

Ψηφιακή Επεξεργασία Ήχου

Μάθημα 8: Τονική Ανάλυση Μουσικού Σήματος

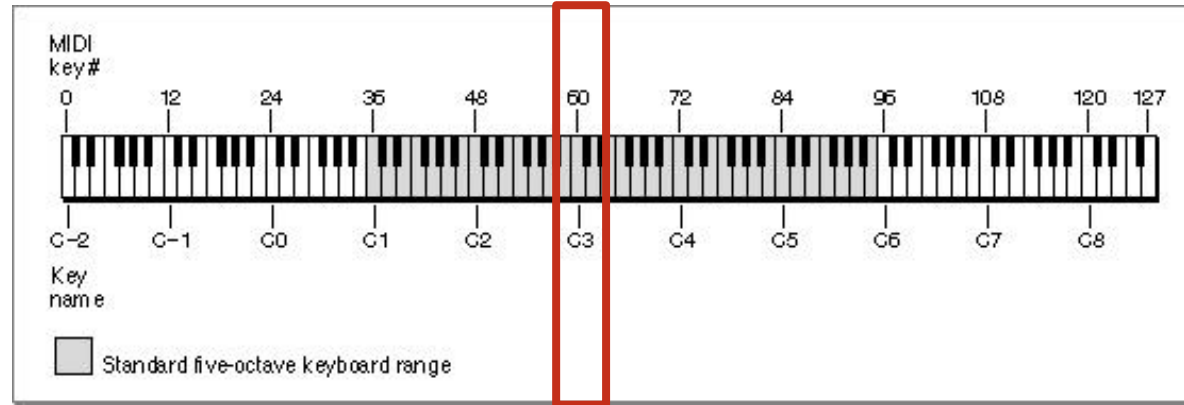
Π.Μ.Σ. «Τεχνολογίες Ήχου και Μουσικής»

Δρ. Χρυσούλα Αλεξανδράκη

Τμήμα Μουσικής Τεχνολογία και Ακουστικής

Ελληνικό Μεσογειακό Πανεπιστήμιο

MIDI Pitch



- ▶ MIDI Pitch (p)
 - ▶ Στο πρωτόκολλο MIDI, η κάθε νότα του συγκεκριμένου συστήματος (σκεφτείτε τα πλήκτρα του πιάνου) αντιστοιχεί σε έναν ακέραιο p (pitch) από 0 έως 127
- ▶ Το σημείο αναφοράς είναι το μεσαίο Ντο (κάτω από το $\Lambda\alpha = 440\text{Hz}$) το οποίο έχει $p = 60$
 - ▶ Ανεβαίνοντας ημιτόνια προσθέτεις τη μονάδα
 - ▶ Π.χ. Το Ντο# έχει $p=61$
 - ▶ Κατεβαίνοντας ημιτόνια αφαιρείς τη μονάδα
 - ▶ Π.χ. το Sib έχει $p=59$
 - ▶ Επομένως το 440Hz έχει $p=69$

Σχέση συχνότητας με pitch

- ▶ Η σχέση που δίνει τη συχνότητα σε Hz για κάθε pitch p είναι:
 - ▶ $F(p) = 440 * 2^{\frac{p-69}{12}} \rightarrow p = 69 + 12 * \log_2 \frac{F(p)}{440}$
 - ▶ Η συχνότητα F έχει εκθετική σχέση με το pitch p , ενώ το p έχει λογαριθμική σχέση με το F
 - ▶ Σε υψηλότερα τονικά ύψη, η συχνότητα του κάθε pitch απέχει περισσότερο από τη συχνότητα του επόμενου ημιτονίου
- ▶ Παράδειγμα:
 - ▶ G#4 αντιστοιχεί σε $f(68) = 415.30$ Hz
 - ▶ A4 αντιστοιχεί σε $f(69) = 440$ Hz
 - ▶ A#4 αντιστοιχεί σε $f(70) = 466.16$

Αλγόριθμοι Pitch tracking

- ▶ Οι περισσότεροι αλγόριθμοι υπολογισμού τονικού ύψους βασίζονται στον υπολογισμό της αυτοσυσχέτισης:
- ▶ **Άσκηση:** Υπολογίστε το τονικό ύψος για το `flute.wav` με τους ακόλουθους τρόπους:
 1. Μέσω STFT: το freq bin στο οποίο μεγιστοποιείται η ενέργεια
 2. Με onset detection και autocorrelation
 3. Με τον αλγόριθμο YIN (`librosa.yin`)
 - ▶ A. de Cheveigné, H. Kawahara, YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*. 111, 1917-1930 (2002).

Flute

$\text{♩} = 80$

mf

3

STFT bin(s) για δεδομένο pitch

- ▶ Μπορούμε να ξέρουμε σε ποιο pitch αντιστοιχεί το κάθε frequency bin του STFT?
 - ▶ $P(p) = \{k: F(p - 0.5) \leq k \frac{f_s}{N} \leq F(p + 0.5)\}$
 - ▶ Παράδειγμα:
 - ▶ $P(69) = \{k: 427.47 \leq k \frac{f_s}{N} \leq 452.89\}$ και $P(57) = \{k: 213.74 \leq k \frac{f_s}{N} \leq 226.44\}$
 - ▶ Ποια είναι τα k για $f_s=44100\text{Hz}$ και $N = 4096$?
 - ▶ Ποιο συχνοτικό εύρος ζώνης (frequency bandwidth) αντιστοιχεί σε κάθε ημιτόνιο?
 - ▶ https://www.audiolabs-erlangen.de/resources/MIR/FMP/C3/C3S1_SpecLogFreq-Chromagram.html

Εύρος ζώνης τονικού ύψους

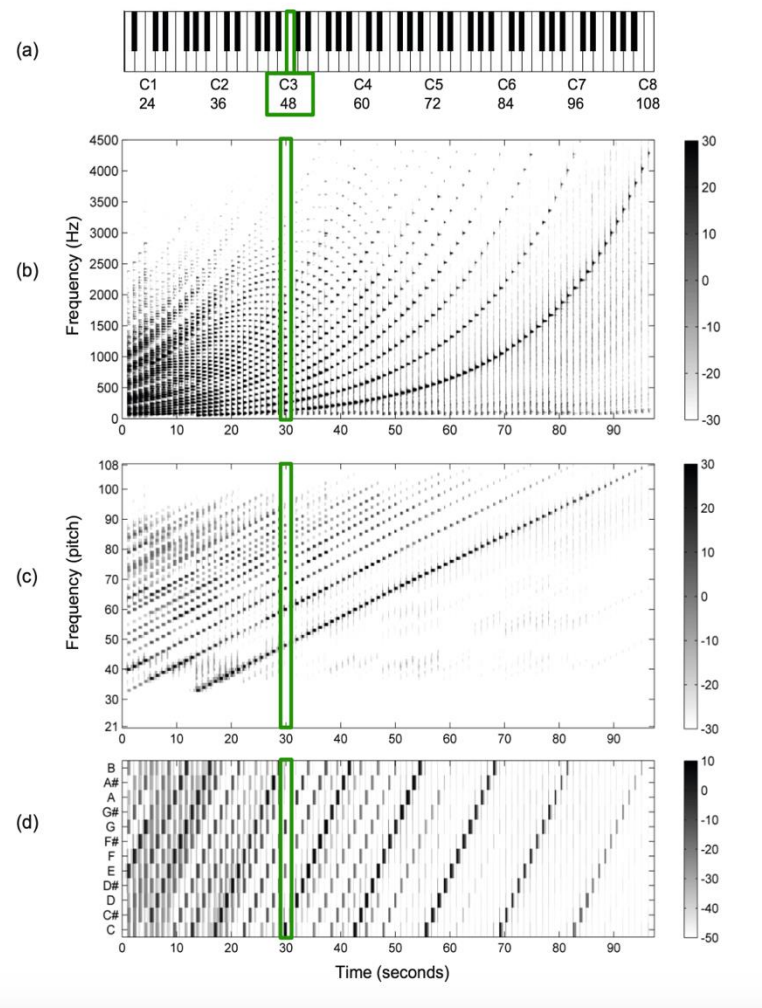
Note	p	$F_{\text{pitch}}(p)$	$F_{\text{pitch}}(p - 0.5)$	$F_{\text{pitch}}(p + 0.5)$	BW(p)
C4	60	261.63	254.18	269.29	15.11
C [#] 4	61	277.18	269.29	285.30	16.01
D4	62	293.66	285.30	302.27	16.97
D [#] 4	63	311.13	302.27	320.24	17.97
E4	64	329.63	320.24	339.29	19.04
F4	65	349.23	339.29	359.46	20.18
F [#] 4	66	369.99	359.46	380.84	21.37
G4	67	392.00	380.84	403.48	22.65
G [#] 4	68	415.30	403.48	427.47	23.99
A4	69	440.00	427.47	452.89	25.42
A [#] 4	70	466.16	452.89	479.82	26.93
B4	71	493.88	479.82	508.36	28.53
C5	72	523.25	508.36	538.58	30.23

$$\text{BW}(p) := F_{\text{pitch}}(p + 0.5) - F_{\text{pitch}}(p - 0.5)$$

Log frequency spectrogram και χρωματογράφημα

- ▶ **Ορισμός:** Log Frequency Spectrogram: Η ενέργεια αντιστοιχεί σε δεδομένο pitch
 - ▶ $\mathcal{Y}_{LF}(n, p) = \sum_{k \in P(p)} |X(k, n)|^2$
- ▶ **Ορισμός:** Χρωματογράφημα το άθροισμα του log frequency spectrogram για το pitch class
 - ▶ *Pitch Class:* Σύνοψη όλων των φθόγγων που έχουν το ίδιο όνομα, ανεξάρτητα από το σε ποια οκτάβα εμφανίζονται
 - ▶ $c = p \% 12 = \{0, 1, 2, \dots, 11\}$ δηλαδή η νότα, χωρίς την πληροφορία οκτάβας
 - ▶ $\mathcal{C}(n, c) = \sum_{\{p: p \% 12 = c\}} \mathcal{Y}_{LF}(n, p)$

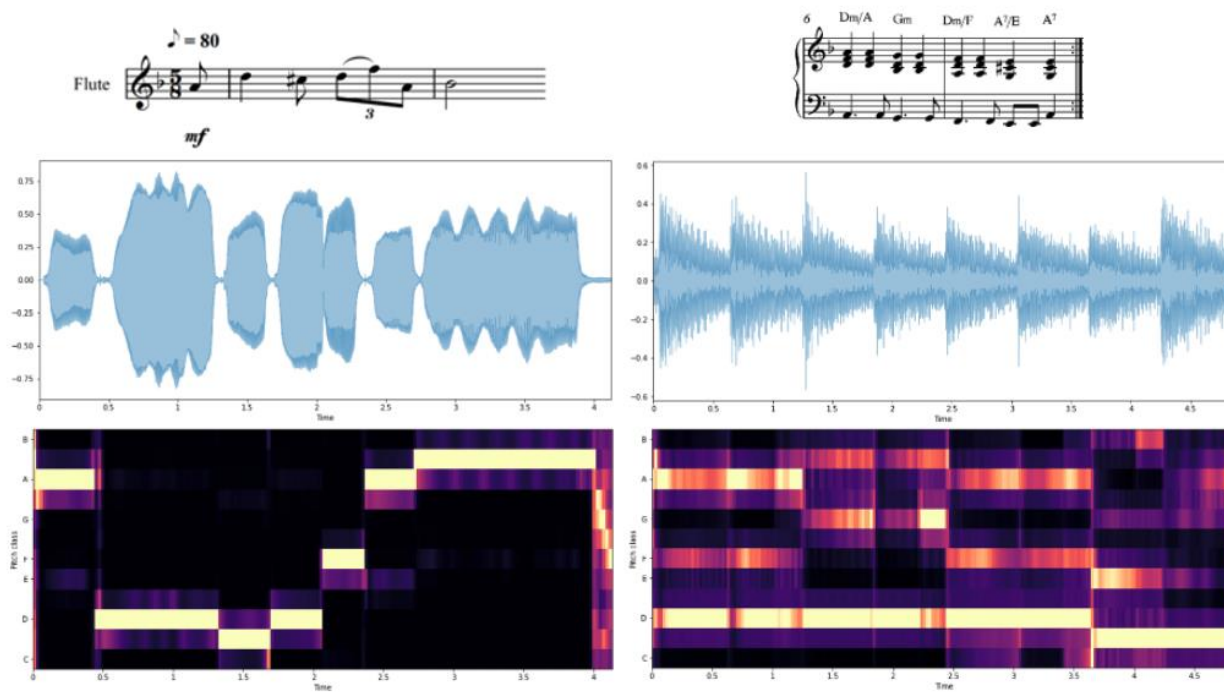
STFT Pitch Information



- ▶ Fig. 3.3 Various representations for a recording of the chromatic scale played on a real piano. The scale ranges from A0 ($p = 21$) to C8 ($p = 108$).
 - ▶ (a) Piano keys representing the chromatic scale.
 - ▶ (b) Magnitude spectrogram.
 - ▶ (c) Pitch-based log-frequency spectrogram.
 - ▶ (d) Chromagram.
 - ▶ For visualization purposes the values are encoded by shades of a logarithmic scale. The C3 ($p = 48$) played at time $t = 30$ sec is highlighted by rectangular frames.
- ▶ Προσέξτε την παρουσία αρμονικών σε κάθε χρώμα (!)
 - ▶ [https://en.wikipedia.org/wiki/Harmonic_series_\(music\)](https://en.wikipedia.org/wiki/Harmonic_series_(music))
- ▶ Σχήμα από:
 - ▶ Müller, M. (2015). *Fundamentals of Music Processing*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-21945-5>, page 119

Χρωματογράφημα

- ▶ Χρησιμοποιείται σε περιπτώσεις που ενδιαφέρει το αρμονικό και όχι το μελωδικό περιεχόμενο, π.χ.
 - ▶ Chord recognition
 - ▶ Key detection (tonic- mode), e.g. C-minor
 - ▶ Audio synchronization
- ▶ Χρησιμοποιείται συνήθως σε **πολυφωνικά σήματα** για τον προσδιορισμό της **αρμονίας** (της ακολουθίας συγχορδιών)
 - ▶ Chord Recognition
- ▶ Σε **μονοφωνικά σήματα** ενδιαφέρει η **μελωδία** για την οποία είναι απαραίτητη η πληροφορία της οκτάβας:
 - ▶ Pitch tracking



Άσκηση Χρωματογράφημα

- ▶ Για το σήμα φλάουτου της προηγούμενης άσκησης
 - ▶ Να απεικονισθεί το χρωματογράφημα με τη συνάρτηση `librosa.feature.chroma_stft()`
- ▶ Κάντε το ίδιο για πολυφωνικά αρχεία, π.χ. `chords.wav`, `Bill_Evans.wav`