EECS 225D
Audio Signal Processing in Humans and Machines

Lecture 18 – Music Signal Analysis

2012-3-21

Professor Nelson Morgan

today's lecture by John Lazzaro

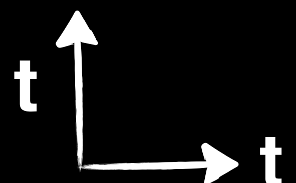
www.icsi.berkeley.edu/eecs225d/spr12/

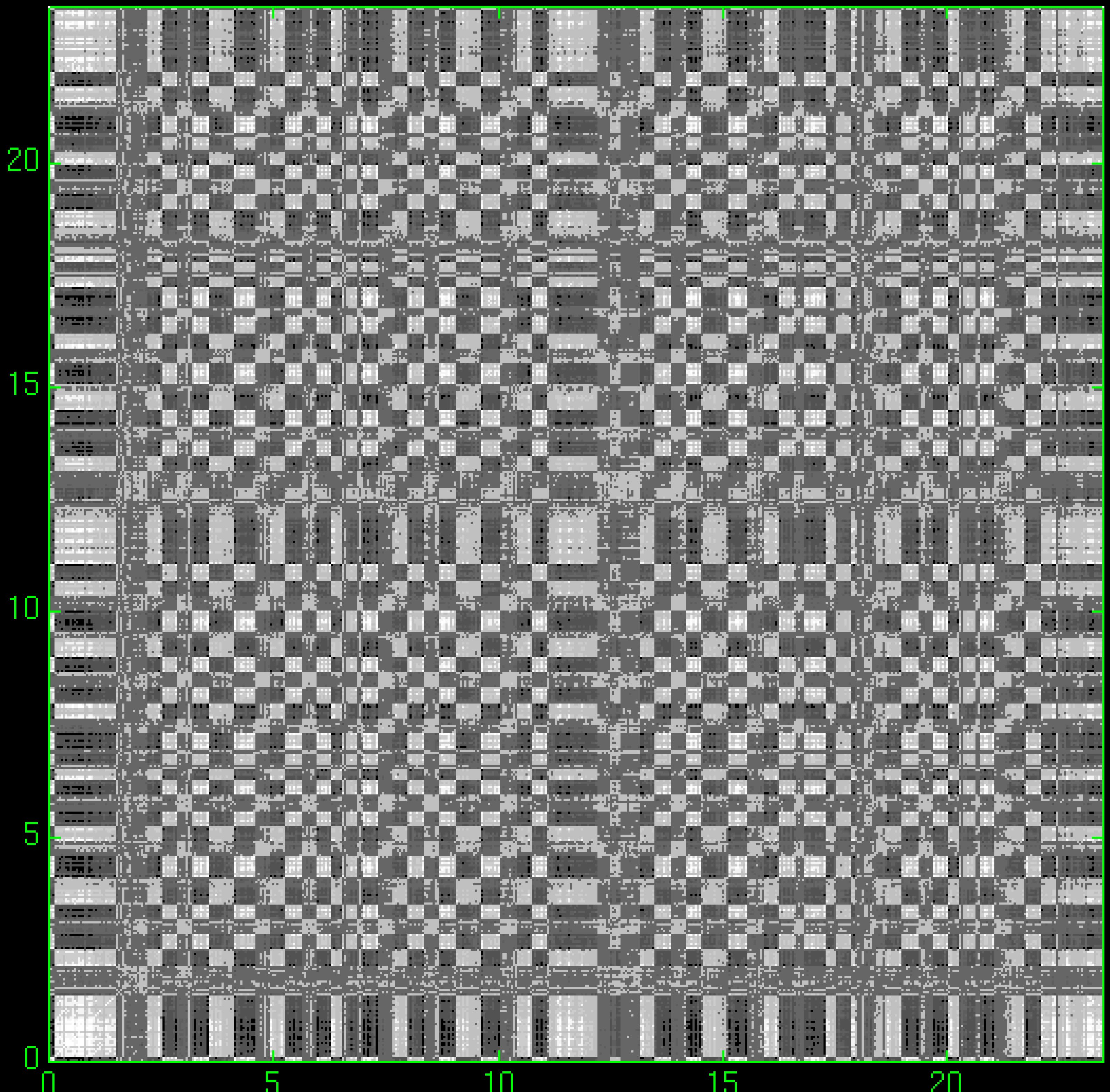
Music appreciation class for computers

- Discovering musical structure
- Chroma: Simple chord detection
- Sheet-music score alignment
- Music transcription

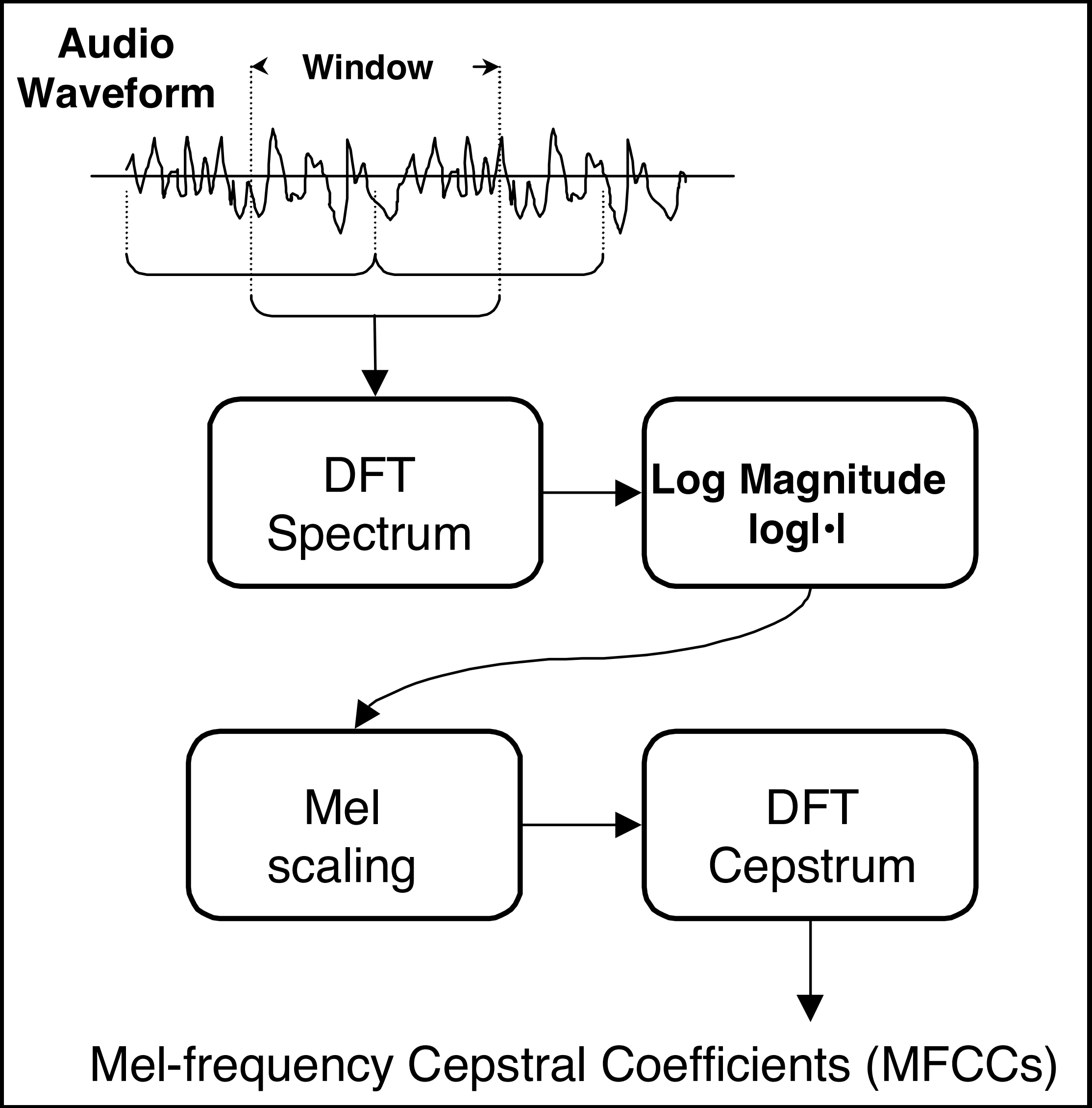
Drum pattern self-similarity.

(Foote, 1999)





Feature vector

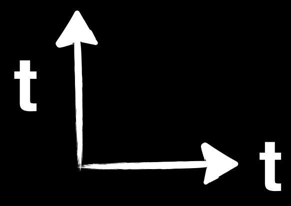


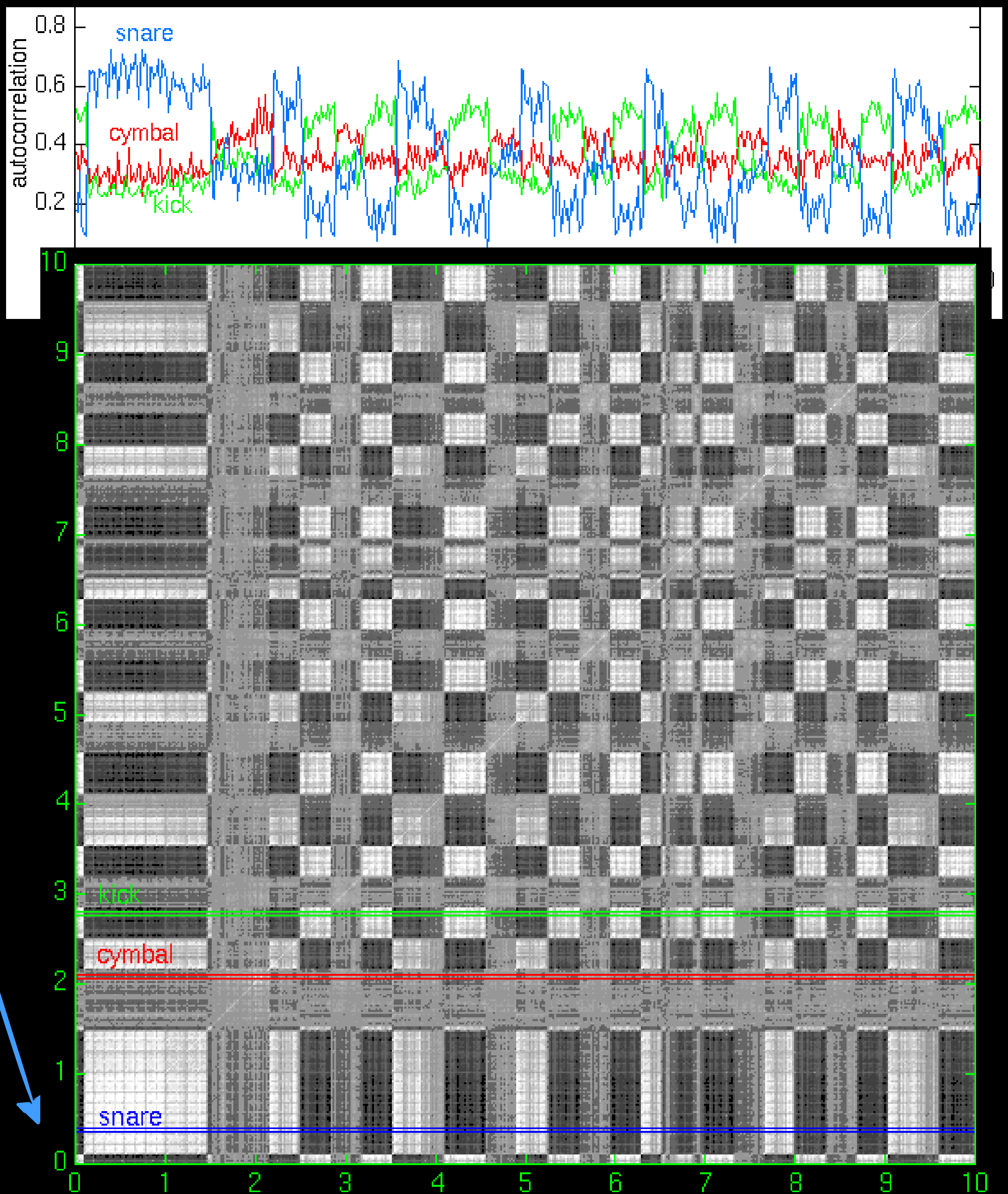
Similarity metric

$$S_w(i,j) \equiv \frac{1}{w}\sum_{k=0}^{w-1}(\boldsymbol{v}_{i+k} \bullet \boldsymbol{v}_{j+k})$$

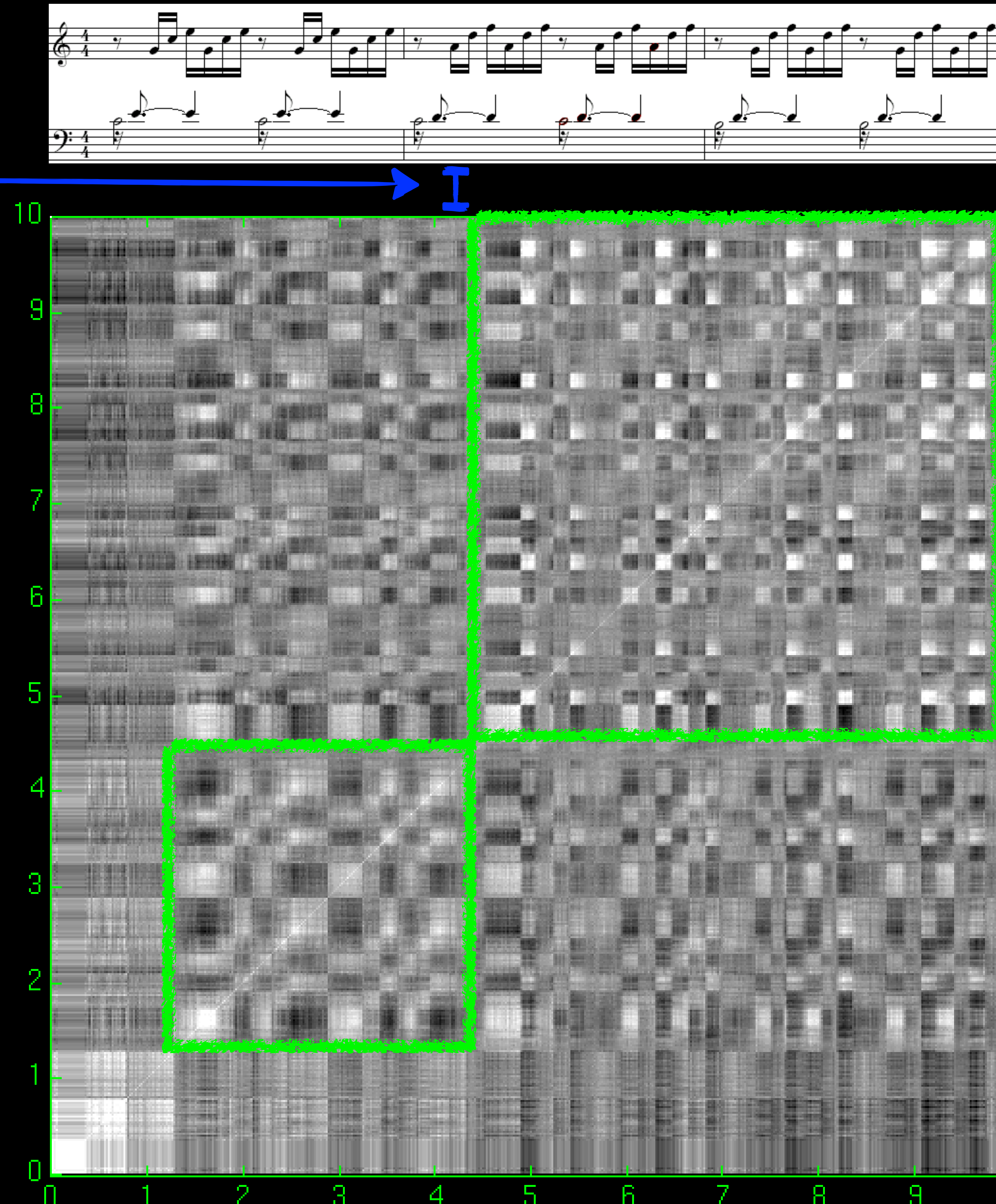
A simple idea, but very effective for pattern discovery ....

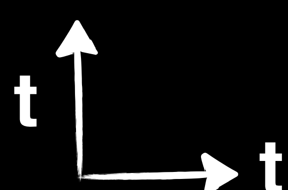
Horizontal **slice** patterns correspond to different drums in the kit.





Chord modulation of repetitive piano pattern revealed by similarities and differences.

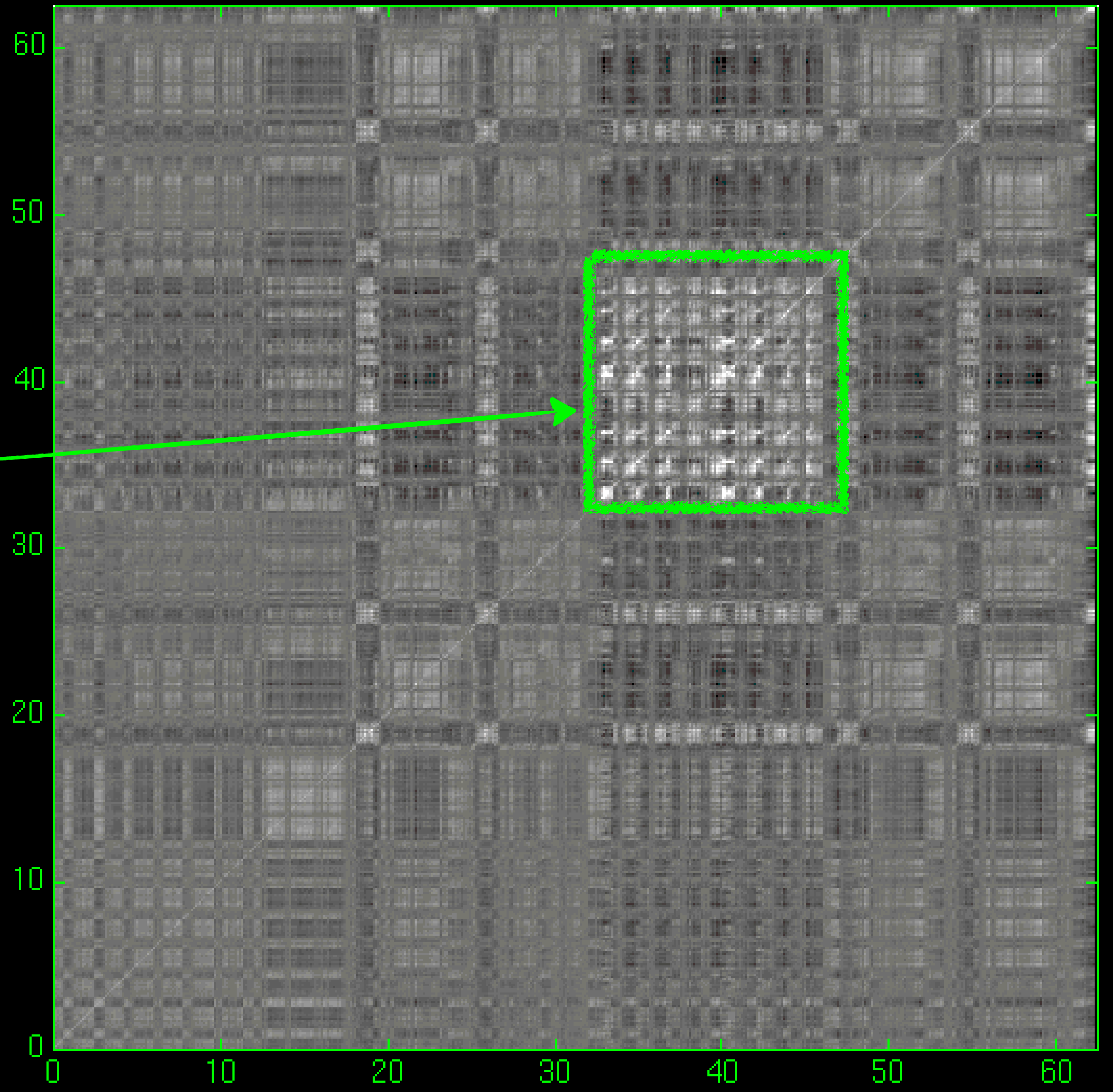




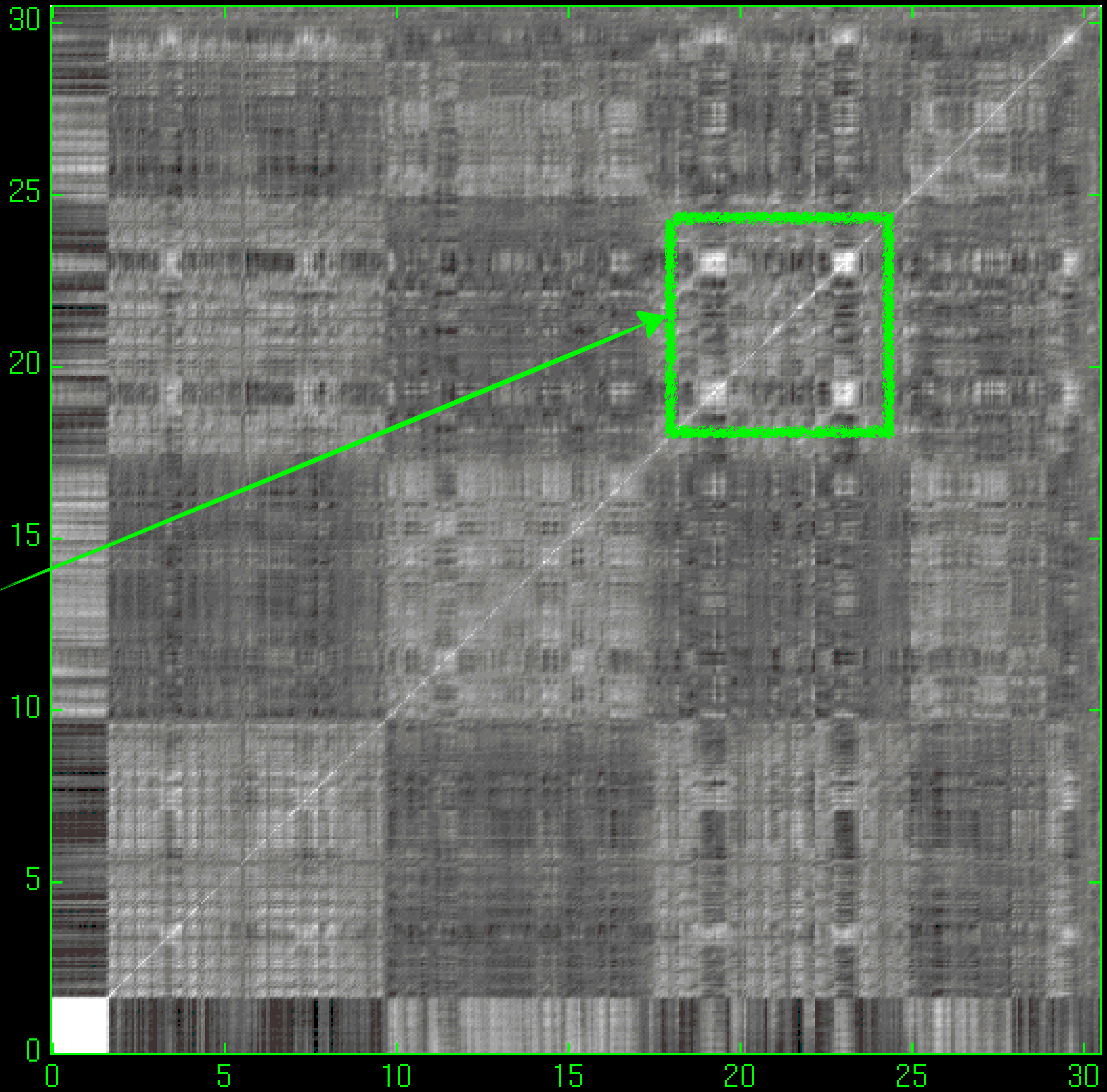
Take 5

Dave Brubeck Quartet.

Alto sax solo.



Horn Concerto No. 4 (Rondo) Mozart.

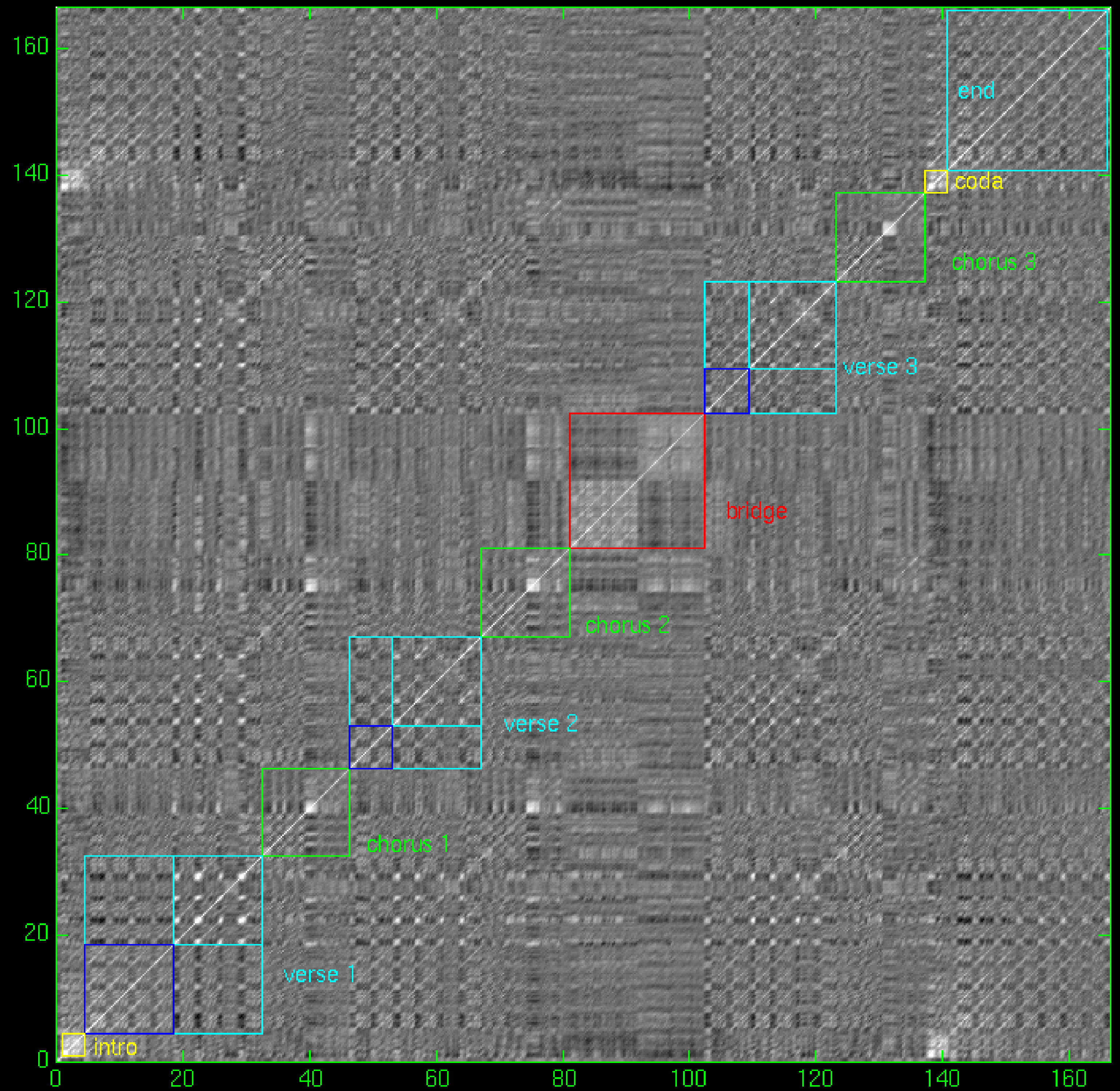


High horn note that starts and ends a section.

Day Tripper

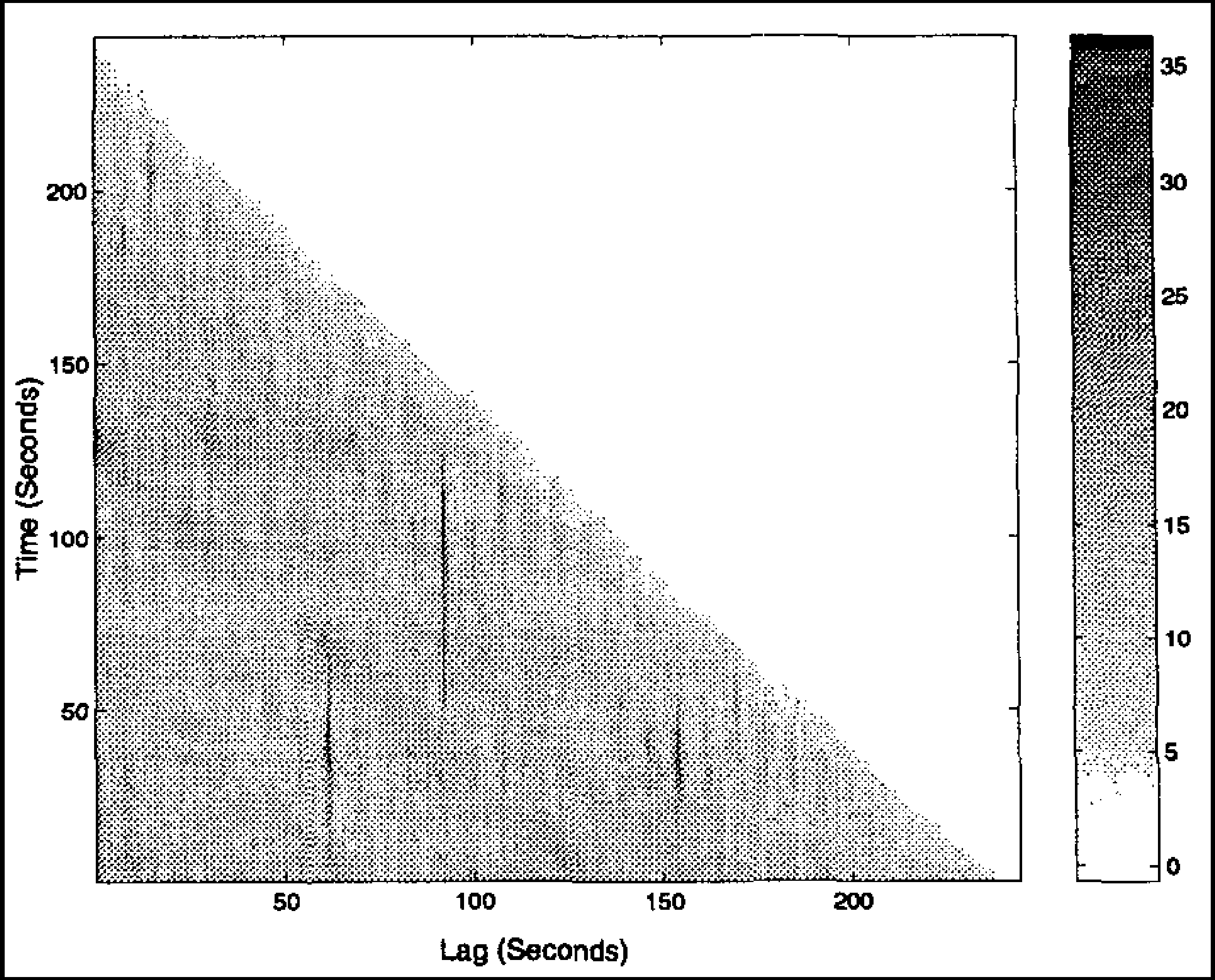
The Beatles

Song sections are easy to segment, identify.



Alternative visualization:

Running time on y axis, correlation lag time on x axis.



Music appreciation class for computers

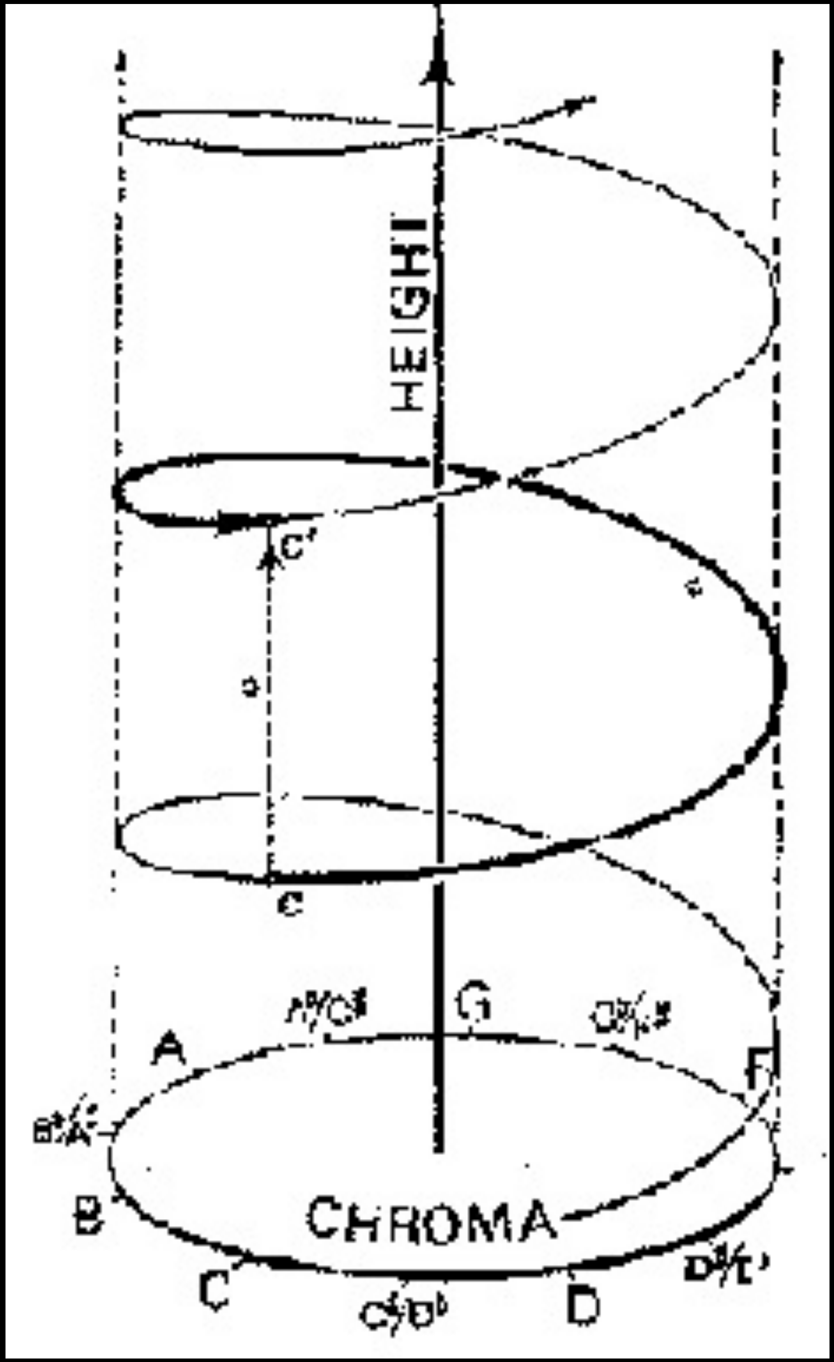
- Discovering musical structure
- Chroma: Simple chord detection
- Sheet-music score alignment
- Music transcription

Musical pitch has two aspects:

Height: Octave

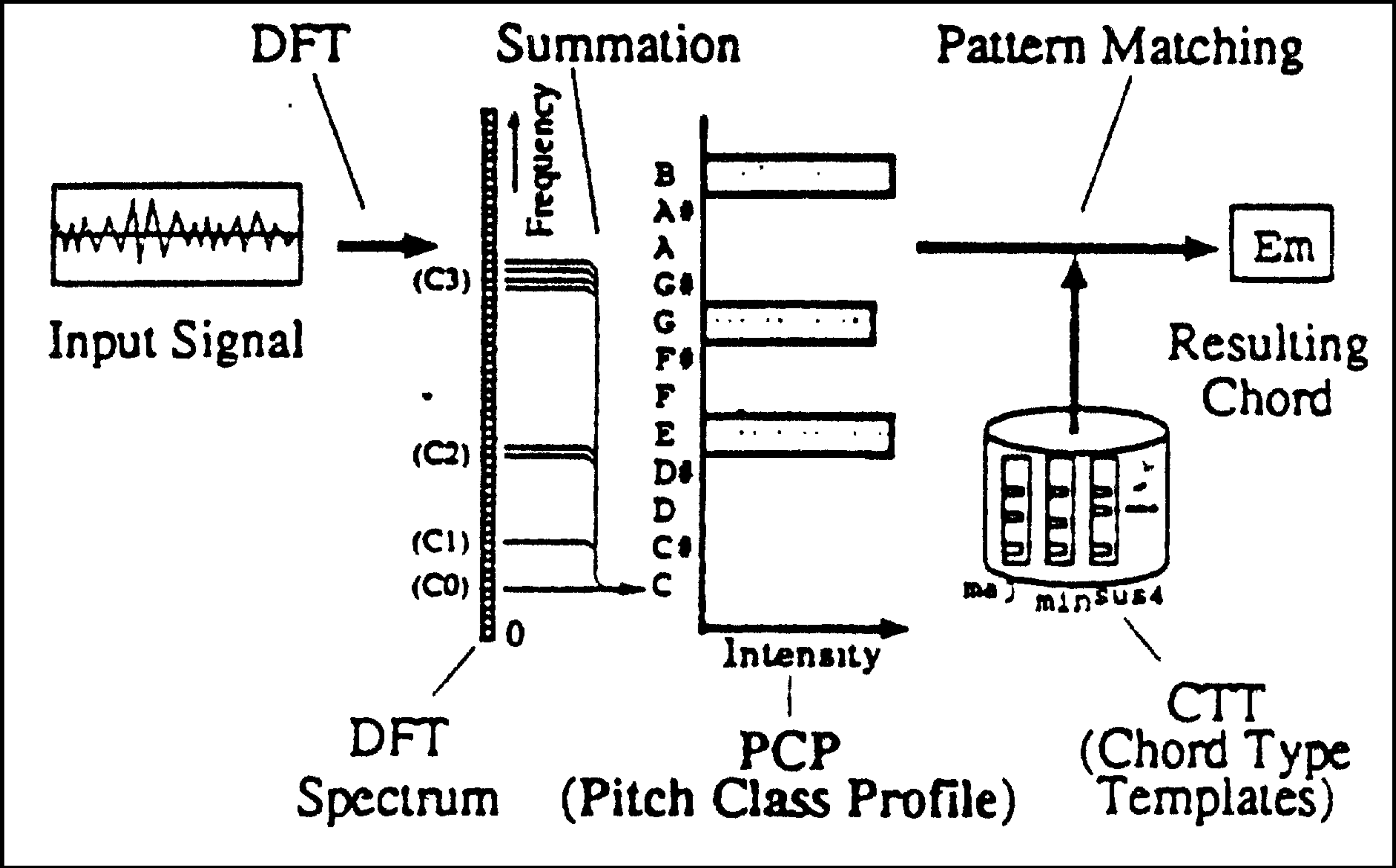
Chroma: Note name

And so, we can draw the notes on a piano keyboard like a spiral ...



If note timbres only has octave harmonics ...

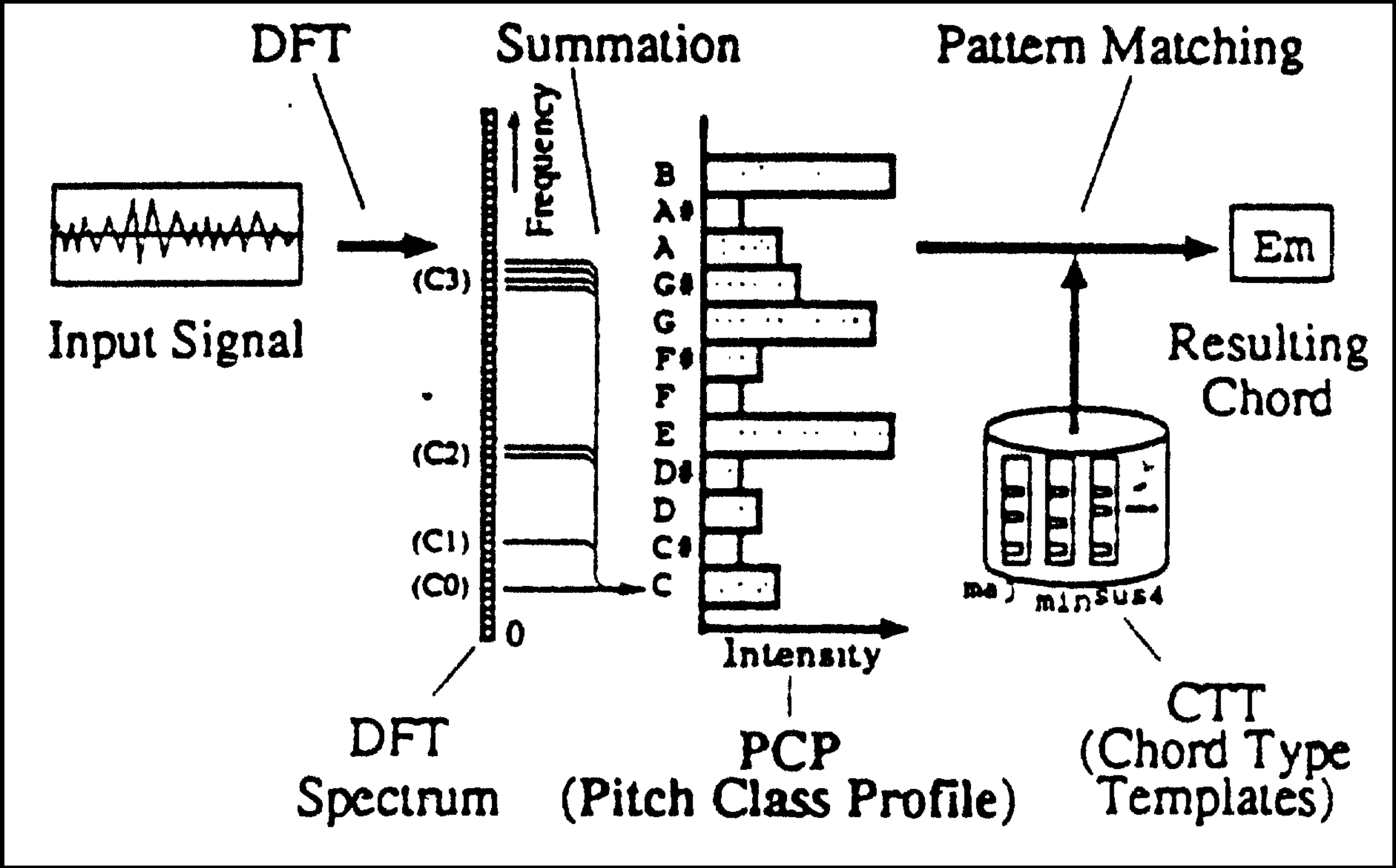
... if notes are in tune to a reference.



A histogram can map waveforms to chords.

(Chroma algorithm, Fujishima, 1999)

Real-world histogram: Note timbres with arbitrary harmonics ... imperfect instrument tuning ...



A histogram can map waveforms to chords.

(Chroma algorithm, Fujishima, 1999)

MIDI		PCH	OCT	CPS (A440)
C	36	6.00	6.000	65.40
C#	37	6.01	6.083...	69.29
D	38	6.02	6.166...	73.41
D#	39	6.03	6.250	77.78
E	40	6.04	6.333...	82.40
F	41	6.05	6.416...	87.30
F#	42	6.06	6.500	92.49
G	43	6.07	6.583...	97.99
G#	44	6.08	6.666...	103.82
A	45	6.09	6.750	110.00
A#	46	6.10	6.833...	116.54
B	47	6.11	6.916...	123.47
C	48	7.00	7.000	130.81
C#	49	7.01	7.083...	138.59
D	50	7.02	7.166...	146.83
D#	51	7.03	7.250	155.56
E	52	7.04	7.333...	164.81
F	53	7.05	7.416...	174.61
F#	54	7.06	7.500	184.99
G	55	7.07	7.583...	195.99
G#	56	7.08	7.666...	207.65
A	57	7.09	7.750	220.00
A#	58	7.10	7.833...	233.08
B	59	7.11	7.916...	246.94
C	60	8.00	8.000	261.62
C#	61	8.01	8.083...	277.18
D	62	8.02	8.166...	293.66
D#	63	8.03	8.250	311.12
E	64	8.04	8.333...	329.62
F	65	8.05	8.416...	349.22
F#	66	8.06	8.500	369.99
G	67	8.07	8.583...	391.99
G#	68	8.08	8.666...	415.20
A	69	8.09	8.750	440.00
A#	70	8.10	8.833...	466.16
B	71	8.11	8.916...	493.88
C	72	9.00	9.000	523.25

Non-octave harmonics?

Example: C0 = 65.40 Hz

3rd harmonic is @ 196.2 Hz, lies between G and G#.

5th harmonic is @ 327 Hz, lies between D# and E.

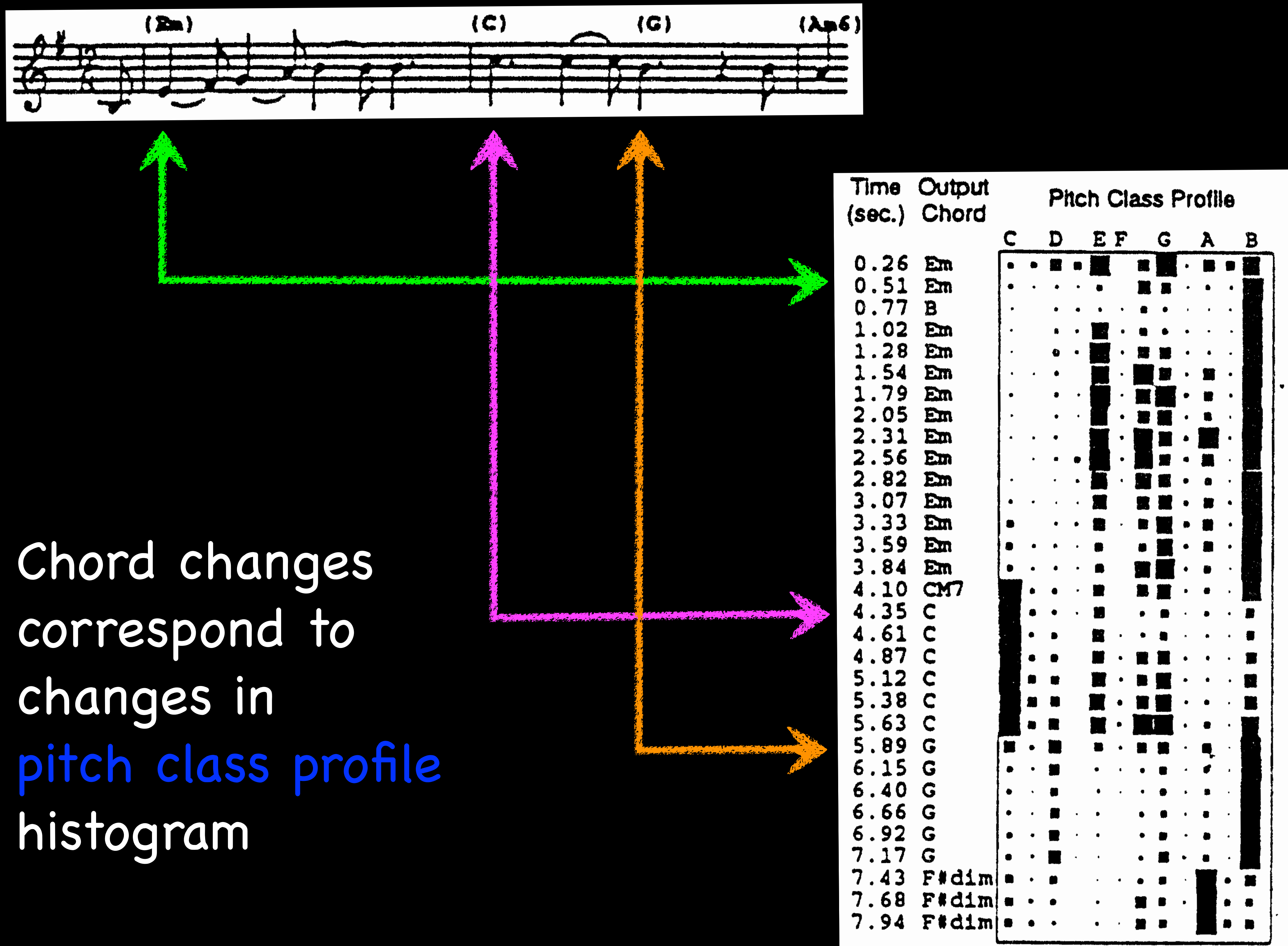
6th harmonic is @ 392.4 Hz, lies between G and G#.

In practice, energy in octave harmonics yields a good SNR.

The first 10 partials ...

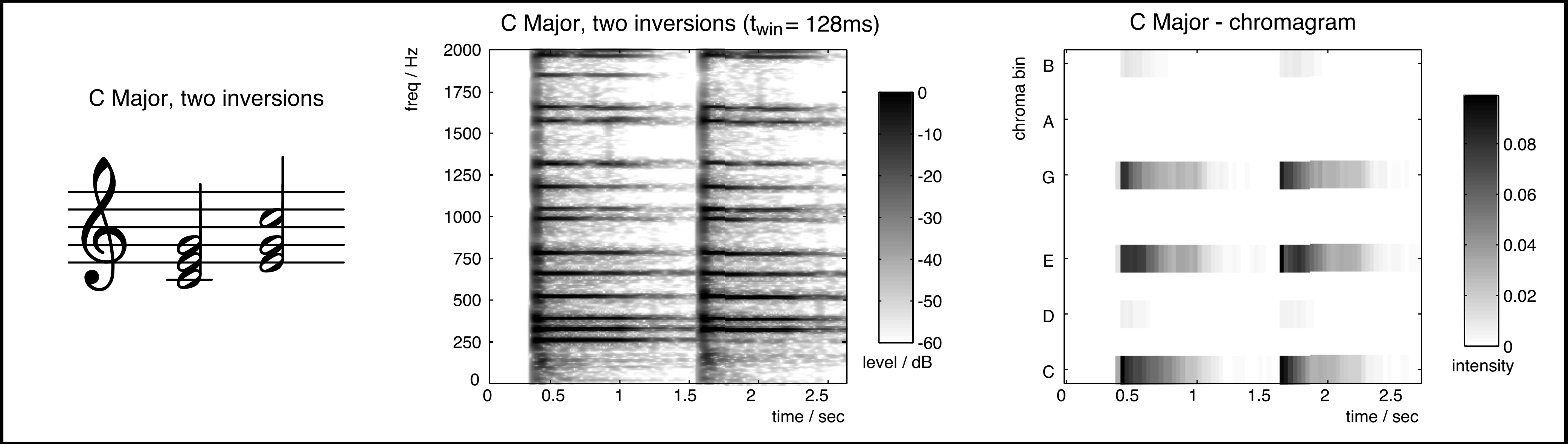
- octaves

	Ratio
The same note (C1)	1:1
One octave higher (C2, 12 semi-tones)	2:1
An octave and a 5^{th} (G2, 19 semi-tones), +2 cents	3:1
Two octaves higher (C3, 24 semi-tones)	4:1
Two octaves higher plus a major 3^{rd} (E3, 28 semi-tones), -14 cents	5:1
Two octaves and a 5^{th} (G3, 31 semi-tones), +2 cents	6:1
Two octaves and a 7^{th} (B3, 34 semi-tones), -31 cents	7:1
Three octaves higher (C4, 36 semi-tones)	8:1
Three octaves and a 2^{nd} (B5, 38 semi-tones), +4 cents	9:1
Three octaves and a major 3^{rd}, (C#5, 40 semi-tones) -14 cents	10:1



Chord changes correspond to changes in pitch class profile histogram

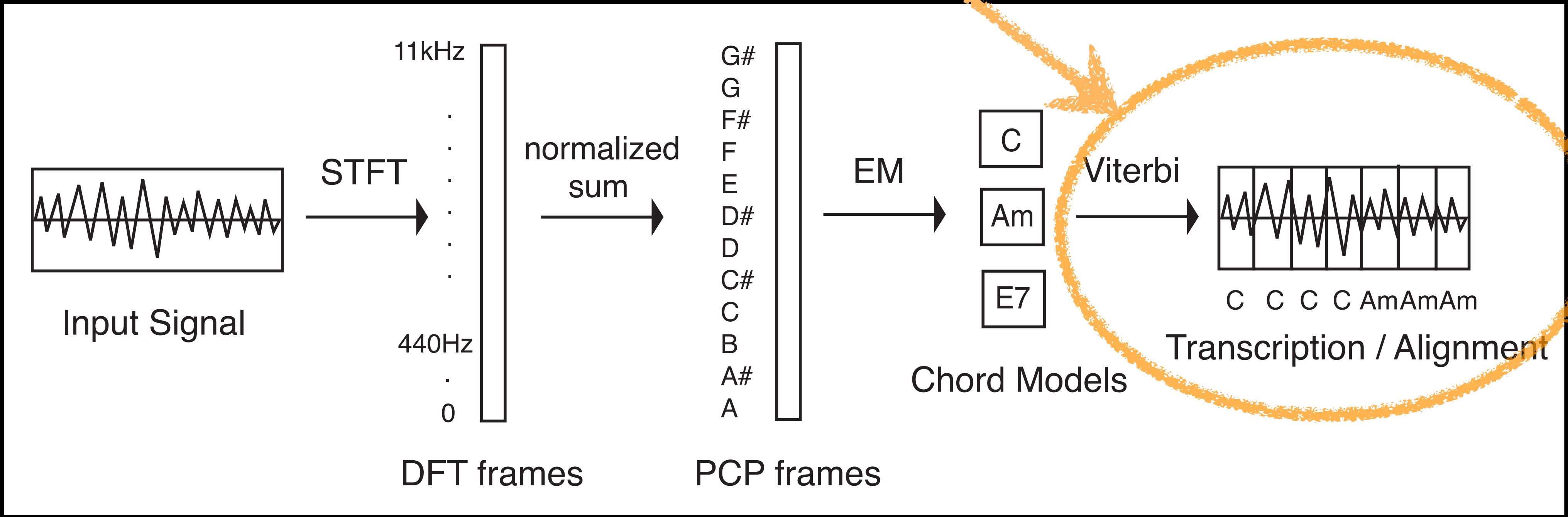
Chroma is invariant to note order of the chord



Usually, a desirable property.

However, for some types of chords, the same note list in different note orders produces different musical effects, and thus are given different chord names.

Chroma + the standard ASR algorithms

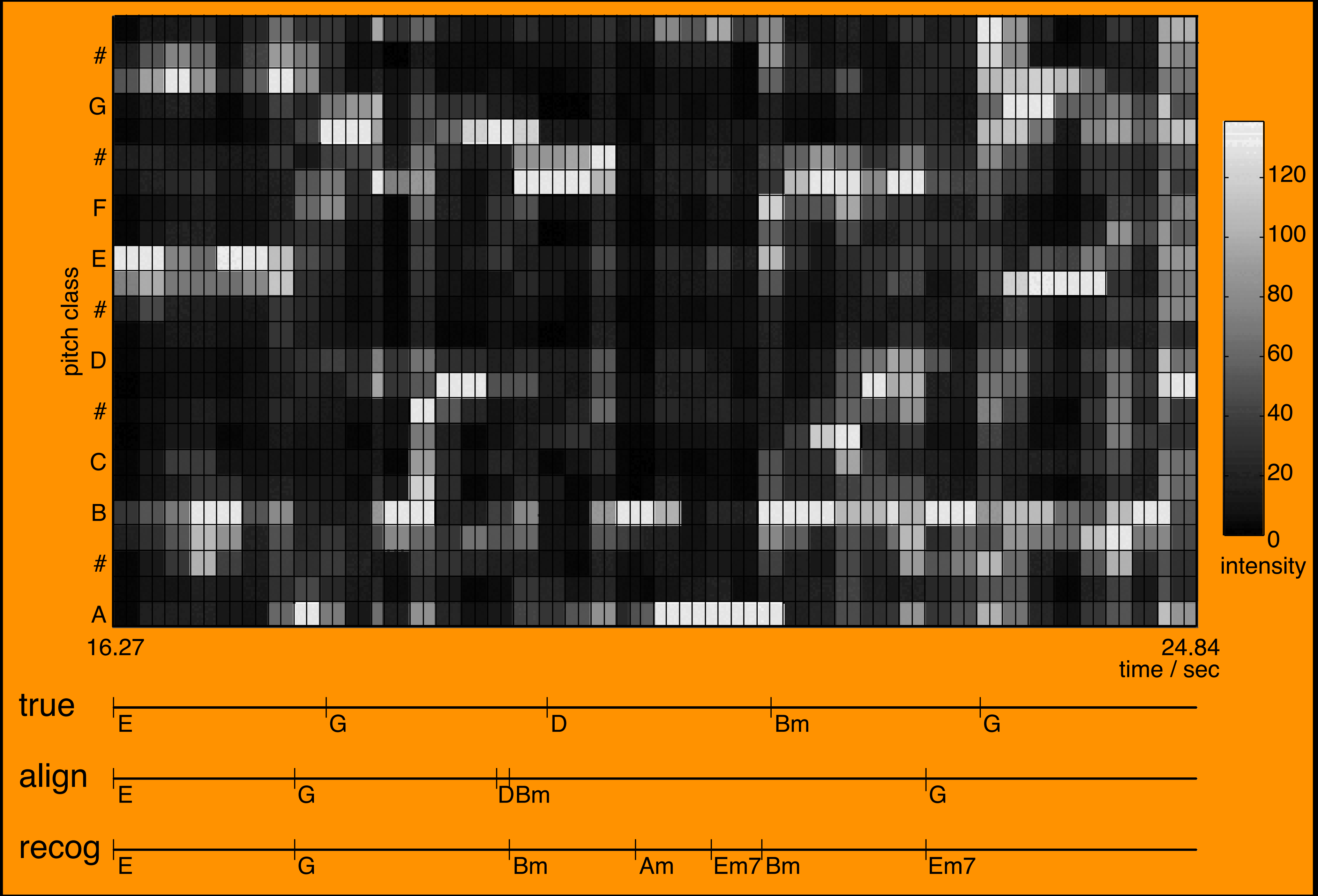


Chord Segmentation and Recognition using EM-Trained Hidden Markov Models

Alexander Sheh and Daniel P.W. Ellis
LabROSA, Dept. of Electrical Engineering,
Columbia University, New York NY 10027 USA
{asheh79,dpwe}@ee.columbia.edu

8 days a week, The Beatles

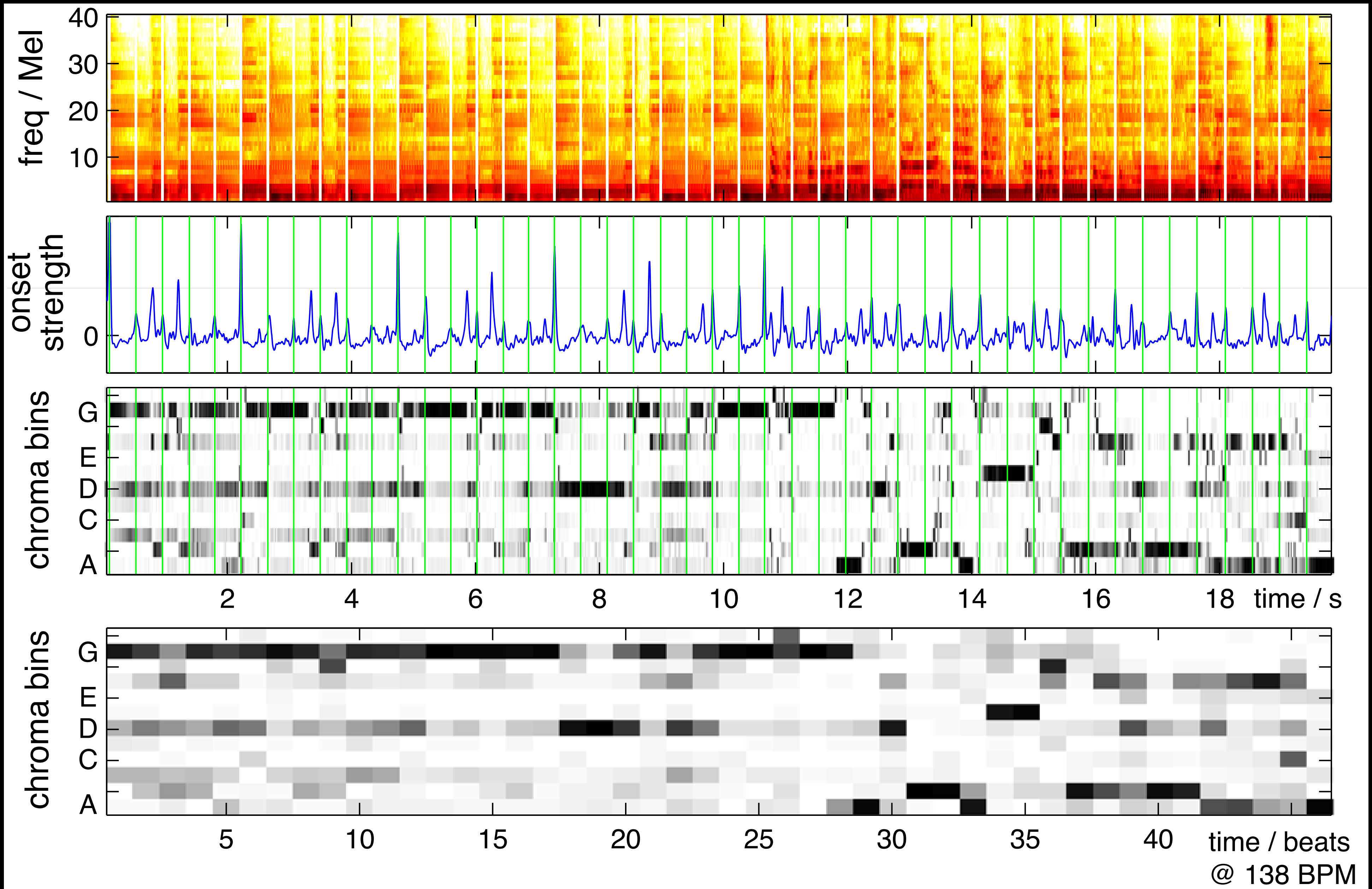
love babe, just like i need you, oh, hold me, love me



Evaluation techniques also borrowed from ASR

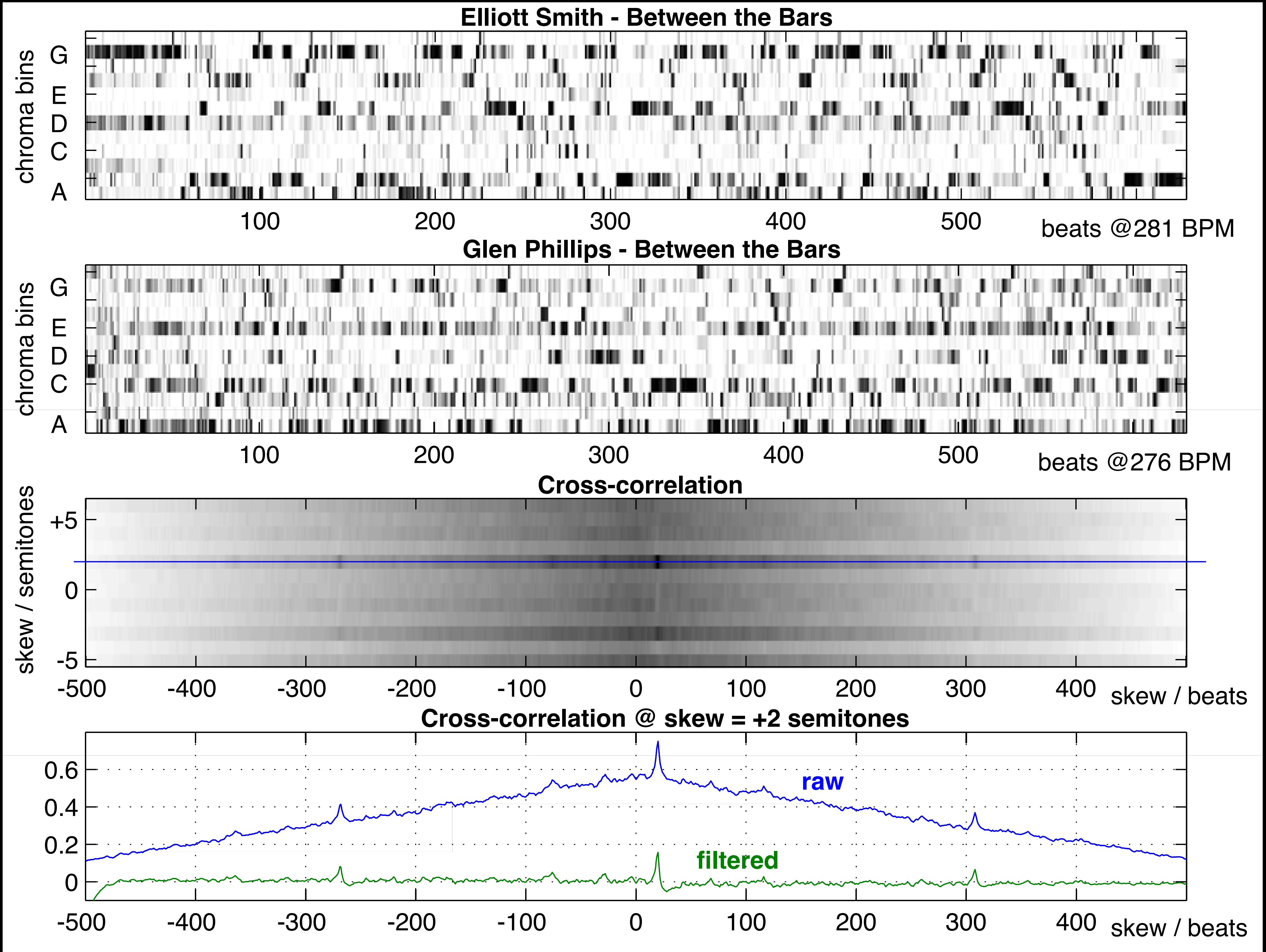
"E MAJOR" CONFUSION MATRIX Eight Days a Week							
	Maj	Min	Maj7	Min7	Dom7	Aug	Dim
E/Fb	158	115	.	9	.	.	.
E#/F	9	.	.	.	.	.	.
F#/Gb	.	.	.	11	.	.	.
G	3	.	.	.	.	.	.
G#/Ab	.	.	.	.	.	.	.
A	9	.	.	.	1	.	.
A#/Bb	.	.	.	.	.	.	.
B/Cb	8	.	.	.	,	.	.
B#/C	.	.	.	.	.	.	.
C#/Db	.	20	.	.	.	.	.
D	.	14	.	.	.	.	.
D#/Eb	.	.	.	.	.	.	.

Per-beat chroma vectors often work better



Elliot Smith, “Between the Bars”.

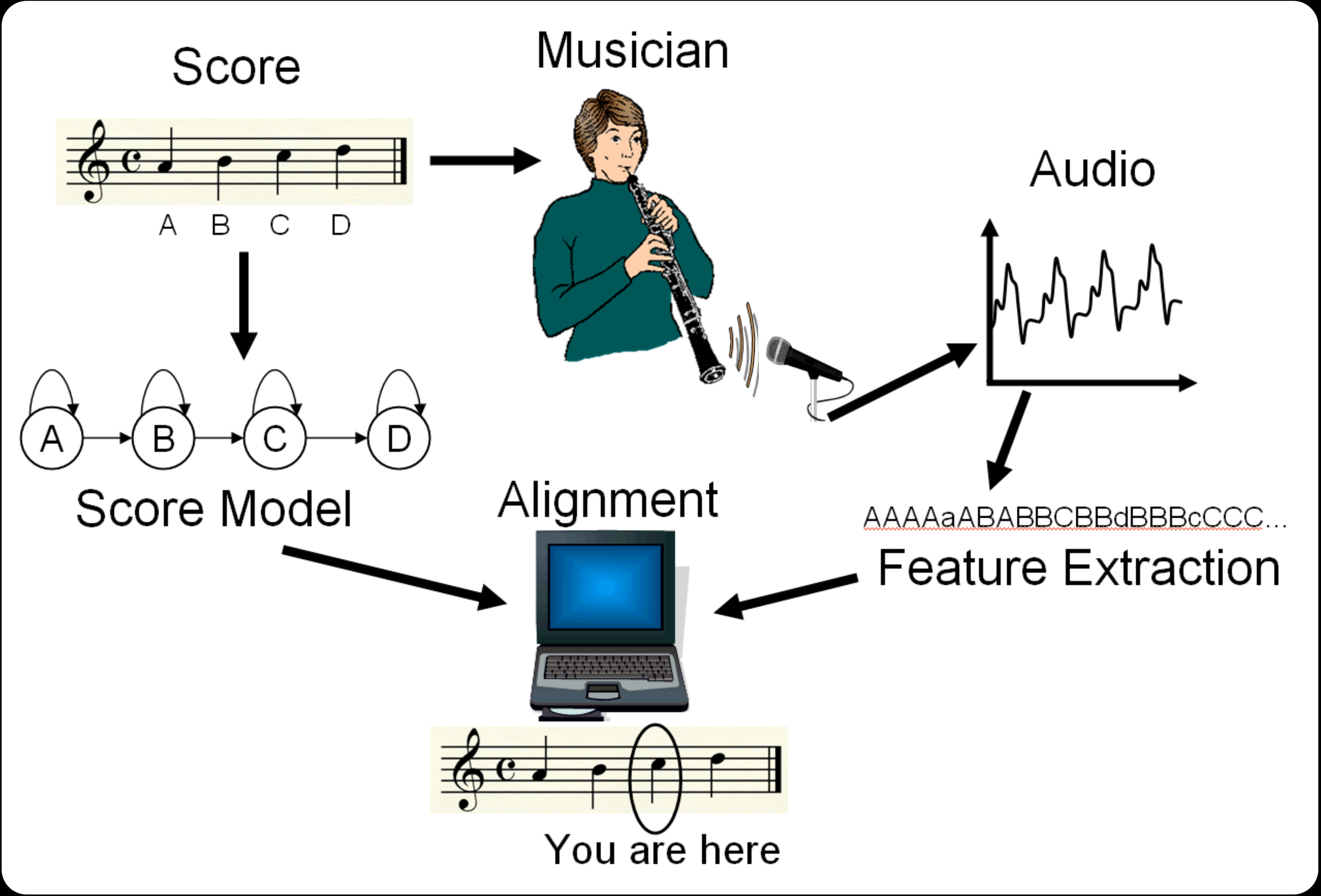
Cover songs: Discover tempo & transposition



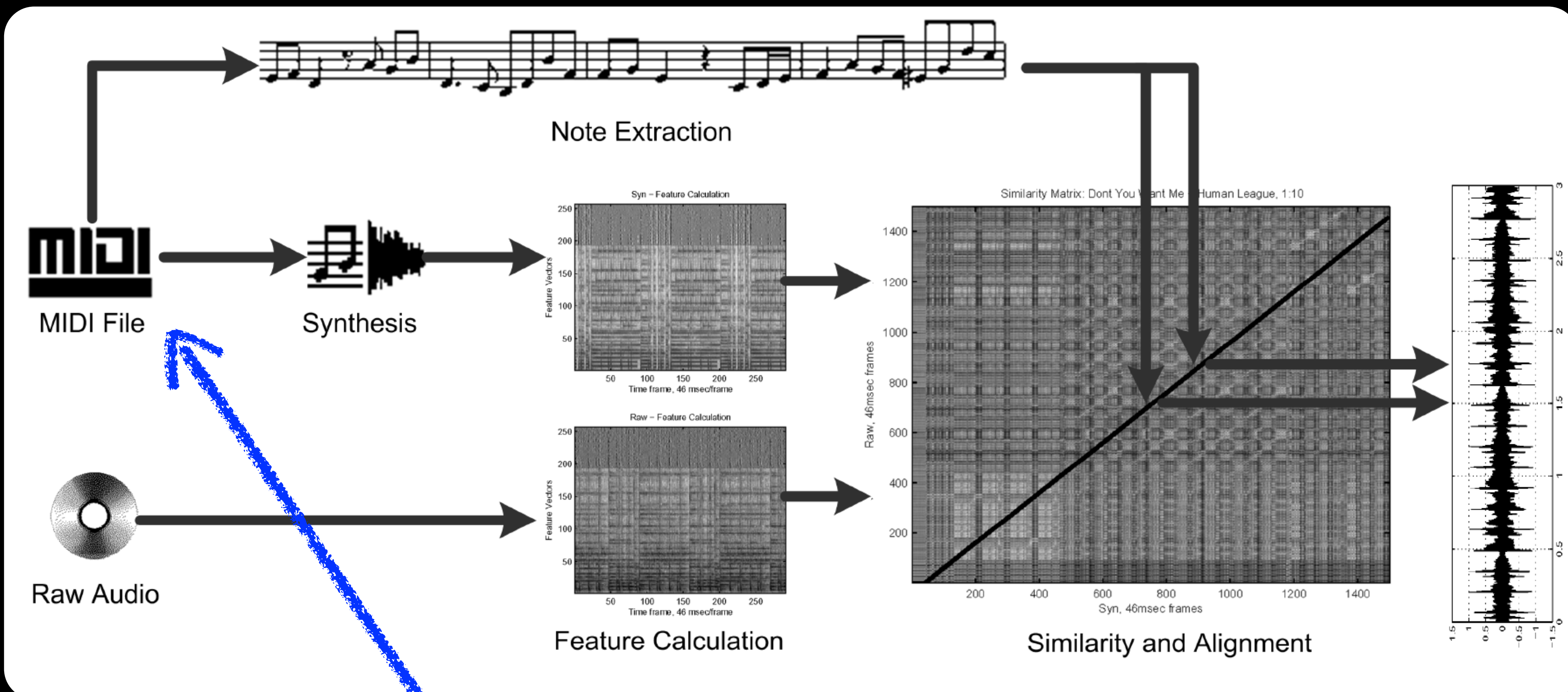
Music appreciation class for computers

- Discovering musical structure
- Chroma: Simple chord detection
- **Sheet-music score alignment**
- Music transcription

There is a "sub-symbolic" approach to this problem.



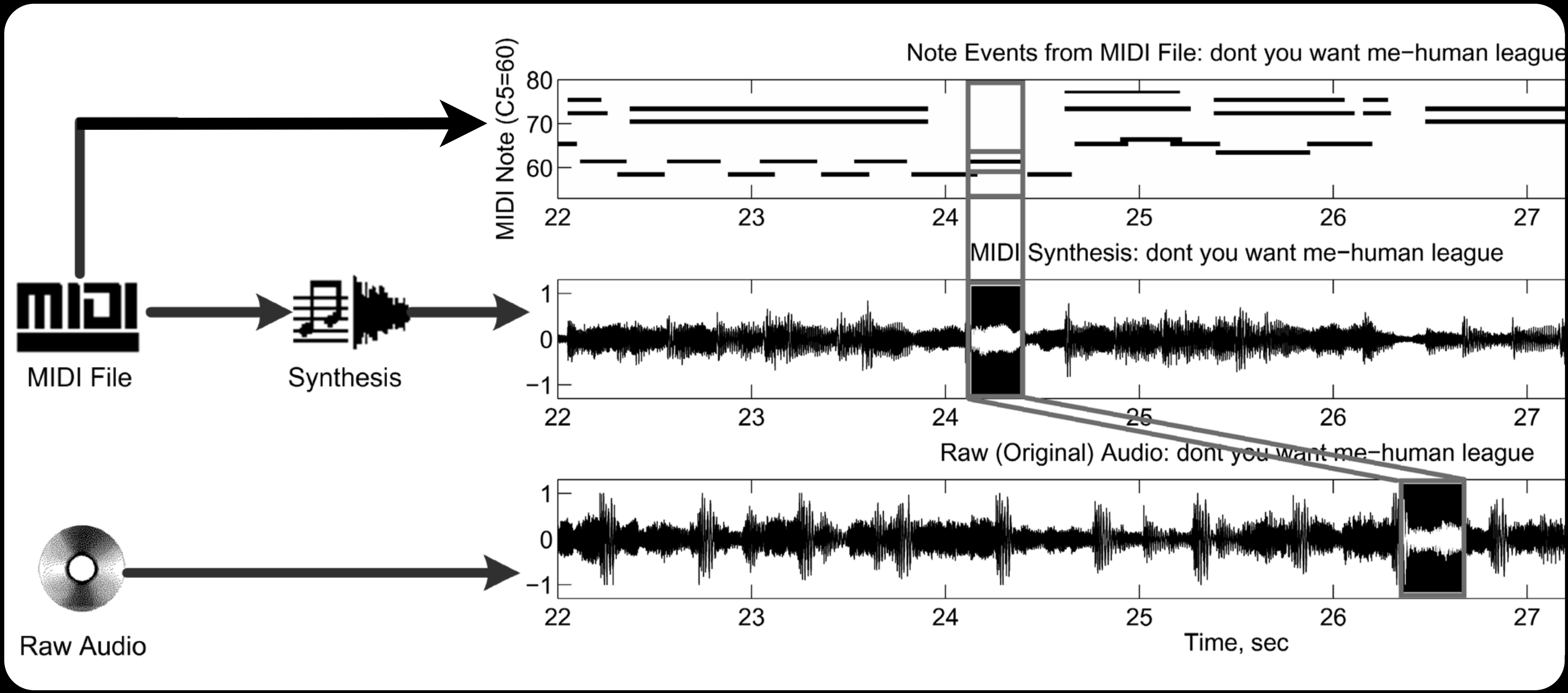
No "transcription" necessary: no need to extract "symbolic" notes from audio.



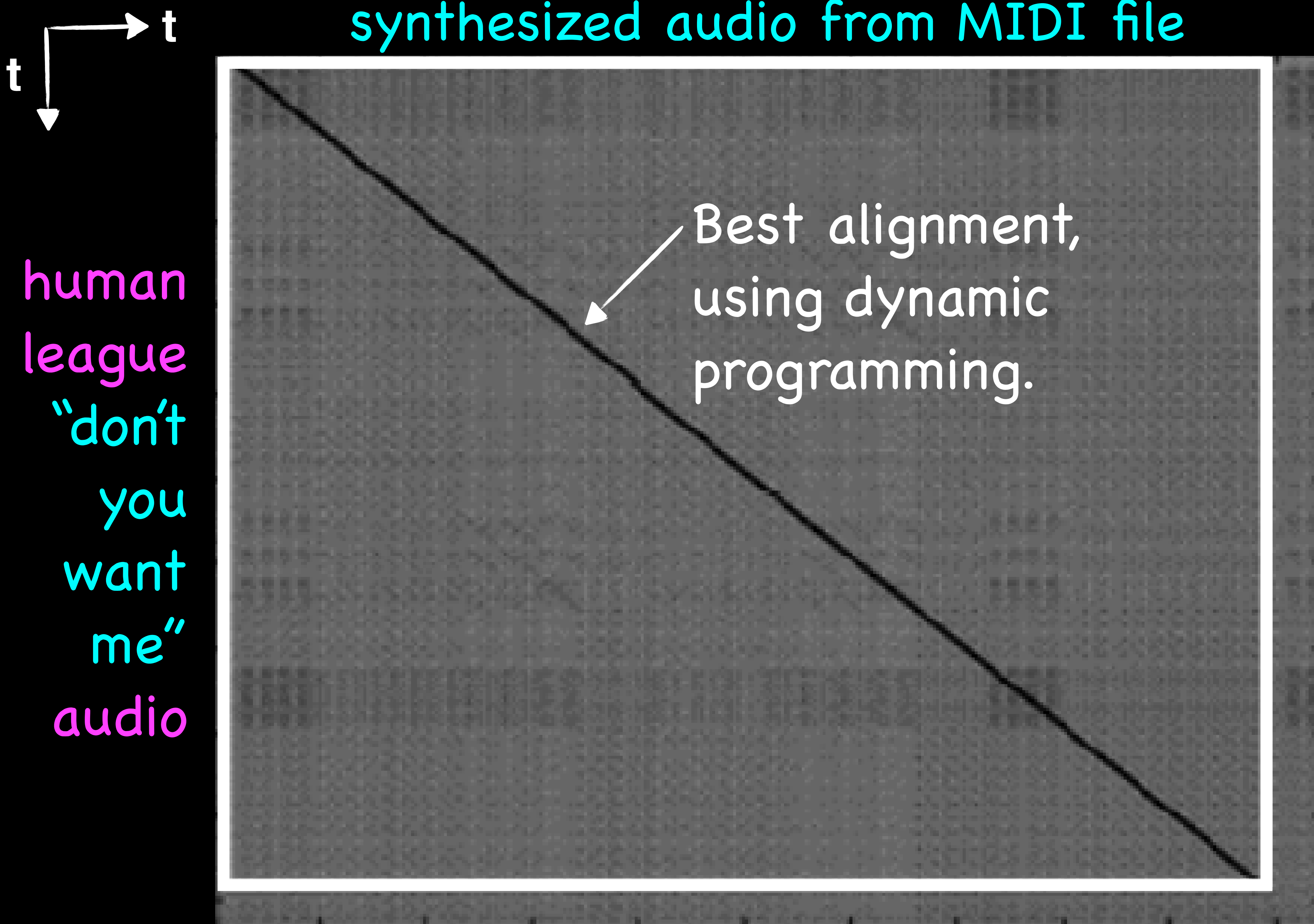
... a "cover song" MIDI file works fine ...

"I was working as a waitress in a cocktail bar"

Notes, synthesized audio, and the record ...

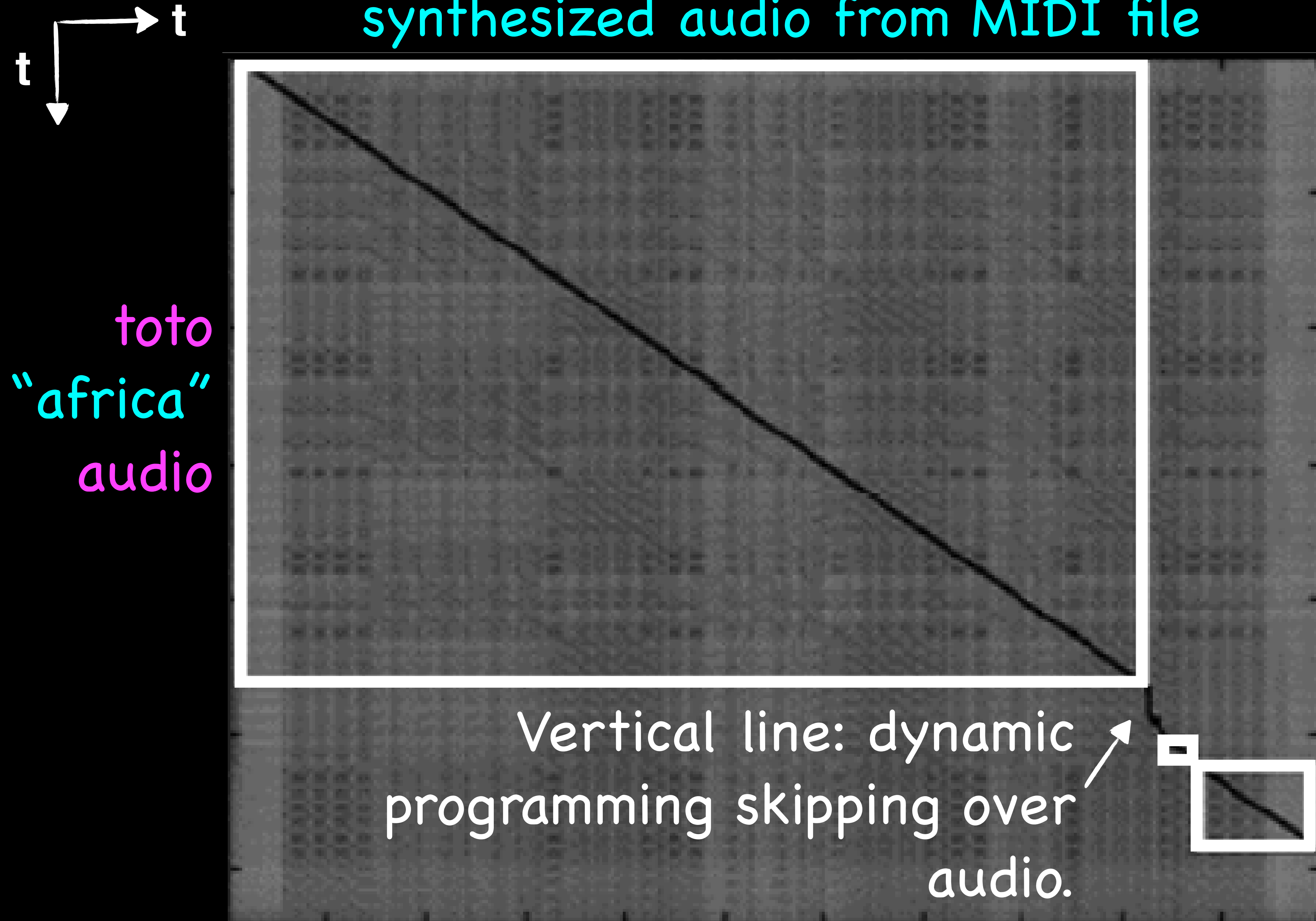


The "alignment" task



Alignment lets us map audio time to score position.

synthesized audio from MIDI file



Alignment recovers from missing MIDI file section.

Audio on a movie set ...

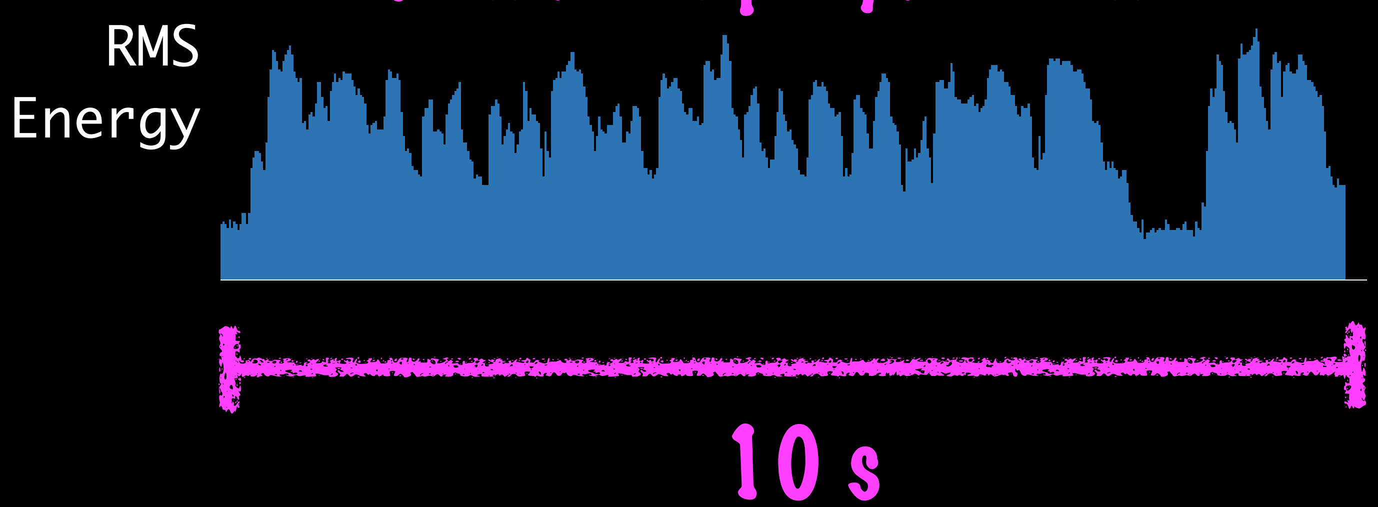
Audio quality may leave something to be desired ...

Redo audio in the studio

Problem: Re-recorded audio must **synchronize** tightly to visuals (lip-sync, footsteps, etc).

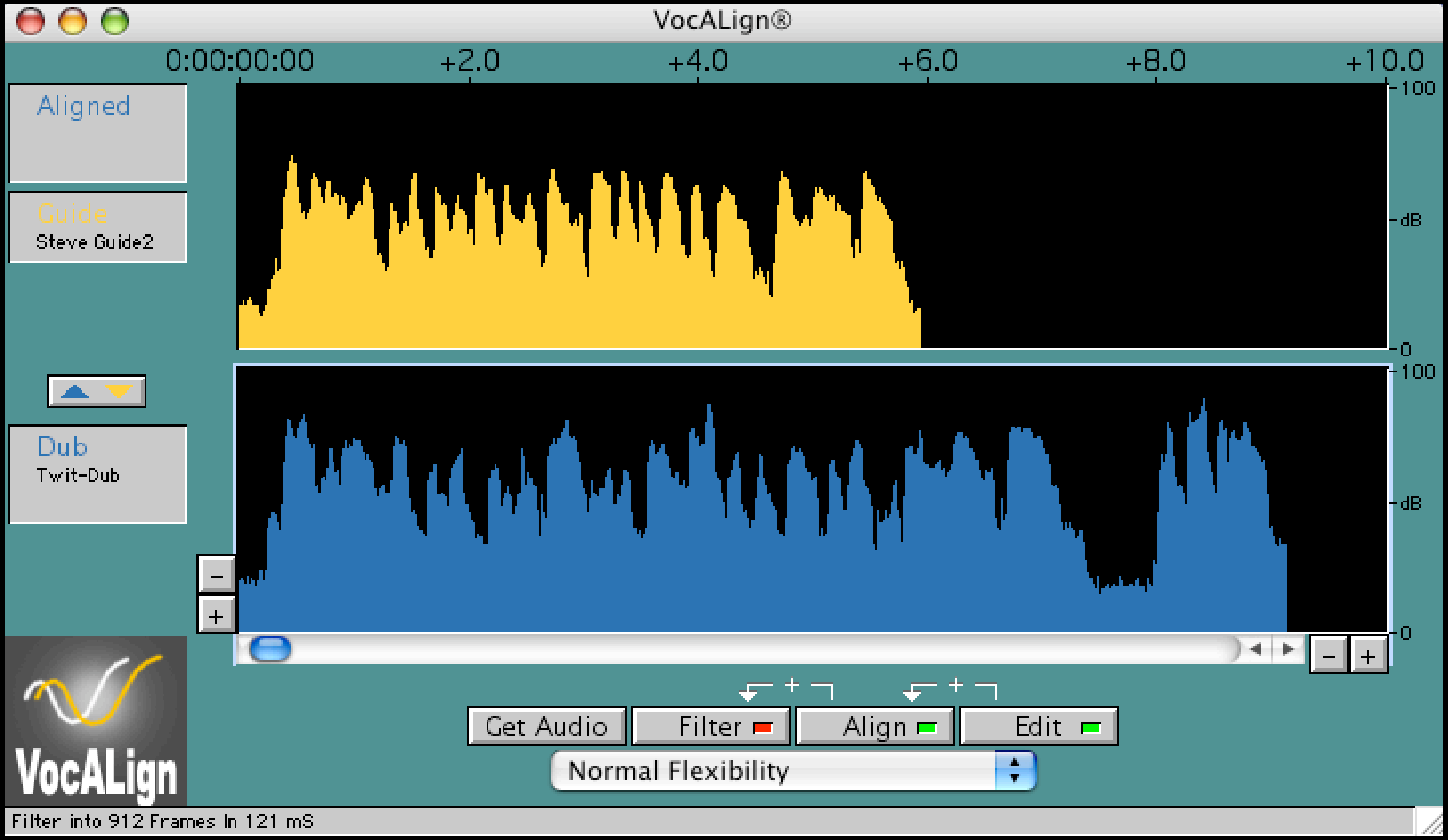
Line spoken on the set.

Will not lip-sync well.



Line re-recorded in the studio.

VocAlign: A plug-in that automatically aligns "dub" audio with "guide" audio.

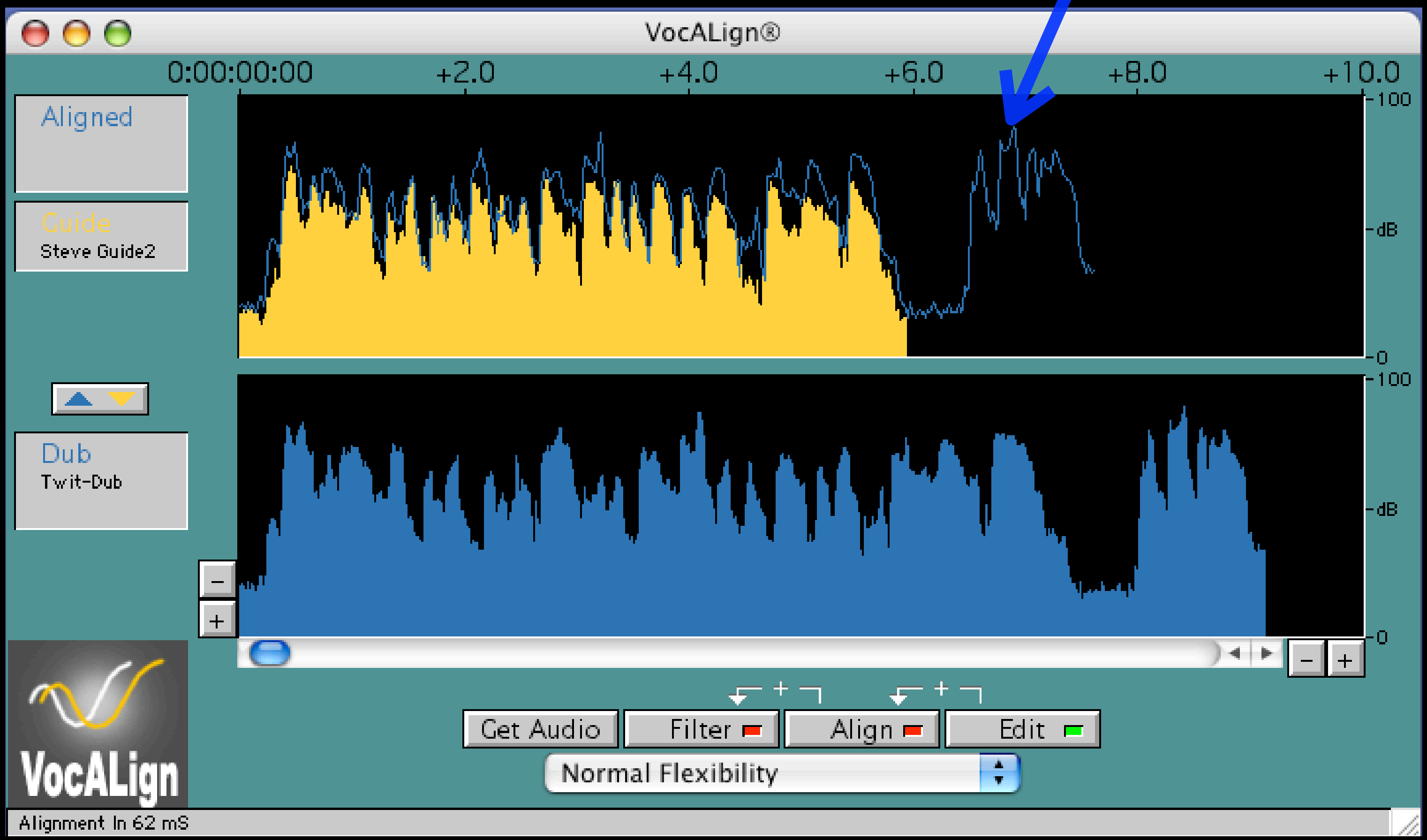


"guide"

"dub"

Setup: User selects segments of dub and guide audio tracks for alignment.

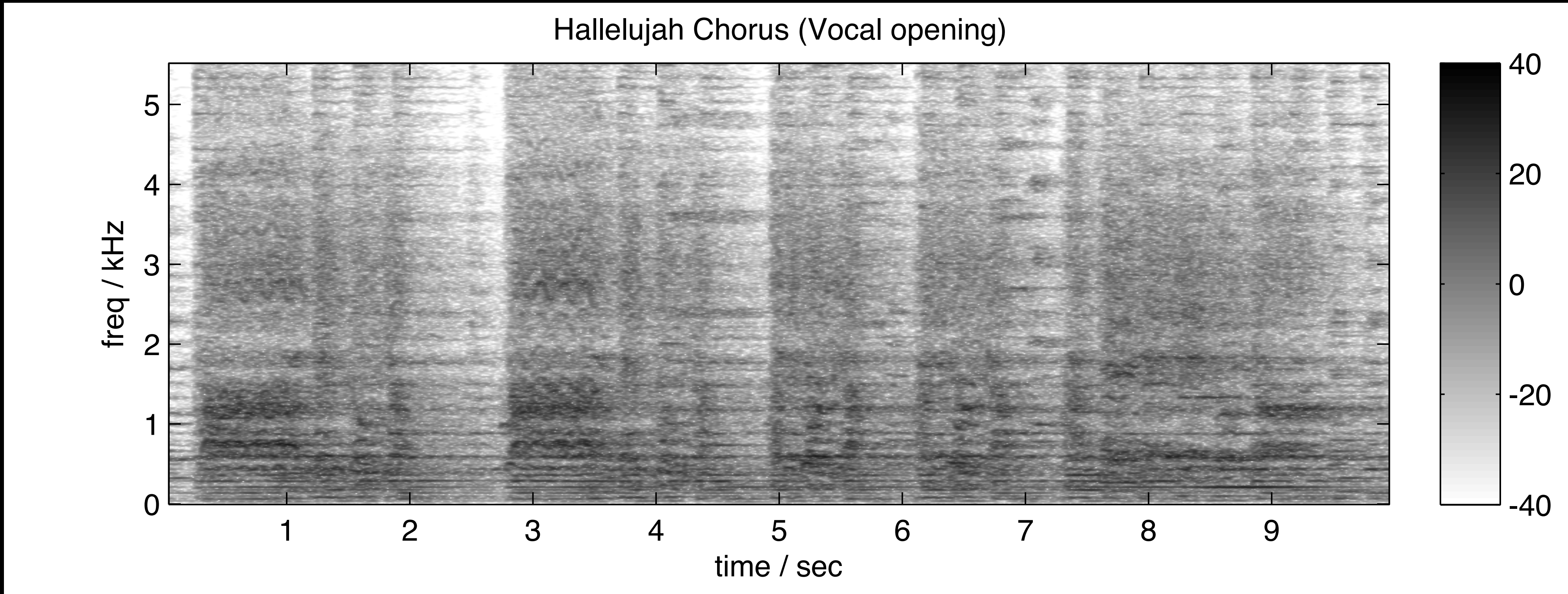
Result: Blue line shows envelope of aligned dub audio (user can also listen).

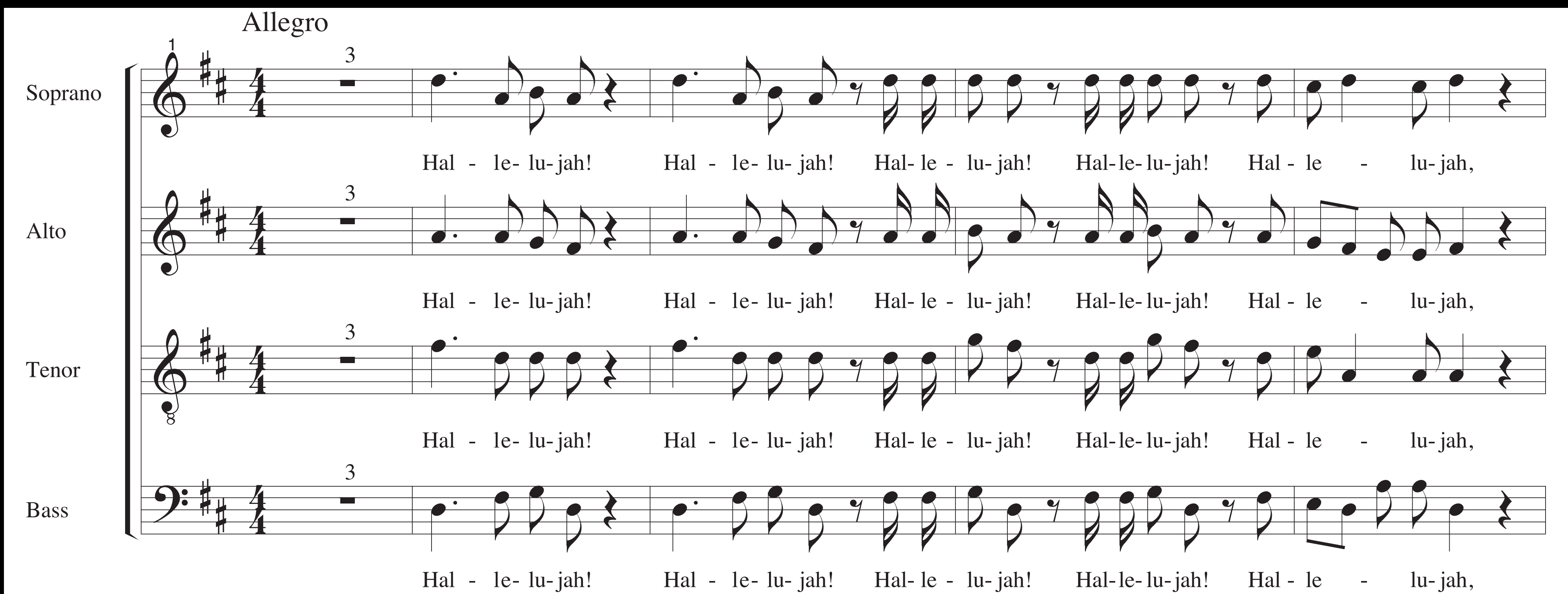


Fine-tuning: User can choose different algorithms to improve fit, then "print" best one.

Music appreciation class for computers

- Discovering musical structure
- Chroma: Simple chord detection
- Sheet-music score alignment
- **Music transcription**

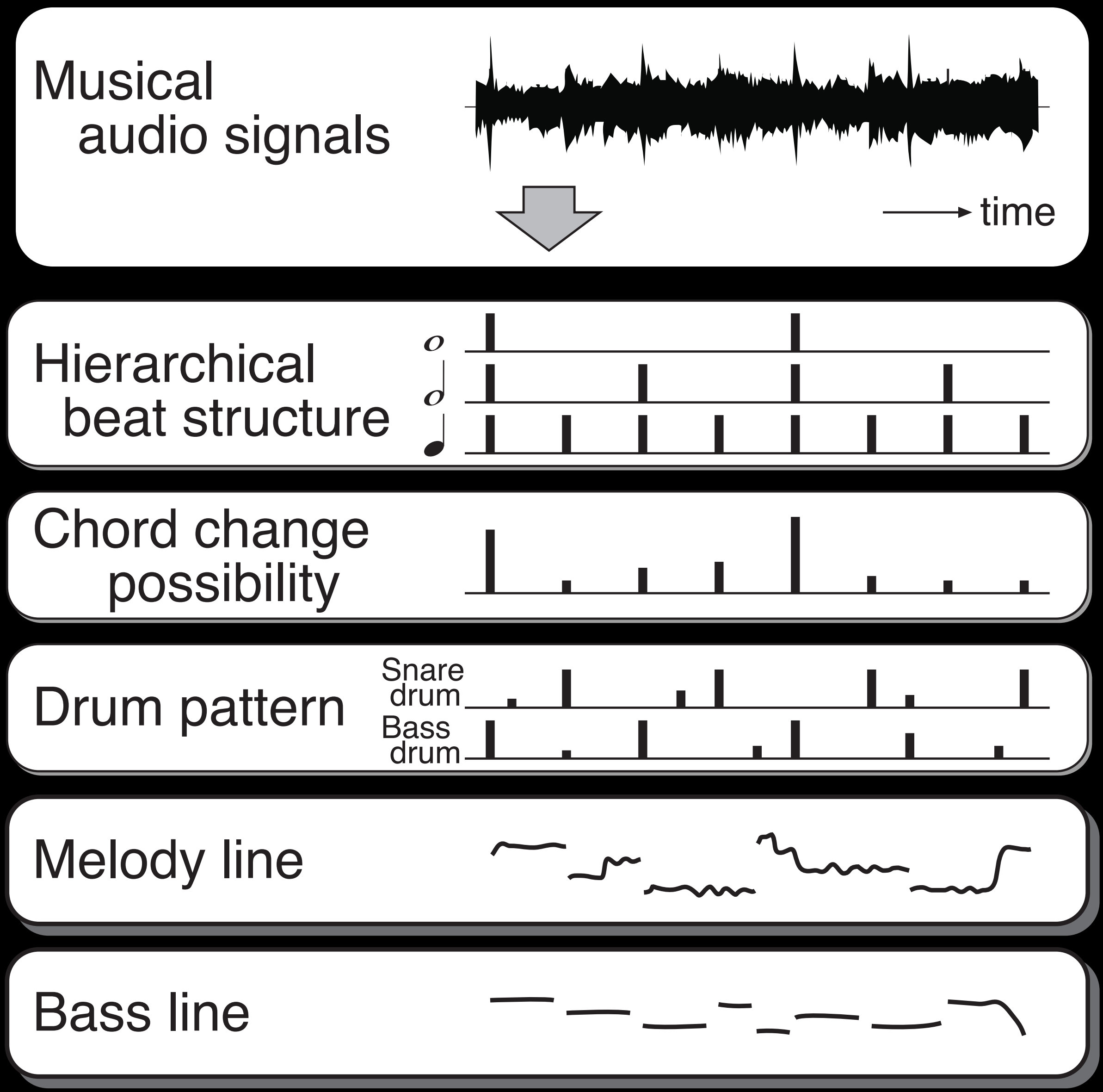

Hallelujah Chorus (Vocal opening)
freq / kHz
0
1
2
3
4
5
time / sec
1
2
3
4
5
6
7
8
9
40
20
0
-20
-40


Allegro
Soprano
Alto
Tenor
Bass
Hal - le- lu- jah! Hal - le- lu- jah! Hal- le - lu- jah! Hal-le-lu-jah! Hal - le - lu- jah,
Hal - le- lu- jah! Hal - le- lu- jah! Hal- le - lu- jah! Hal-le-lu-jah! Hal - le - lu- jah,
Hal - le- lu- jah! Hal - le- lu- jah! Hal- le - lu- jah! Hal-le-lu-jah! Hal - le - lu- jah,
Hal - le- lu- jah! Hal - le- lu- jah! Hal- le - lu- jah! Hal-le-lu-jah! Hal - le - lu- jah,

Masataka Goto's transcription system

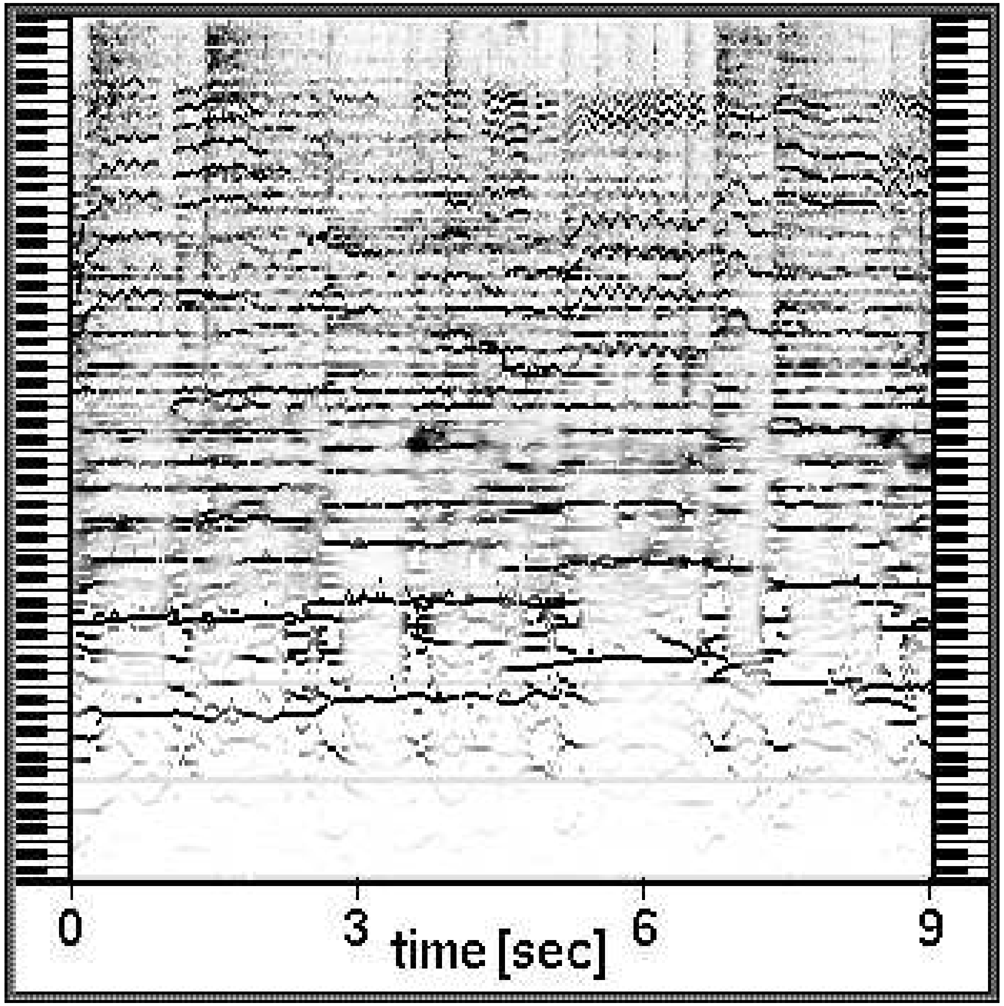
Similar to "similarity" & "chroma" techniques

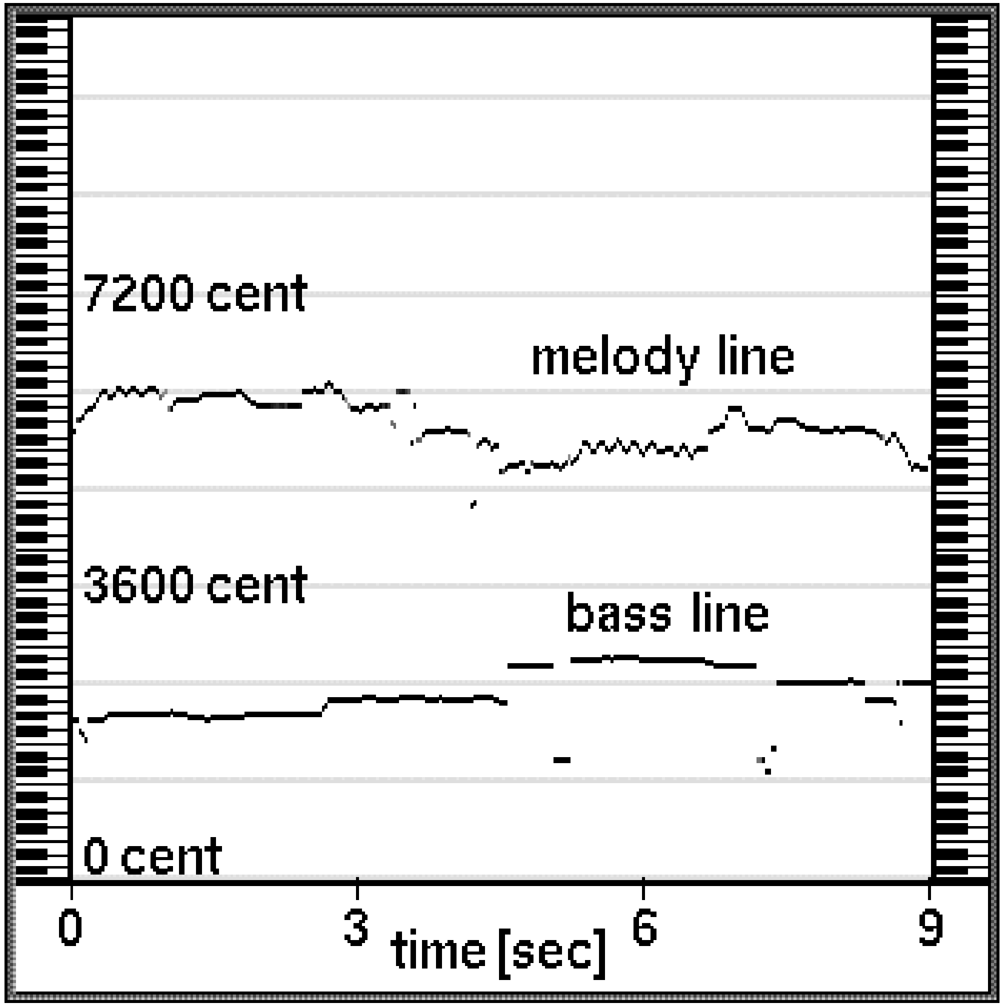
Monophonic transcription



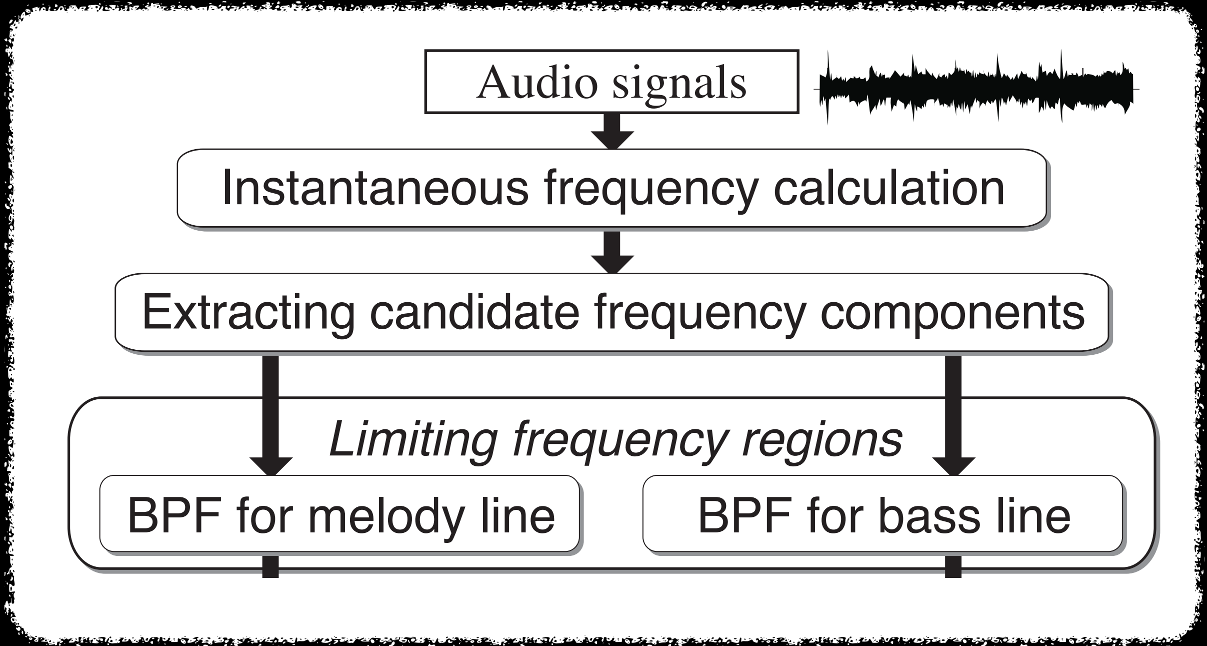
Screenshot of the system in action

Harmonic partials → Pitch tracks

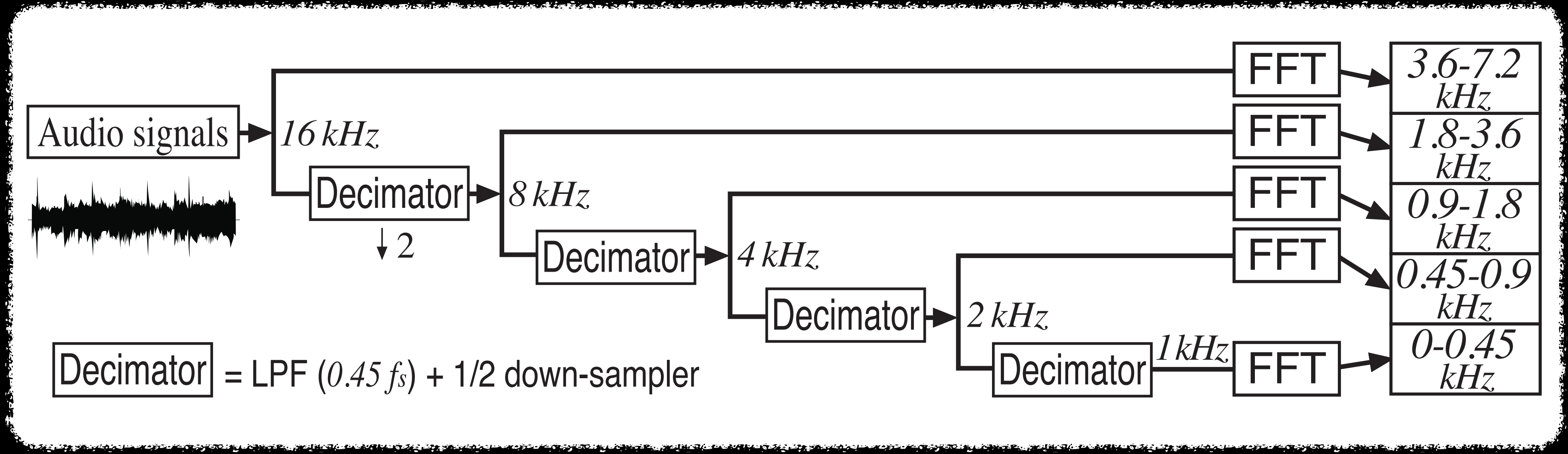




The front end

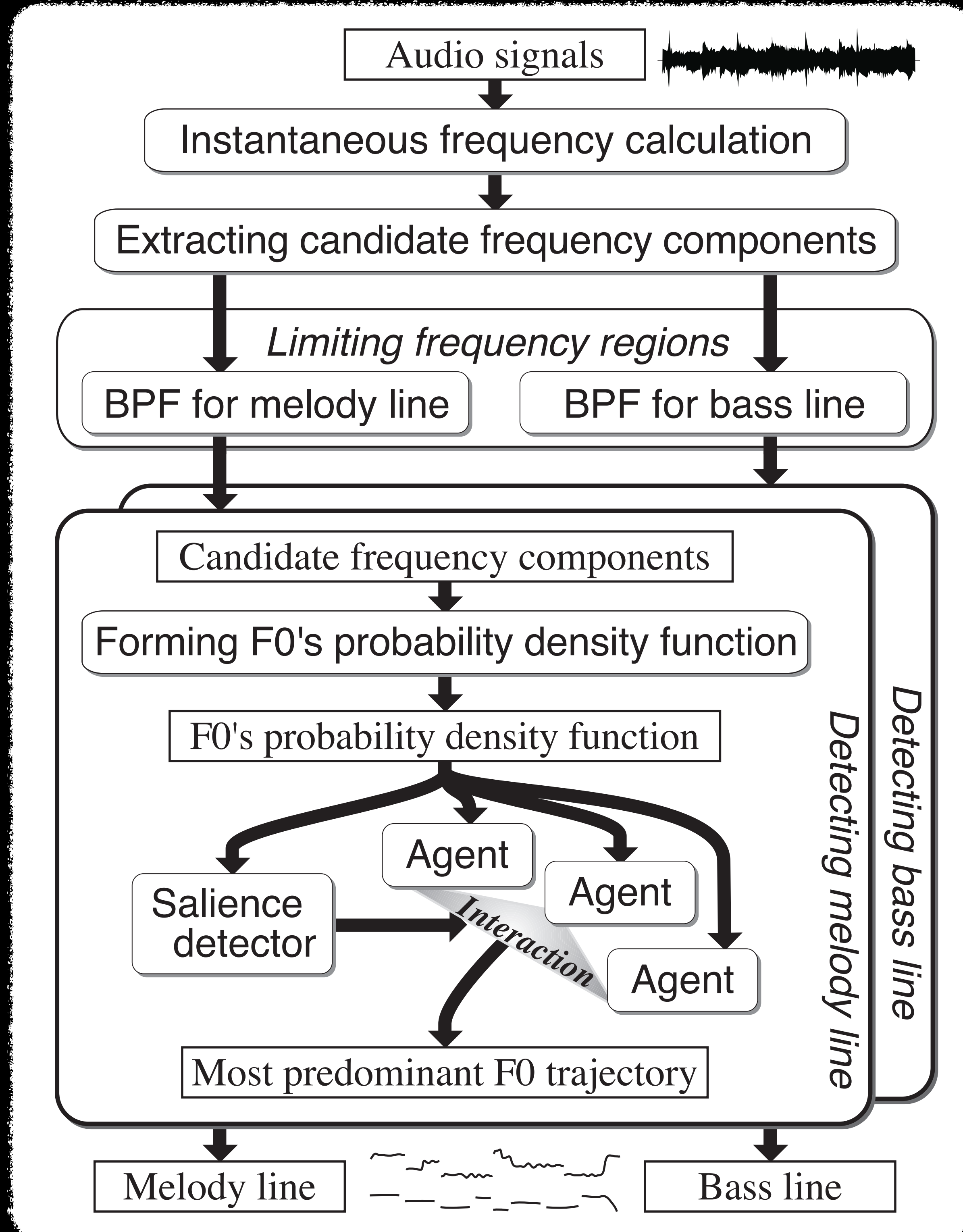


Filterbank



"Agent" system

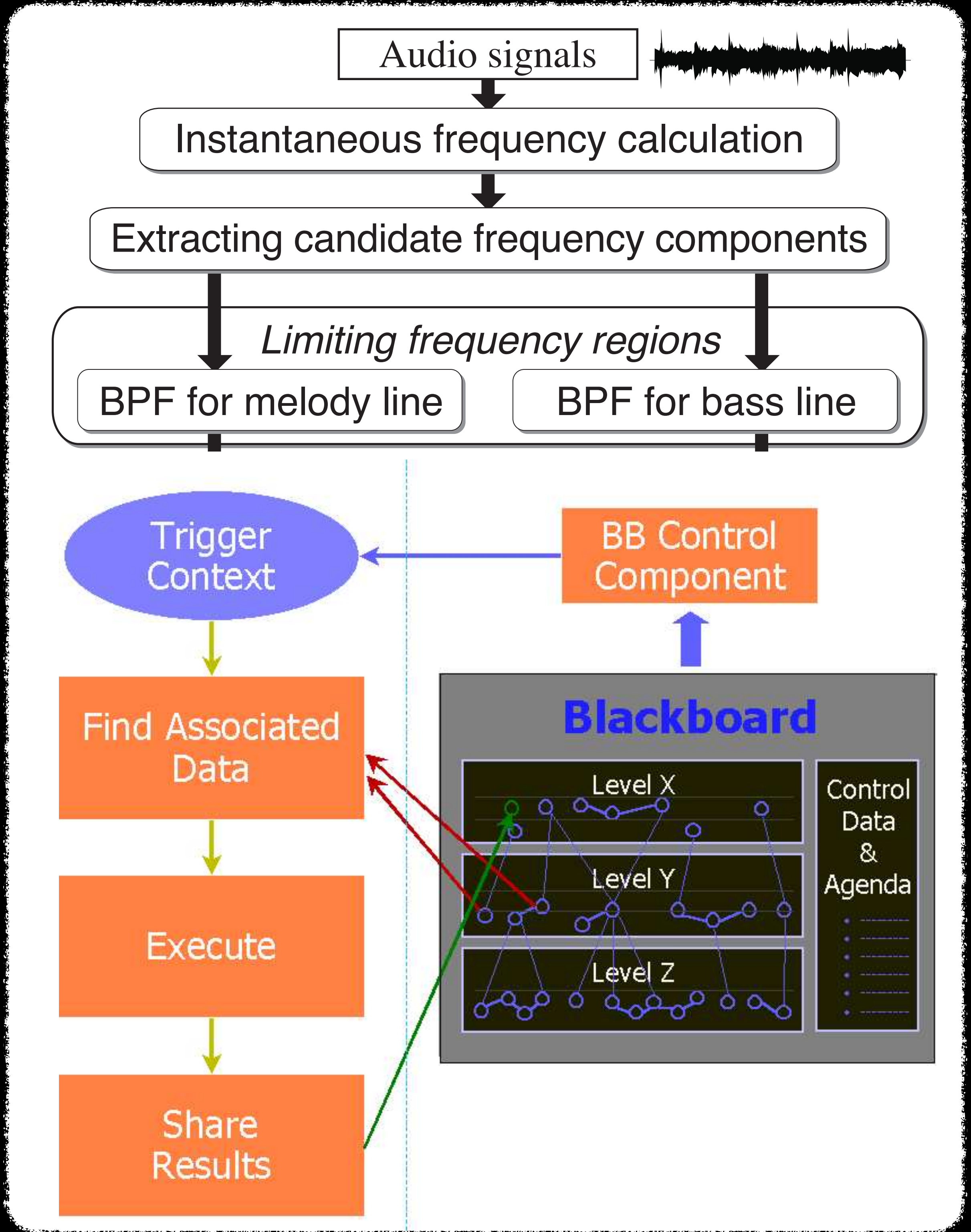
Agent system decides the temporal extant and spectral makeup of each note in a melody line.



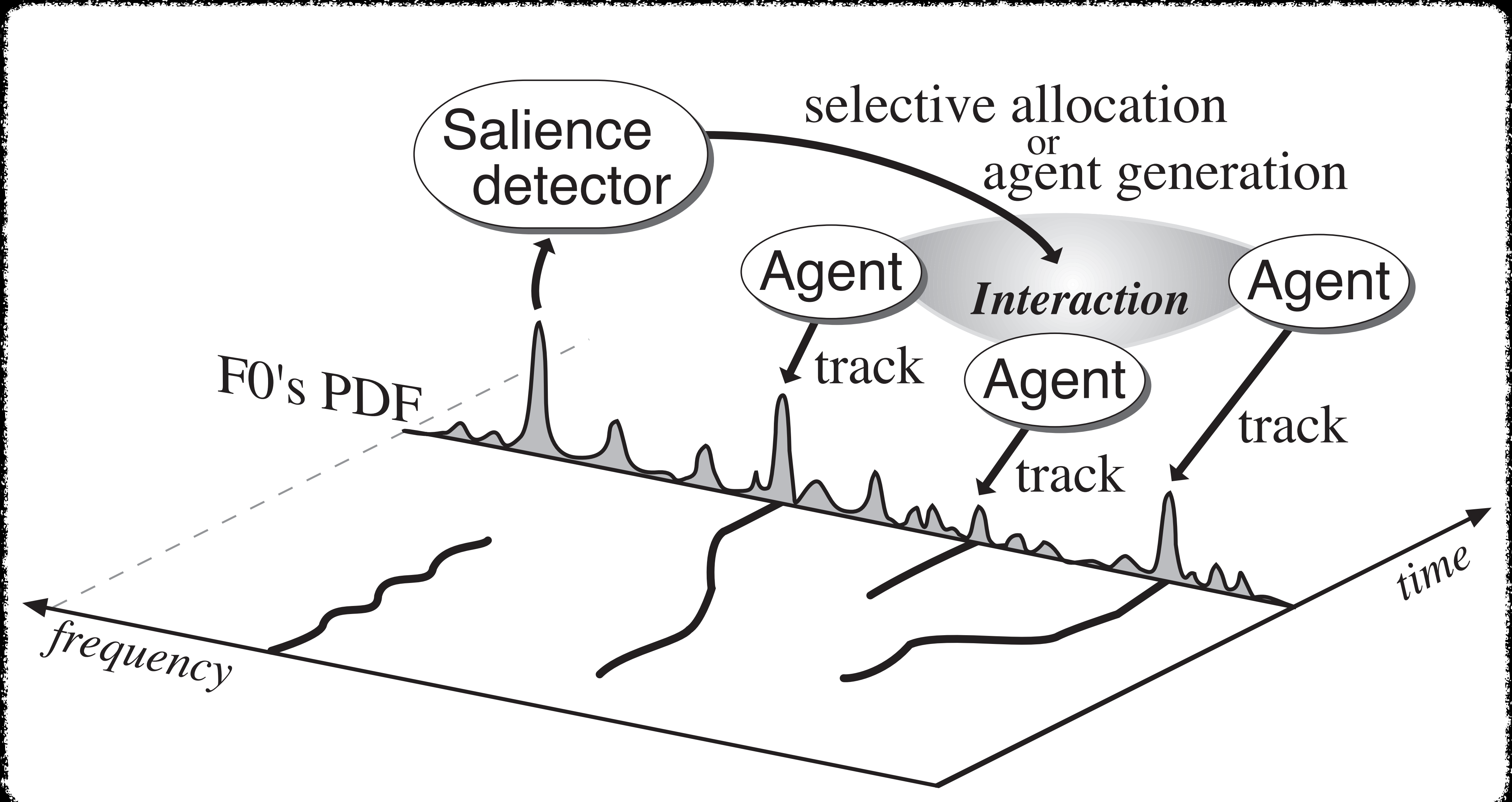
"Agent" systems

Model: Humans collaborating on solving a math program by taking turns reading, writing, and erasing equations on a black board.

Agents are small programs that play the "human" role.

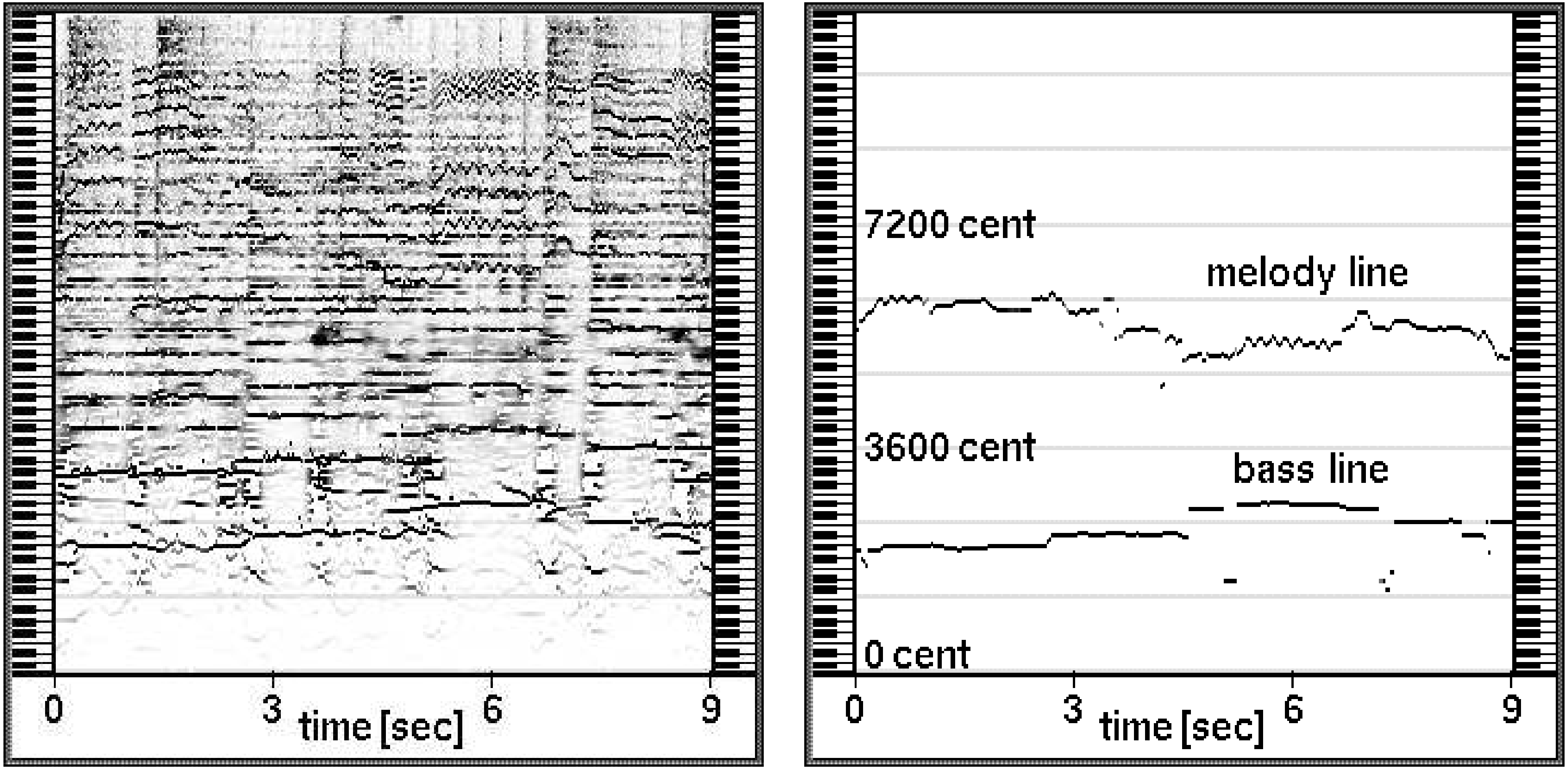


"Agents" in action

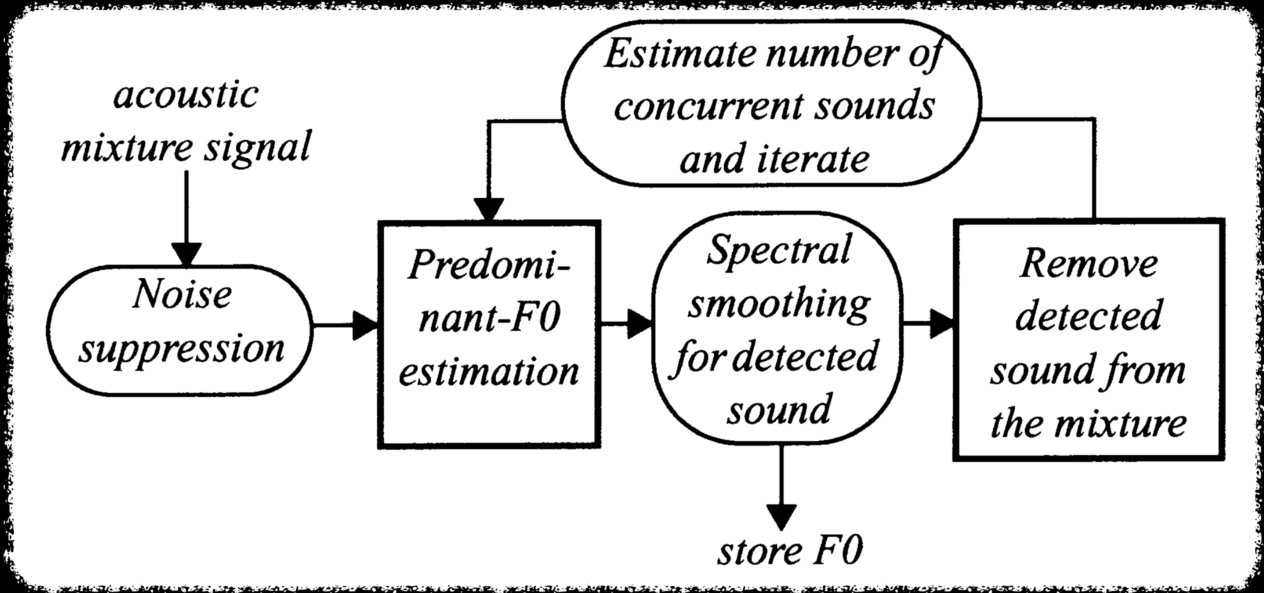


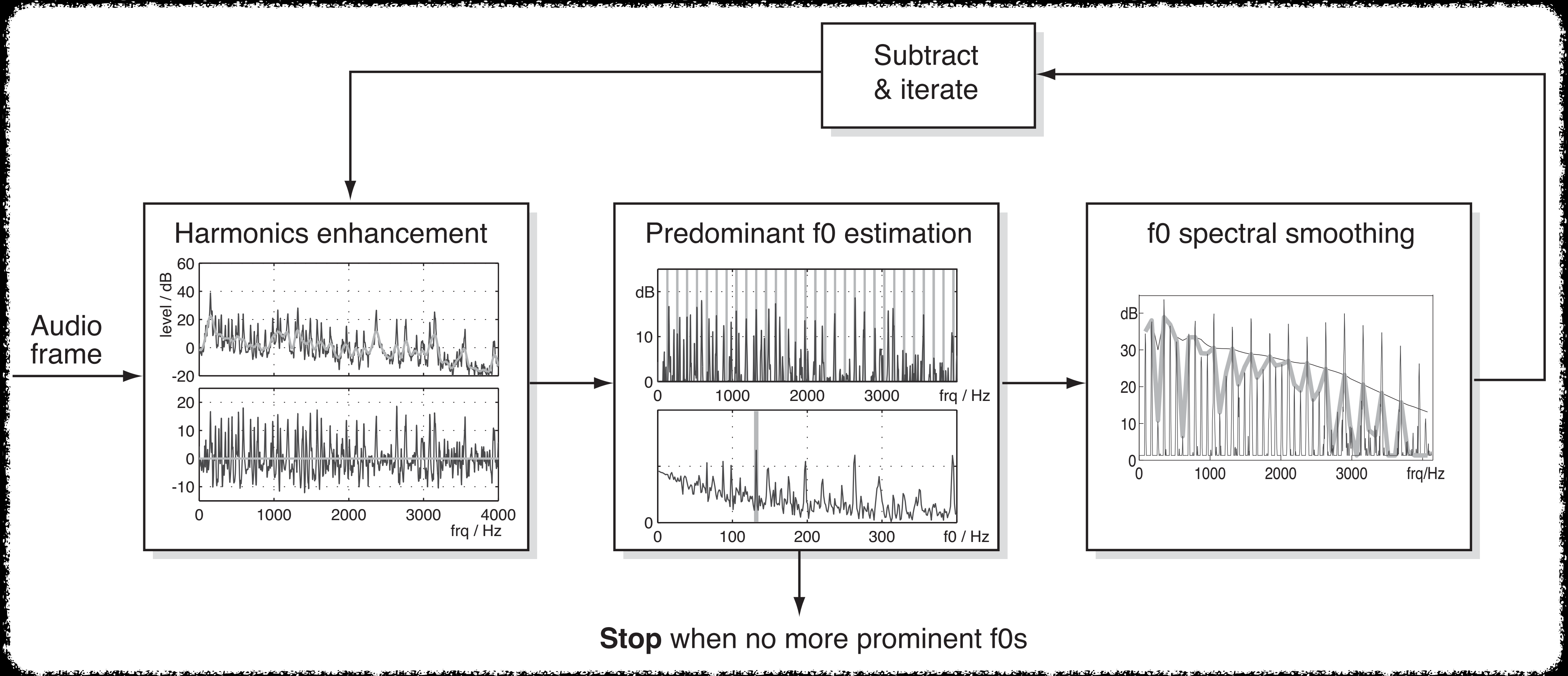
Screenshot of the system in action

Harmonic partials → Pitch tracks



Klapuri's subtractive system



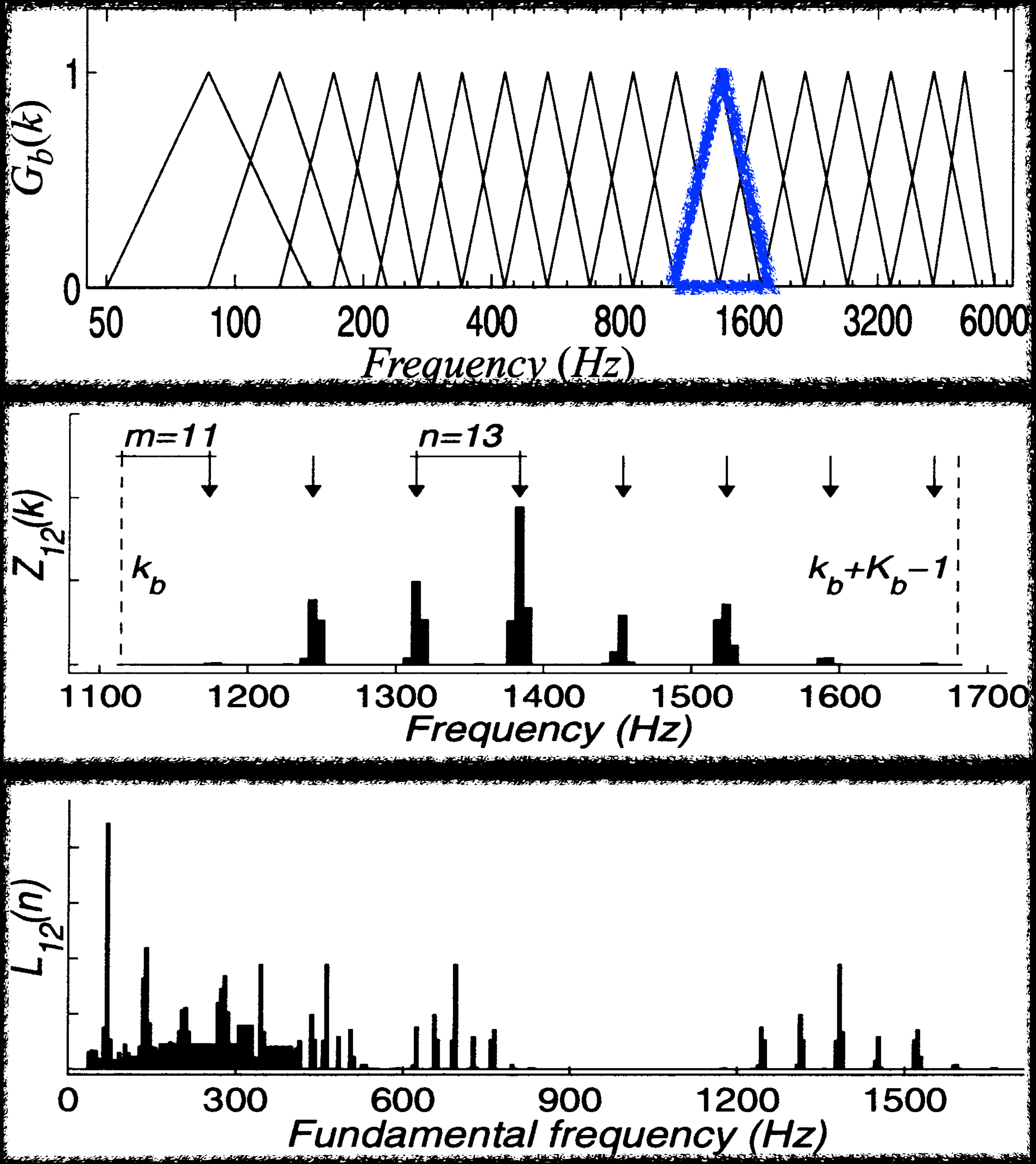


Estimate most likely F0 for each band

Band 12:
1.1–1.6 kHz

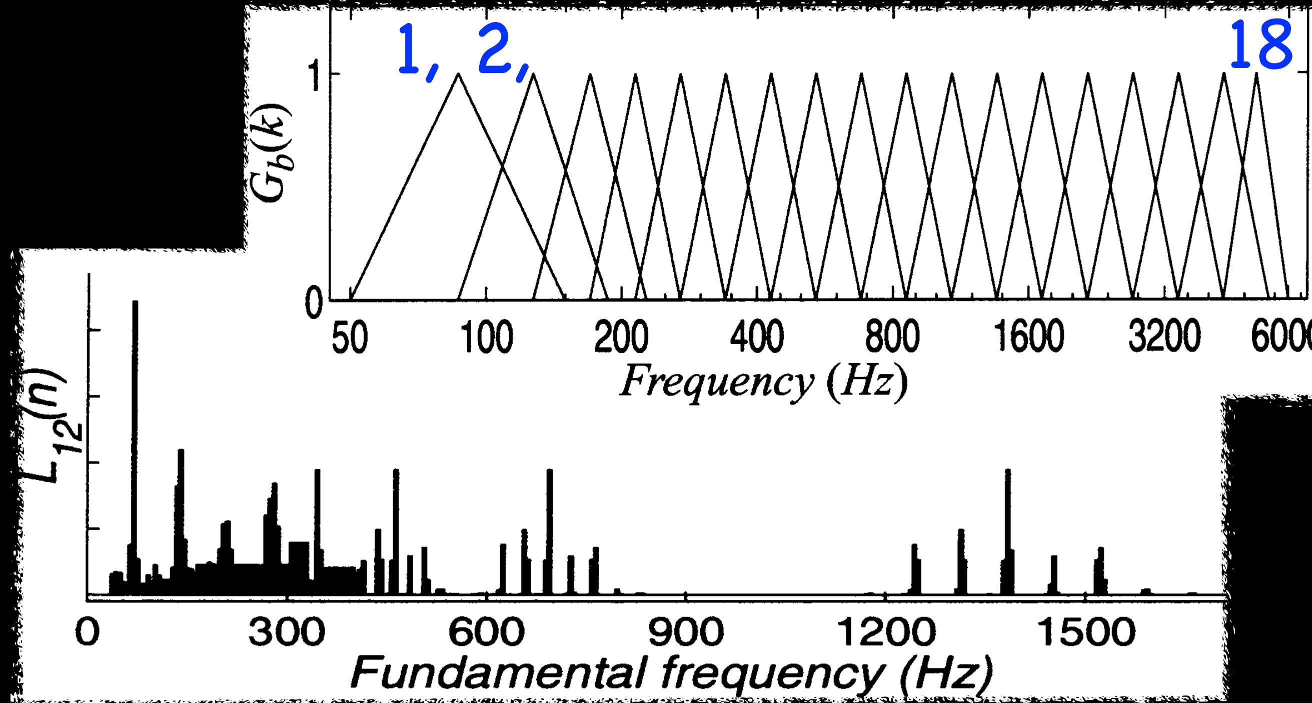
Energy peaks within Band 12.

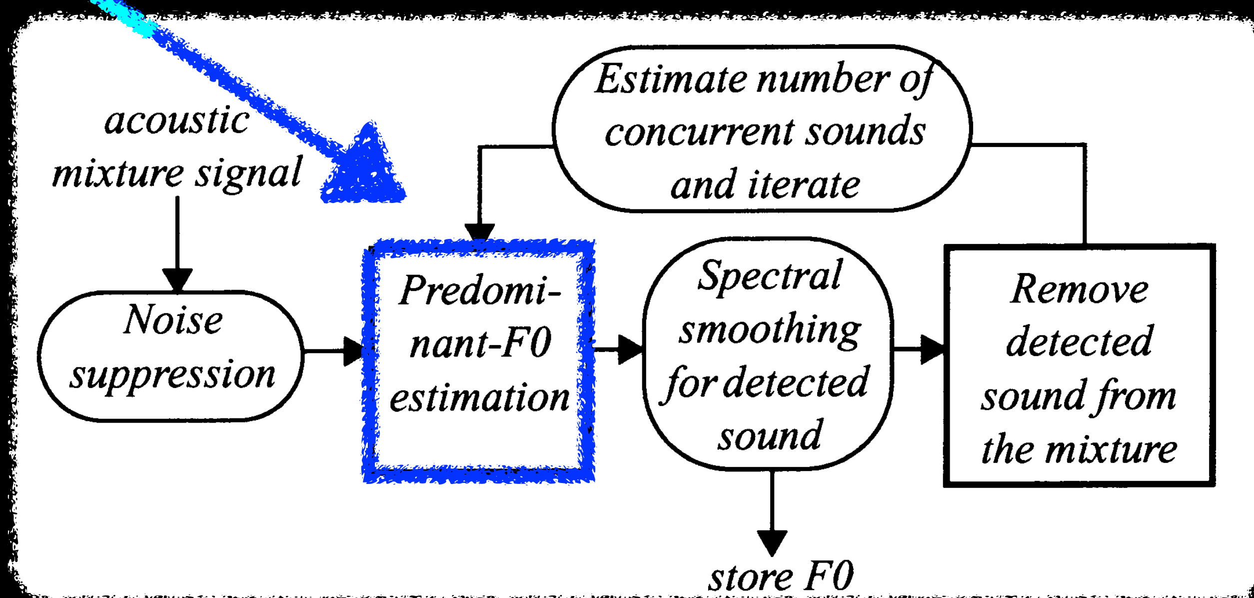
Weights $L_{12}(n)$. Horizontal "n" axis plots F0s.



n

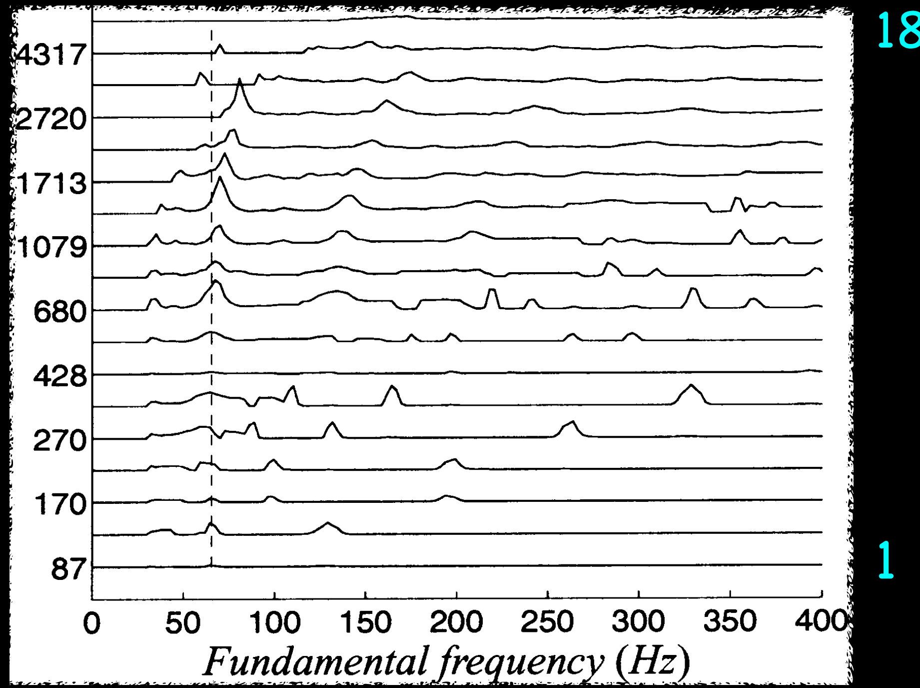
Use $L_1(n)$ … $L_{18}(n)$ to compute a global F0 value.



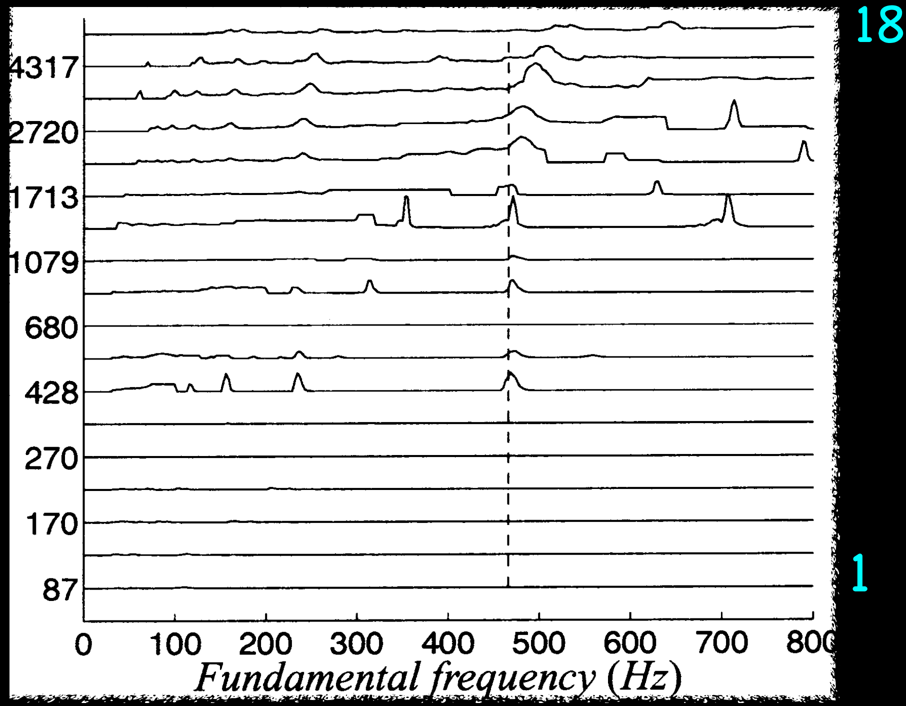


$L_1(n)$ ... $L_{18}(n)$ for two piano notes

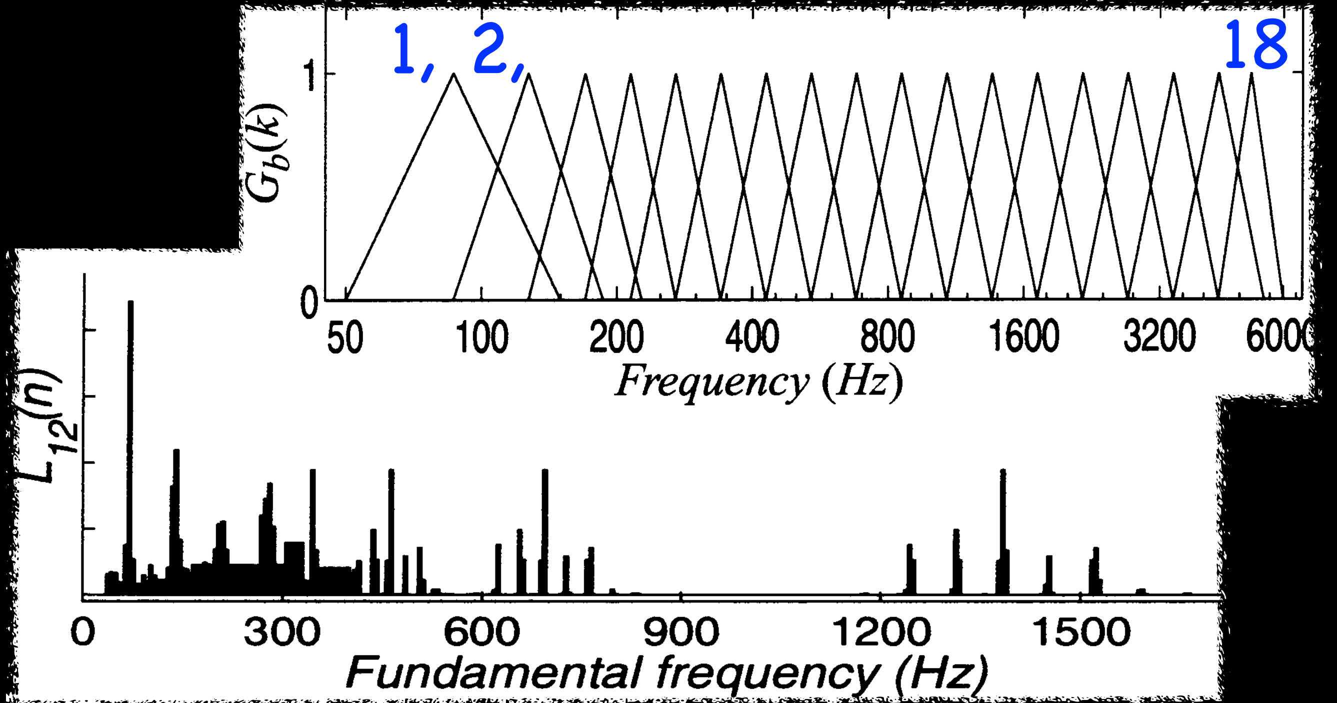
F0 = 65 Hz

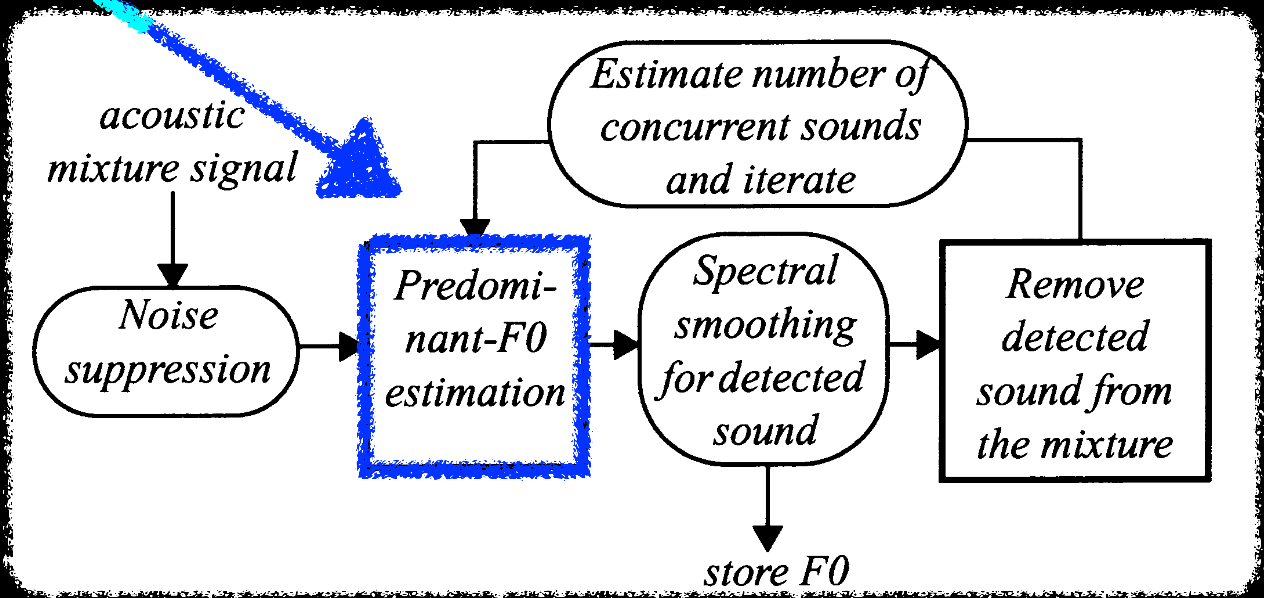


F0 = 470 Hz



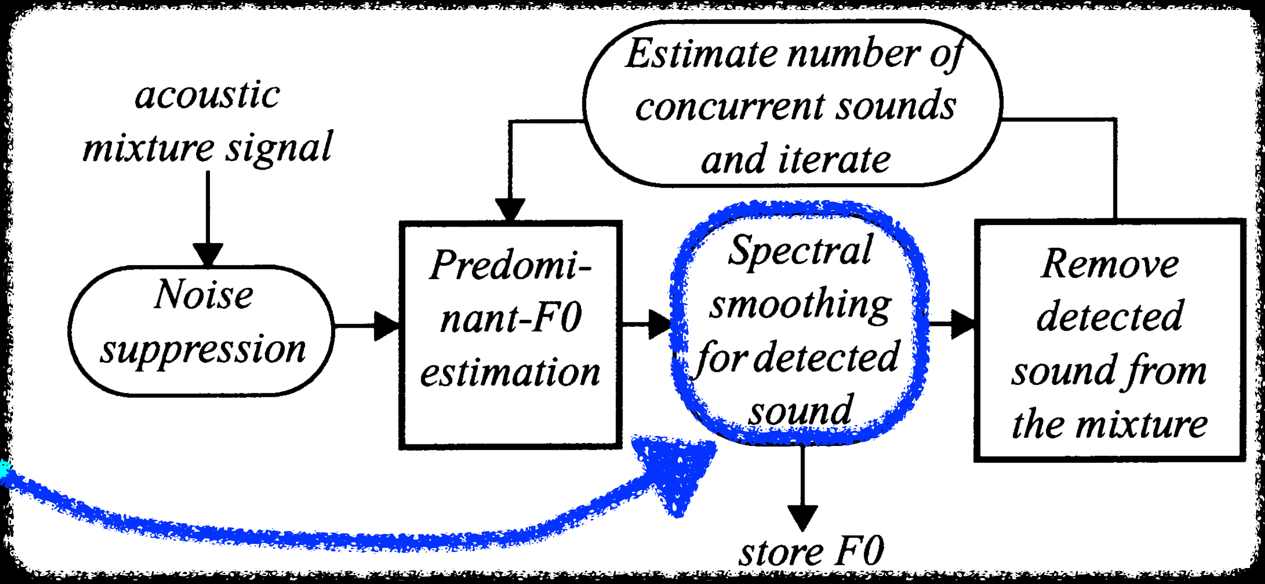
Use $L_1(n)$ … $L_{18}(n)$ to compute a global F0 value.



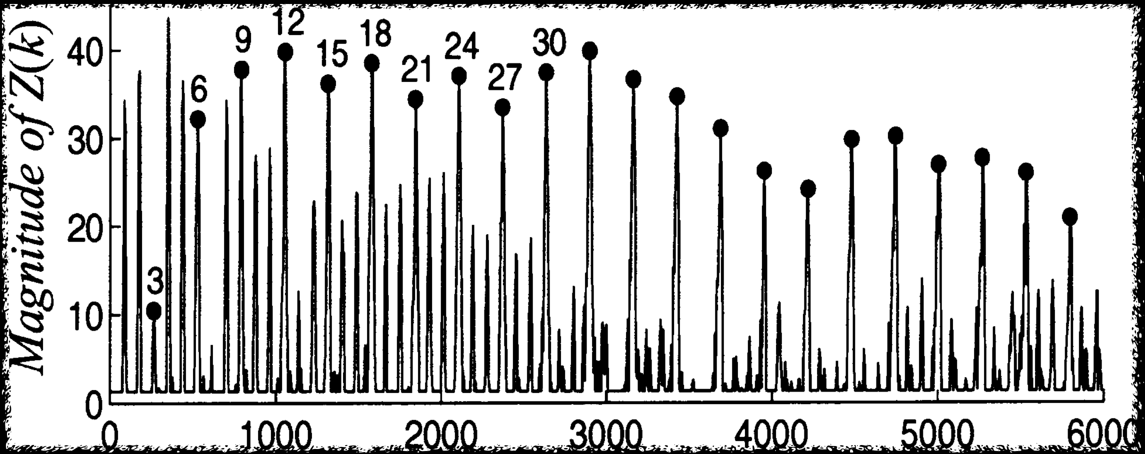


And then …

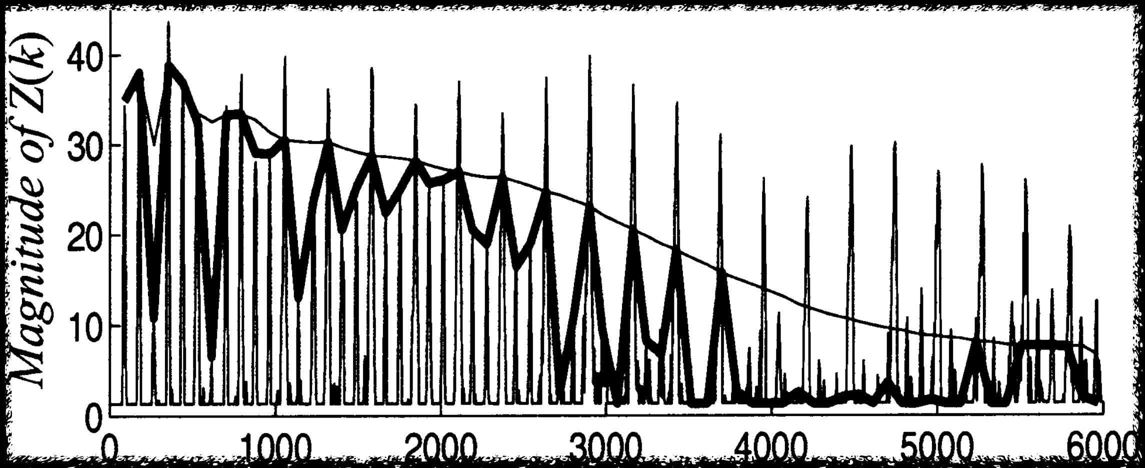
Make a smooth version of the note, and remove.



Black dots are partials of the note for F0.



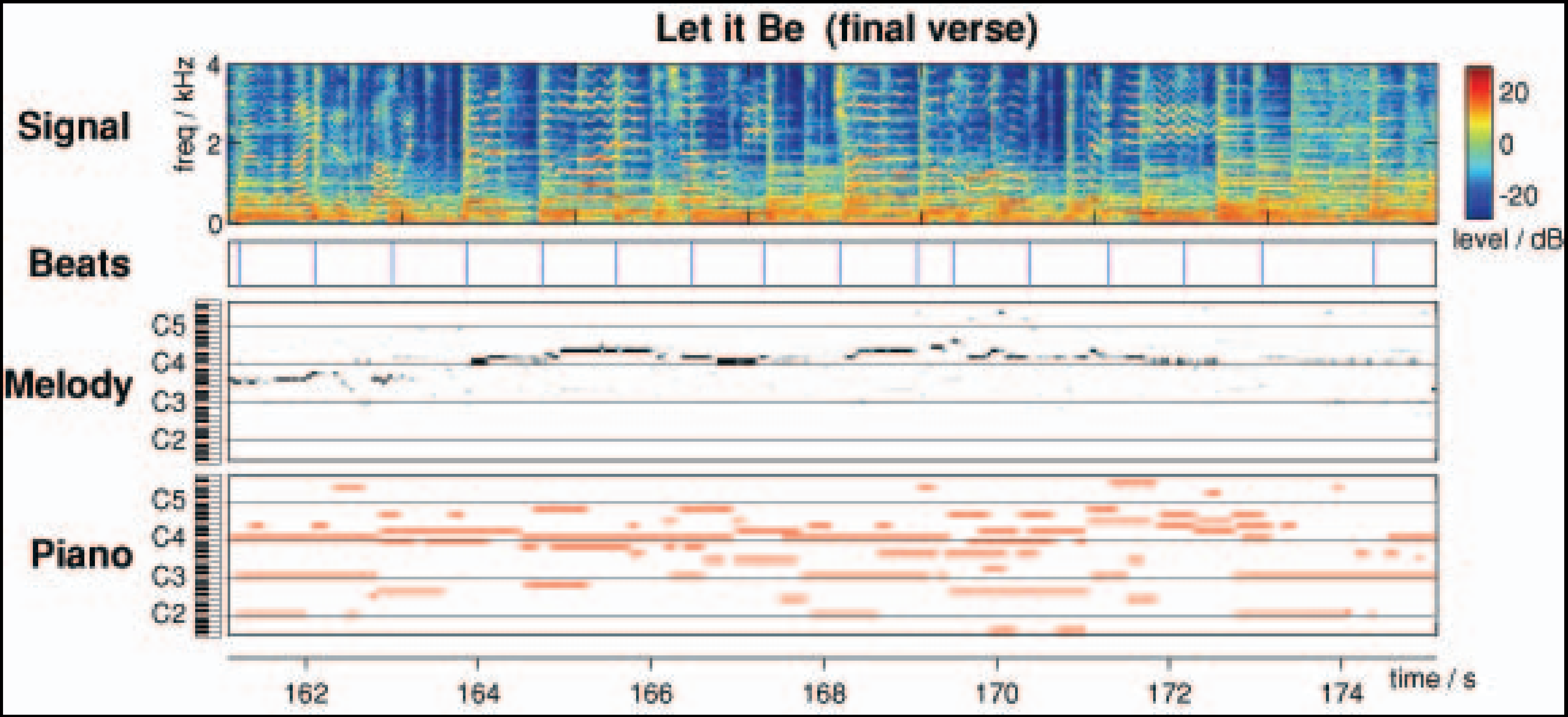
Thick/thin lines are smoothing algorithms.



Music appreciation class for computers

- Discovering musical structure
- Chroma: Simple chord detection
- Sheet-music score alignment
- Music transcription

And, in the end ...



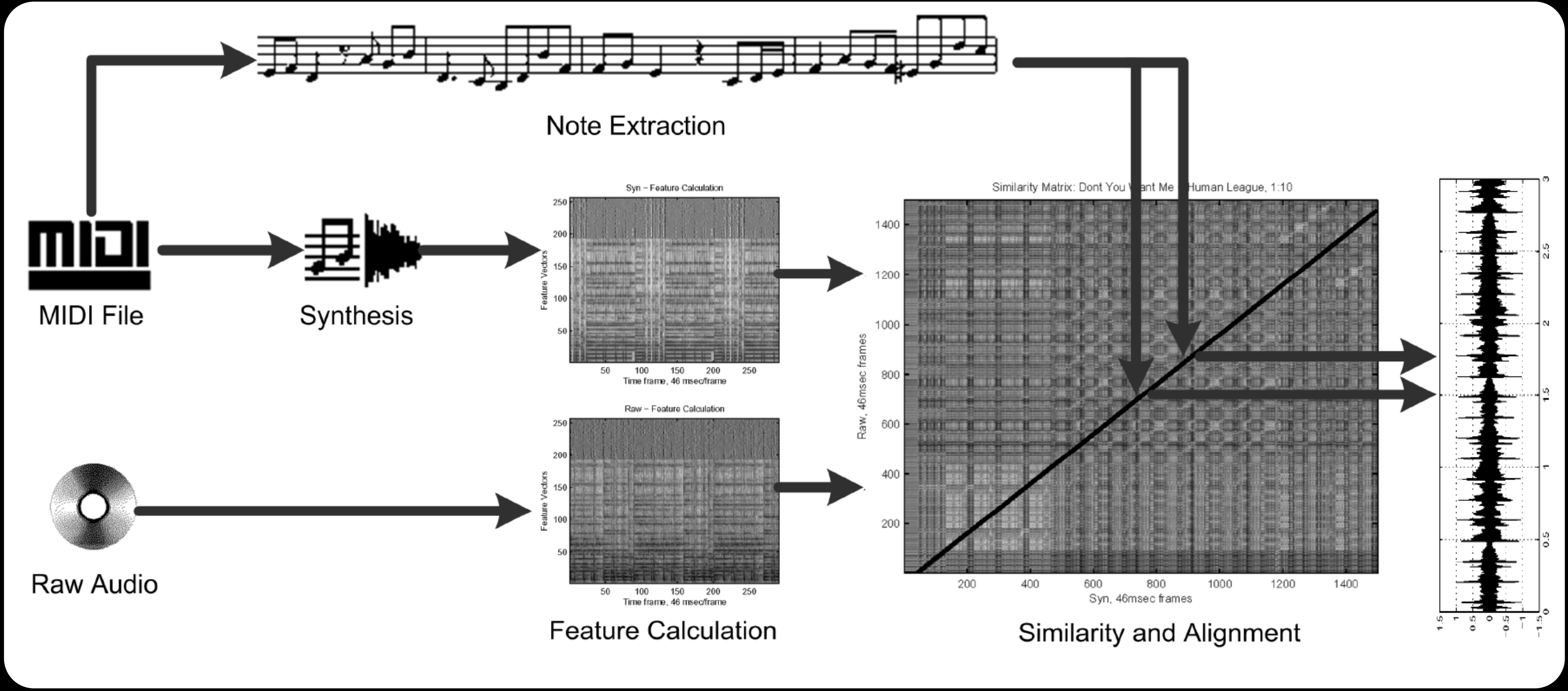
Research topics ...

A chroma replacement

- octaves

	Ratio
• The same note (C1)	1:1
• One octave higher (C2, 12 semi-tones)	2:1
An octave and a 5th (G2, 19 semi-tones), +2 cents	3:1
• Two octaves higher (C3, 24 semi-tones)	4:1
Two octaves higher plus a major 3rd (E3, 28 semi-tones), -14 cents	5:1
Two octaves and a 5th (G3, 31 semi-tones), +2 cents	6:1
Two octaves and a 7th (B3, 34 semi-tones), -31 cents	7:1
• Three octaves higher (C4, 36 semi-tones)	8:1
Three octaves and a 2nd (B5, 38 semi-tones), +4 cents	9:1
Three octaves and a major 3rd, (C#5, 40 semi-tones) -14 cents	10:1

Apply this architecture to other applications



No "transcription" necessary: no need to extract "symbolic" notes from audio.