

DSP - Project 3

Task 1

Έχοντας τα 3 ηχητικά σήματα¹ που ηχογραφήθηκαν για προηγούμενες ασκήσεις (φωνήματα /a/, /o/ και /e/), έγιναν τα εξής βήματα:

1. Για κάθε σήμα εφαρμόστηκε ο **FFT** (Fast Fourier Transform) σε ένα μέρος του (αποκοπή ενός μέρους των δειγμάτων από τη μέση του σήματος)
2. Υπολογίστηκε το **magnitude (polar representation)**:

$$\text{Mag}_{\text{signal}} = \sqrt{\text{fft_signal}_{\text{real}}^2 + \text{fft_signal}_{\text{imaginary}}^2}$$

3. Κρατώντας μόνο το πρώτο μισό των τιμών του magnitude, μέσα από μια επαναληπτική διαδικασία με χρήση κώδικα, σχεδιάζουμε τα 4 σημεία / bins με τη μεγαλύτερη ενέργεια, όπως φαίνεται και στα **Figures 1, 2 και 3**.

Λόγω της φύσης των παραπάνω ηχητικών σημάτων ως **σύνθετοι** ήχοι (ανθρώπινη φωνή σε μη - επαγγελματική ηχογράφηση) και όχι καθαροί τόνοι, το γράφημα που αναπαριστά το **magnitude** των 4 πιο “ισχυρών” ενεργειακά bins (**peaks**), δεν αποτελεί τον καταλληλότερο τρόπο εύρεσης της θεμελιώδους συχνότητας αυτών των σημάτων. Ο ήχος που αποτελεί φυσική ομιλία, ξεκινά από τους παλμούς των φωνητικών μας **χορδών**, οι οποίοι καθορίζουν την f_0 (fundamental frequency) του και εξέρχεται από το φωνητικό κανάλι, περνώντας από φιλτράρισμα ανάλογα με τις θέσεις και το σχήμα των **αρθρωτών**, αποκτώντας και **αρμονικές** συχνότητες (formants λόγω αντίχησης στην κοιλότητα). Αποτελεί **ψυχοακουστικό** χαρακτηριστικό, καθώς αντίληψή της διέπεται από **υποκειμενικότητα**, η οποία εξαρτάται πολλές φορές από διάφορα άλλα χαρακτηριστικά όπως η **ένταση** ή η **φυσιολογία** του ανθρώπου που την αντιλαμβάνεται. Στα γραφήματα που ακολουθούν στο πεδίο των **συχνοτήτων**, θα δούμε πως η θεμελιώδης συχνότητα (**pitch**) ενός ηχητικού σήματος, δεν αποτελεί πάντα την κορυφή με τη μεγαλύτερη **ενέργεια**, καθώς υπάρχουν και οι **αρμονικές** αυτής, που βρίσκονται σε συχνότητες με τιμή, ακέραιο πολλαπλάσιο της f_0 (κυρίως components f_1 και f_2). Επιπλέον, σε περιπτώσεις που το input sample του αλγορίθμου FFT δεν είναι ακέραιο πολλαπλάσιο της περιόδου του, παρατηρείται το φαινόμενο “**spectral leakage**”, παρουσιάζοντας παραποιημένες πληροφορίες για την f_0 . Αντ’ αυτού, υπάρχουν άλλες μέθοδοι όπως για παράδειγμα ο αλγόριθμος “**autocorrelation**” (αυτοσυσχέτισης), που είναι αρκετά αποδοτικός σε περιπτώσεις σημάτων με μεγάλη περιοδικότητα, όπως τα ηχητικά σήματα της άσκησης (έχει αναπτυχθεί κώδικας για την εύρεση του pitch μέσω της συνάρτησης **librosa.autocorrelation()**).

Το αποτέλεσμα της εφαρμογής της συνάρτησης **autocorrelation()** της βιβλιοθήκης **librosa** ως προς τη θεμελιώδη συχνότητα των σημάτων, ήταν τα παρακάτω:

- f_0 του σήματος με το φώνημα / a / : **233.33 Hz**
- f_0 του σήματος με το φώνημα / o / : **219.40 Hz**
- f_0 του σήματος με το φώνημα / e / : **297.97 Hz**

¹ Χρησιμοποιήθηκαν μόνο τα **φωνήεντα**, καθώς τα σύμφωνα που είχαν επιλεγεί ανήκουν στην κατηγορία των μη ηχηρών φθόγγων, των οποίων τα σήματα δεν έχουν περιοδικότητα συνεπώς ούτε και θεμελιώδη συχνότητα, εφόσον κατά την εκφώνησή τους δεν πάλλονται οι φωνητικές χορδές.

Αξιοποιώντας το λογισμικό Praat, έγιναν επιπλέον μετρήσεις για τα formants και την f_0 των σημάτων:

- φώνημα / **a** /
 - f_0 : **233.33 Hz**
 - f_1 :
- φώνημα / **o** /
 - f_0 : **219.40 Hz**
 - f_1 :
- φώνημα / **e** /
 - f_0 : **297.97 Hz**
 - f_1 :

Τα σήματα των φωνημάτων **a**, **o** και **e**, είχε υπολογιστεί “με το μάτι” πως είχαν αντίστοιχα συχνότητες **~230 Hz**, **~217 Hz** και **~294 Hz** . Τα γραφήματα **1**, **2** και **3**, μας οδηγούν στα εξής συμπεράσματα:

- Σε όλες τις περιπτώσεις εντοπίζεται η θεμελιώδης συχνότητα (ή τουλάχιστον μια προσεγγιστική τιμή της)

Task 2

Αρχικά δημιουργήθηκε το τεχνητό σήμα

$$s = s_1 + s_2$$

όπου

$$s_1 = \sin(2\pi f_1 t), \quad f_1 = 50Hz$$

$$s_2 = \sin(2\pi f_2 t), \quad f_2 = 70Hz$$

με συχνότητα δειγματοληψίας $44100Hz$.

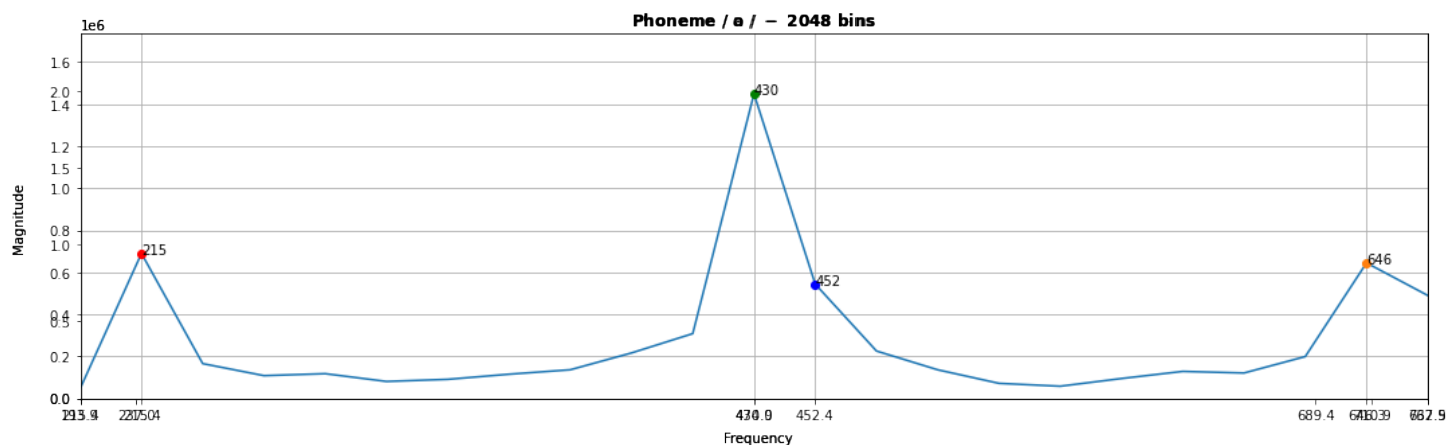


Figure 2

Task 3

Notes

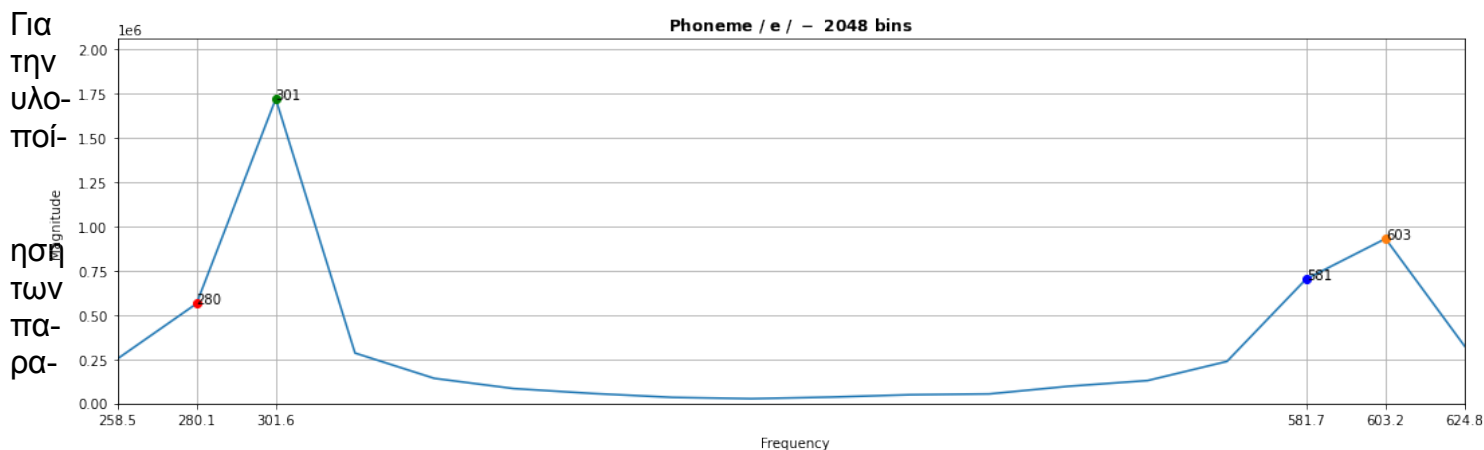


Figure 3

πάνω ζητούμενων, αναπτύχθηκε κώδικας που θα σταλεί μαζί με το παρόν αρχείο.