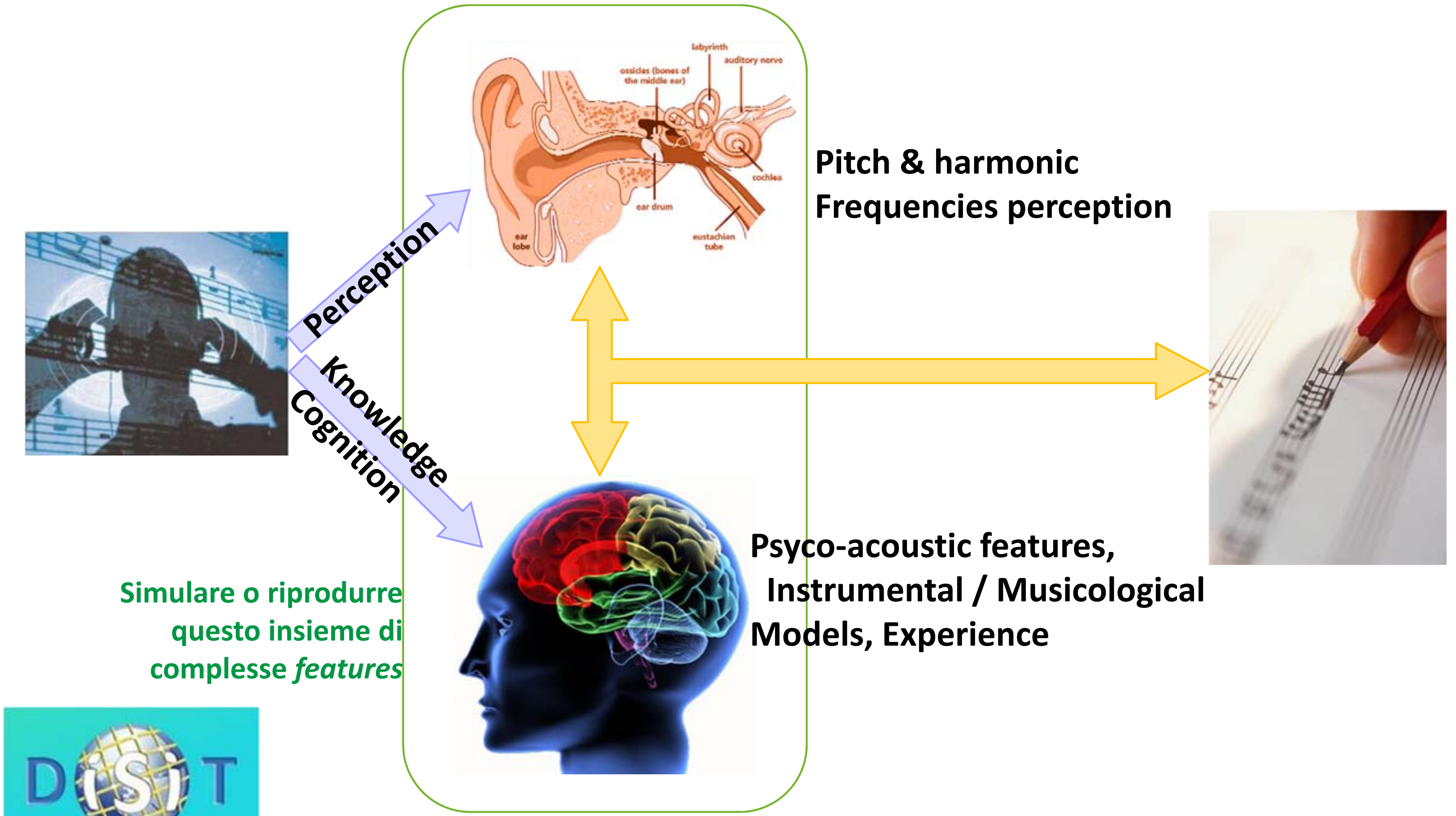




AUTOMATIC MUSIC TRANSCRIPTION – 1. *Introduzione*

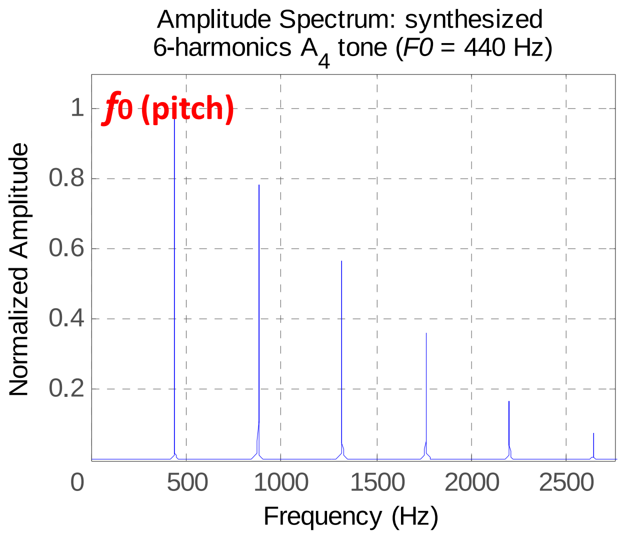
Automatic Transcription VS Human Ear-Trained Transcription

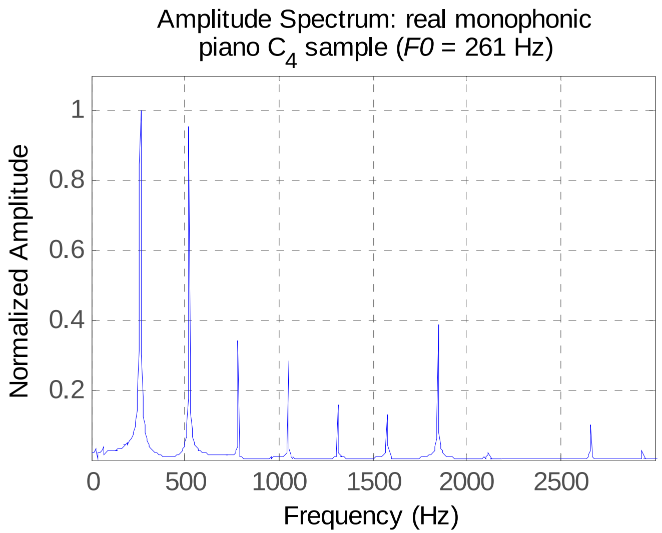


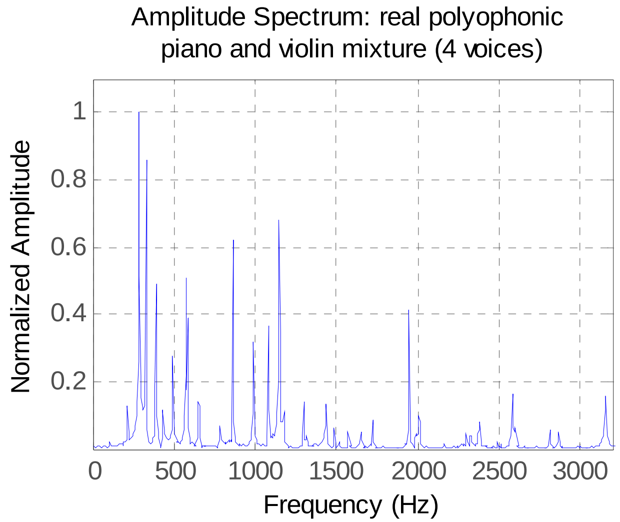


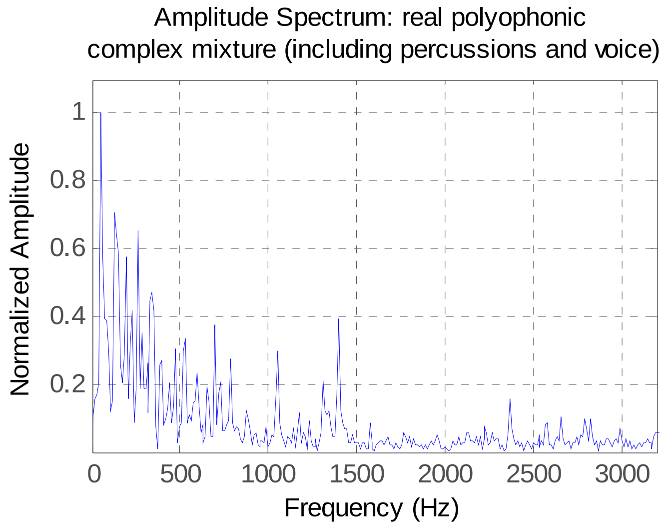
Rappresentazione dei segnali musicali

- Serie di Fourier.
- Spettro in frequenza.
- La frequenza fondamentale f_0 è il *pitch* percepito del suono.
- Il numero di armoniche e la loro ampiezza influisce sul timbro del suono.

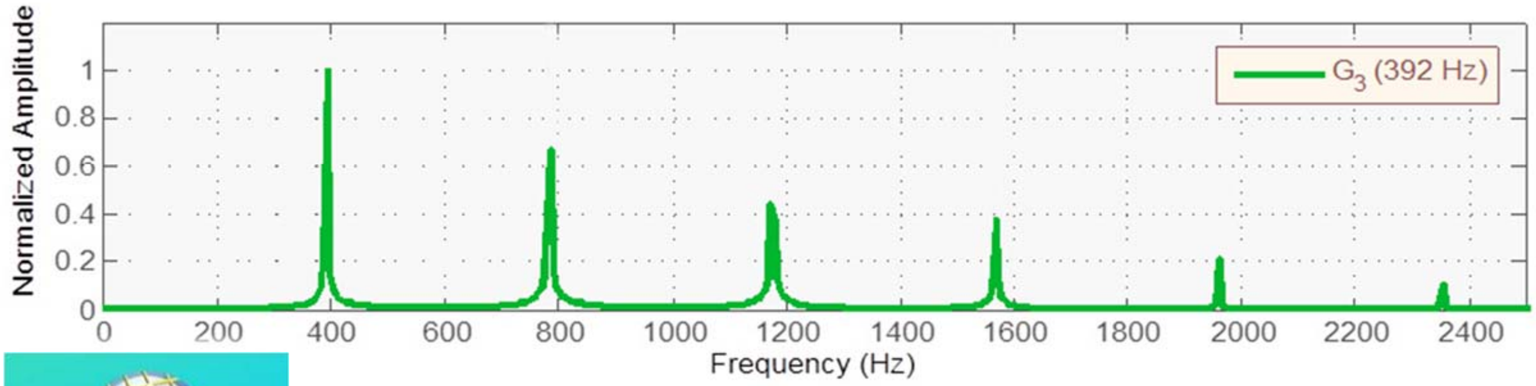




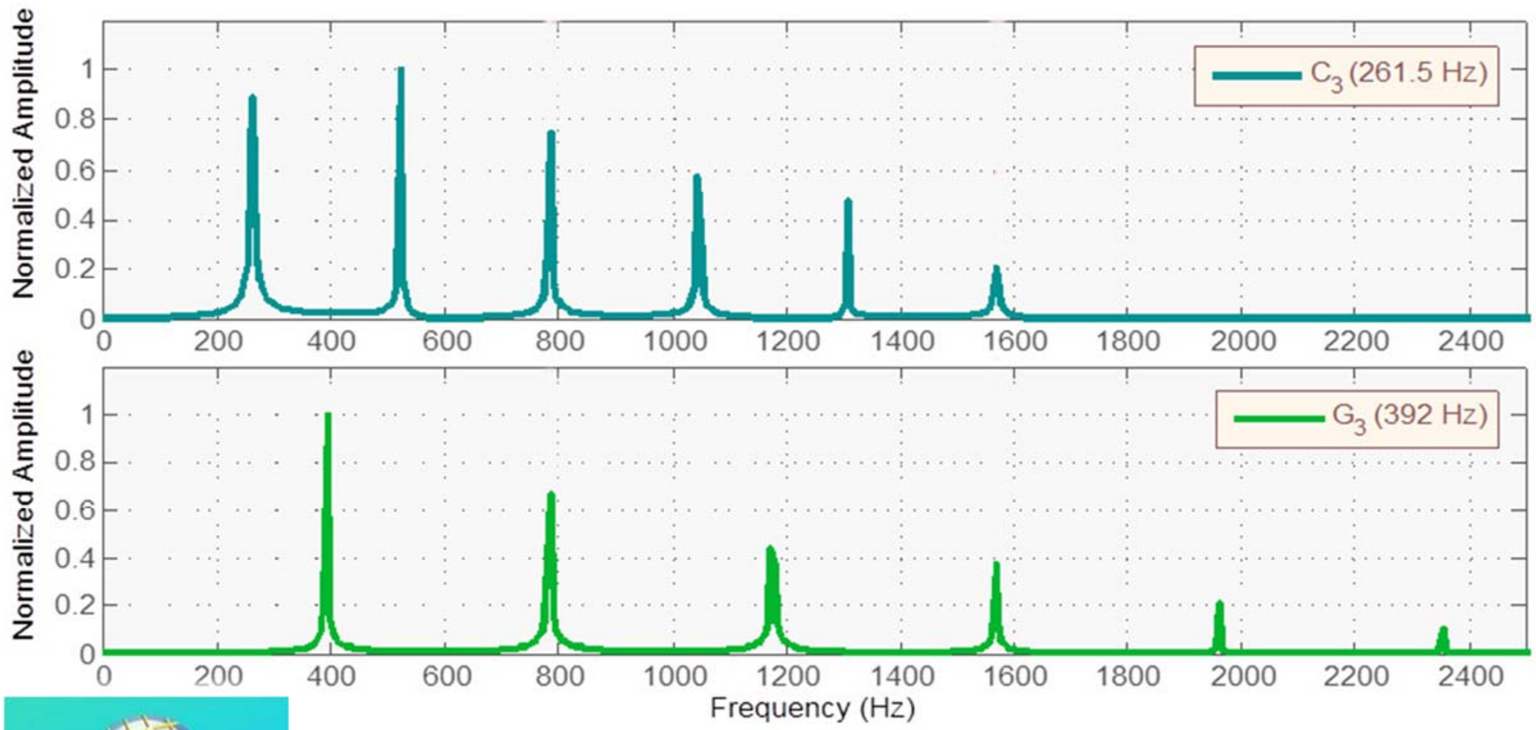




Il Problema della Polifonia

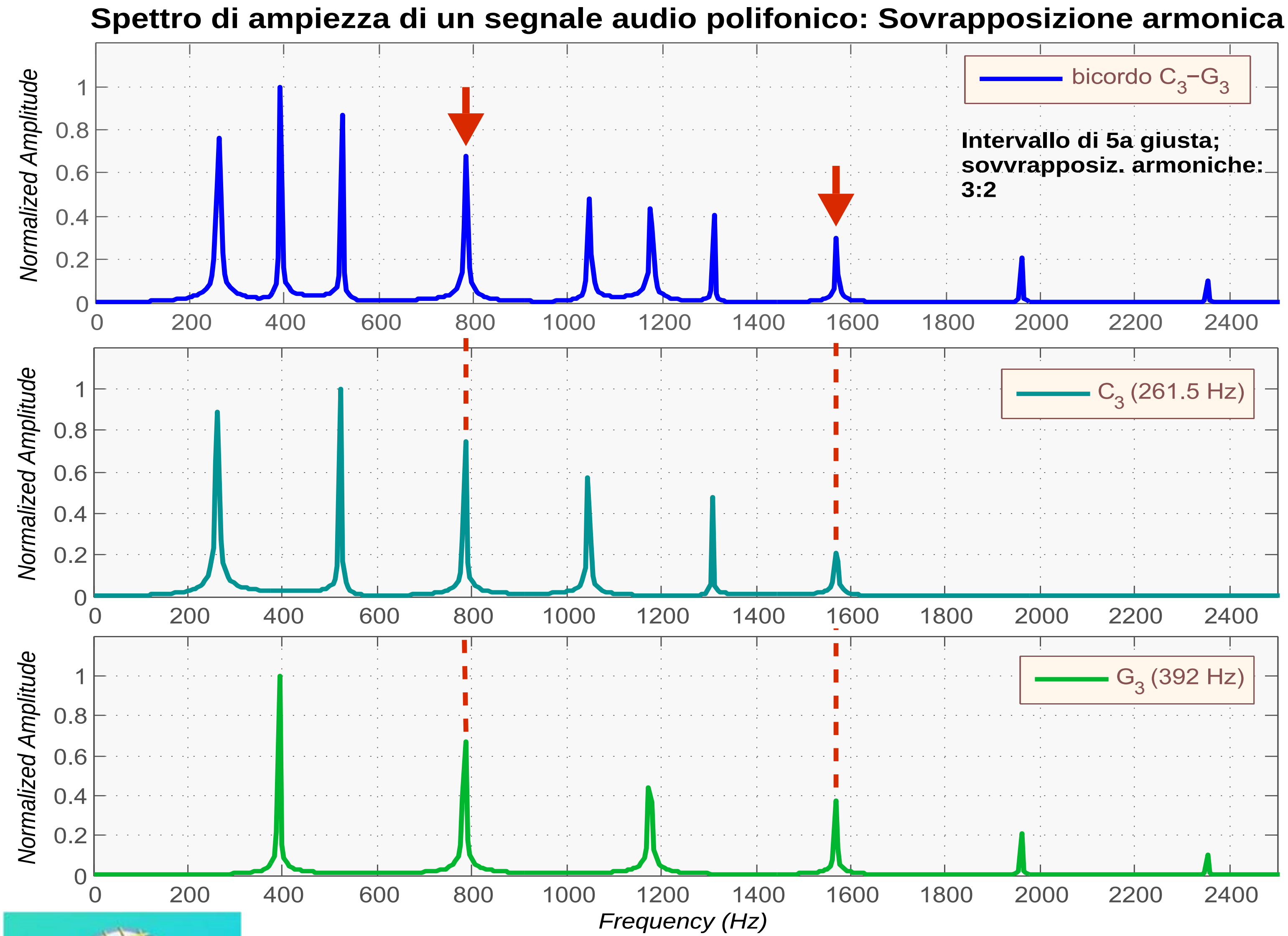


Il Problema della Polifonia



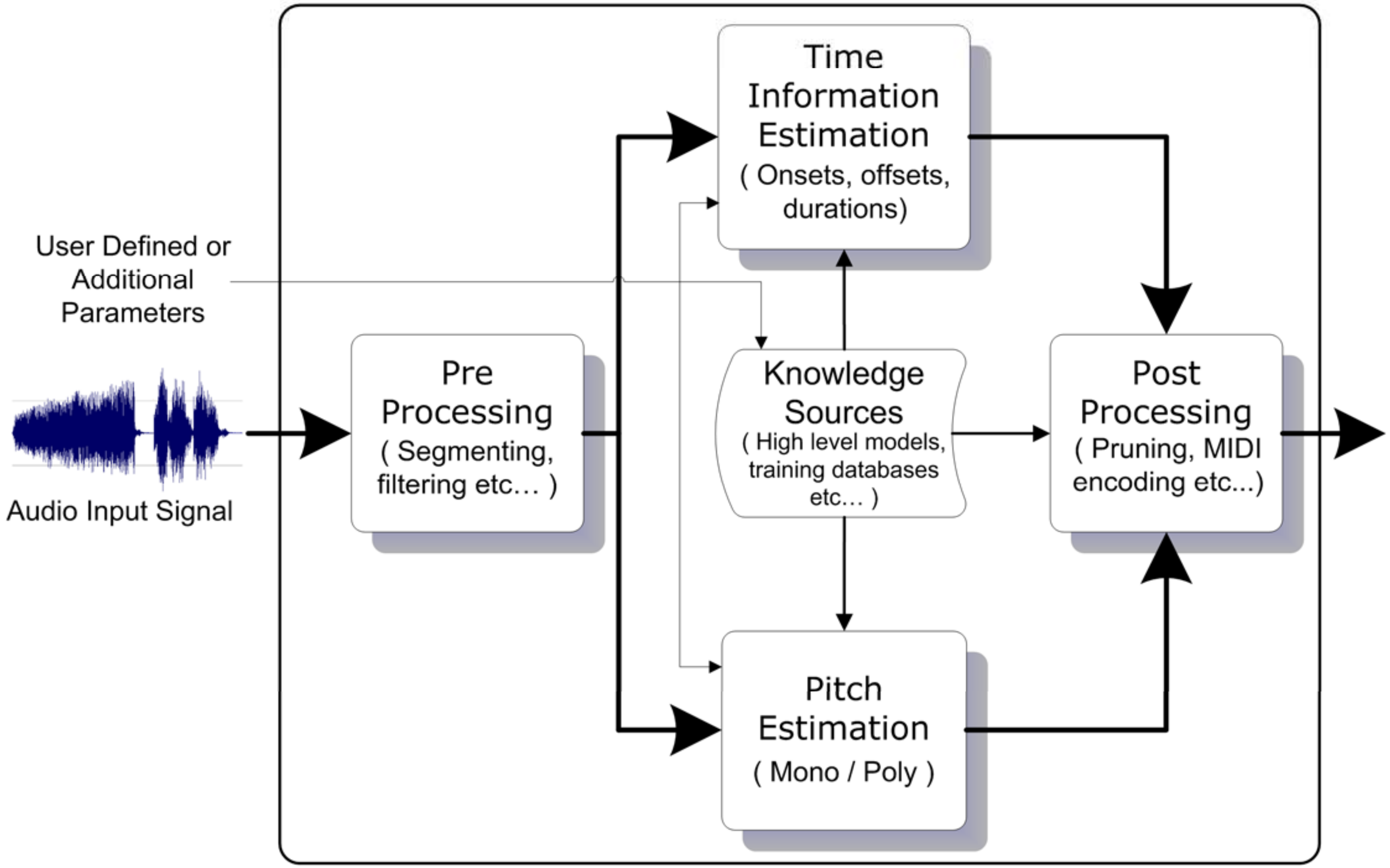
http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

Il Problema della Polifonia

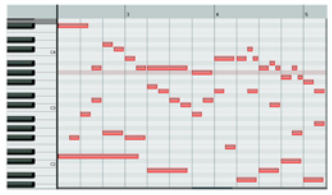


AUTOMATIC MUSIC TRANSCRIPTION – 1. *Introduzione*

Automatic Music Transcription System Architecture



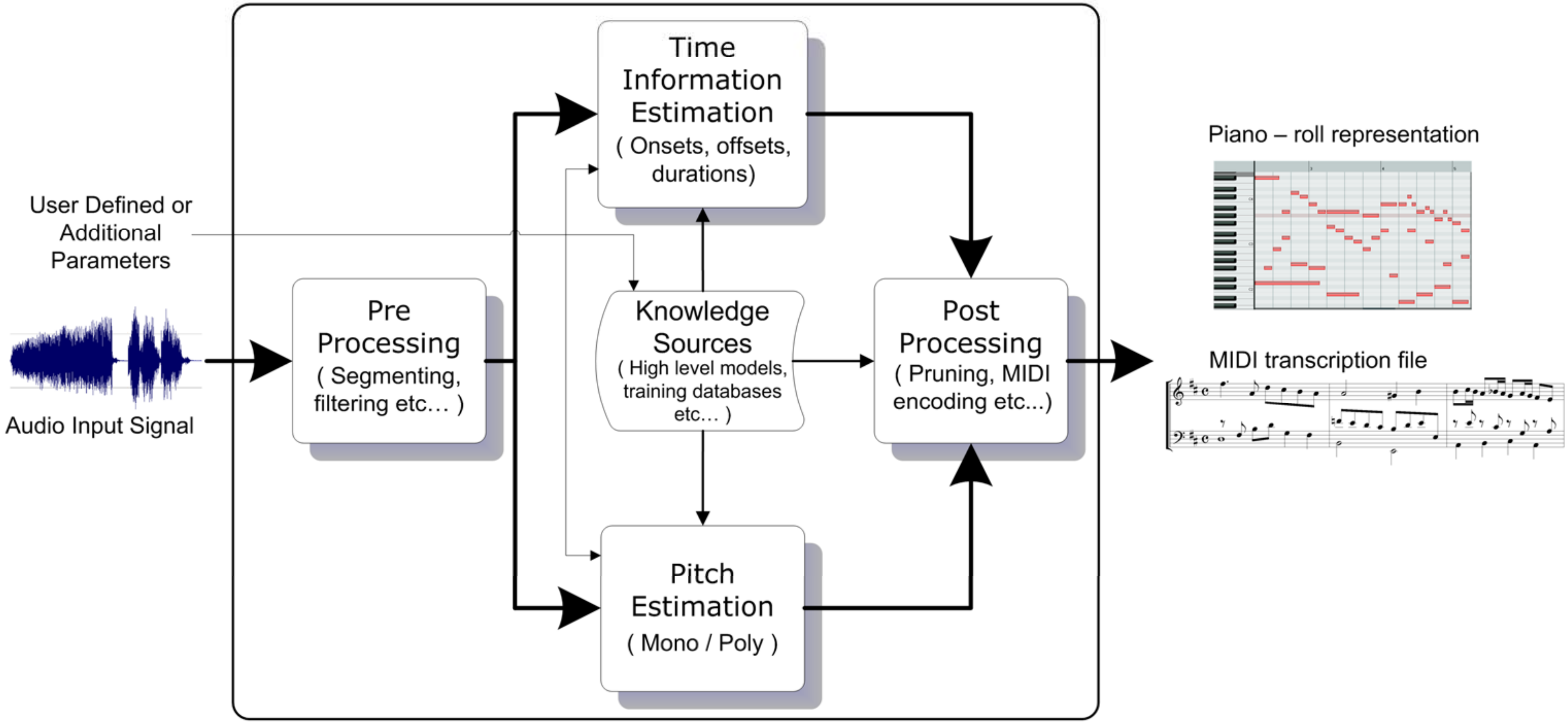
Piano – roll representation



http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

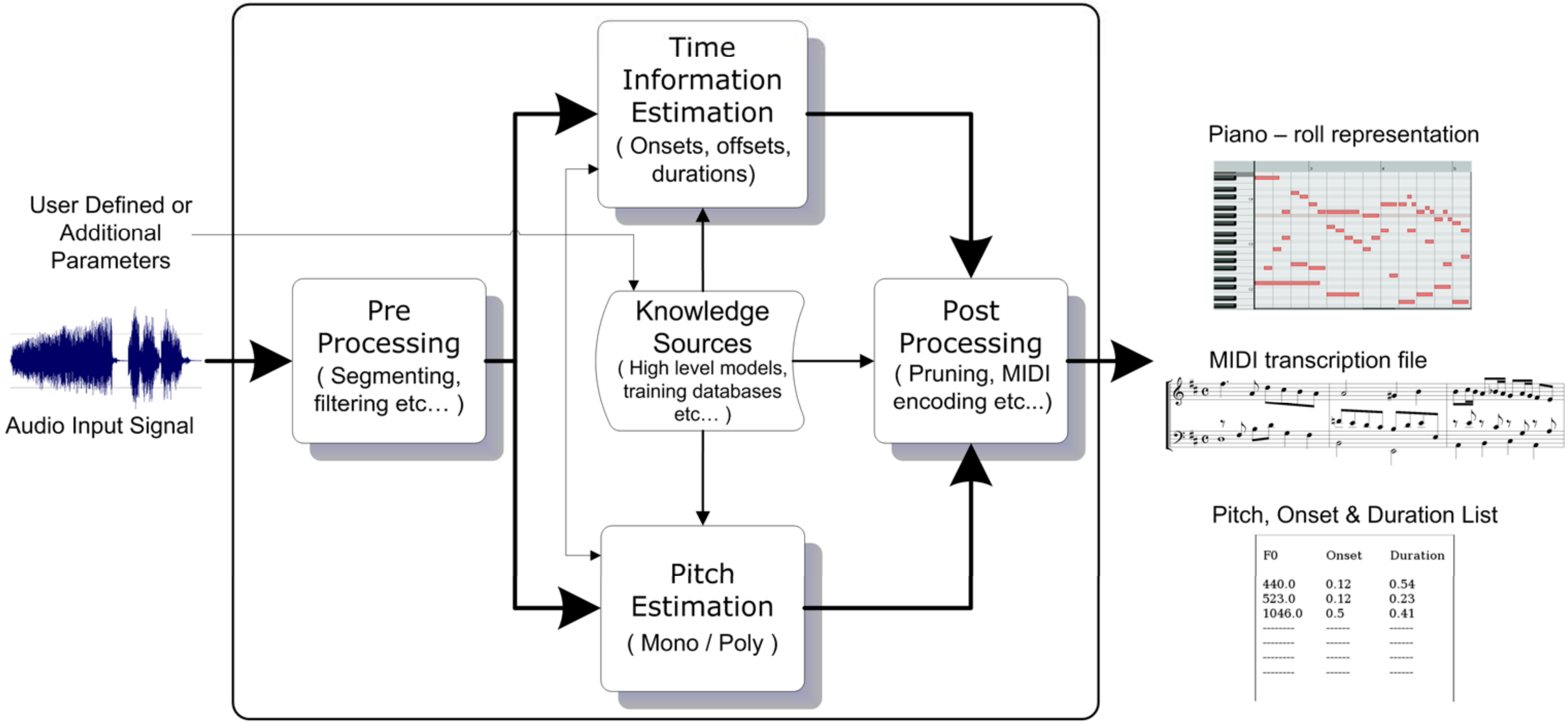
AUTOMATIC MUSIC TRANSCRIPTION – 1. *Introduzione*

Automatic Music Transcription System Architecture



AUTOMATIC MUSIC TRANSCRIPTION – 1. *Introduzione*

Automatic Music Transcription System Architecture

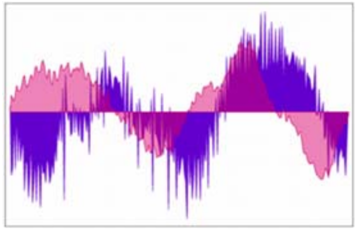


2. *Stato dell'Arte*

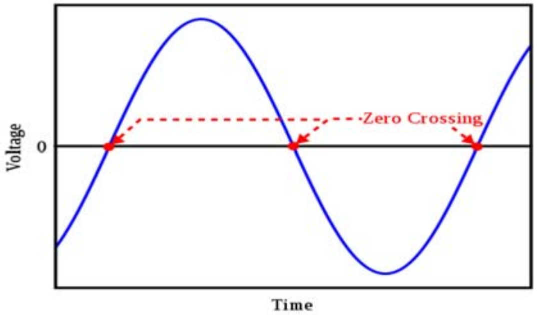
AUTOMATIC MUSIC TRANSCRIPTION – 2. *Stato dell'Arte*

PCM Audio File

(44.1 KHz, 16 bits, mono/stereo)



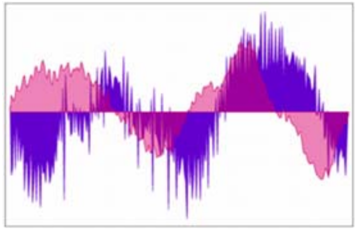
Tecniche di Ptich Estimation nel dominio del ***tempo*** (zero-crossing rate detection, autocorrelazione temporale ecc…)



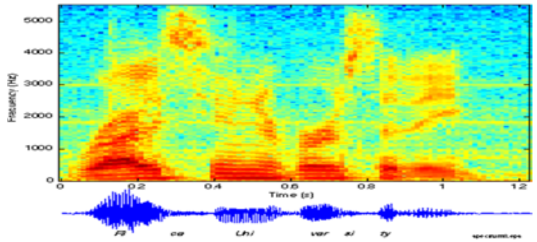
AUTOMATIC MUSIC TRANSCRIPTION – 2. *Stato dell'Arte*

PCM Audio File

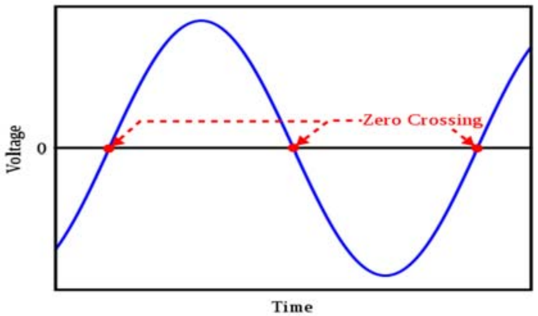
(44.1 KHz, 16 bits, mono/stereo)



Tecniche di Ptich Estimation nel dominio della ***frequenza*** (constant Q, Spettrogramma ecc…)



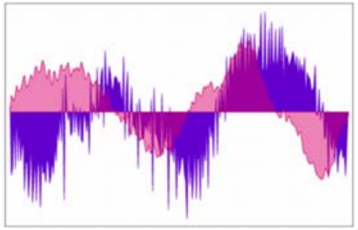
Tecniche di Ptich Estimation nel dominio del ***tempo*** (zero-crossing rate detection, autocorrelazione temporale ecc…)



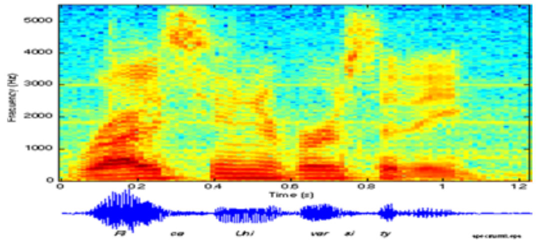
AUTOMATIC MUSIC TRANSCRIPTION – 2. *Stato dell'Arte*

PCM Audio File

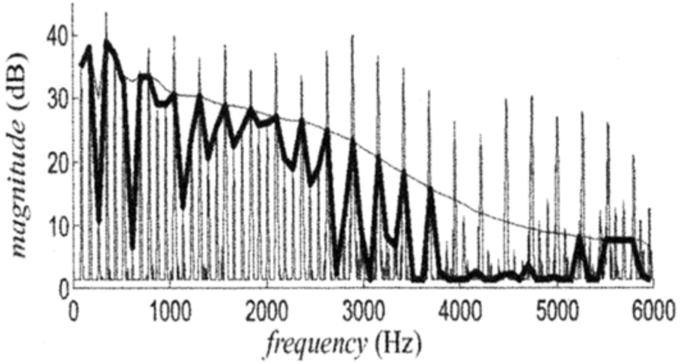
(44.1 KHz, 16 bits, mono/stereo)



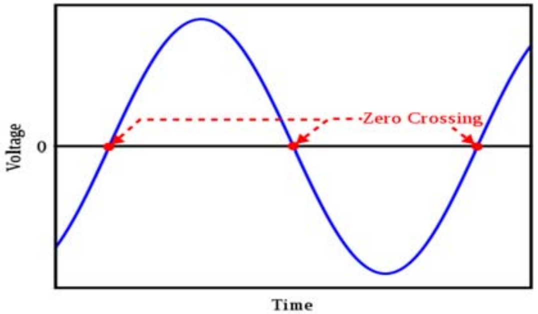
Tecniche di Ptich Estimation nel dominio della ***frequenza*** (constant Q, Spettrogramma ecc…)



Pattern Matching Armonico



Tecniche di Ptich Estimation nel dominio del ***tempo*** (zero-crossing rate detection, autocorrelazione temporale ecc…)

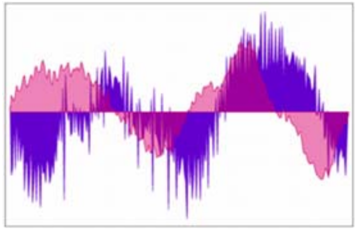


http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

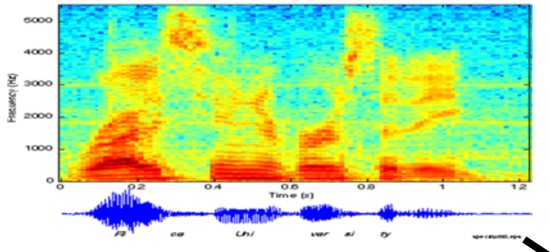
AUTOMATIC MUSIC TRANSCRIPTION – 2. *Stato dell'Arte*

PCM Audio File

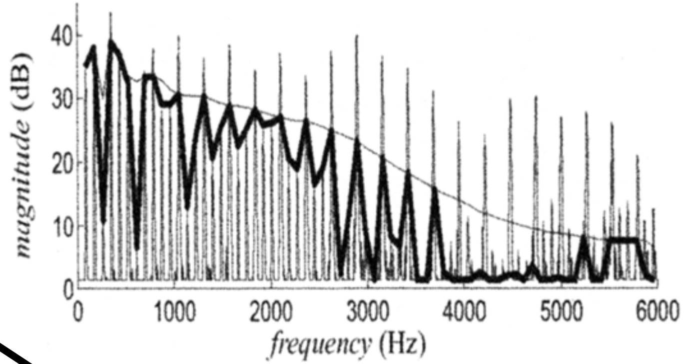
(44.1 KHz, 16 bits, mono/stereo)



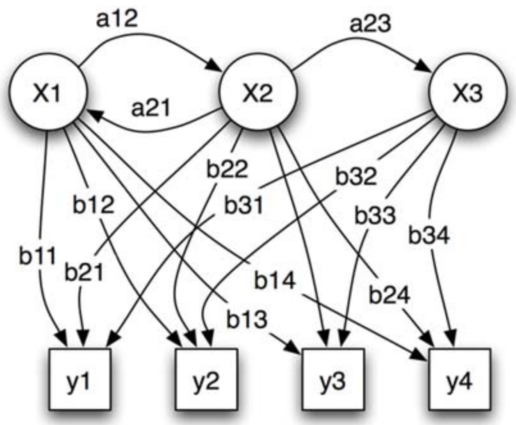
Tecniche di Ptich Estimation nel dominio della ***frequenza*** (constant Q, Spettrogramma ecc…)



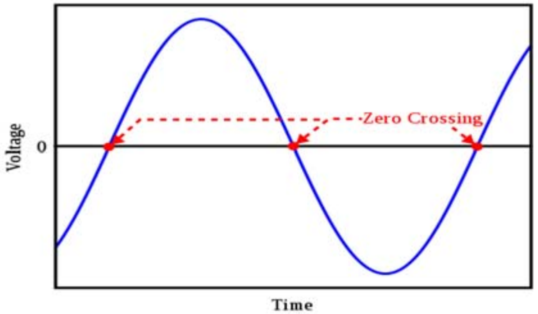
Pattern Matching Armonico



Modelli Statistici (HMM, Bayesian Networks)

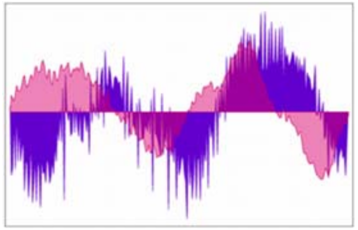


Tecniche di Ptich Estimation nel dominio del ***tempo*** (zero-crossing rate detection, autocorrelazione temporale ecc…)

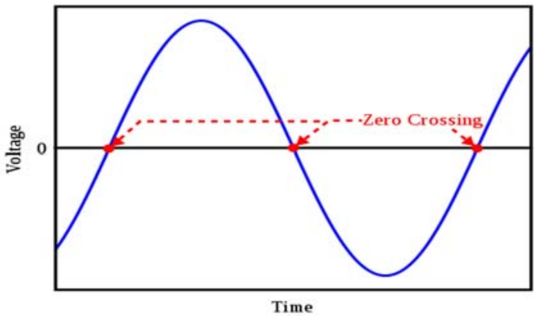


AUTOMATIC MUSIC TRANSCRIPTION – 2. *Stato dell'Arte*

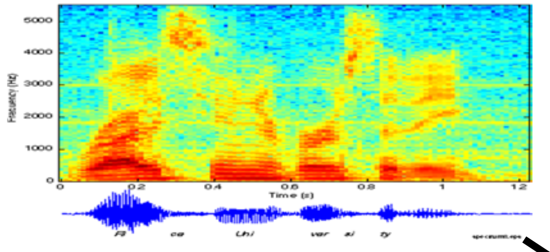
PCM Audio File
(44.1 KHz, 16 bits, mono/stereo)



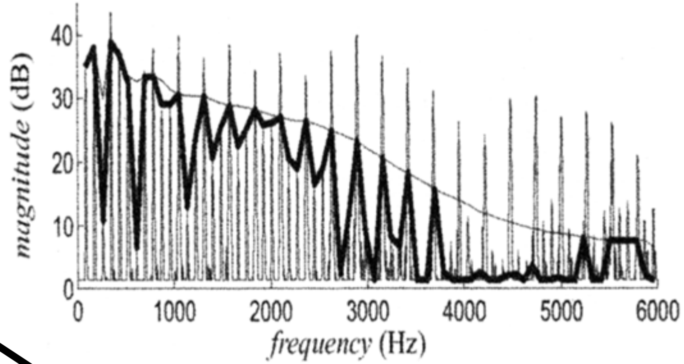
Tecniche di Ptich Estimation nel dominio del ***tempo*** (zero-crossing rate detection, autocorrelazione temporale ecc…)



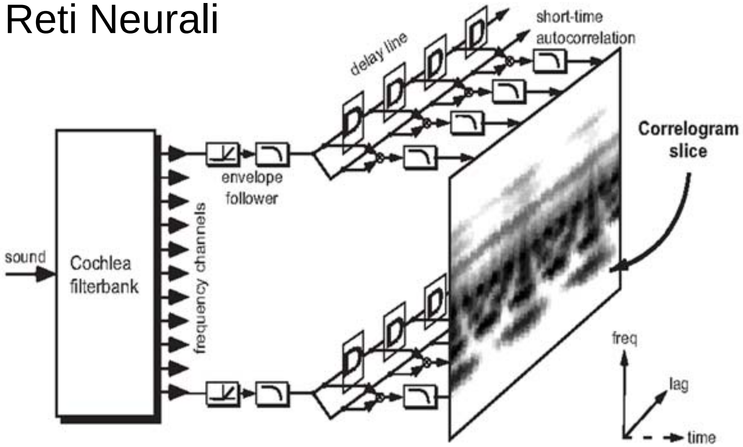
Tecniche di Ptich Estimation nel dominio della ***frequenza*** (constant Q, Spettrogramma ecc…)



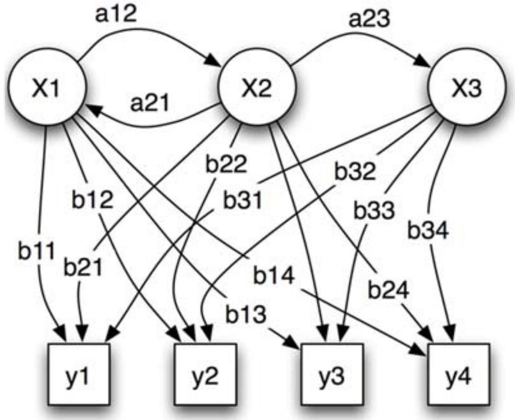
Pattern Matching Armonico



Modelli Uditivi (psico-acustici) + Reti Neurali



Modelli Statistici (HMM, Bayesian Networks)



Review approfondita dello Stato dell'Arte (1 di 5)

Reference (Group)	Year	System Input / Output	Pre-Processing & Mid-Level	Real time / Offline	Source Avail.	Mono / Poly	Time / Freq.	Pitch Estimation Knowledge	Rhythm Info Extraction: Onsets	Rhythm Info Extraction: Durations	Addiotional Notes	Evaluation Material
[Moorer77] (CCRMA)	1975 1977	I: N.A. O: N.A.	Band-pass filter bank (optimum comb filter) Short - time FFT	N.A.	No	Poly	F	Time periodicity research by detecting sinusoidal components	No	No	Max # voices: 2 Limited freq. range F0s ratio can't have an integer relationship	Synthesized violin and real guitar duets
[PisGal79] (U-M)	1979	I: N.A. O: F0s & amplitudes list	Spectral equalization to enhance partials	Offline	No	Poly	F	Evaluation of harmonic relations among spectral peaks	No	No	Robust with missing F0 & inharmonic partials	Synth. and real signals (carillon bells)
[Friedman79]	1979	I: Speech signal, fs=10 kHz O: 24 ms pitch frames	Band-Pass FIR Lo-Freq emphasis	Suitable for Real time		Poly	T	Zero-crossing rate detection upon processed waveform	No	No	Pitch estimation for speech	3-second speech sample
[ChaJaf86] (CCRMA)	1986	I: Digital audio recording O: Hihg-level MIDI score	Bounded Q Frequency Transform	N.A.	No	Poly	F	Grouping partials in sinusoidal analysis	Detect changes in spectral energy over time	No	Knowledge of source acoustic applied	N.A.
[KatIno89] (KANSEI)	1989	I: N.A. O: N.A.	Time - frequency map obtained with the interpolation method using complex spectra.	N.A.	No	Poly	F	Peaks extraction in the frequency domain and matching procedure	No	No	Heuristic rules implemented to group detected frequency peaks into notes.	System developed for piano, guitar and shamisen. Results are not reported.
[SlaLyo90]	1990	I: N.A. O: Time-frequency pitch representation	Correlogram: cochlear model, 2nd order filter bank, HWR and Automatic Gain Control	N.A.	Partially	Poly	F	Periodicities research in the correlogram by use of the autocorrelation function	No	No	—	Acoustical Society of America Database Qulitative, not clear results
[Maher90]	1990	I: Digital audio signals < 20 s O: Chains of peaks for partials tracking	Short-time FFT (512 and 1024 sample window) Hi-freq. pre-emphasys	Offline	No	Poly	F	McAulay-Quatieri sinusoidal & two-way mismatch analysis Several strategies to resolve colliding sinusoidal partials	No	No	Limited to duets, non ovrlapping frequency ranges; nearly harmonic sounds	Synth samples and real signals (basson/ clari-net and trumpet/tuba) Qualitative results

AUTOMATIC MUSIC TRANSCRIPTION – 2. *Stato dell'Arte*

Review approfondita dello Stato dell'Arte (2 di 5)

Reference (Group)	Year	System Input / Output	Pre-Processing & Mid-Level	Real time / Offline	Source Avail.	Mono / Poly	Time / Freq.	Pitch Estimation Knowledge	Rhythm Info Extraction: Onsets	Rhythm Info Extraction: Durations	Addiotional Notes	Evaluation Material
[KasTan92] (UT)	1992	I: Monaural signals O: multi-channel MIDI	A/D conversion and spectrogram representation	N.A.	No	Poly	F	Peaks - peaking in STFT (segre-gation); statistic rules for partials grouping (integration)	No	No	Timbre models to detect different instruments	Synthesized Vivaldi Concerto (op. 3, no. 6)
[KasTan93] (UT)	1993	I: Monaural signals (48 kHz / 16 bits) O: multi-channel MIDI	Frequency Analysis: band-pass 2-order IIR filters	N.A.	No	Poly	F	Frequency content extraction by pinching planes thresholds and bottom-up clustering	No	No	Automatic timbre modelling based on perceptual rules	Synth. random chords Good recognition up to 3 voices
[Hawley93] (MIT)	1993	I: N.A. O: N.A.	Short-time spectral analysis	N.A.	No	Poly	F	Spectral comb filtering for note identification	Hi-freq. energy con-tent; bilinear time-domain filtering	No	—	Bach piano excerpts Non extensive tests
[Kas_et_al.95] (UT)	1995	I: Monaural signals O: multi-channel MIDI	STFT	N.A.	No	Poly	F	Frequency content extraction by pinching planes and Bayesian network integration	No	No	Many knowledge sources applied (timbre, chord type...)	2-3 voices synthesized MIDI chords with real instruments samples
[Martin96a] (MIT)	1996	I: CD quality audio input O: Transcription in counterpoint style	STFT Blackboard architecture front-end	Offline	No	Poly	F	Knowledge-based source (KS) applied to sinusoidal track extraction	Peaks picking on squared and low-pass filtered signal energy	No	—	4-voices Bach Corales Bad in octave detection Good recognition in B2 - A4 notes interval
[Martin96b] (MIT)	1996	I: N.A. O: MIDI file, symbloic score or piano-roll	Log-lag correlogram (Auditory model of pitch perception)	N.A.	No	Poly	F	Periodicities research in the correlogram (autocorrelation) Knowledge sources applied (Blackboard framework)	Energy maxima of signal envelope	No	Advantages of auditory models in detecting octave intervals	Mono & Poly test on Bach piano pieces Qualitative results
[FerCas98]	1998	I: N.A. O: MIDI file	Multi scale sinusoidal model (constant-Q) filter bank	Not true real time	No	Poly	F	Prominent harmonic pattern search in synth. spectrum (ampltudes of peaks are set after a *quality-of-fit* measure).	No	No	General source models Masking effect test Post processing to kill too short notes.	Not specified dataset High error rate for typical musical signals are revealed

Review approfondita dello Stato dell'Arte (3 di 5)

Reference (Group)	Year	System Input / Output	Pre-Processing & Mid-Level	Real time / Offline	Source Avail.	Mono / Poly	Time / Freq.	Pitch Estimation Knowledge	Rhythm Info Extraction: Onsets	Rhythm Info Extraction: Durations	Addiotional Notes	Evaluation Material
[TolKar2000] (HUT)	2000	I: N.A. O: N.A.	Pre-whitening filtering two channels - filter bank (Hp & LP, crossover @ 1 kHz)	Real time	No	Poly	ACF Transf. domain	Periodicity estimation by the *summary autocorrelation* function on both channels	No	No	F0 estimation examples available on Web	2-4 Clarinet tones mixed to form various chords
[Goto2000] [Goto04] (AIST)	2000 2004	I: 16-bit PCM signal fs = 16 kHz O: N.A.	STFT obtained with a multi-rate filter bank Band-pass filter to split the spectrum in 2 freq. regions	Real time	No	Poly	F	Frequency-to-instantaneous frequency mapping Maximize probability function for each candidate F0	No	No	Melody and bass line detection from real-world audio signals	10 excerpts from commercial CD recordings.
[Marolt01] (SONIC)	2001	I: PCM signal, fs=44.1 kHz O: MIDI file	Auditory model 200 *gammatone* filters	N.A.	Yes	Poly	F	Network of adaptative oscillators	Multy-layer perceptron NN	Time observation of NN activity	Piano transcription Note range: A1-C8	120 synthesized piano pieces
[CheKaw02] (YIN)	2002	I: N.A. O: N.A.	Autocorrelation (ACF) Method	Suitable for real time	No	Mono	F	Difference function (similar to autocrrelation function) computed via FFT, to find periodicities in signals	No	No	Cumulative mean function and parabolic interpolation to reduce sub-harmonic errors	Speech databases Informal evluation on music
[Raphael02]	2002	I: N.A. O: MIDI file	Fourier analysis of input audio signal	N.A.	No	Poly	F	Generative probabilistic HMM (Baum-Welch training)	Signal "burstiness" measures for attack, steady and silence states		Method for piano music transcription	Excerpts from Mozart's piano sonata 18, K570
[GodDav03] (CAM)	2003	I: PCM, 22.05 kHz O: frame by frame F0 list	sinusoidal model differentiated for mono/poly	N.A.	No	Mono	T	Bayesian Models MCMC harmonic inference	Frame by frame F0 tracking		Audio examples on the Web	Solo flute extract (Debussy's *Syrinx*)
[Klapuri03] (TUT)	2003	I: 16-bit PCM signal fs = 44.1 kHz O: Frame by frame F0 list	DFT on Hamming windowed signal frames Signal plus noise model	N.A.	Algorithm only	Poly	F	Analysis of harmonic relationships between partials on 18 overlapping bands	No	No	Iterative estimation and harmonic pattern cancellation from the spectrum	Mixed samples from McGill, Iowa and IRCAM database
[BruMoNe03]	2003	I: 16 bits PCM, 44.1 kHz (mono or multi-ch.) O: List of note parameters	Patterson-Meddis auditory model	N.A.	No	Mono Poly	F	Neural Network tracking of pitches detected by the onset detection algorithm	Peak-Picking algorithm on signal envelope	Offset detected by Neural Networks	Different instrument models are used Training Mode' available	Piano, guitar and violin samples Bach corale excerpt

Review approfondita dello Stato dell'Arte (4 di 5)

Reference (Group)	Year	System Input / Output	Pre-Processing & Mid-Level	Real time / Offline	Source Available	Mono / Poly	Time / Freq.	Pitch Estimation Knowledge	Rhythm Info Extraction: Onsets	Rhythm Info Extraction: Durations	Addiotional Notes	Evaluation Material
[RyyKla05] (TUT)	2005	I: PCM stereo, 44.1 kHz O: MIDI file	70 Ch. band pass filter HWR, STFT for all bands	N.A.	No	Poly	F	Comb filters bank estimates periodicity in Freq. domain	Positive changes in F0s strength	Note events tracking by HMM	Evaluated in MIREX 2007-08 Accuracy ≈ 61% (MF0E) F-measure ≈ 34% (NT)	Excerpts from WC Database
[BeDaSa06] (QMUL)	2006	I: PCM files fs = 22.05 kHz O:	STFT & spectral smoothing Signal frames modeled as weighted sum of an internal database piano waveforms			Poly	F	Generate F0 hypotheses for revelant amplitude partials Heuristic rules for partials grouping	Temporal parameters estimated integratin frame estimations over time		Method for transcription of recorded piano music. Use of a hybrid method combining time-freq. info	Disklavier played piano MIDI files. Error rate increases for high harmonic reate chords
[CemKap06]	2006	I: N.A. O: Piano-roll of note parameters	Modified sinusoidal model (state space form) to obtain a piano-roll like representation	Real time	No	Poly	T	Use of bayesian networks, switching Kalman filters and a generative model to esti-mate note parameters.	Onsets and offsets are detected by transitions of the states *mute-sound* of the generators		The approach used allows to remove the frame by frame assumption for audio analysis	Own recordings of 2-3 voices chords. Qualitative results, especially offset errors
[PolEll07]	2007	I: 8 kHz sampled audio from MIDI files O: N.A.	STFT	N.A.	No	Poly	F	87 One-versus-all SVM classifier for each piano note trained using *Sequential Minimal Optimization*	Note tracking by two state (on/off) HMM	No	Method for piano music transcription; Evaluation test are available	Synthesized, recorded and Disklavier played MIDI files
[KaNiSa07] (UT)	2007	I: PCM files fs = 44.1 kHz O: F0s, onset & offsets list	Multi-resolution power spectrum obtained via Gabor wavelet transform	Offline	No	Poly	F	Harmonic temporal clustering (HTC) model for source separation	Joint estimation by using HTC model		Evaluated in MIREX 2007, 2008 and 2009 editions: Accuracy ≈ 49% (MF0E) F-measure ≈ 32% (NT)	Excerpts from RWC-Classical and RWC-Jazz databases
[Klapuri08] (TUT)	2008	I: N.A. O: N.A.	Auditory (*gammatone*) filter bank (two types of 2nd order IIR resonators) IHC model, HWR	Real time	No	Poly	F	Periodicities search in the *Summary spectrum*. Detect F0 as peaks of a salience function	Energy peaks detection on signal envelope	No	Iterative estimation and harmonic pattern cancellation from the summary spectrum	Mixed samples from McGill, Iowa and IRCAM database

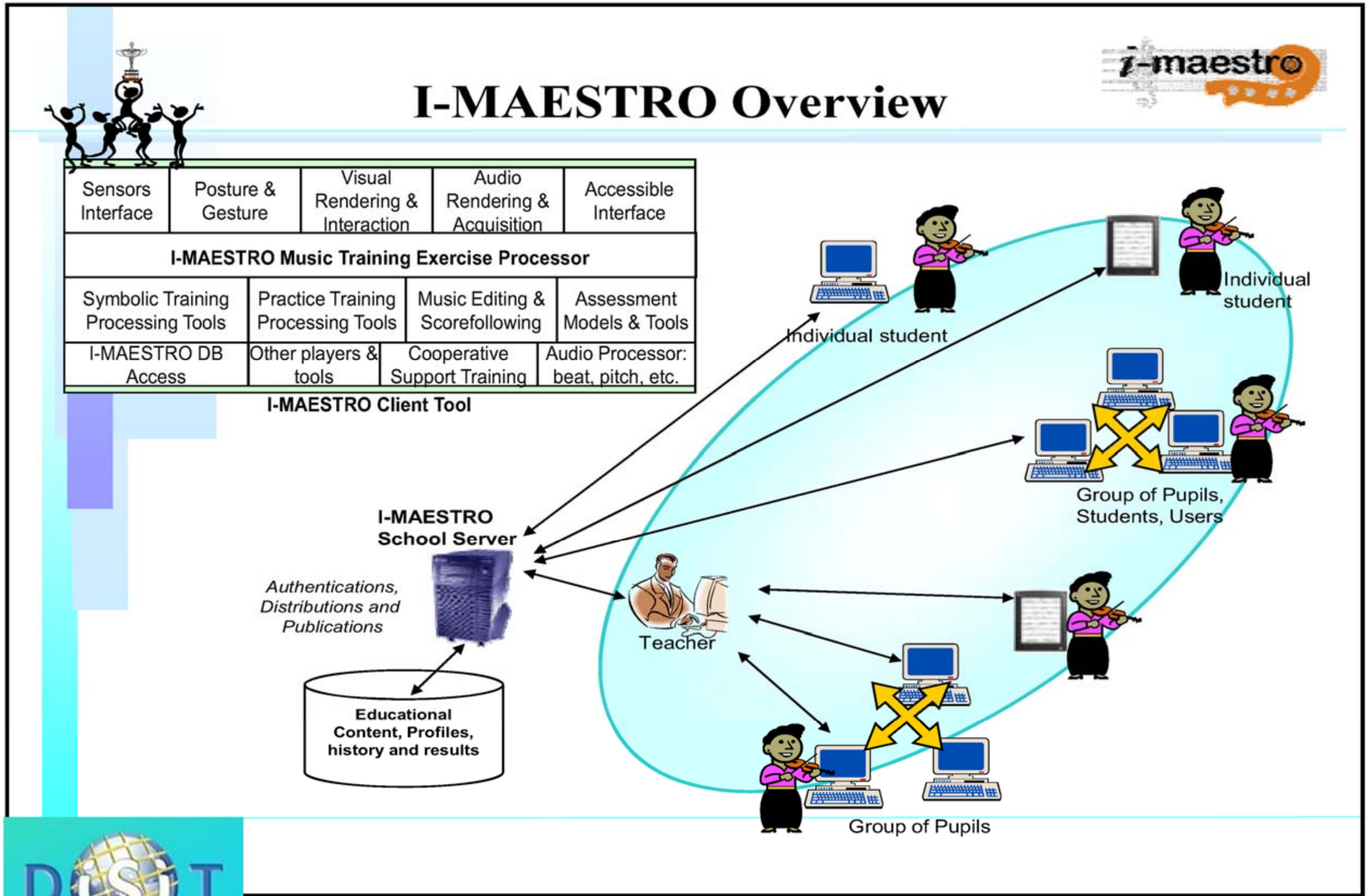
Review approfondita dello Stato dell'Arte (5 di 5)

Reference (Group)	Year		System Input / Output	Pre-Processing & Mid-Level	Real time / Offline	Source Available	Mono / Poly	Time / Freq.	Pitch Estimation Knowledge	Rhythm Info Extraction: Onsets	Rhythm Info Extraction: Durations	Addiotional Notes	Evaluation Material
[DuaZha08]	2008	I	PCM audio signals	STFT	N.A.	No	Poly	F	Maximum Likelihood F0 est. Average Harmonic Strucutre (AHS) extraction for source separation	No	No	The system performs bad F0 recognition for inharmonic sounds	Synthesized, real instru-ments & singing voice. Non standard evaluation metrics used for F0 est.
		O	Piano-roll (F0 tracking)										
[PerIñe08]	2008	I	PCM mono audio signals, fs = 44.1 kHz	STFT. Gaussian smoothing of spectral patterns	N.A.	No	Poly	F	Candidates F0 with best salience function, calculated by considering partial ampli-tude and spectral smoothness	No	No	Evaluated in MIREX 2007 (previous version) & 2008. Accuracy ≈ 62% (2008 MF0E). F-measure ≈ 25% (2008 NT)	4000 chords; random mixtures of various samples (1 to 4 voices polyphony)
		O	Sequence of MIDI notes										
[ViBeBa08]	2008	I	N.A.	ERB - scale time/freq. representation (similar to STFT)	Offline	No	Poly	F	NMF methods using harmonic / inharmonic constraints on the basis spectra with fixed/adapta-tive tuning + spectral smoothness	Thresholding amplitude se-quence of each detected pitch	Offset est. is the same as for onsets	Evaluated in MIREX 2007, improved in MIREX 2008. Accuracy ≈ 54% (2008 MF0E). F-measure ≈ 20% (2008 NT)	43 Disklavier 30 seconds excerpts
		O	N.A.										
[Cha_et_al.08] (IRCAM)	2008	I	N.A.	Spectral analysis based on sinusoidal and noise model	N.A.	No	Poly	F	Spectral matching, spectral smoothing and synchronous amplitude evolution of single sources	F0 tracking using a high-order HMM model with two states: attack and sustain		Evaluated in MIREX 2007, improved in MIREX 2008-09. Accuracy ≈ 69% (2009 MF0E). F-measure ≈ 36% (2008 NT)	Samples from McGill, Iowa, IRCAM database
		O	N.A.										
[DuHaPa09]	2009	I	PCM audio signals fs = 44.1 kHz	Spectral analysis. Spectrum divided into *peaks* and non-peaks regions	N.A.	No	Poly	F	Maximum Likelihood parameter estimation in the frequency domain, using also neighboring frames estimates	Build pitch trajectories by constraint clustering problem with two class: *must-link* and cannot-*link*		Evaluated in MIREX 2009. Accuracy ≈ 57% (MF0E). F-measure ≈ 22% (NT)	10 real music performances (4-parts Bach's chorales)
		O	N.A.										
[ArNePa09]	2009	I	PCM mono or stereo 16 bits, fs = 44.1 kHz	Joint constant-Q and Bispectral (higher order spectra) analysis	Offline	No	Poly	F	Iterative 2D harmonic pattern matching (in the bispectrum domain) and subsequent cancellation	Peaks-picking in Kullback-Leibler divergence (over spectral frames)	Tracking of note events on STFT	Evaluated in MIREX 2009. 1st ranked in piano NT task. Accuracy ≈ 48% (MF0E). F-measure ≈ 23% (NT)	Excerpts from RWC Classical datbase;
		O	MIDI file; pitches, onsets & offsets list										

3. Architettura del Sistema

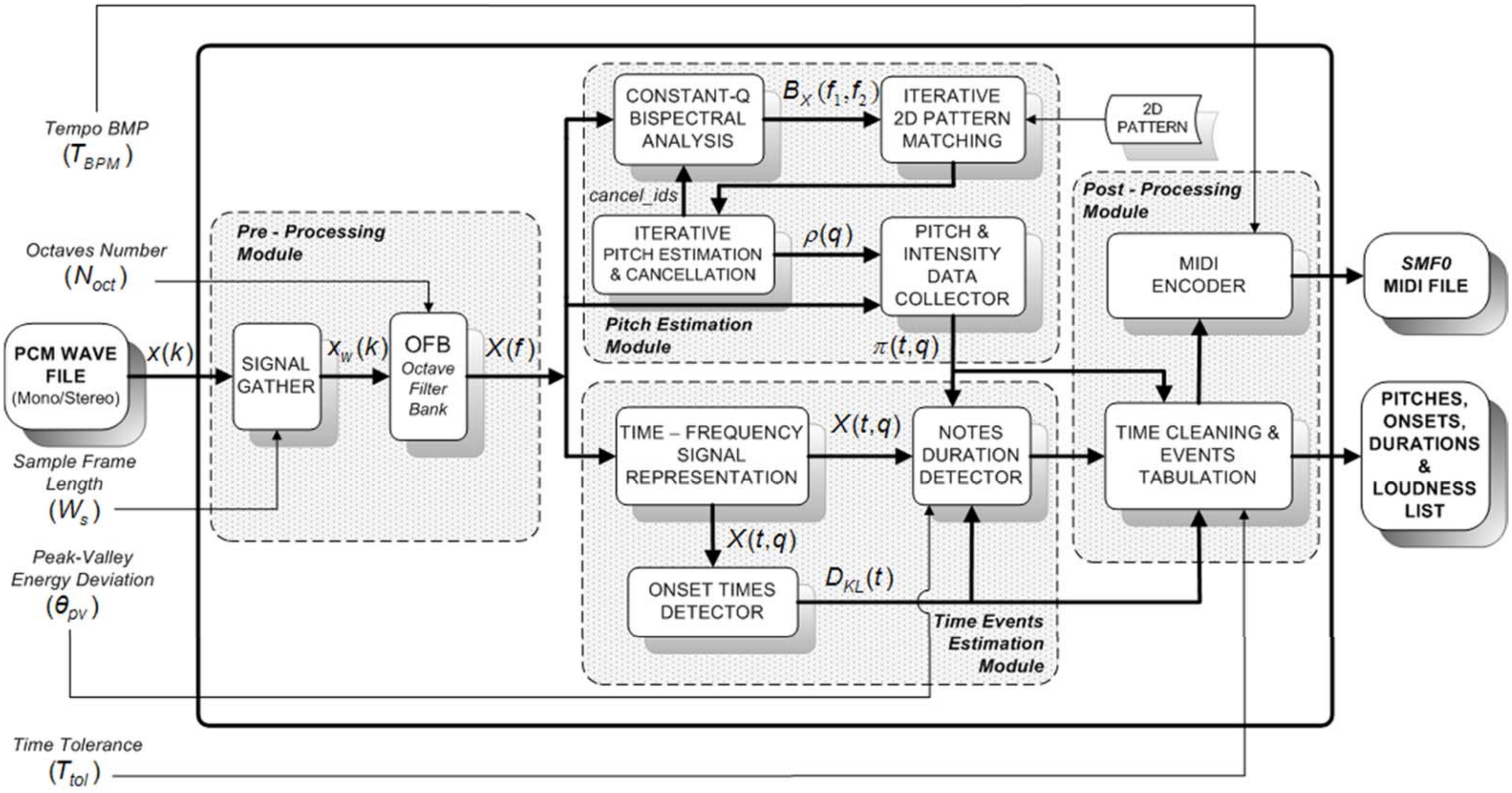
AUTOMATIC MUSIC TRANSCRIPTION – 3. *Architettura del Sistema*

- Sistema ideato e proposto nell'ambito del progetto europeo ***i-maestro***



AUTOMATIC MUSIC TRANSCRIPTION – 3. *Architettura del Sistema*

System Block Architecture



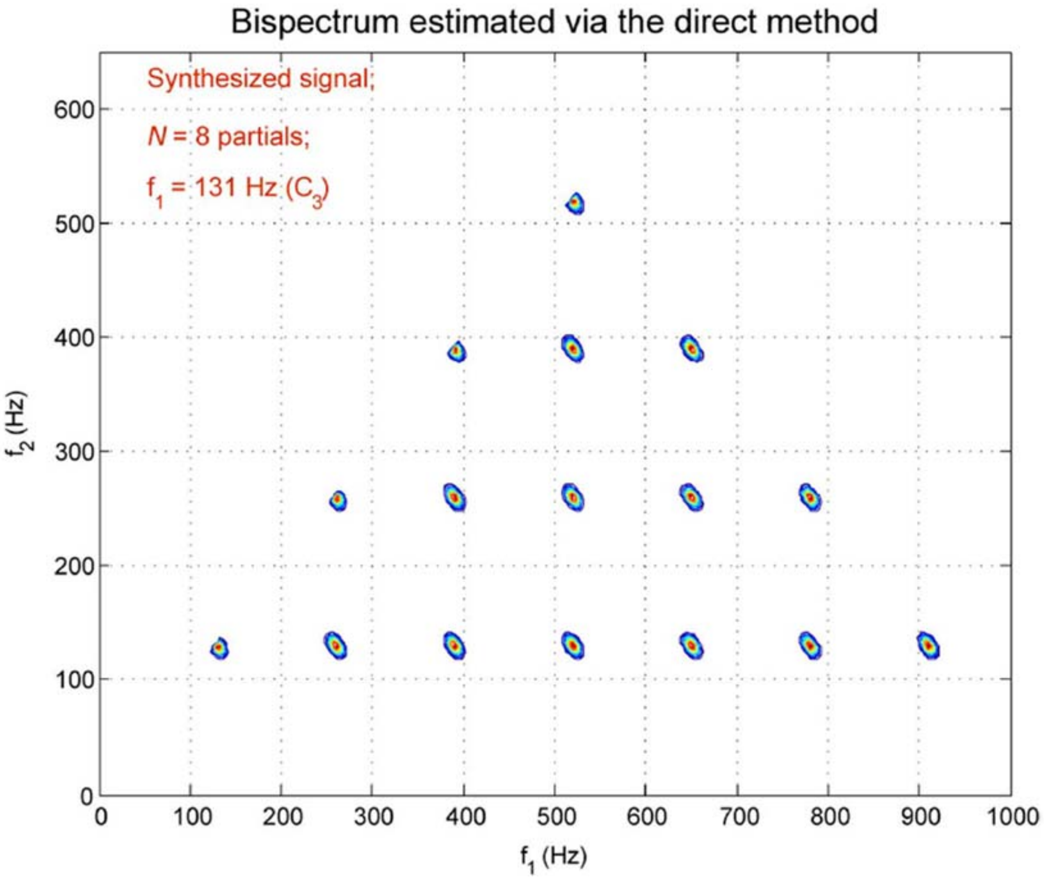
Pitch Estimation Module: Constant -Q Bispectral Analysis

- Analisi di spettri di ordine superiore (*HOSA*)*:* **Bispettro**

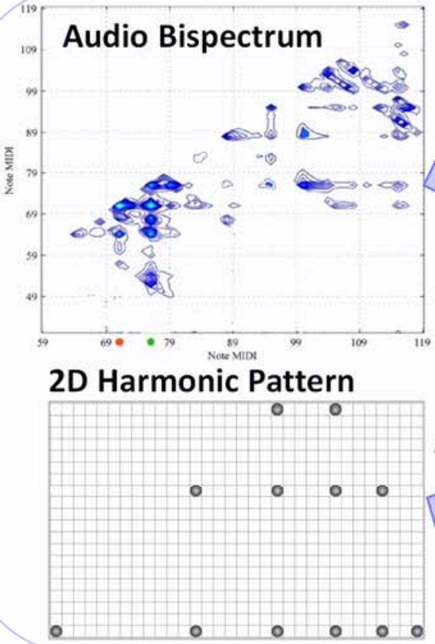
$$B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2),$$

dove *X(f)* è lo spettro del segnale audio in ingresso.

- Rappresentazione 3D in frequenza. Ogni suono genera un pattern armonico bidimensionale, la cui geometria è più utile per risolvere la sovrapposizione delle armoniche concorrenti.



Pitch Estimation Module: Multiple F0 – Estimation Algorithm

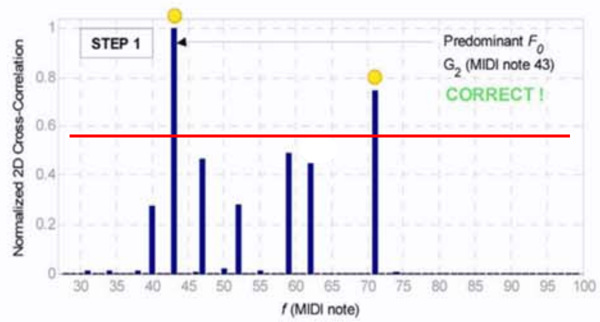


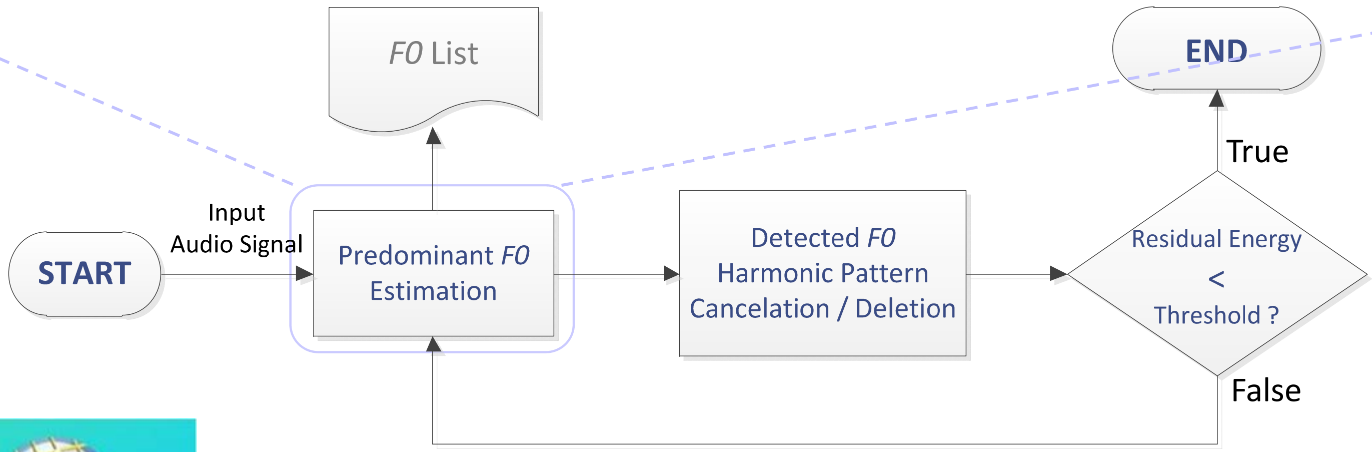
$B_x(f_1,f_2)$

$P(f_1,f_2)$

$$\rho(q) = \sum_{m_1=0}^{C_P-1} \sum_{m_2=0}^{R_P-1} P(m_1,m_2)|B_x(q+m_1,q+m_2)|$$

2D Correlation (Pattern Matching)

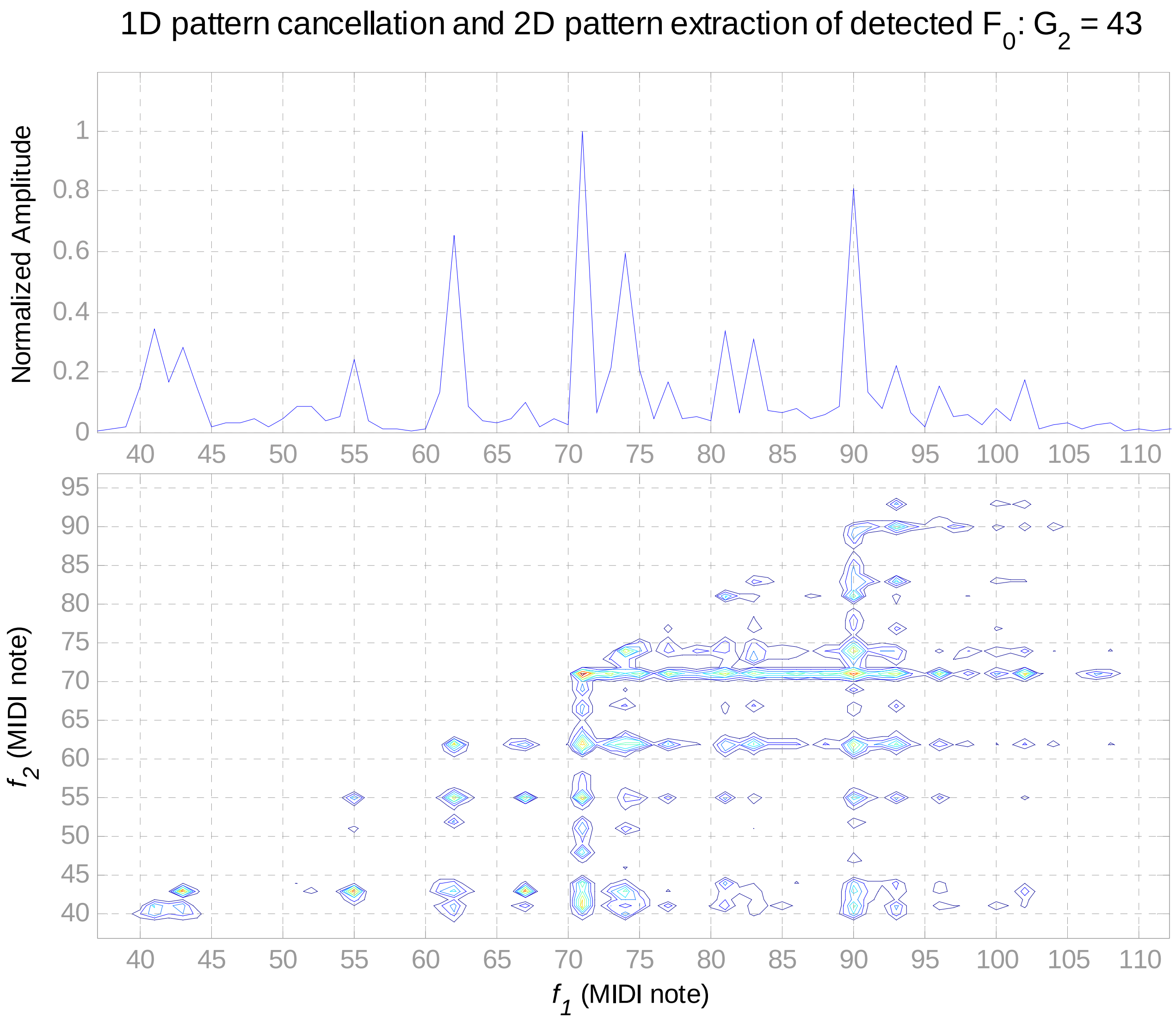




Pitch Estimation Module: Constant -Q Bispectral Analysis

Aspetti innovativi:

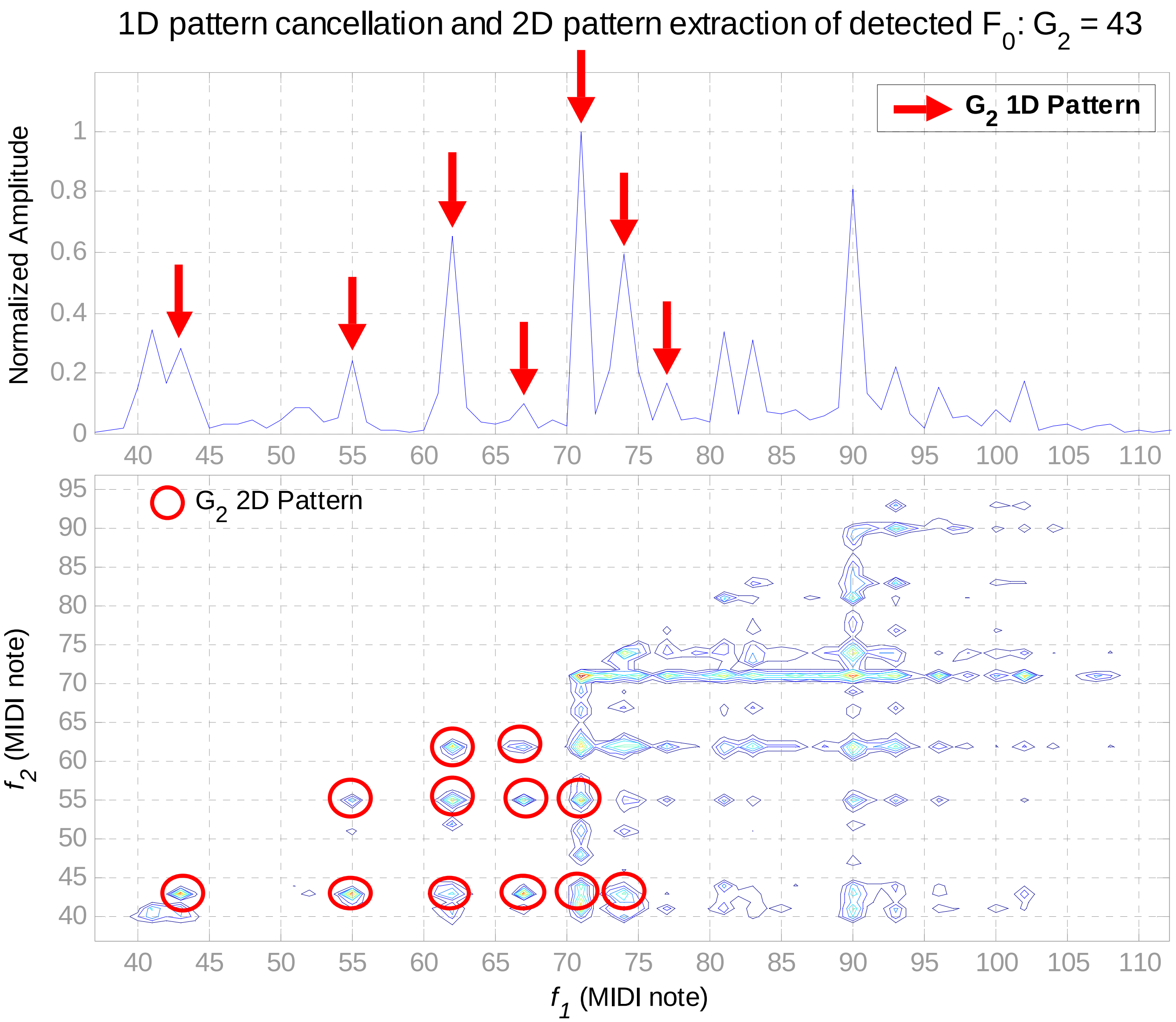
- Non linearità del bispettro e vantaggi nell'algoritmo di *Iterative Pitch Estimation and 2D Pattern Cancelation / Subtraction*.
- Rappresentazione tridimensionale in frequenza.
- Esempio a fianco: polifonia a tre suoni, tricordo G_2-D_4-B_4.



Pitch Estimation Module: Constant -Q Bispectral Analysis

Aspetti innovativi:

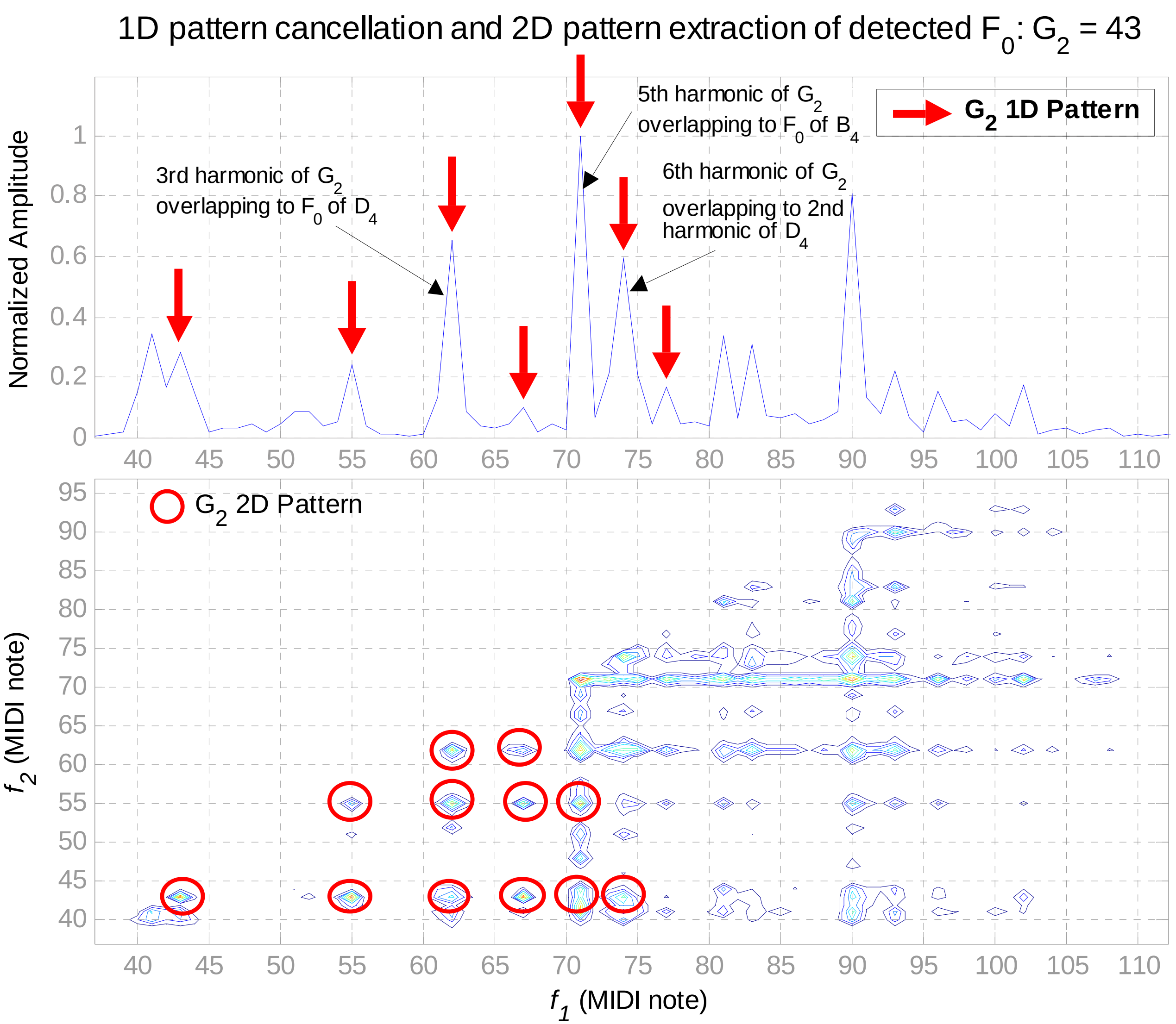
- Non linearità del bispettro e vantaggi nell'algoritmo di *Iterative Pitch Estimation and 2D Pattern Cancelation / Subtraction.*
- Rappresentazione tridimensionale in frequenza.
- Esempio a fianco: polifonia a tre suoni, tricordo G2-D4-B4.



Pitch Estimation Module: Constant -Q Bispectral Analysis

Aspetti innovativi:

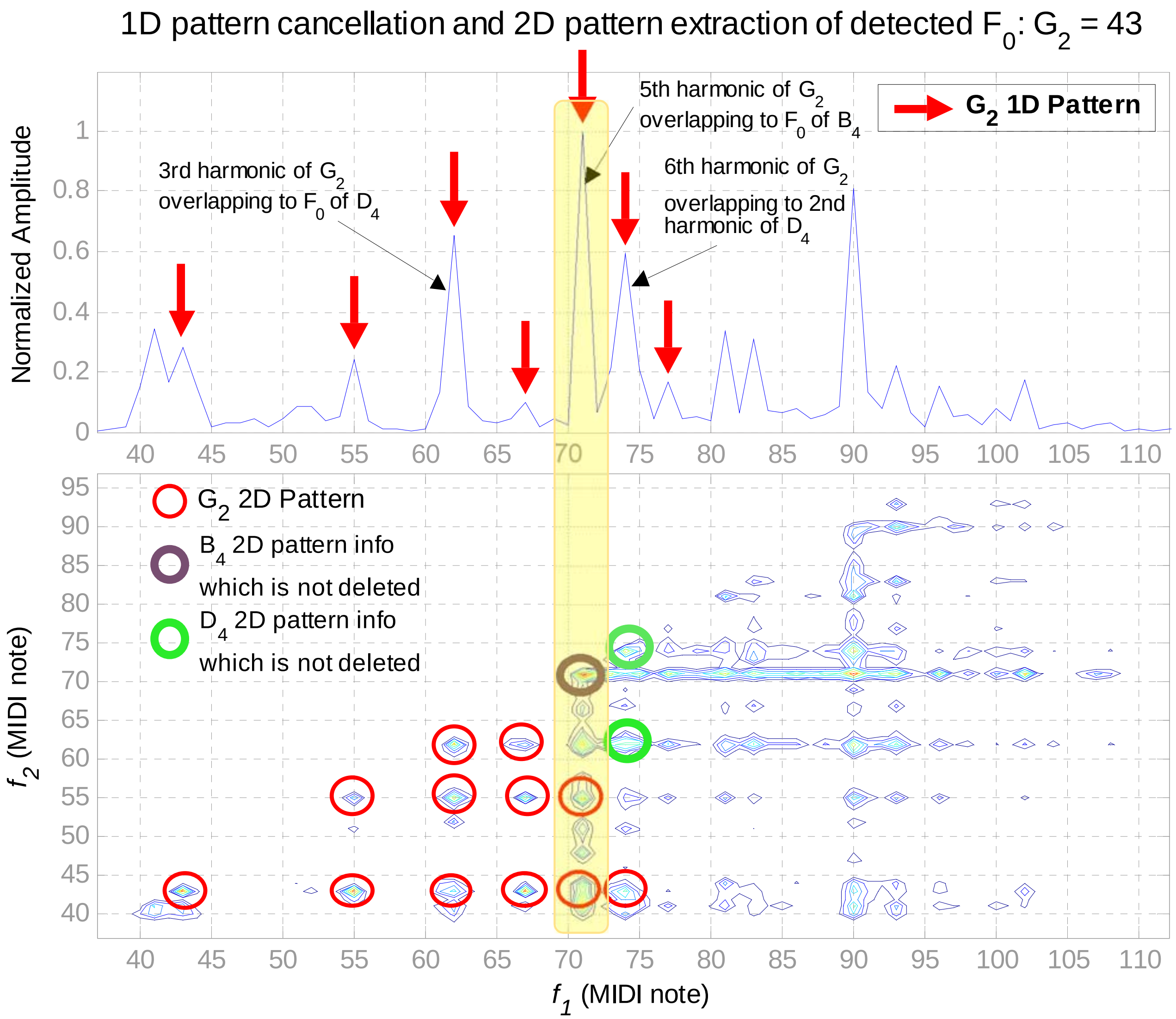
- Non linearità del bispettro e vantaggi nell'algoritmo di *Iterative Pitch Estimation and 2D Pattern Cancelation / Subtraction*.
- Rappresentazione tridimensionale in frequenza.
- Esempio a fianco: polifonia a tre suoni, tricordo G2-D4-B4.



Pitch Estimation Module: Constant -Q Bispectral Analysis

Aspetti innovativi:

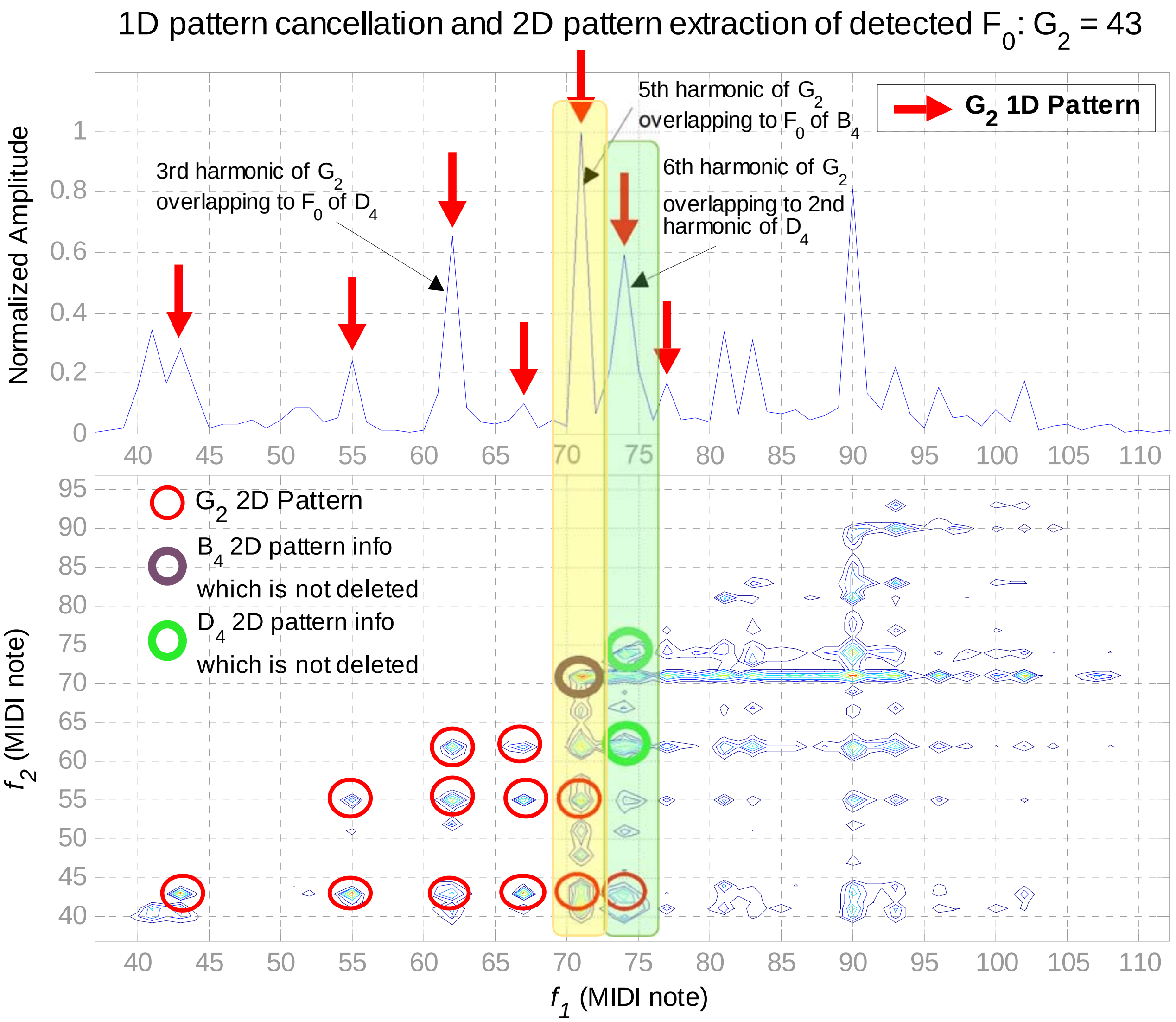
- Non linearità del bispettro e vantaggi nell'algoritmo di *Iterative Pitch Estimation and 2D Pattern Cancelation / Subtraction*.
- Rappresentazione tridimensionale in frequenza.
- Esempio a fianco: polifonia a tre suoni, tricordo G2-D4-B4.



Pitch Estimation Module: Constant -Q Bispectral Analysis

Aspetti innovativi:

- Non linearità del bispettro e vantaggi nell'algoritmo di *Iterative Pitch Estimation and 2D Pattern Cancelation / Subtraction*.
- Rappresentazione tridimensionale in frequenza.
- Esempio a fianco: polifonia a tre suoni, tricordo G2-D4-B4.

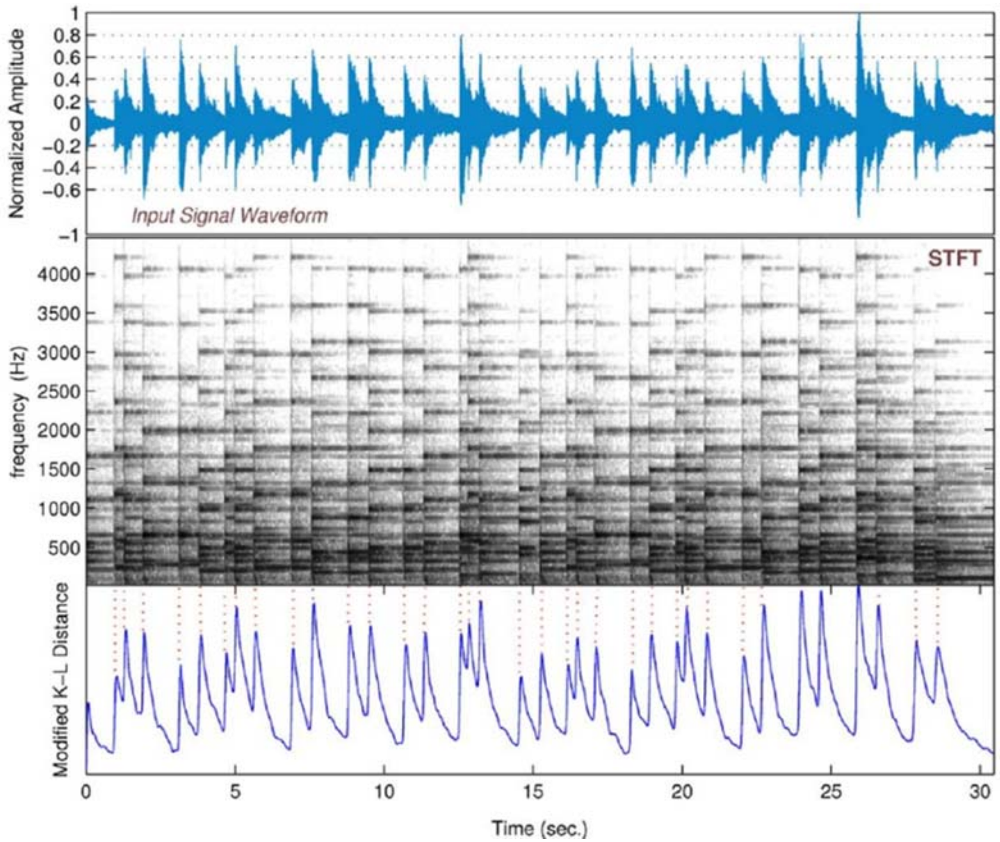


Time Estimation Module: Kullback-Leibler Divergence e Analisi Tempo-Frequenza

- **ONSET Detector** (istante di attacco delle note): utilizzo della *Kullback-Leibler Divergence*:

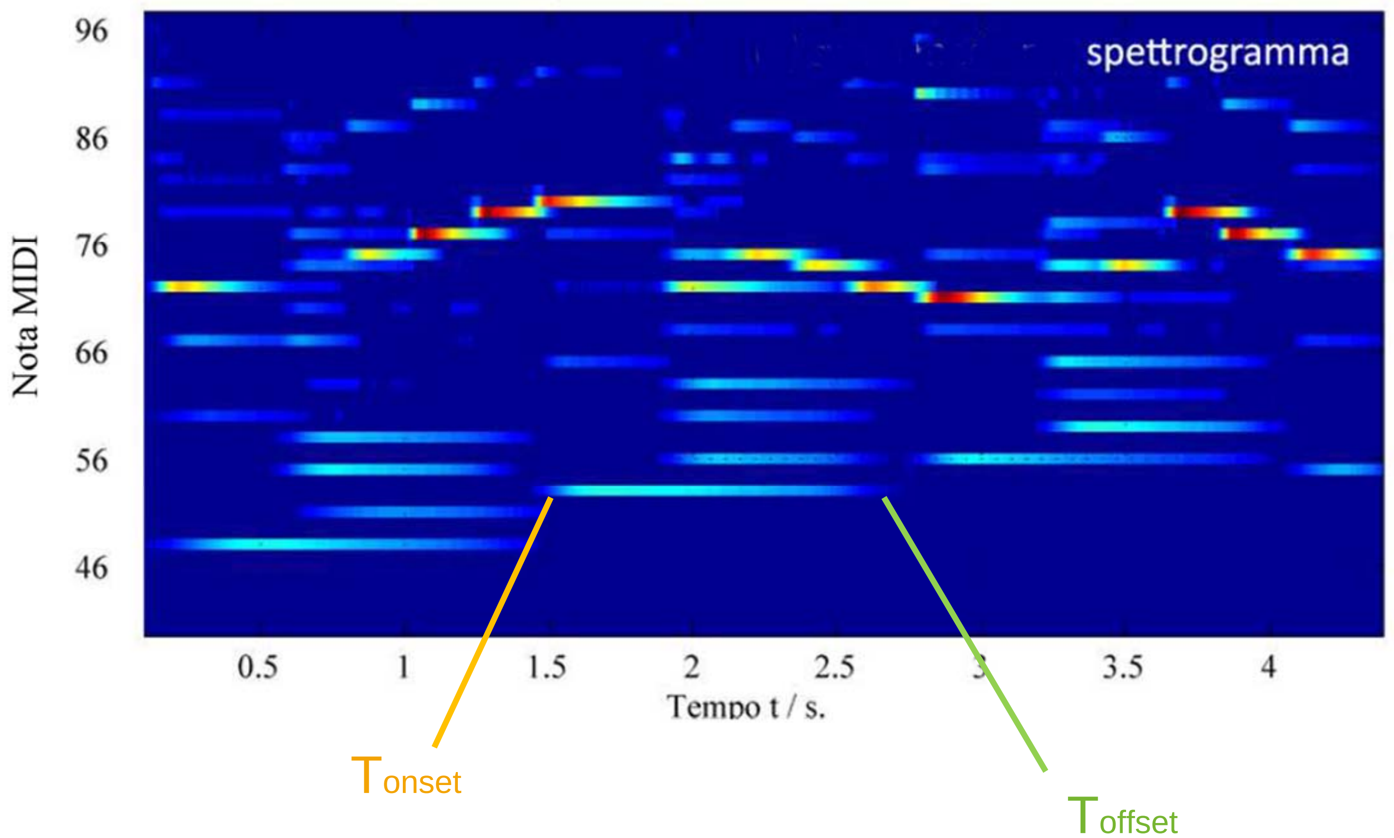
$$D_{\mathrm{KL}}(t) = \sum_{q=1}^{12N_{\mathrm{oct}}} \log\left(1 + \frac{|X(t,q)|}{|X(t-1,q)| + \varepsilon}\right)$$

per stimare le variazioni significative di energia tra frame consecutivi dello spettrogramma del segnale audio.



Time Estimation Module: Kullback-Leibler Divergence e Analisi Tempo-Frequenza

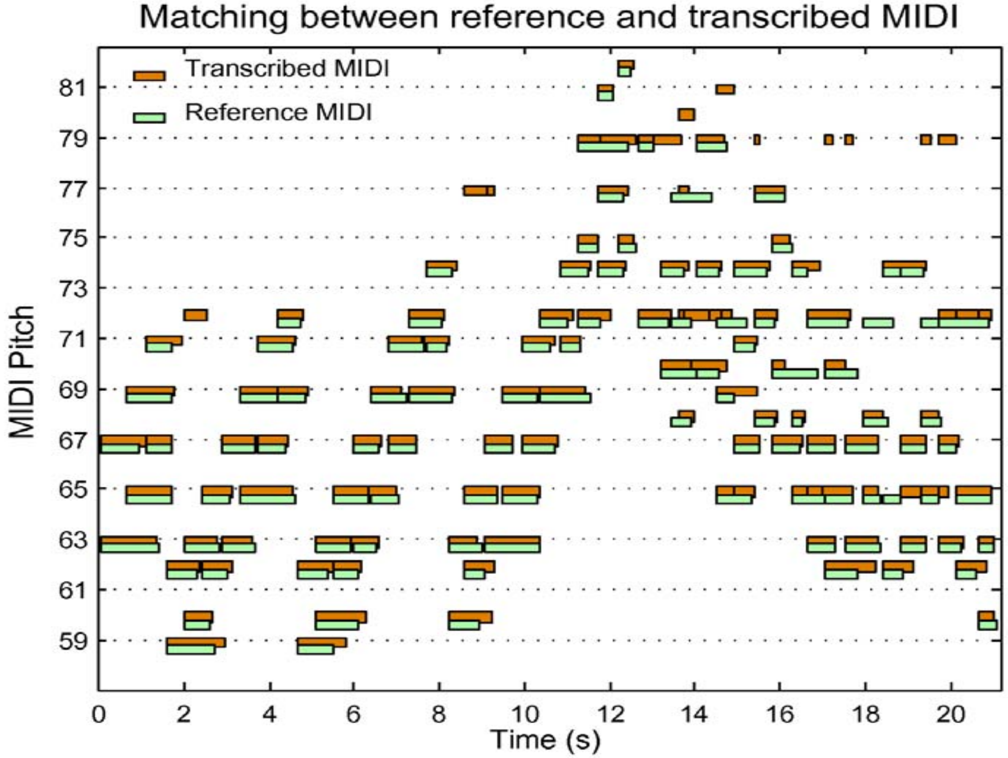
- **DURATION Detector** (durata delle note): analisi tempo – frequenza del segnale audio.



$$Duration = T_{offset} - T_{onset}\ .$$

4. Validazione e Risultati

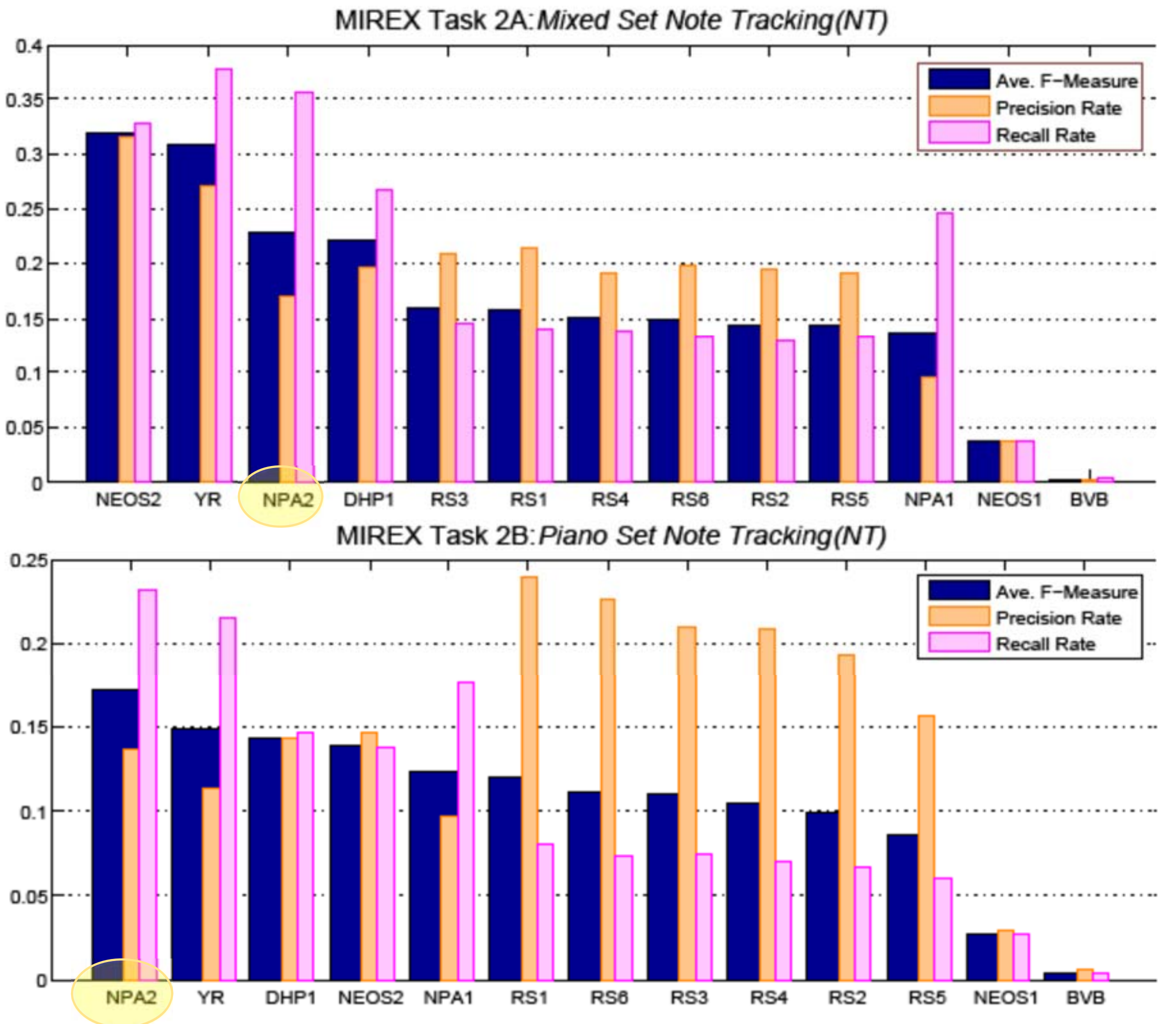
AUTOMATIC MUSIC TRANSCRIPTION – 4. *Validazione e Risultati*



AUTOMATIC MUSIC TRANSCRIPTION – 4. *Validazione e Risultati*

MIREX International Contest (*Music Information Retrieval Evaluation eXchange*)

ISMIR Conference



Criteri di valutazione:

(a) **Nota** rilevata correttamente se:

1) *F0* stimata a ± 3% (in Hz) rispetto alla *F0* corrispondente nel *reference*;

2) Onset stimato a ± 50 ms rispetto al *reference*;

3) Durata ± 20% rispetto al *reference*.

(b) : Parametri di valutazione delle prestazioni:

Precision = $\frac{Tp}{Tp + Fp}$;

Recall = $\frac{Tp}{Tp + Fn}$;

F – Measure = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$.

http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

A1. Altri Progetti Svolti

Altri progetti svolti nel corso del dottorato di ricerca:

- **OSIM** (**Open Space Innovative Mind**) – Studio e sviluppo di una piattaforma web, e costruzione di una base di conoscenza semantica (web-ontology) per la navigazione e la ricerca di risorse, competenze e pubblicazioni dell'Ateneo fiorentino.

- **Palamede** – Studio e modellazione di soluzioni di gestione di archivi culturali ed educazionali in relazione a media complessi e collezioni.

  Reingegnerizzazione della piattaforma open source OJS (per la pubblicazione e gestione ondine di riviste scientifiche) per permettere una gestione multi-Press;

- Sviluppo di un innovativo sistema di rilevamento della presenza di persone all'interno di un accesso protetto (in collaborazione con **MESA** s.r.l.)

  Tecniche di signal processing e applicati alla risposta di sensori passivi a microonde. Uso di modelli statistico-predittivi per implementare le regole di decisione del sistema (regressione multilineare e logistica).

A2. Pubblicazioni

Pubblicazioni:

1. Argenti F., Nesi P., Pantaleo G.: ***Automatic music transcription: from monophonic to polyphonic***, Chapter n. 3 in: Musical Robots and Interactive Multimodal Systems, Springer Ed., ISBN 978-3-642-22290-0, 2011.

2. Argenti F., Nesi P., Pantaleo G.: ***Automatic Transcription of Polyphonic Music Based on the Constant-Q Bispectral Analysis***, IEEE Transactions on Audio, Speech and Language Processing, vol. 19(6), pp. 1610-1630, Aug. 2011.

3. Bellini P., Nesi P., Pantaleo G.: ***Palamede: a Multi-Press Open Journal System***, Proc. Of the 17th International Conference on Distributed Multimedia Systems. (DMS 2011), vol. , pp. 58-63, Florence 18-20 Aug 2011.

4. Bellini P., Cenni D., Fuzier A., Nesi P., Pantaleo G., Paolucci M.: ***Metrics for Best Practice Networks Analysis***, The 16th Int. Conf. on Distributed Multimedia Systems, Hyatt Lodge at McDonald's Campus, Oak Brook, Illinois, USA, 2010.

5. Argenti F., Nesi P. and Pantaleo G.: ***Multiple F0-Estimation & Tracking for MIREX 2009***, in Proc. of ISMIR 2009 International Conference, 2009.

http://www.disit.dsi.unifi.it/
pantaleo@dsi.unifi.it

~ FINE ~