Επεξεργασία Φωνής και Ήχου

Project

2/8/2016

Θεμελής Κωνσταντίνος – Καποδίστρια Αγγελική

Μέρος Α – Βασικοί Αλγόριθμοι Εκτίμησης Παραμέτρων Φωνής

Α-1. Ηχογράφηση

Στόχος του πρώτου μέρους ήταν η μελέτη κάποιων βασικών αλγορίθμων εκτίμησης παραμέτρων. Για τον σκοπό αυτό, ηχογραφήθηκε ένα σήμα φωνής, συγκεκριμένα το όνομα «Αγγελική Καποδίστρια» (αρχείο voicename_female.wav), με συχνότητα δειγματοληψίας τα 16kHz. Η ηχογράφηση του και η αποθήκευση του έγιναν με τον παρακάτω κώδικα ενώ ακολουθεί και η γραφική παράσταση του σήματος:

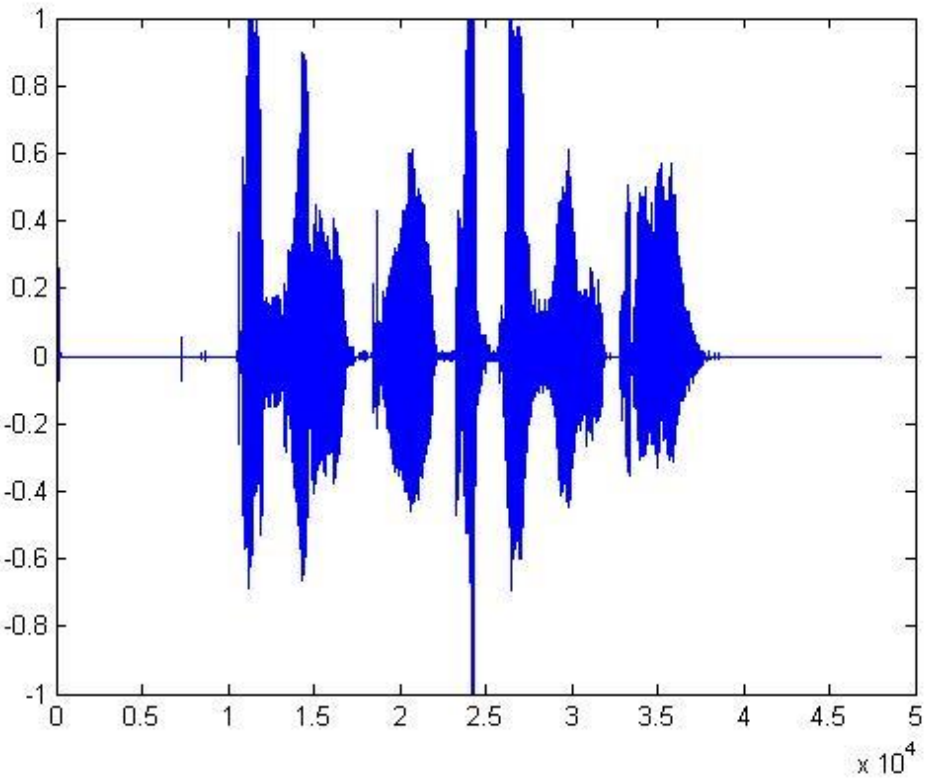
```
pause(1);
recObj = audiorecorder(16000,8,1);

disp('Start speaking.')
recordblocking(recObj, 3);
disp('End of Recording.');

data = getaudiodata(recObj);

audiowrite(' voicename_female.wav ', data, 16000);
```

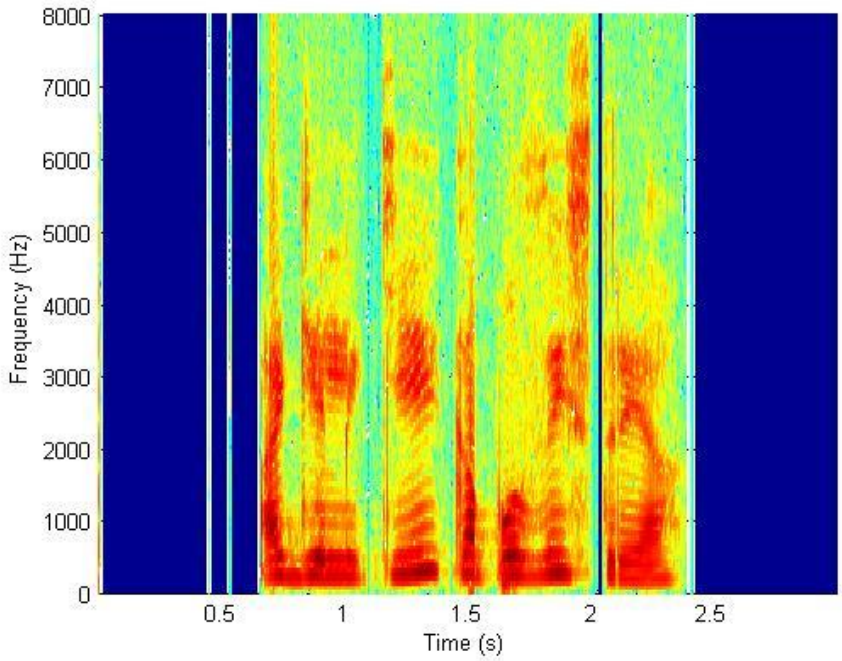
(Το τμήμα αυτό είναι σε σχόλια έτσι ώστε να μην χρειάζεται ξανά και ξανά ηχογράφηση, αρκεί να διαβάσουμε το αποθηκεύμενο σήμα ήχου.)



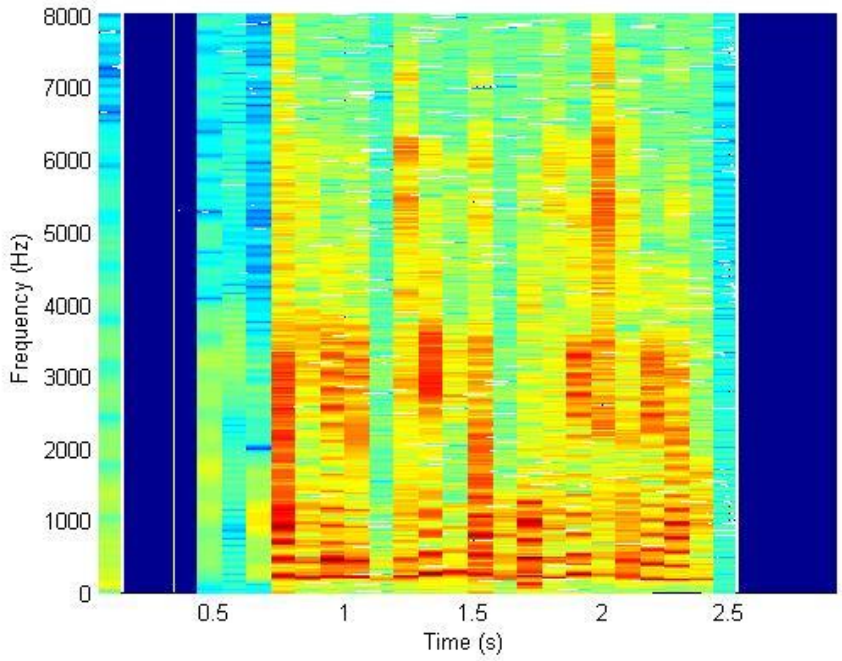
Εικόνα 1: Signal Name "Angeliki Kapodistria"

Α-2. Φασματόγραμμα Σήματος Φωνής

Για τον υπολογισμό του φασματογράματος, χρησιμοποιήθηκαν παράθυρα Hamming των 10msec και των 100msec με μετατόπιση 5msec και στις δύο περιπτώσεις και συχνότητα δειγματοληψίας όπως και πριν 16KHz. Τα spectrograms και για τα δύο παράθυρα είναι τα ακόλουθα:



Εικόνα 2: Φασματόγραμμα για παράθυρο 10msec



Εικόνα 3: Φασματόγραμμα για παράθυρο 100msec

Όπως φαίνεται και από τα σχήματα, στον άξονα των x έχουμε το χρόνο, στον άξονα των y τη συχνότητα και η ποσότητα της ενέργειας αναπαρίσταται με χρώματα. Κόκκινες περιοχές υποδηλώνουν υψηλή ενέργεια, για παράδειγμα κλείσιμο φωνητικών πτυχών, ενώ περιοχές με γαλάζιο ή πράσινο χρώμα είναι περιοχές με λιγότερη ενέργεια ή περιοχές σιωπής.

Παρατηρώντας τα δύο φασματογράμματα, βλέπουμε πως η χρήση διαφοερτικού μήκους παραθύρων μας βοηθάει στη μελέτη διαφορετικών χαρακτηριστικών. Στην Εικόνα 2, έχουμε ένα wideband spectrogram, δηλαδή χρησιμοποιούμε μικρά τμήματα του σήματος τα οποία μας επιτρέπουν να διερευνήσουμε τα χαρακτηριστικά της φωνητικής οδού. Από την άλλη, στην Εικόνα 3, έχουμε ένα narrowband spectrogram και μεγάλα τμήματα του σήματος που διευκολύνουν την διερεύνηση χαρακτηριστικών πηγής.

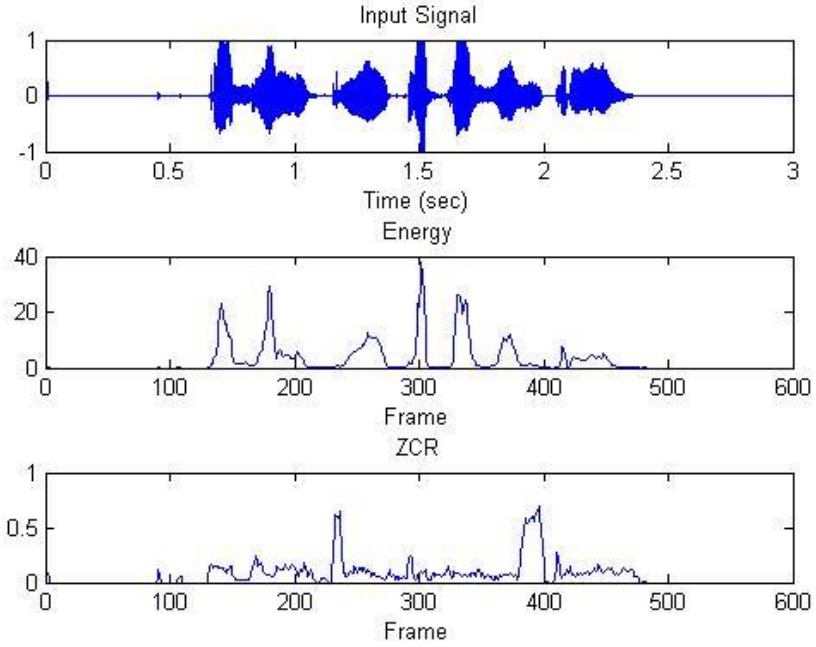
Α-3. Έκτίμηση Έμφωνων/Άφωνων Τμημάτων Σήματος

Για τον διαχωρισμό των τμημάτων του σήματος σε voiced, unvoiced και silence χρησιμοποιήθηκαν η ενέργεια του σήματος και το zero crossing rate. Η συνάρτηση detectVUS ορίζει τιμές κατωφλίου τόσο για την ενέργεια όσο και για το zero crossing rate και με βάσει τις τιμές αυτές κατατάσσει τα τμήματα ως εξής:

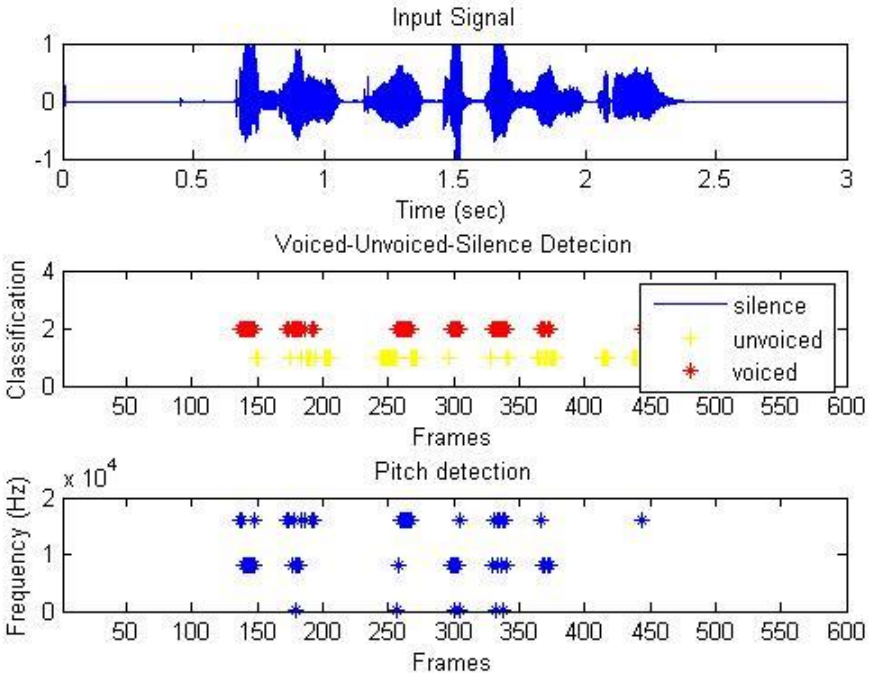
Energy (Short - Time)	**Zero Crossing Rate**	**Decision**
High	Low	Voiced
High	Close to 0	Voiced
High	High	Voiced
Low	Low	Voiced
Low	High	Unvoiced
Low	Close to 0	Unvoiced
Close to 0	Close to 0	Silence
Close to 0	Low	Silence

Στην Εικόνα 4 παρουσιάζεται το σήμα, η ενέργεια του και το zero crossing rate ενώ στην Εικόνα 5 έχει σχεδιαστεί ηαπόφαση του αλγορίθμου σε voiced, unvoiced και silence τμήματα.

(Τα silence τμήματα είναι με μπλε γραμμή στην γραμμή του άξονα και δεν είναι ευδιάκριτα)



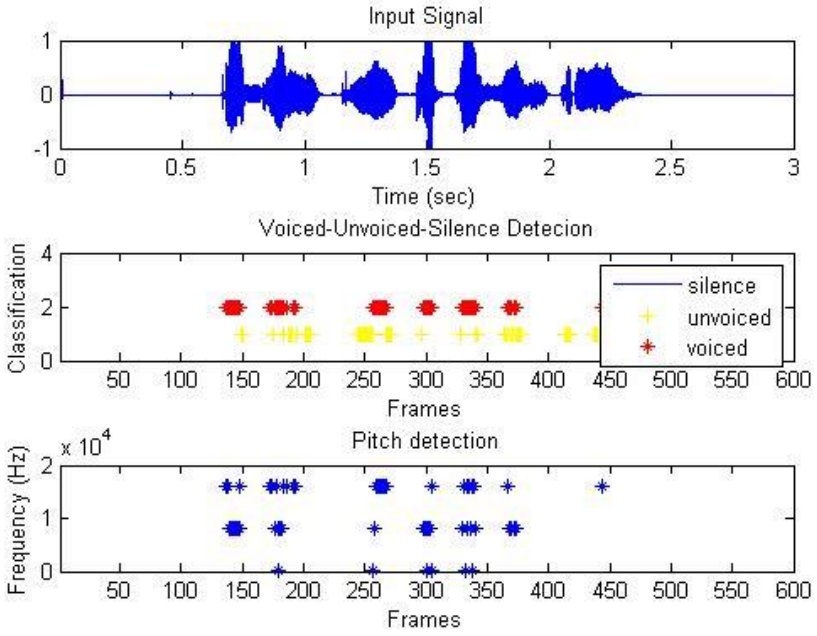
Εικόνα 4: Signal - Energy – ZCR



Εικόνα 5: Voiced-Unvoiced-Silence Parts and Pitch

Α-4. Εκτίμηση Θεμελιώδους Συχνότητας Διέγερσης

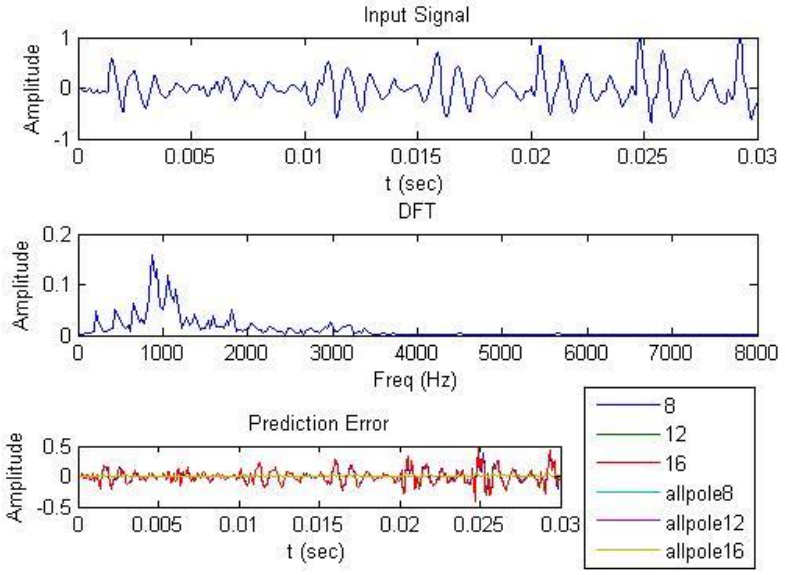
Η θεμελιώδης συχνότητα βρέθηκε με τη χρήση φάσματος και τα αποτελέσματα της μεθόδου φαίνονται στην Εικόνα 6 που ακολουθεί:



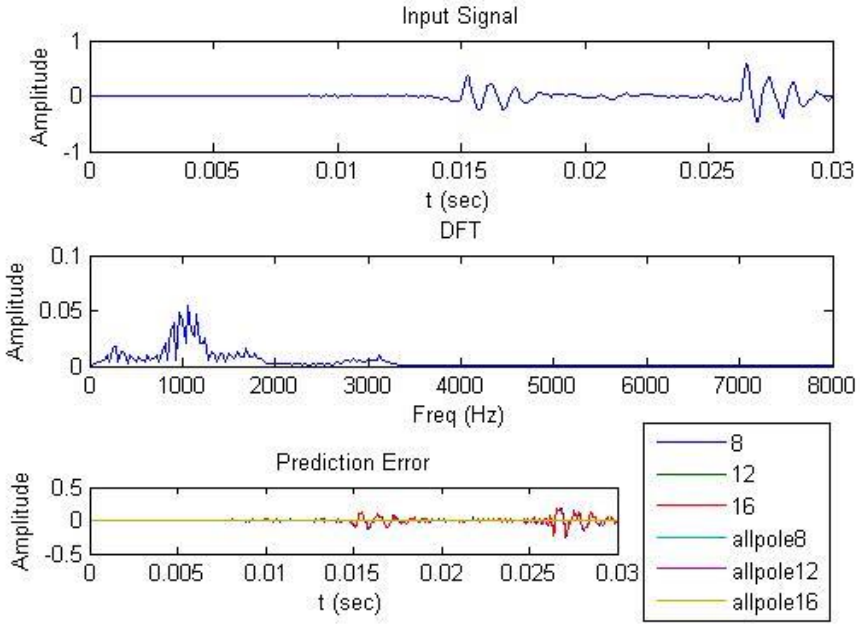
Εικόνα 6: Pitch

Α-5. Γραμμική Πρόβλεψη Σήματος

Έχοντας απομονώσει τα ζητούμενα τμήματα του σήματος, υπολογίζουμε το διάνυσμα χαρακτηριστικών LPC για τάξη φίλτρου 8, 12, 16 με τη χρήση της συνάρτησης LPC και εν συνεχεία τον διακριτό μετασχηματισμό Fourier των τμημάτων αυτών. Στις Εικόνα 7 και Εικόνα 8 παρουσιάζονται το λάθος εκτίμησης, ο διακριτός μετασχηματισμός Fourier και το μέτρο της all-pole συνάρτησης μεταφοράς για το voiced και unvoiced κομμάτι αντίστοιχα.



Εικόνα 7: DFT - Predicton Error - All-pole for Voiced

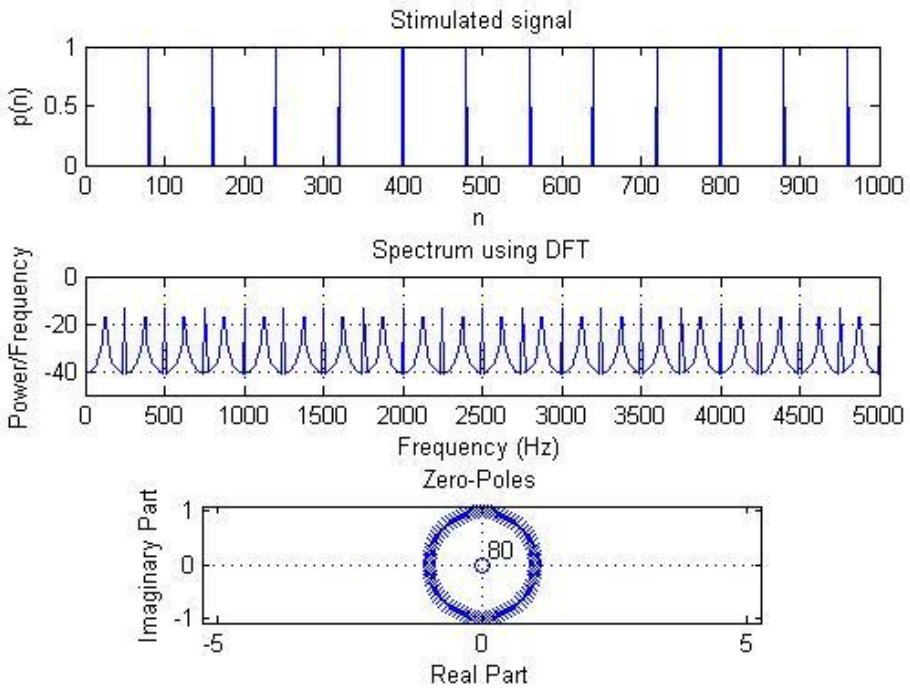


Εικόνα 8: DFT - Prediction Error - All-Pole for Unvoiced

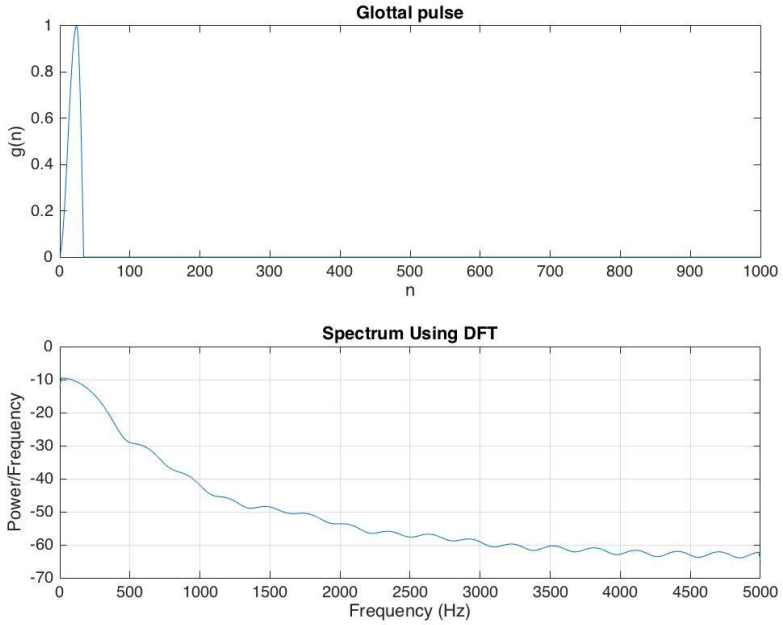
Μέρος Β – Απλοϊκό Σύστημα Σύνθεσης Φωνήματος

(Για το τρέξιμο του part B χρειάζεται να τρέξει το partB_ao.m, στο οποίο δίνεται σαν όρισμα εξαρχής το N_p. Αρχικά, η τιμή του είναι 80 όπως αναφέρεται. Η συνάρτηση αυτή καλεί όλες τις υπόλοιπες για τη δημιουργία του φωνήματος, απλώς πρέπει να είναι σε σχόλια οι εντολές για τις γραφικές παραστάσεις και για την αναπαραγωγή του κάθε φωνήεντος. Οι υπόλοιπες συναρτήσεις τρέχουν μεμονωμένα κανονικά και δείχνουν την παραγωγή του κάθε φωνήεντος αν χρειαστεί.)

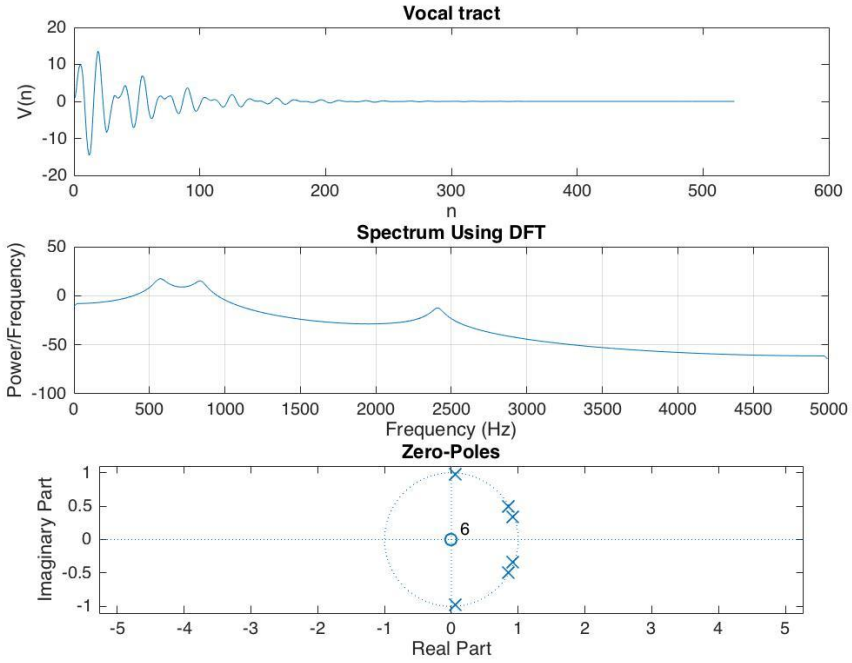
Έχοντας κάνει τα βήματα που περιγράφονται, δημιουργήθηκε το φωνήνεν /AO/. Στις παρακάτω εικόνες (Εικόνα 9 –Εικόνα 13) παρουσιάζονται με την σειρά το σήμα διέγερση, ο γλωτιδιακός παλμός, το σύστημα της φωνητικής οδού, το φορτίο ακτινοβολίας και το σήμα για το φωνήεν /AO/ μαζί με το φάσμα και τα διαγράμματα πόλων και μηδενικών κάθε φορά.



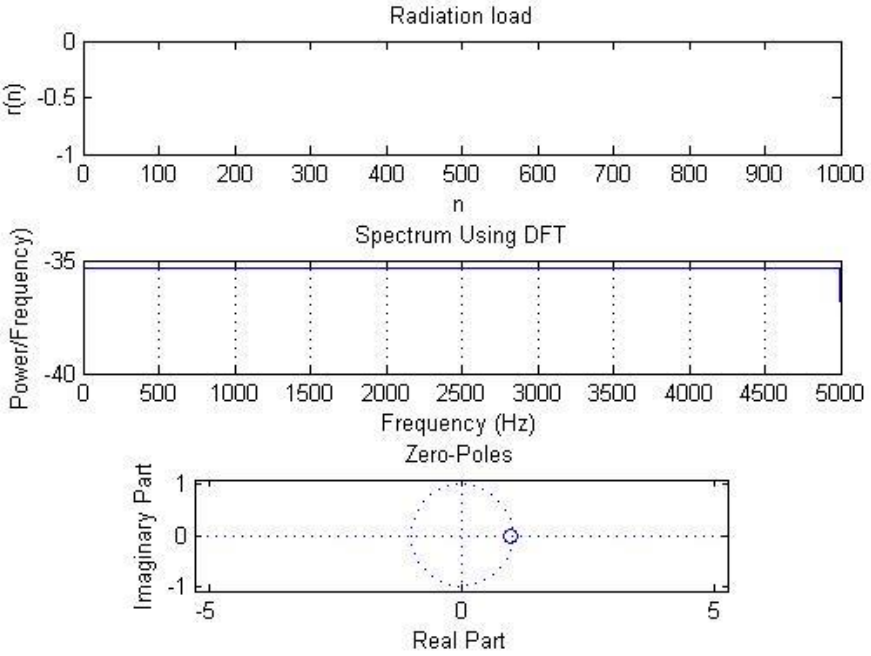
Εικόνα 9: Stimulated Signal



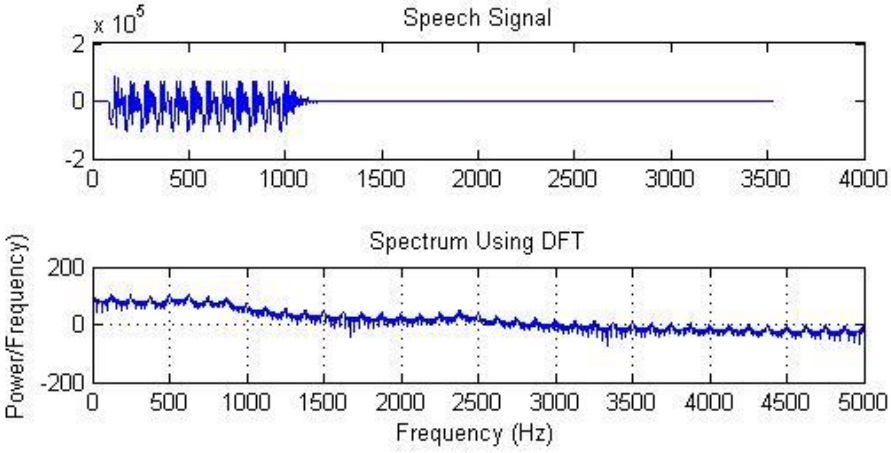
Εικόνα 10: Glottal Pulse



Εικόνα 11: Vocal Tract

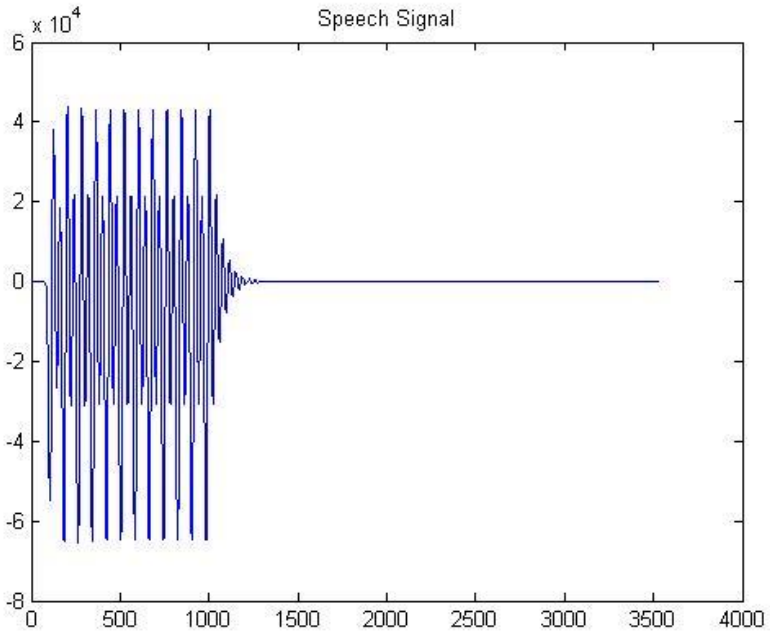


Εικόνα 12: Radiation Load

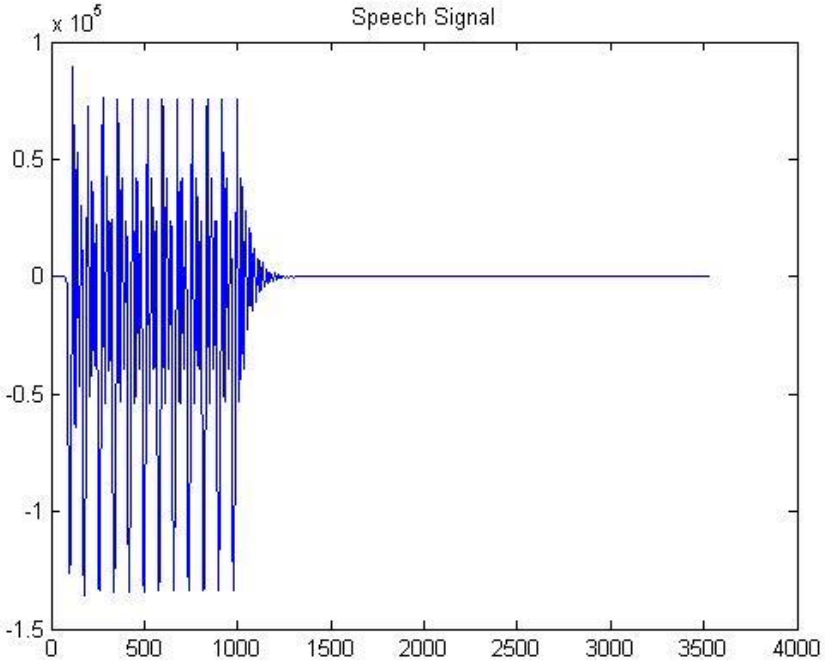


Εικόνα 13: /AO/

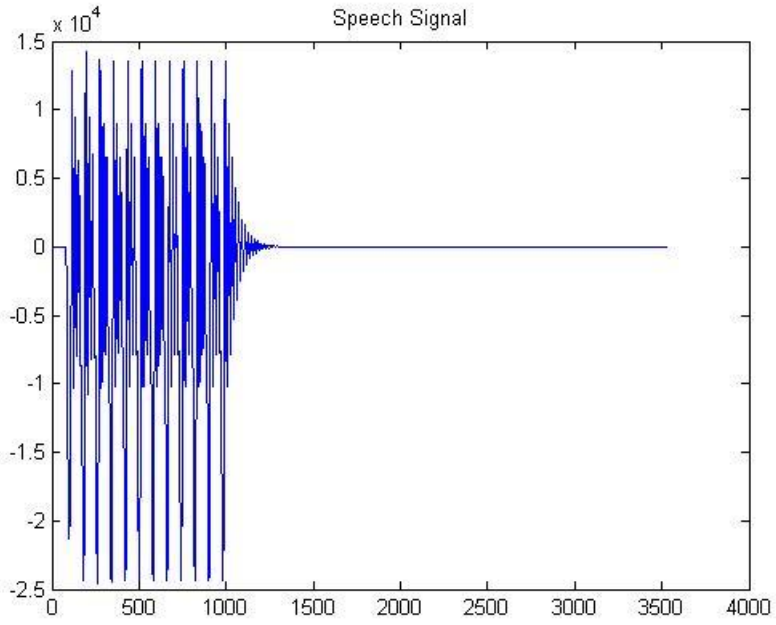
Χρησιμοποιώντας τα κατάλληλα F_1, F_2, F_3 δημιουργήθηκαν και τα υπόλοιπα φωνήεντα, τα s[n] των οποίων είναι τα ακόλουθα:



Εικόνα 14: /IY/



Εικόνα 15: /UH/



Εικόνα 16: /EH/

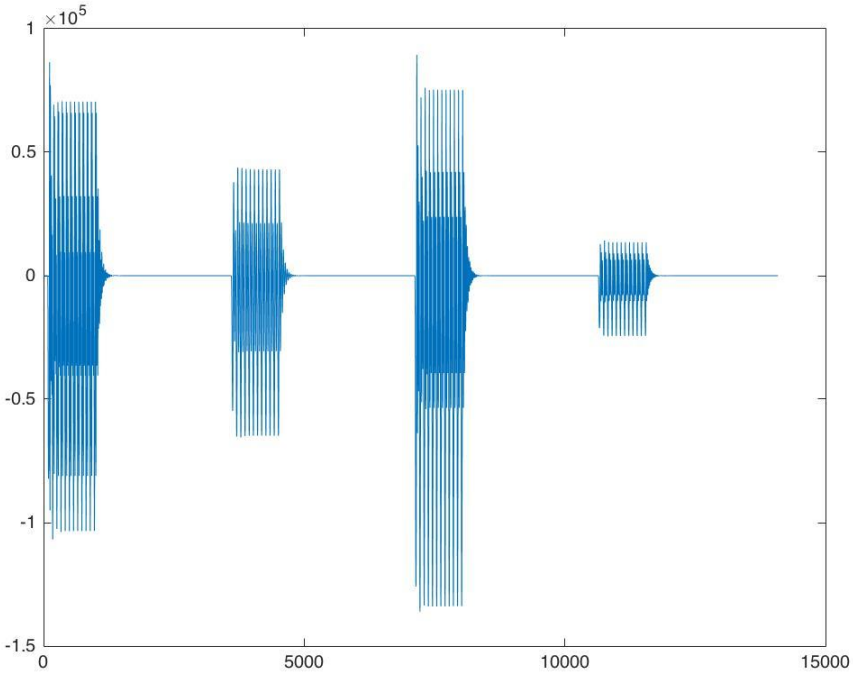
Η διαφορά που παρατηρείται είναι το πόσο είναι το πλάτος του κάθε σήματος. Για παράδειγμα τα /IY/ και /EH/ είναι της τάξης 10^4, ενώ τα άλλα δύο είναι της τάξης 10^5.

Για τη δημιουργία και την εγγραφή σε αρχείο ήχου του τελικού σήματος, χρησιμοποιήθηκε ο παρακάτω κώδικας:

```
% Signal concatenation
final_signal = [s_ao  s_iy  s_uh  s_eh];
k = audioplayer(final_signal, 16000);
play(k);

% Plot the final signal and create a .wav file
n=1:length(final_signal);
figure('name', 'Final Signal');
plot(n, final_signal)
audiowrite('final_signal.wav', final_signal, 16000);
```

Ο παραπάνω κώδικας είχε ως αποτέλεσμα την δημιουργία του σήματος:



Εικόνα 17: Final Signal

Όσον αφορά το τελευταίο ερώτημα, ο διπλασιασμός της συχνότητας, που θα οδηγήσει σε $N_p = 40$, έχει ως αποτέλεσμα διαφορετική ποιότητα και καθαρότητα στον ήχο. Από όσα μπορέσαμε να συμπεράνουμε, ο ήχος έτσι μας φάνηκε περισσότερος κατανοητός και ξεκάθαρος με το παραπάνω N_p.