Ψηφιακή Επεξεργασία Ήχου
Μάθημα 4: Μετάβαση στο πεδίο Συχνότητας
για χρονικά μεταβαλλόμενα φάσματα

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

Δρ. Χρυσούλα Αλεξανδράκη

Τμήμα Μουσικής Τεχνολογία και Ακουστικής

Ελληνικό Μεσογειακό Πανεπιστήμιο

Short Term Fourier Transform (STFT)

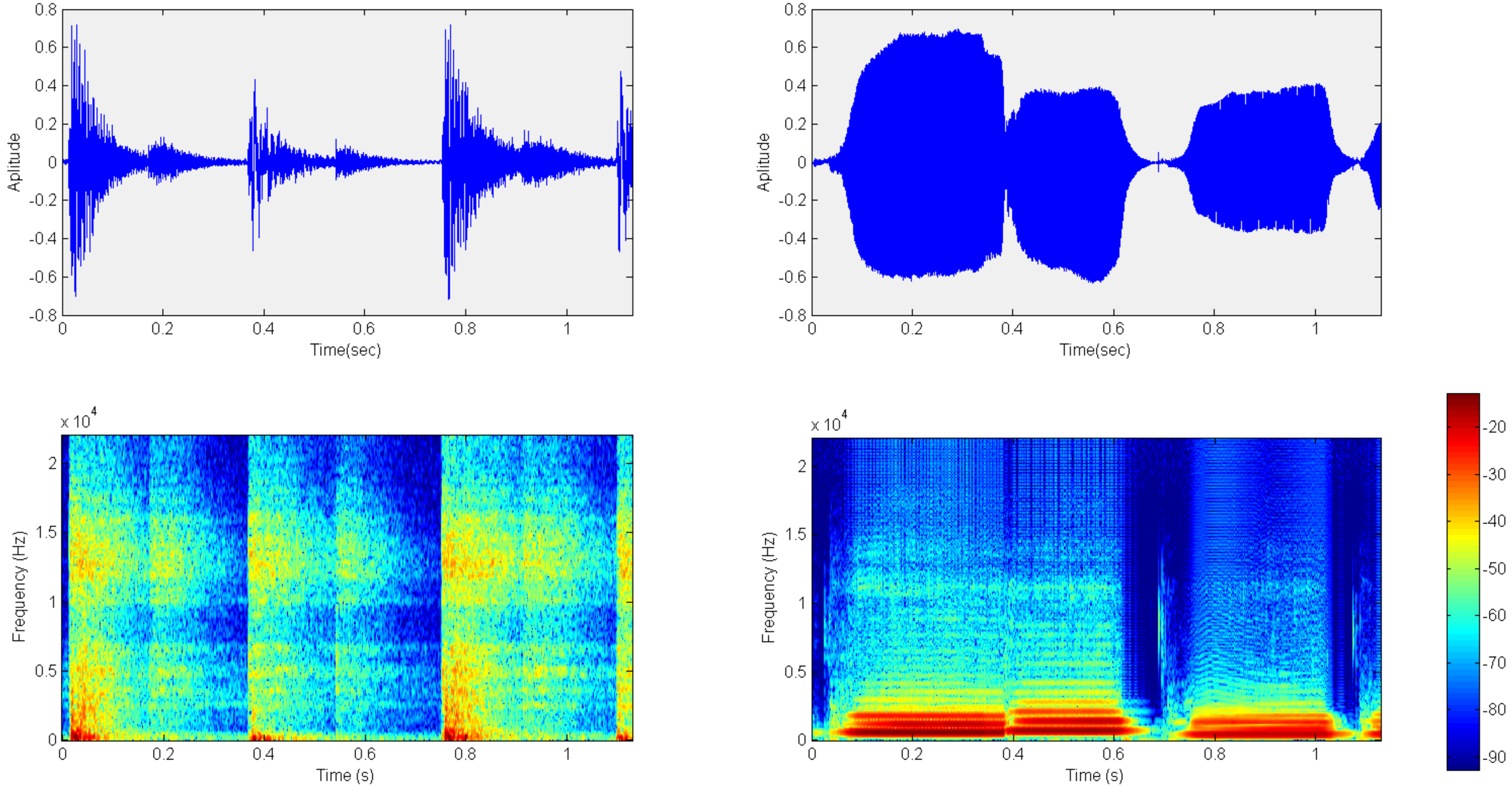
For non-stationary processes…

Για σήματα των οποίων το φάσμα μεταβάλλεται στο χρόνο..

Επιμερισμός Σήματος σε παράθυρα

- Ο μετασχηματισμός Fourier θεωρεί ότι το σήμα έχει σταθερή περιοδικότητα κι επομένως το **φάσμα του δε μεταβάλλεται στο χρόνο**
- Σταθερή περιοδικότητα έχουν μόνο τα **συνθετικά σήματα** (κατασκευασμένα με μαθηματικό υπολογισμό)
- Μπορεί εντούτοις να θεωρηθεί ότι κάποια από τα πραγματικά σήματα παρουσιάζουν κατά τόπους και για περιορισμένο χρονικό διάστημα σταθερή περιοδικότητα και άρα σταθερό φάσμα.
- Προκειμένου να βρεθούν ***οι συχνότητες που υπάρχουν στο σήμα σε κάθε χρονική στιγμή***, το σήμα πρέπει να επιμερισθεί σε παράθυρα
  - Για κάθε τέτοιο παράθυρο υπολογίζεται ο Fourier Transform (DFT)
  - Τα παράθυρα αυτά, παρατεταγμένα στο χρόνο μας δίνουν το φασματογράφημα (spectrogram)

Φασματογράφημα – Ένταση όλων των συχνοτήτων σε κάθε χρονικό παράθυρο



Παραθυροποίηση

- Παράθυρα σήματος x(n) με μήκος N, ισοδυναμεί με πολ/σμο σήματος με ένα τετραγωνικό παλμό:
  - $\mathrm{w}(n) = \begin{cases} 1, & 0 \leq n < N \\ 0, & otherwise \end{cases}$
  - Το προκύπτον φάσμα X(k) ισούται με τη συνέλιξη του φάσματος του αρχικού σήματος με το φάσμα του τετραγωνικού παλμού
- Οι ασυνέχειες που προκύπτουν από την παραθυροποίηση του σήματος έχουν ως συνέπεια τη διαρροή φασματικής ενέργειας σε πλευρικές συχνοτικές ζώνες, φαινόμενο γνωστό ως διαρροή φάσματος (spectral leakage)

Παραθυροποίηση => Ασυνέχεια => Φασματική Διαρροή

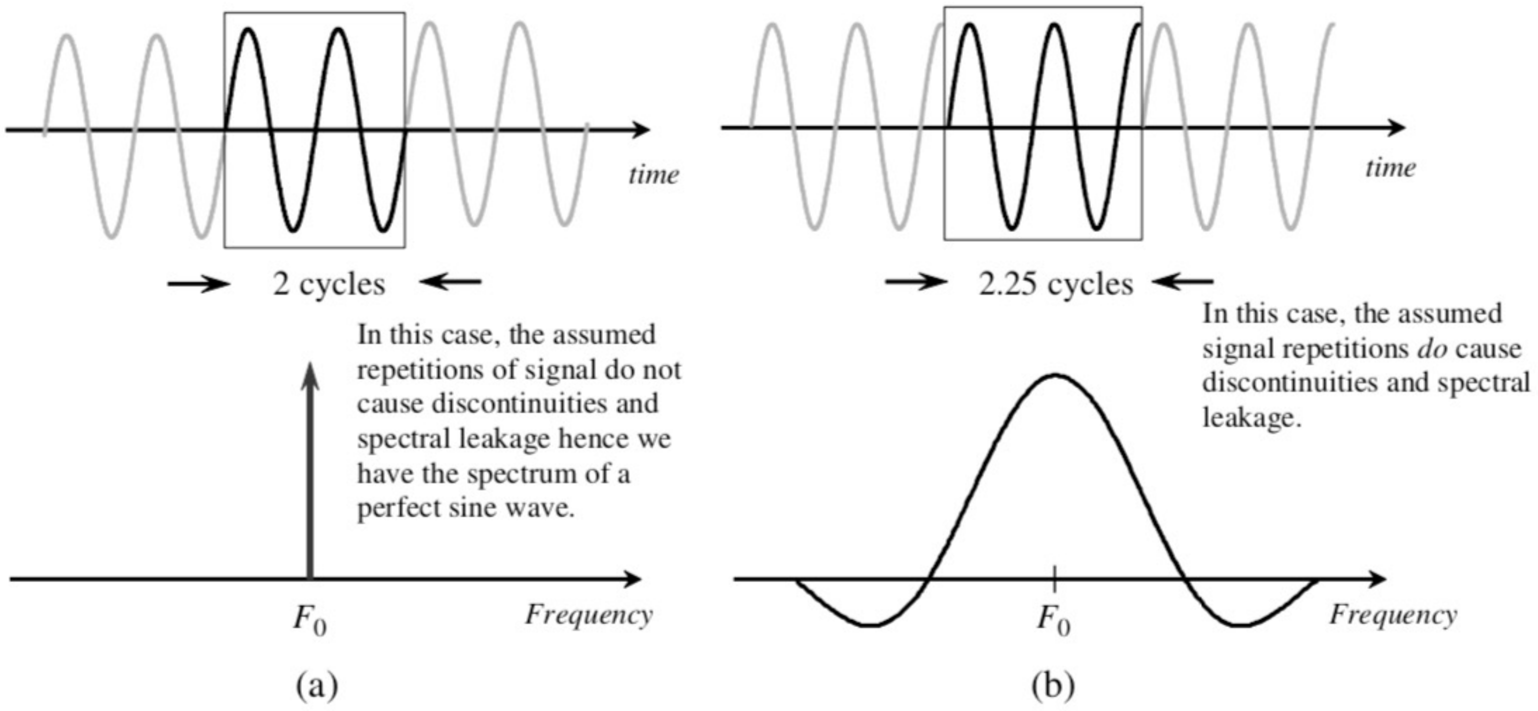
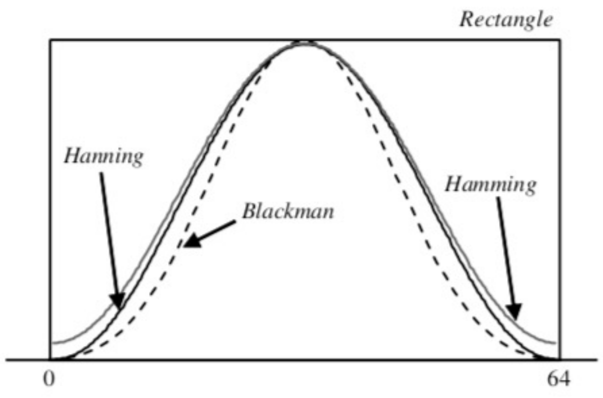
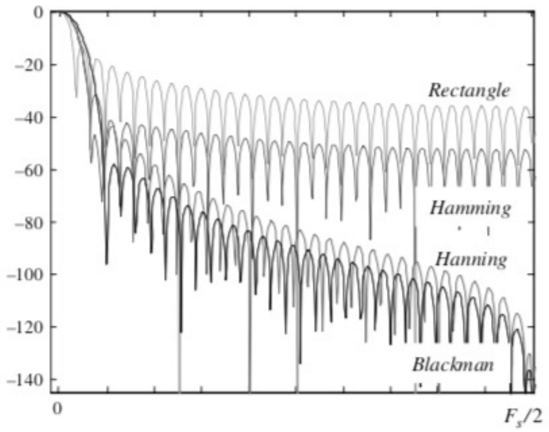


Figure 2.17 The DFT spectrum of $\exp(j2\pi fm)$: (a) an integer number of cycles within the N-sample analysis window, (b) a non-integer number of cycles in the window. Note that DFT theory assumes the signal is periodic hence outside the window the signal is assumed to repeat itself with a periodicity equal to the window length.

Συνήθεις παραθυρικές συναρτήσεις





- Για να αμβλυνθεί η διαρροή φάσματος (spectral leakage) αντί για τετραγωνικό παλμό συνήθως πολ/ζουμε με κάποια παραθυρική συνάρτηση που δεν εμφανίζει ασυνέχειες, αλλά πέφτει σταδιακά στο μηδέν, όπως:

- $$w(m) = \begin{cases} a - (1-a)\cos\frac{2\pi n}{N}, & 0 \le n < N \\ 0, & othewise \end{cases}$$

- Όπου:
  - *Rectangle, a =1*
  - *Hanning, a=0.5*
  - *Hamming, a=0.54*
  - *Blackman*
    - $w(m) = 0.42 - 0.5\cos\frac{2\pi n}{N+1} - 0.08\cos\frac{4\pi n}{N+1} 0 \le n < N$

Επικάλυψη παραθύρων

- Διαδοχικά παράθυρα του σήματος επικαλύπτονται, ώστε να μη χαθεί πληροφορία που οφείλεται στο γεγονός ότι η παραθυρική συνάρτηση εξασθενίζει το σήμα στα άκρα του παραθύρου.
- Το πλήθος δειγμάτων κατά το οποίο διαφέρουν διαδοχικά παράθυρα μεταξύ τους λέγεται Hop Size
  - Συνήθως h = N/2 ή N/4, όπου N το συνολικό εύρος του παραθύρου (σε πλήθος δειγμάτων)

Συχνοτική / Χρονική Ανάλυση

- Σε πολλές περιπτώσεις το μήκος του παραθύρου στο οποίο γίνεται μετ/σμος Fourier οφείλει να είναι μικρό:
  - Είτε για μεγαλύτερη ανάλυση στο χρόνο (time resolution $\Delta t = N/Fs$)
  - Είτε επειδή το σήμα λαμβάνεται και πρέπει να αναλυθεί σε πραγματικό χρόνο κι έτσι απαιτείται ελαχιστοποίηση του delay Δt
- Όμως, φάσμα σήματος N σημείων δίνει freq resolution $\Delta f = fs/N$
  - Π.χ. για fs=44100, δίνει
    - N=64 => Δt = 1.4ms και Δf = 689Hz
    - N=512 => Δt = 11.6ms και Δf = 86Hz
    - N=1024 => Δt = 23ms και Δf = 43Hz

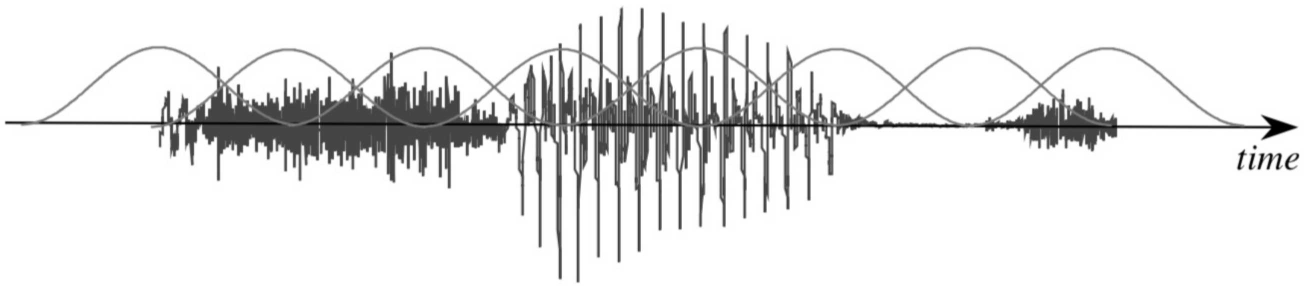
Φασματική Λείανση (Spectral Smoothing)

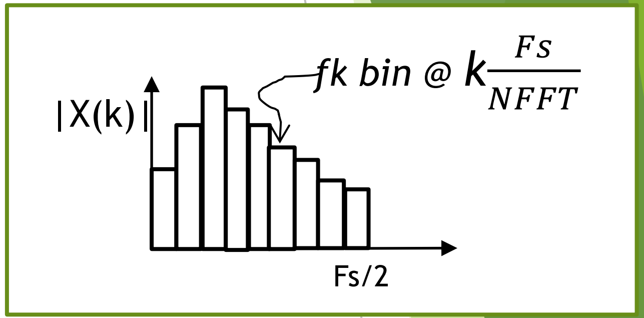
- Σε αυτές τις περιπτώσεις υπάρχει σημαντικό όφελος εάν το N παραμείνει μικρό αλλά το παράθυρο επεκταθεί με μηδενικά, ώστε το πλήθος σημείων του φάσματος να είναι μεγαλύτερο. Πράγματι:
  - Θεωρήστε σήμα [x(0), x(1), …, x(N-1)]. Δίνει N σημεία φάσματος [X(0), X(1), …, X(N-1)].
  - Εάν το σήμα επεκταθεί σε 2N σημεία με τα σημεία N, N+1, …, 2N να είναι όλα μηδενικά, τότε το φάσμα του θα είναι:
  - $\mathrm{X(k)} = \sum_{n=0}^{2N-1} x(n) e^{-j\frac{2\pi}{2N}nk} = \sum_{n=0}^{N-1} x(n) e^{-j\frac{\pi}{N}nk}$ με *k=0,1, …2N-1*
  - Τα άρτια σημεία του φάσματος, δηλ. *k = 0, 2, 4, ..2N-2* είναι τα ίδια σημεία που θα περνάμε αν ο μετ/σμος ήταν μήκους N, ενώ τα τα περιττά είναι ενδιάμεσα σημεία που παρεμβάλλονται ανάμεσα στα άρτια.
  - Η τεχνική αυτή λέγεται ***zero-padding***

Zero padding Theorem

- Αποδεικνύεται ότι:
  - Zero-padding in the time-domain corresponds to ideal interpolation in the frequency domain
- https://ccrma.stanford.edu/~jos/dft/Zero_Padding_Theorem_Spectral.html
- https://math.stackexchange.com/questions/26432/discrete-fourier-transform-effects-of-zero-padding-compared-to-time-domain-inte

STFT (Short Term Fourier Transform)





- Το αποτέλεσμα του STFT είναι, για κάθε τμήμα (block) του σήματος:
  - Ένας πίνακας τιμών, όπου κάθε μια από αυτές τις τιμές αντιστοιχεί σε μια διαφορετική ***συχνοτική μπάντα (frequency bin)***
  - Παράδειγμα:
    - fs=44100Hz, N = 512, NFFT=1024, hop_size=128
    - Θα δώσει για κάθε Δt=h/fs = 11.6ms audio
    - ένα φάσμα που θα αποτελείται από 1024 bins πλάτους **Δf = fs/NFFT** = 43Hz ομοιόμορφα κατανεμημένα στο διάστημα 0-fs (Hz), δηλαδή ***512 bins στο από 0- fs/2=22100Hz***

STFT - Algorithm

1. Τεμαχιοποίηση (segmenation) σήματος σε τμήματα (blocks) μήκους N
   - Όπου N – δύναμη κάποια του 2
2. Πολ/σμος κάθε block με κάποια παραθυρική συνάρτηση (Hamming, Hanning, κλπ.)
3. Προσθήκη μηδενικών για zero padding
   - Ώστε το μήκος του νέου zero-padded block NFFT να αντιστοιχεί και πάλι σε κάποια δύναμη του 2
4. Υπολογισμός fft για το zero-padded block
5. Κάθε επόμενο block εμφανίζει επικάλυψη με το προηγούμενο δηλαδή προκύπτει με μετακίνηση του δείκτη τρέχοντος δείγματος κατά κάποιον ακέραιο που λέγεται hop-size
6. Επανάληψη βήματος 2 έως 5 έως ότου τελειώσουν τα blocks του ηχητικού δείγματος

Εργαστηριακό Μέρος

- Άσκηση 5.1 - Αρχή Απροσδιοριστίας
  - Να διερευνήσετε την αρχή της απροσδιοριστίας χρησιμοποιώντας ένα σήμα δειγματοληψίας fs=1500Hz το οποίο προκύπτει από το άθροισμα δύο ημιτόνων με συχνότητες f1=250Hz και f2=270 Hz. Το φάσμα να απεικονισθεί για:
    - N=1024 δείγματα του σήματος
    - N=64 δείγματα
    - N = 64 δείγματα με zero-padding έως τα 512 δείγματα
- Άκηση 5.2 - Παραθυροποίηση και Διαρροή Φάσματος
  - Απεικονίστε το φάσμα ενός ημιτονοειδούς σήματος:
    - Όταν στη διάρκειά του περιλαμβάνεται ακέραιο πλήθος περιοδων (επαναλήψεων)
    - Όταν στη διαρκειά του δεν περιλαμβάνεται μη ακέραιο πλήθος περιόδων
    - Αφότου πολλαπλασιάσετε το σήμα με κάποια παραθυρική συνάρτηση
- Άσκηση 5.3 - Υλοποίηση STFT
  - Υλοποιήστε με δικό σας κώδικα τον αλγόριθμο STFT για ένα ημιτονοειδές σήμα με fs = 1000 f0 = 15 N = 256
- Άσκηση 5.4
  - Υλοποιήστε με δικό σας κώδικα τον αλγόριθμο STFT για το αρχείο piano.wav
- Άσκηση 5.5
  - Απεικονίστε το φασματογράφημα του αρχείου piano.wav χρησιμοποιώντας τη συνάρτηση matplotlib.pyplot.specgram