Ψηφιακή Επεξεργασία Ήχου
Μάθημα 6: Εξαγωγή Ηχητικών Χαρακτηριστικών

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

Δρ. Χρυσούλα Αλεξανδράκη

Τμήμα Μουσικής Τεχνολογία και Ακουστικής

Ελληνικό Μεσογειακό Πανεπιστήμιο

Audio Frame vs. Audio Block

- Audio Block
  - Ένα τμήμα του ήχου, π.χ. 1024 δείγματα ήχου
- Audio Frame
  - Μία φέτα του ήχου
  - 2- καναλια ήχου => δείγματα ανά χρονική στιγμή
  - 4 –καναλια ήχου
  - Frame -> 4 δείγματα ήχου

Μαθηματική Έκφραση STFT

- Ο STFT ενός σήματος πραγματικών τιμών είναι η μιγαδική συνάρτηση δύο μεταβλητών:
  - **X(k, n) :=** $\sum_{l=0}^{N-1} x(l+nh)w(l)e^{-2\pi jkl/N}$
- Όπου
  - x σήμα πραγματικών τιμών
  - $w(l)$ παραθυρική συνάρτηση της οποίας οι μη-μηδενικές τιμές περιορίζονται στο διάστημα $l \in [0, N-1]$, με *N* ακέραιο (δύναμη του 2)
  - N το μήκος του ***audio block*** στο οποίο εφαρμόζεται ο STFT
  - n αύξον δείκτης του παραθύρου/audio block
  - h hop size (δηλαδή η χρονική ολίσθηση του παραθύρου σε δείγματα)
- Καθώς το ο δείκτης συχνότητας k είναι ομοιόμορφα κατανεμημένος στο διάστημα [0, f_s] (f_s – συχνότητα δειγματοληψίας):
  - X(k, n) αντιπροσωπεύει το φάσμα στη συχνότητα $f(k) = k*f_s/N$ (Hz) και στο χρονικό σημείο $T(n) = n*h/f_s$(sec)

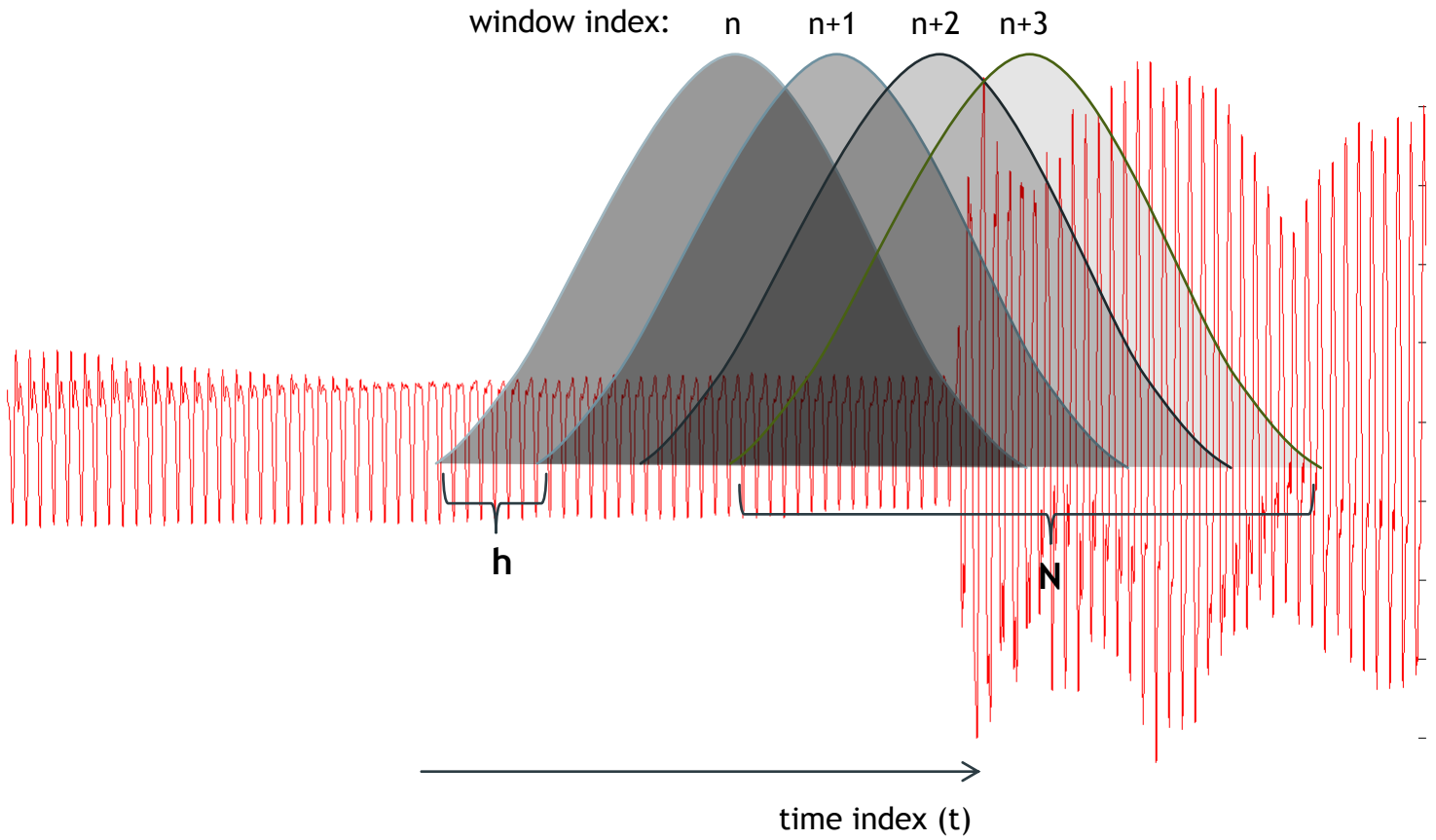
Αποτέλεσμα STFT

- Τόσο η συνάρτηση **matplotlib.pyplot.specgram** όσο και η **scipy.signal.stft** για σήμα x(n) πραγματικών τιμών συνολικής διάρκειας ΔΤ, επιστρέφουν ένα πίνακα διαστάσεων (N/2+1) x (ΔΤ /h), ή αλλιώς αν η συνολική διάρκεια του σήματος σε δείγματα είναι A, τότε ο πίνακας έχει διάσταση:
  - (N/2+1) x A/h -1
  - Το πλήθος των στηλών (audio blocks) μπορεί να διαφοροποιείται κατά +1 ή -1 ανάλογα με την υλοποίηση και τα ορίσματα των παραπάνω συναρτήσεων

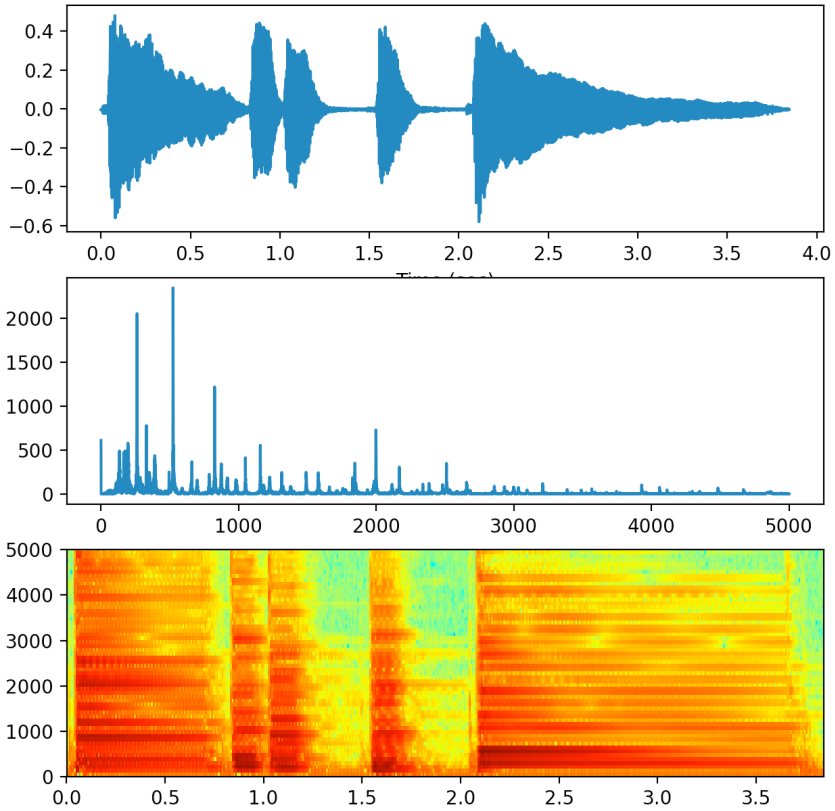
- Στον πίνακα αυτό, διαφορετικές **γραμμές** αντιστοιχούν σε διαφορετικές ζώνες συχνοτήτων (frequency bins) ενώ διαφορετικές **στήλες** αντιστοιχούν σε διαφορετικά audio blocks.

$$X_{kn} = \begin{bmatrix} X_{1,1} & \cdots & X_{1,A/(N-h)} \\ \vdots & \ddots & \vdots \\ X_{\frac{N}{2}+1,1} & \cdots & X_{\frac{N}{2}+1,,A/(N-h)} \end{bmatrix}$$

STFT χωρίς zero-padding



Αναπαραστάσεις Σήματος



- Κυματομορφή:
  - Μπορούμε να δούμε τα χρονικά συμβάντα αλλά δε γνωρίζουμε το περιεχόμενό τους
- Φάσμα
  - Μπορούμε να δούμε το φασματικό περιεχόμενο, αλλά δεν μπορούμε να ξέρουμε σε ποια συμβάντα αντιστοιχεί
- Φασματογράφημα
  - Μπορούμε να δούμε και τα συμβάντα και το φασματικό περιεχόμενό τους

Φασματογράφημα -> Περιοδικότητα και χρονική ακρίβεια

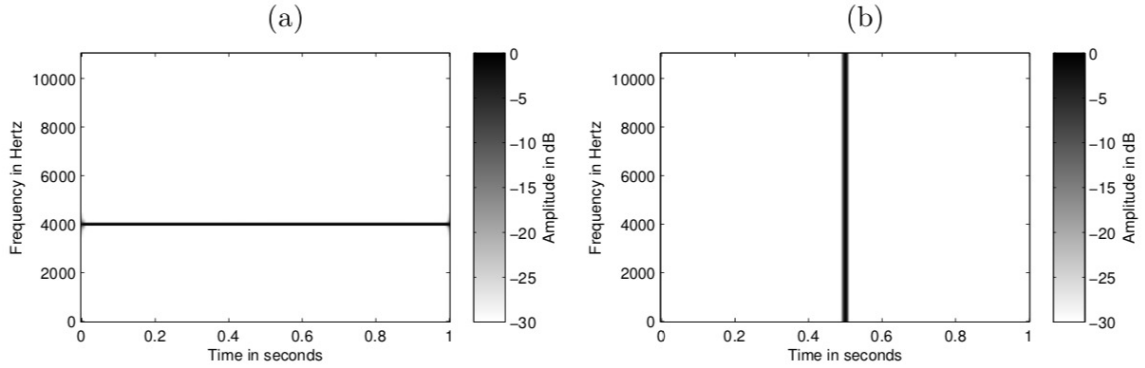


Figure 1: **(a):** Spectrogram of an ideal harmonic signal. **(b):** Spectrogram of an ideal percussive signal.

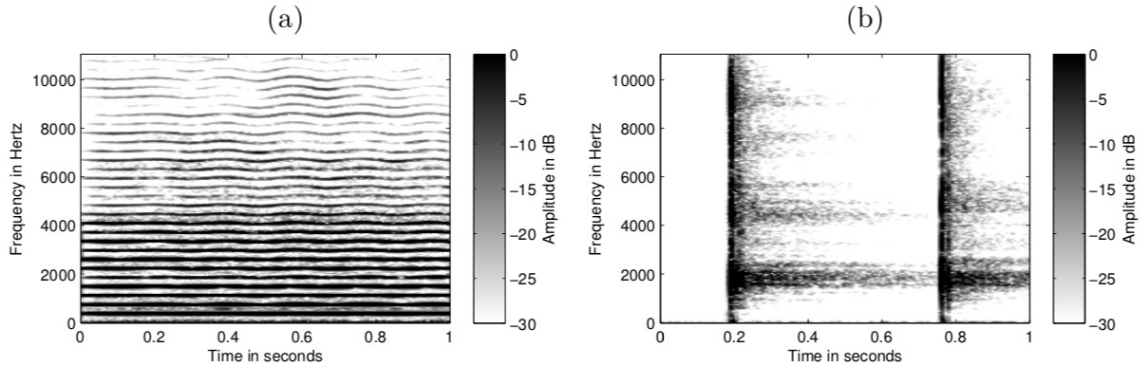
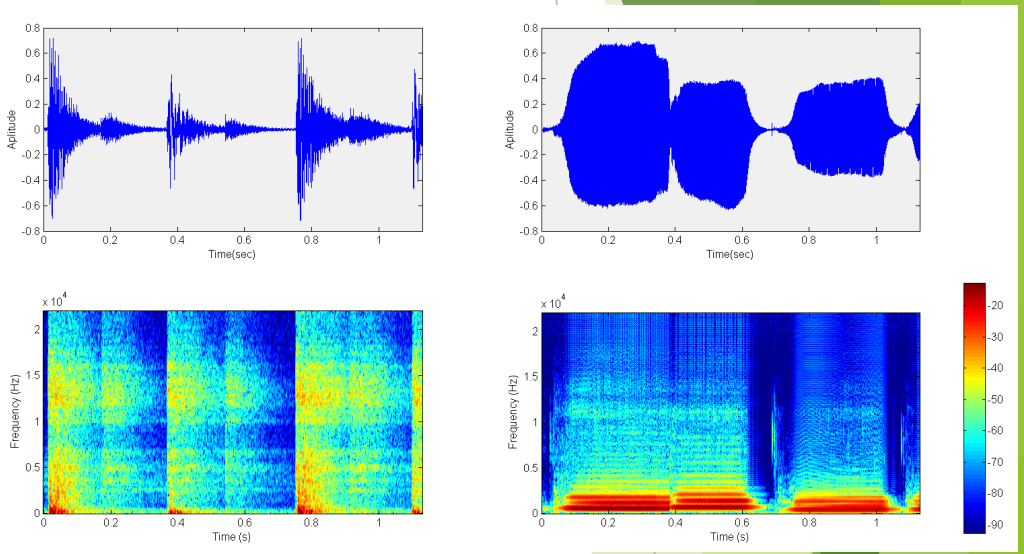


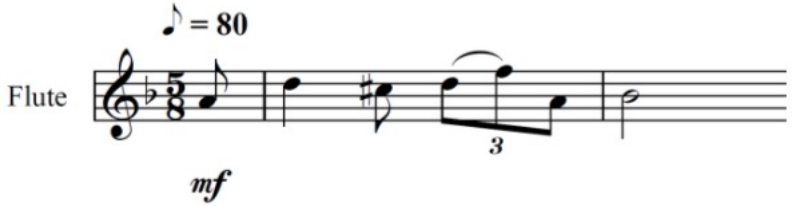
Figure 2: **(a):** Spectrogram of a recording of a violin. **(b):** Spectrogram of a recording of a castanets.



Drum Signal=> Salient Onsets
Flute Signal => Salient Pitches

Άσκηση

- Ξεκινώντας από το αρχείο flute.wav του οποίου η παρτιτούρα είναι η ακόλουθη:



1. Να απεικονίσετε το φασματογράφημα
    a) Αλλάξτε το αρχείο ήχου και χρησιμοποιήστε το drum.wav, και piano.wav
2. Να βρείτε τα frequency bins στα οποία ο STFT μεγιστοποιείται, για κάθε block
3. Σε τι συχνότητες αντιστοιχούν?
4. Πώς σχετίζονται με τις συχνότητες που αναπαρίστανται στην παραπάνω παρτιτούρα?
5. Μπορείτε να βρείτε τα χρονικά σημεία στα οποία εμφανίζονται τα note onsets? Σε ποιο δείγμα αντιστοιχούν?
6. Διαπιστώστε την αρχή της απροσδιοριστίας.

Audio Features

- Numerical values describing the contents of a signal according to different kinds of inspection: temporal, spectral, perceptual, etc.
- Αναφέρονται στον υπολογισμό κάποιας τιμής η οποία αφορά συνολικά ένα audio block (π.χ. από N=1024 δείγματα). Η τιμή αυτή μπορεί να υπολογίζεται σε κάποια αναπαράσταση σήματος, π.χ.
  - Στο χρόνο => Temporal Features
  - Στο φάσμα => Spectral Features
  - Στο χρώμα => Chroma Features
  - Κ.ο.κ.
- Η εξέλιξη αυτής τιμής για διαδοχικά audio blocks μου δίνει μακροσκοπική πληροφορία για το περιεχόμενο του σήματος.
- Dimensionality Reduction:
  - Μειώνουν τη διάσταση του διανύσματος ήχου (δηλαδή του σήματος)
  - Π.χ. από 44100 δείγματα (1 sec) σε 44100/1024 = 43 δείγματα
- Παρέχουν πληροφορία για το περιεχόμενο του Σήματος

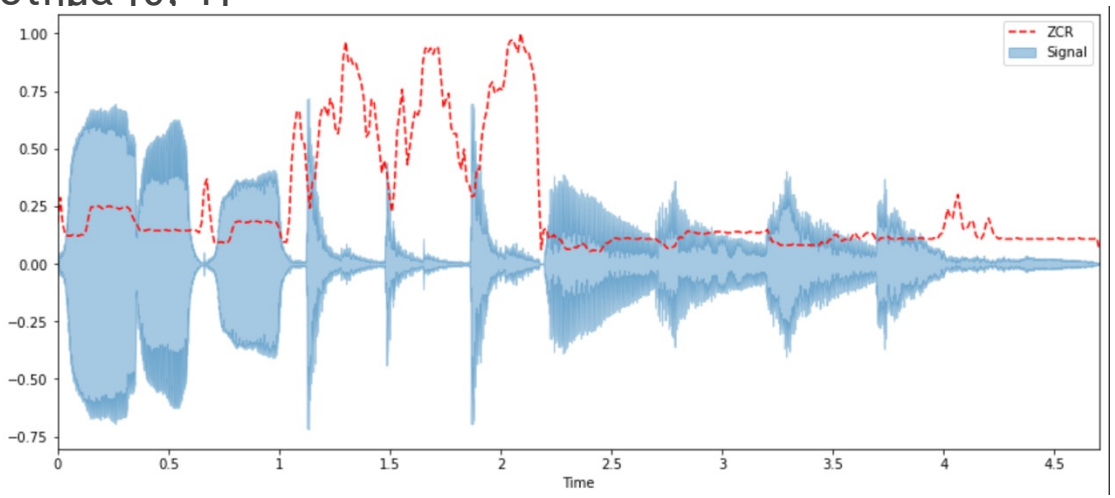
Χρονικά Χαρακτηριστικά

Energy, ZCR

Autocorrelation

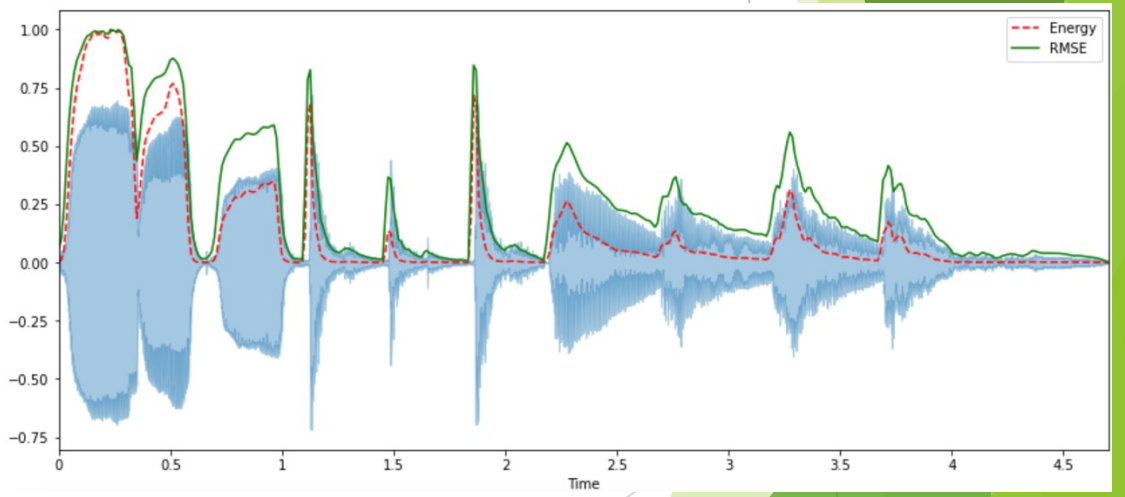
Zero Crossing Rate - Ρυθμός διέλευσης μηδενικού άξονα

- Μετράει τη συχνότητα με την οποία η κυματομορφή αλλάζει πρόσημο
  - $\mathrm{ZCR}(n) = \frac{1}{N}\sum_{t=n}^{n+N} 1_{\mathbb{R}_{<0}}\left[x(t)x(t-1)\right]$
- Περιοδικά σήματα αναμένεται να έχουν χαμηλές τιμές, ενώ θορυβώδη με αρμονικά σήματα παρουσιάζουν υψηλότερες τιμές
- Υπολογίζεται για κάθε audio block στη χρονική αναπαράσταση του σήματος
- Για την αναπαράσταση στο ακόλουθο σήμα το ZCR έχει κανονικοποιηθεί στο διάστημα [0, 1]



Χαρακτηριστικά Ενέργειας

- Ενέργεια
  - Επιτρέπει την παρακολούθηση διακυμάνσεων του πλάτους μιας κυματομορφής (envelop follower)
  - Μπορεί βέβαια να υπολογισθεί είτε στην κυματομορφή είτε στο φάσμα (θεώρημα Parseval)
  - $E(n) = \sum_{t=n}^{n+N}(x(t))^2 = \frac{1}{N}\sum_{k=0}^{N/2-1}|X(n,k)|^2$
- Ενεργός τιμή Ενέργειας (Root Mean Square)
  - $\mathrm{RMSE(n)} = \sqrt{\mathrm{E(n)}}$
- Log Energy
  - Μετράει την ηχητική πίεση σε dB
  - $LE(n) = 10\log_{10} E(n)$



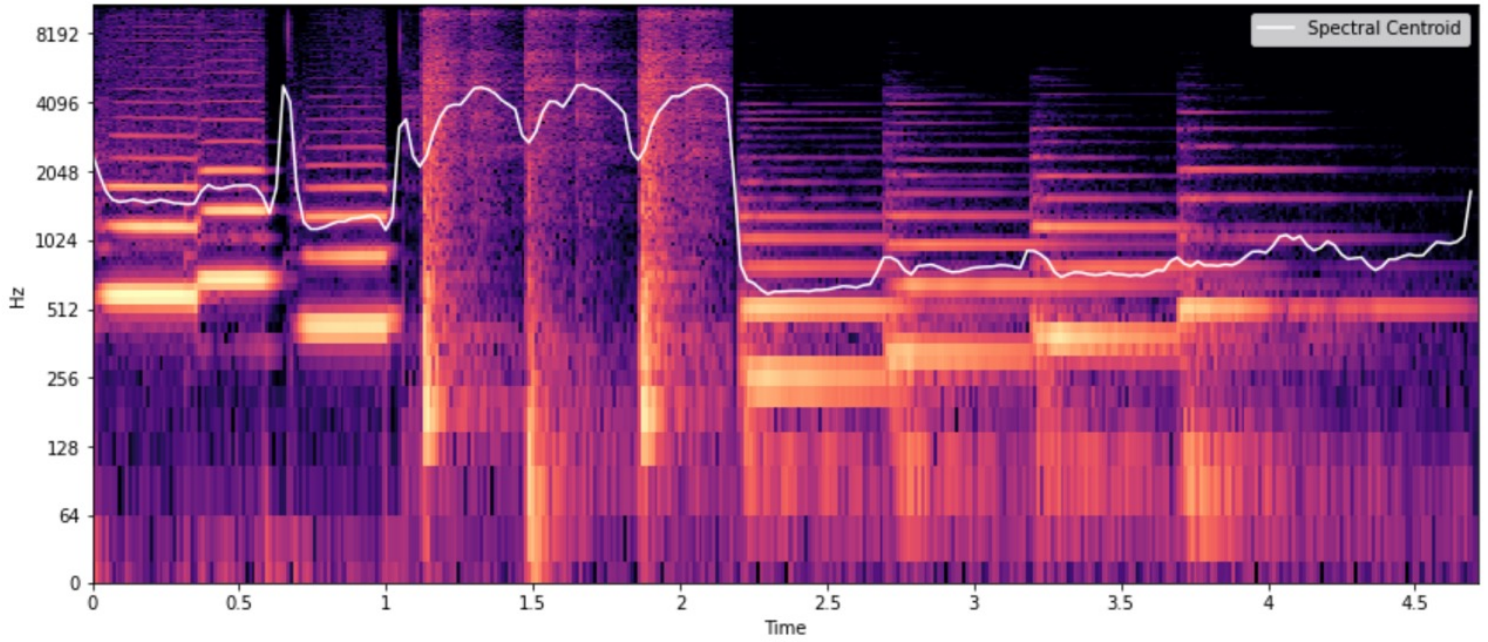
Κανονικοποιημένες τιμές ενέργειας στο [0, 1]

Φασματικά Χαρακτηριστικά

Spectral Centroid,, Spectral Flux

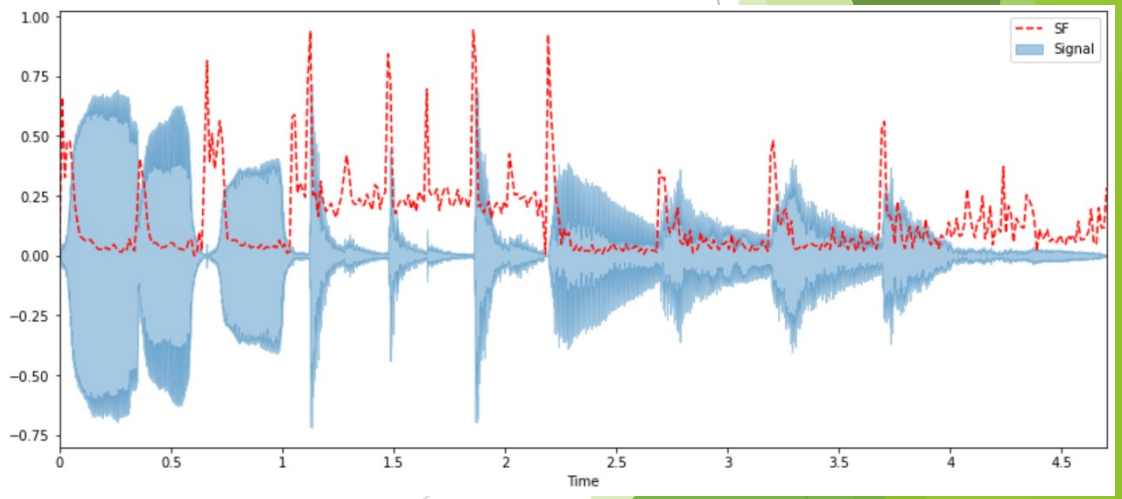
Specral Centroid - Φασματικός Κεντροειδής

- Φασματικός Κεντροειδής
  - Υποδεικνύει πού βρίσκεται το «κέντρο μάζας» του φάσματος. Αντιληπτικά, έχει μια ισχυρή σύνδεση με την αντίληψη της «φωτεινότητας (brightness)» ενός ήχου
  - SC(n) $=\frac{\sum_{k=0}^{N/2-1} f(k)|X(k,n)|}{\sum_{k=0}^{N/2-1} |X(k,n)|}$, όπου f(k) η κεντρική συχνότητα του k-bin



Spectral Flux – Φασματική Ροή

- Μετράει την αλλαγή στο φάσμα του ενός block από το επόμενο
- Αναμένεται να παρουσιάζει υψηλές τιμές για ξαφνικές αλλαγές του σήματος
- Έχει χρησιμοποιηθεί κατά καιρούς σε διάφορες παραλλαγές:
  - $SF_1(n) = \sum_{k=0}^{K-1} H(|X(k,n)| - |X(k,n-1)|)$,
    - using $\mathrm{L}-1$ norm , $H(x) = \frac{x+|x|}{2}$ is the half-wave rectifier (ημιανόρθωση)
  - $SF_2(n) = \sum_{k=0}^{K-1} \{H(|X(k,n)| - |X(k,n-1)|)\}^2$,
    - using the L-2 norm (overemphasizing differences)
  - $SF_3(n) = \frac{\sum_{k=0}^{K-1} H(|X(k,\mathrm{n})| - |X(k,n-1)|)}{\sum_{k=0}^{K-1} |X(k,n)|}$



Άσκηση

- Ξεκινώντας από το αρχείο flute.wav
  1. Να υπολογίσετε (με δικό σας κώδικα) και να αποδώσετε γραφικά τα χαρακτηριστικά logEnergy, Spectral Centroid και Spectral Flux
  2. Προσπαθήστε να εντοπίσετε τα σημεία έναρξης νοτών επιλέγοντας τις κορυφές (τοπικά μέγιστα) της φασματικής ροής (spectral flux) με τη συνάρτηση
     - https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html