[Εξηγείστε περιεκτικά και επαρκώς την εργασία σας. Επιτρέπεται προαιρετικά η συνεργασία εντός ομάδων των 2 ατόμων. Κάθε ομάδα 2 ατόμων υποβάλλει μια κοινή αναφορά που αντιπροσωπεύει μόνο την προσωπική εργασία των μελών της. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός του βιβλίου και του εκπαιδευτικού υλικού του μαθήματος, πρέπει να το αναφέρετε. Η παράδοση της αναφοράς και του κώδικα της εργασίας θα γίνει ηλεκτρονικά στο mycourses.ntua.gr και επιπλέον η αναφορά της εργασίας θα παραδίδεται τυπωμένη και προσωπικά στην γραμματεία του εργαστηρίου Ρομποτικής (2.1.12, παλαιό Κτ.Ηλεκ.), ώρες 09.00-14.30.

Θέμα: Εισαγωγή στην Ψηφιακή Επεξεργασία Σημάτων με MATLAB και Εφαρμογές σε Ακουστικά Σήματα

Μέρος 1ο - Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής (Ενέργεια και Ρυθμός Εναλλαγής Προσήμου)

Οι μετρήσεις βραχέος χρόνου είναι μετρήσεις που γίνονται σε ένα μετακινούμενο παράθυρο του σήματος και είναι ιδιαίτερα χρήσιμες όταν θέλουμε να παρατηρήσουμε τοπικά χαρακτηριστικά αυτού του σήματος. Πιο συγκεκριμένα, η ενέργεια βραχέος χρόνου ορίζεται ως:

$$E_n = \sum_{m=-\infty}^{\infty} [x[m]w[n-m]]^2 \tag{1}$$

όπου $w[n]$ ένα παράθυρο της επιλογής μας, το οποίο συνήθως είναι το Hamming παράθυρο (συνάρτηση `hamming` στο MATLAB). Αντίστοιχα, ο ρυθμός εναλλαγής προσήμου (Zero Crossing Rate) ορίζεται ως:

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x[m]] - sgn[x[m-1]]|w[n-m] \tag{2}$$

1.1. Θεωρήστε το σήμα φωνής της πρότασης 'Όλα αυτά ήταν η άμυνα μες στο μυαλό μου' που περιέχεται στο αρχείο "speech_utterance.wav" (συχνότητα δειγματοληψίας: $f_s = 16$ kHz) του συμπληρωματικού υλικού "dsp18_lab1_Data.zip" της άσκησης στο mycourses. Προαιρετικά, ηχογραφήστε στον υπολογιστή σας (π.χ. με χρήση του εργαλείου `praat`) την εκφώνηση μιας πρότασης και διαβάστε το σήμα στο MATLAB (με χρήση της συνάρτησης `audioread`). Ο στόχος είναι να μετρήστε την ενέργεια βραχέος χρόνου και το ρυθμό εναλλαγής προσήμου. Χρησιμοποιήστε παράθυρο Hamming, μήκους 20-30 ms. Τί παρατηρείτε μεγαλώνοντας το μήκος του παραθύρου; Μπορείτε να χρησιμοποιήσετε αυτές τις μετρήσεις για να διαχωρίσετε φωνή από σιωπή ή έμφωνους (π.χ. /aa/,/ih/) από άφωνους ήχους (π.χ. /f/,/p/); Προαιρετικά, χρησιμοποιήστε το `praat` ως εργαλείο επισκόπησης του σήματος φωνής.

1.2. Επαναλάβετε την ίδια διαδικασία για το σήμα μουσικής "music.wav" που επίσης βρίσκεται στο συμπληρωματικό υλικό.

Σημειώνεται ότι σε κάθε περίπτωση οι απαντήσεις σας πρέπει να συνοδεύονται με τις σχετικές γραφικές παραστάσεις και σχόλια ώστε να είναι όσο το δυνατό τεκμηριωμένες.

Χρήσιμες Συναρτήσεις του MATLAB:

- **help**: Όλες οι συναρτήσεις του MATLAB έχουν αναλυτική επεξήγηση που μπορεί να εμφανιστεί με την εντολή >> *help* <*function*> στο MATLAB command prompt.
- **buffer**: Η MATLAB συνάρτηση **buffer()** είναι χρήσιμη για την κατάτμηση ενός σήματος σε παράθυρα δεδομένων διαστάσεων και ολισθήσεων.

Matlab Tutorial: http://cvsp.cs.ntua.gr/courses/dsp/Material/MATLAB_Tutorial_ntua_2010.pdf

Μέρος 2ο - Ανάλυση και Σύνθεση Σήματος με τον Μετ/σμό Fourier Βραχέος Χρόνου (STFT)

Ο STFT είναι ο πιο διαδεδομένος μετ/σμός για την μελέτη του συχνοτικού περιεχομένου χρονικά-μεταβαλλομένων σημάτων, αναλύοντάς τα κατά μικρά χρονικά διαστήματα με υπολογισμό του Μετ/σμού Fourier Διακριτού Χρόνου (DTFT) σε καθένα από αυτά. Αναλυτικά:

$$\mathrm{STFT}(\tau,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-\tau]e^{-j\omega n},\ \text{όπου } \tau \in Z.$$

Στην MATLAB υλοποίηση του STFT χρησιμοποιείται ο DFT (και όχι ο DTFT), οπότε η συχνότητα ω δειγματοληπτείται στα $\omega_k = 2\pi k/N$, όπου N το μήκος του DFT μεγαλύτερο ή ίσο από το μήκος L του παραθύρου που χρησιμοποιείται. Για την υλοποίηση του STFT, επιλέγουμε το είδος του κυλιόμενου παραθύρου (π.χ. τετραγωνικό ή Hamming κλπ.), το μήκος του παραθύρου (L), καθώς και το βήμα ολίσθησης του παραθύρου (R). Μια χρήσιμη ιδιότητα του Μετ/σμού STFT, είναι ότι μπορούμε να ανασυνθέσουμε το αρχικό σήμα, δεδομένου ότι τηρούνται κάποιες συνθήκες μεταξύ του L και του R για το συγκεκριμένο είδος παραθύρου (βλ. Παράρτημα Α).
Σε αυτήν την άσκηση θα υλοποιήσουμε τον STFT με την βοήθεια του οποίου θα παρατηρήσουμε πώς μεταβάλλεται το συχνοτικό περιεχόμενο του φωνητικού σήματος ‘speech_utterance.wav’ με το πέρασμα του χρόνου. Επίσης, θα υλοποιήσουμε τον αντίστροφο Μετ/σμό ISTFT, για να ανακτήσουμε πίσω στο χρόνο το αρχικό σήμα και να το ακούσουμε.

2.1. Υπολογισμός STFT: Χρησιμοποιήστε παράθυρο Hamming, με μήκος 40msec και βήμα ανάλυσης 20msec. Για κάθε χρονικό παράθυρο, υπολογίστε τον DFT του παραθυροποιημένου σήματος κάνοντας χρήση της συνάρτησης `fft`. Ο διδιάστατος πίνακας που προκύπτει στο τέλος, αντιστοιχεί στον Μετ/σμό STFT(τ, ω_k). Υλοποιείστε την συνάρτηση με όνομα `mySTFT.m` η οποία παίρνει σαν είσοδο το σήμα και τις μεταβλητές παραθύρου και επιστρέφει τον Μετ/σμό STFT.

2.2 Αναπαραστήστε το πλάτος $|\mathrm{STFT}(\tau, f)|$ με τις κατάλληλες τιμές στους άξονες του χρόνου και των συχνοτήτων, όπου f οι συχνότητες συνεχούς χρόνου (με $f_k = \omega_k f_s/2\pi$), με χρήση της συνάρτησης **surf**. Απομονώστε δύο χρονικά τμήματα του $|STFT(\tau, f)|$ τα οποία να αντιστοιχούν στο φωνήεν /α/ και δύο που να αντιστοιχούν στο φωνήεν /ο/ της δοσμένης πρότασης. Τι παρατηρείτε;

2.3. Ανακατασκευή αρχικού σήματος από τον STFT: Για την ανακατασκευή του αρχικού σήματος, ακολουθείται η αντίστροφη διαδιακασία. Για την ανακατασκευή του κάθε παραθύρου, εφαρμόζουμε αντίστροφο IDFT μετ/σμό με χρήση της συνάρτησης `ifft`. Για να είναι επιτυχής η ανακατασκευή του αρχικού σήματος από τα επιμέρους ανακατασκευασμένα γειτονικά παράθυρα, χρησιμοποιείται η τεχνική OverLap-Add (OLA, βλ. Παράρτημα A), που θεωρεί πως τα ανακατασκευασμένα πλαίσια εμφανίζουν την ίδια επικάλυψη με αυτή που θεωρήθηκε κατά τον υπολογισμό του STFT, οπότε το τελικό σήμα (με ακρίβεια μιας πολλαπλασιαστικής σταθεράς) προκύπτει με την κατάλληλη (αφού το κάθε πλαίσιο τοποθετηθεί σωστά στο χρόνο) πρόσθεση των επικαλυπτόμενων πλαισίων. Αφού ανακατασκευάσετε το αρχικό σήμα φωνής, ακούστε το με τη βοήθεια της συνάρτησης `sound` και αποθηκεύστε το σαν αρχείο ήχου με το όνομα 'speech_utterance_rec.wav', κάνοντας χρήση της συνάρτησης `audiowrite`.

2.4. Με χρήση της συνάρτησης `ola.m` που βρίσκεται στο συμπληρωματικό υλικό, επαληθεύστε ότι η συνθήκη OLA ικανοποιούνταν για τις παραμέτρους παραθύρου που χρησιμοποιήσατε προηγουμένως. Εναλλακτικά, δοκιμάστε να επαναλάβετε την διαδικασία ανάλυσης/ανακατασκευής, χρησιμοποιώντας Hamming παράθυρο με μήκος 40msec και βήμα ανάλυσης 30msec. Τι παρατηρείτε στο άκουσμα του νέου ανακατασκευασμένου σήματος;

Μέρος 3ο - Φασματική Ανάλυση Ημιτονοειδών και Ανίχνευση Απότομων Μεταβάσεων με τον Μετ/σμό Fourier Βραχέος Χρόνου (STFT) και τον Μετ/σμό Wavelets (διακριτοποιημένο DT-CWT)

Στην προηγούμενη άσκηση είδαμε ότι με τον STFT μπορούμε να αναλύσουμε το συχνοτικό περιεχόμενο των σημάτων σε μικρά χρονικά διαστήματα κάνοντας χρήση ενός κυλιόμενου παραθύρου. Το μήκος L του παραθύρου $w[n]$ που θα χρησιμοποιηθεί καθορίζει την σχέση μεταξύ της διακριτικής ικανότητας στη συχνότητα και της ανάλυσης στο χρόνο. Μικρό παράθυρο πετυχαίνει καλή ανάλυση στο χρόνο με όμως χειρότερη διακριτική ικανότητα στην συχνότητα, ενώ αντίστροφα, μεγάλο παράθυρο στο χρόνο πετυχαίνει καλή διακριτική ικανότητα στο πεδίο συχνοτήτων χάνοντας σε ανάλυση στον χρόνο. Ένα από τα χαρακτηριστικά του $STFT$ είναι ότι το μήκος L του παραθύρου w είναι σταθερό και επιλέγεται εξαρχής. Συνήθεις τιμές για εφαρμογές narrow-band επεξεργασίας φωνής είναι L=30-50msec (για wide-band $\approx$ 5msec).
Ο STFT είναι ο πιο διαδεδομένος χρονο-συχνοτικός μετ/σμός, ωστόσο υπάρχουν και άλλοι μετ/σμοί κατάλληλοι για την μελέτη του συχνοτικού περιεχομένου των σημάτων, οι οποίοι μάλιστα σε κάποιες περιπτώσεις μπορεί να υπερτερούν του STFT.
Ένας εναλλακτικός τρόπος αναπαράστασης των σημάτων είναι με τον Μετ/σμό των Wavelets. Στην περίπτωση σημάτων συνεχούς χρόνου ορίζεται ως:

$$\mathrm{CWT}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-\tau}{s}\right) dt,$$

όπου $\psi(t)$ είναι η βασική-μητρική συνάρτηση η οποία μπορεί να μετατοπιστεί κατά τ και να μεγεθυνθεί ή σμικρυνθεί κατά s. Στην περίπτωση διακριτών σημάτων, μπορεί να εφαρμοστεί η διακριτοποιημένη μορφή του DT-CWT (Discrete-time Continuous Wavelet Transform) (ρουτίνες cwt,cwtft στην MATLAB), η οποία αποτελεί παραλλαγή του γνωστου διακριτού μετ/σμού DWT. Στην περίπτωση των Wavelets δεν υπάρχει ο περιορισμός του σταθερού μήκους παραθύρου όπως συνέβαινε στον STFT.

3.1. Έστω ότι δειγματοληπτούμε με $F_s = 1000$Hz το σήμα $x(t)$ στο διάστημα $[0, 2]$ sec:

$$x(t) = 1.5\cos(2\pi 80t) + 2.5\sin(2\pi 150t) + 0.15v(t)$$

όπου $v(t)$ λευκός Gaussian θόρυβος μηδενικής μέσης τιμής (υλοποιείται με την συνάρτηση **randn** της MATLAB). Ισχύει ότι $x[n] = x(nT_s)$.

(α) Υπολογίστε το σήμα $x[n]$ και κάνετε την γραφική του παράσταση με χρήση της ρουτίνας **plot**.
(β) Υπολογίστε τον STFT με χρήση της ρουτίνας **spectrogram**. Χρησιμοποιείστε μήκος παραθύρου ίσο με 0.04sec και επικάλυψη ίση με 0.02sec. Αναπαραστήστε το πλάτος $|\mathrm{STFT}(\tau, f)|$ με τις κατάλληλες τιμές στους άξονες του χρόνου και των συχνοτήτων, όπου f οι συχνότητες συνεχούς χρόνου (με $f_k = \omega_k f_s / 2\pi$), με χρήση της συνάρτησης **surf**.
(γ) Υπολογίστε τον DT-CWT με χρήση της ρουτίνας **cwtft**, επιλέγοντας το "Morlet" wavelet. Για τον υπολογισμό των κλιμάκων s και των αντίστοιχων συχνοτήτων f χρησιμοποιείστε την συνάρτηση **wavescales**[1] (βρίσκεται στο συμπληρωματικό υλικό). Αναπαραστήστε το πλάτος $|\mathrm{DT\text{-}CWT}(\tau, f)|$ με τις κατάλληλες τιμές στους άξονες του χρόνου και των συχνοτήτων, με χρήση της συνάρτησης **surf**.
(δ) Τί παρατηρείτε;

3.2. Έστω ότι δειγματοληπτούμε με $F_s = 1000\mathrm{Hz}$ το σήμα $x(t)$ στο διάστημα $[0, 2]$ sec:

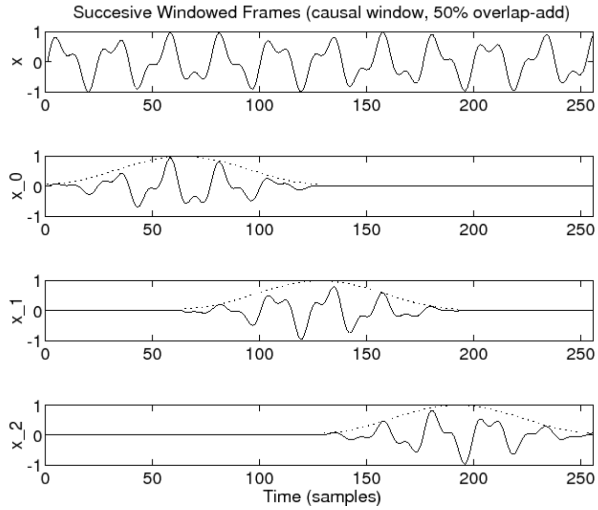
$$x(t) = 1.5\cos(2\pi 40t) + 1.5\cos(2\pi 100t) + 0.15v(t) + 5(\delta(t - 0.625) + \delta(t - 0.650))$$

όπου $v(t)$ λευκός θόρυβος μηδενικής μέσης τιμής όμοια με πριν. Ισχύει ότι $x[n] = x(nT_s)$. Στις χρονικές στιγμές 625msec και 650msec το σήμα είχε απότομες μεταβολές σε σύντομο χρονικό διάστημα, τις οποίες θέλουμε να εντοπίσουμε. Παράλληλα θέλουμε στην άναλυση μας να μελετήσουμε το συχνοτικό περιεχόμενο του σήματος (2 κύριες ημιτονικές συχνότητες).
(α) Υπολογίστε το σήμα $x[n]$ και κάνετε την γραφική του παράσταση με χρήση της ρουτίνας **plot**.
(β) Υπολογίστε τον STFT με χρήση της ρουτίνας **spectrogram**. Χρησιμοποιείστε τα εξής μήκη παραθύρου: i) 0.06sec, ii) 0.04sec, iii) 0.02sec, και επικάλυψη ίση με 50% κάθε φορά. Για κάθε περίπτωση παραθύρου, αναπαραστήστε το πλάτος $|\mathrm{STFT}(\tau, f)|$ με τις κατάλληλες τιμές στους άξονες του χρόνου και των συχνοτήτων, με χρήση της συνάρτησης **contour**.
(γ) Υπολογίστε τον DT-CWT με χρήση της ρουτίνας **cwtft**, επιλέγοντας το "Morlet" wavelet. Για τον υπολογισμό των κλιμάκων s και των αντίστοιχων συχνοτήτων f χρησιμοποιείστε την συνάρτηση **wavescales**. Αναπαραστήστε το πλάτος $|\mathrm{DT\text{-}CWT}(\tau, f)|$ με τις κατάλληλες τιμές στους άξονες του χρόνου και των συχνοτήτων, με χρήση της συνάρτησης **contour**.
(δ) Τί παρατηρείτε; Υπάρχει κάποιος μετ/σμός που να πετυχαίνει και τα δύο ζητούμενα δηλ. τον εντοπισμό και των δύο διακριτών απότομων μεταβολών καθώς και τις βασικές συχνότητες του σήματος;

ΠΑΡΑΔΟΤΕΑ (α) Συνοπτική αναφορά που θα απαντάει στα δοθέντα ερωτήματα και θα περιλαμβάνει τις ζητούμενες γραφικές αναπαραστάσεις (β) Ηλεκτρονική παράδοση του κώδικα MATLAB για όλες τις ασκήσεις.

[1] Η κλήση της συνάρτησης $[s, f]$= wavescales('morl',F_s) επιστρέφει το διάνυσμα κλιμάκων s και τις αντίστοιχες ψευδο-συχνότητες f για καλύτερη σύγκριση του CWT με τον STFT.

ΠΑΡΑΡΤΗΜΑ Α: Overlap-Add Ανάλυση



Σχήμα 1: Ανάλυση σε επικαλυπτόμενα πλαίσια.

Θεωρήστε την ανάλυση ενός σήματος $x[n]$ σε πλαίσια χρησιμοποιώντας ένα μηδενικής φάσης παράθυρο $w[n]$ πεπερασμένου μήκους L. Τότε μπορούμε να εκφράσουμε το m παραθυρωμένο πλαίσιο δεδομένων ως:

$$x_m[n] \triangleq x[n]w[n-mR], \quad n \in (-\infty, \infty)$$

όπου

$$R \triangleq \text{χρονικό βήμα ανάλυσης}, \; m \triangleq \text{αύξων δείκτης πλαισίου}$$

Το χρονικό βήμα ανάλυσης είναι ο αριθμός των δειγμάτων μεταξύ των χρόνων έναρξης διαδοχικών πλαισίων. Συγκεκριμένα, είναι ο αριθμός των δειγμάτων κατά τον οποίο μετακινούμε κάθε επόμενο παράθυρο. Στο Σχήμα 1 φαίνεται το σήμα εισόδου και τρία διαδοχικά παραθυρωμένα πλαίσια ανάλυσης χρησιμοποιώντας ένα αιτιατό παράθυρο Hamming μήκους $L = 128$ με 50% επικάλυψη ($R = L/2 = 64$).
Για να δουλέψει η ανάλυση σε πλαίσια θα πρέπει να μπορούμε να ανακατασκευάσουμε το σήμα $x[n]$ από τα επιμέρους επικαλυπτόμενα παράθυρα, ιδανικά με απλή πρόσθεσή τους στις αρχικές χρονικές τους θέσεις. Αυτό μπορεί να γραφτεί ως:

$$x[n] = \sum_{m=-\infty}^{\infty} x_m[n] = x[n] \sum_{m=-\infty}^{\infty} w[n-mR]$$

Οπότε, $x[n] = \sum_m x_m[n]$ αν και μόνο αν

$$\sum_{m \in \mathbf{Z}} w[n-mR] = 1, \forall n \in \mathbf{Z}.$$

Η συνθήκη αυτή θα πρέπει να ελέγχεται πριν την ανάλυση σε παραθυρωμένα επικαλυπτόμενα πλαίσια αν είναι αναγκαία η ανασύνθεση του συνολικού σήματος. Ελέγξτε για παράδειγμα με το script ola.m που σας δίνεται, ότι στην περίπτωση του παραθύρου Hamming, όπως ορίζεται στο MATLAB, υπάρχει πρόβλημα για $R = (L-1)/2$ και L περιττό.