

Πληροφορίες

Μάθημα: ΜΔΕ671 Προχωρημένα Θέματα Επεξεργασίας Φωνής και Λόγου

Υπολογιστική Εργασία

Καθηγητής: Ποταμιάνος Γεράσιμος

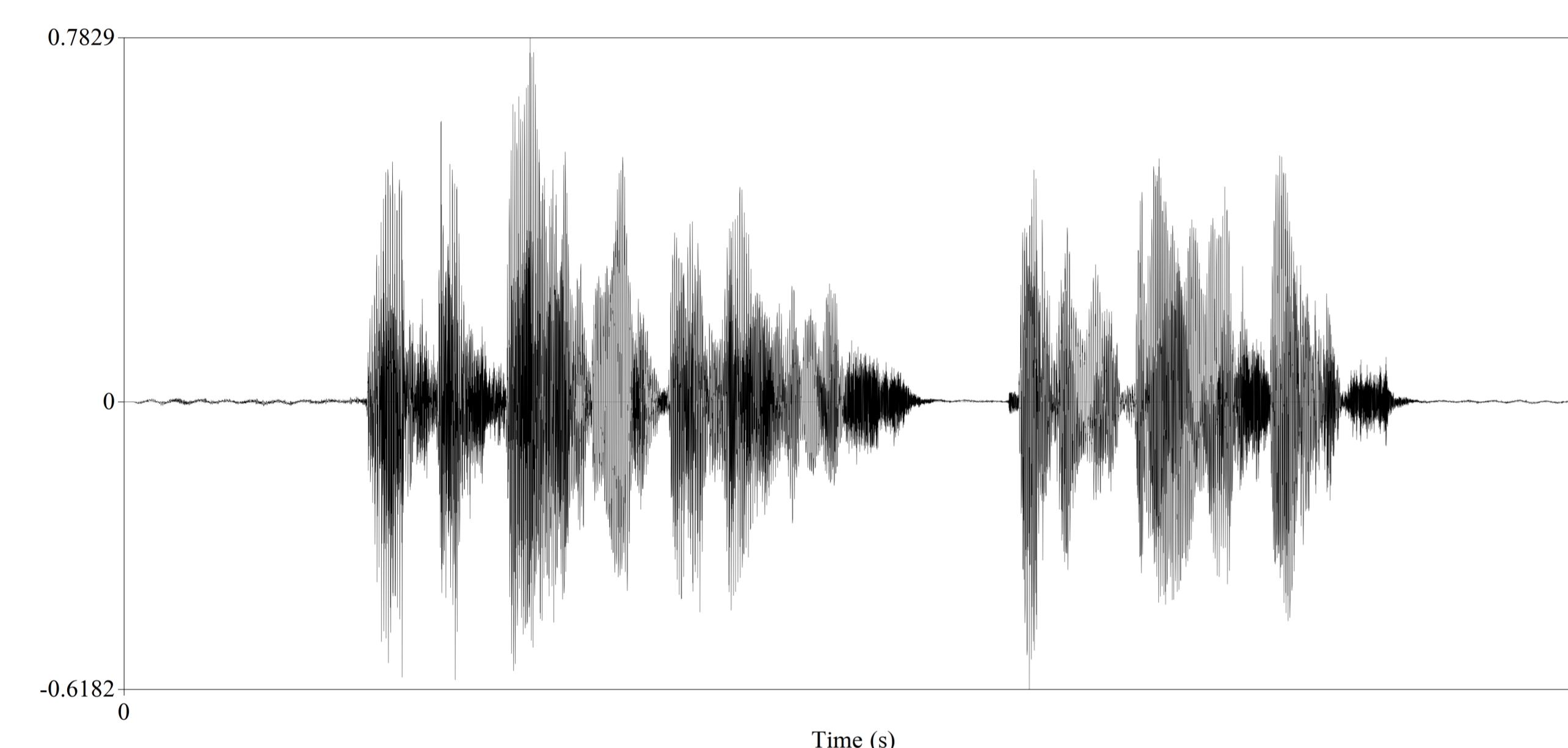
Ακαδημαϊκό Έτος: 2023-24

Ημερομηνία υποβολής: 25 Φεβρουαρίου 2024

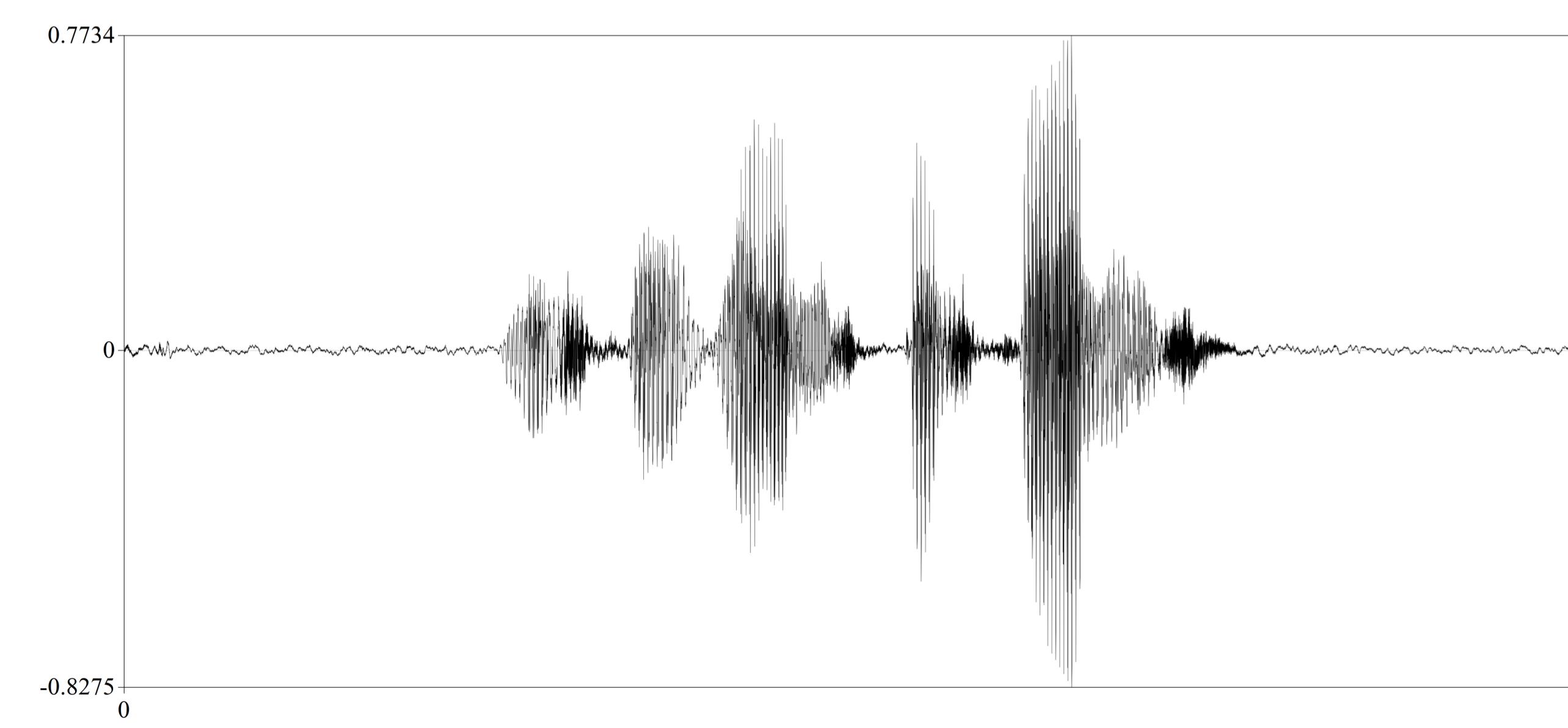
Ηχογράφηση

Για την ηχογράφηση του ονοματεπώνυμου μας με χρήση του εργαλείου **SoX** και συχνότητα δειγματοληψίας τα 16kHz και βάθος δείγματος 16 bits, διάρκειας 5 δευτερολέπτων:

```
sox -t alsa default -c 1 -r 16000 -b 16 -e signed-integer -t wav  
recorded_audio.wav remix 1 trim 0 5
```



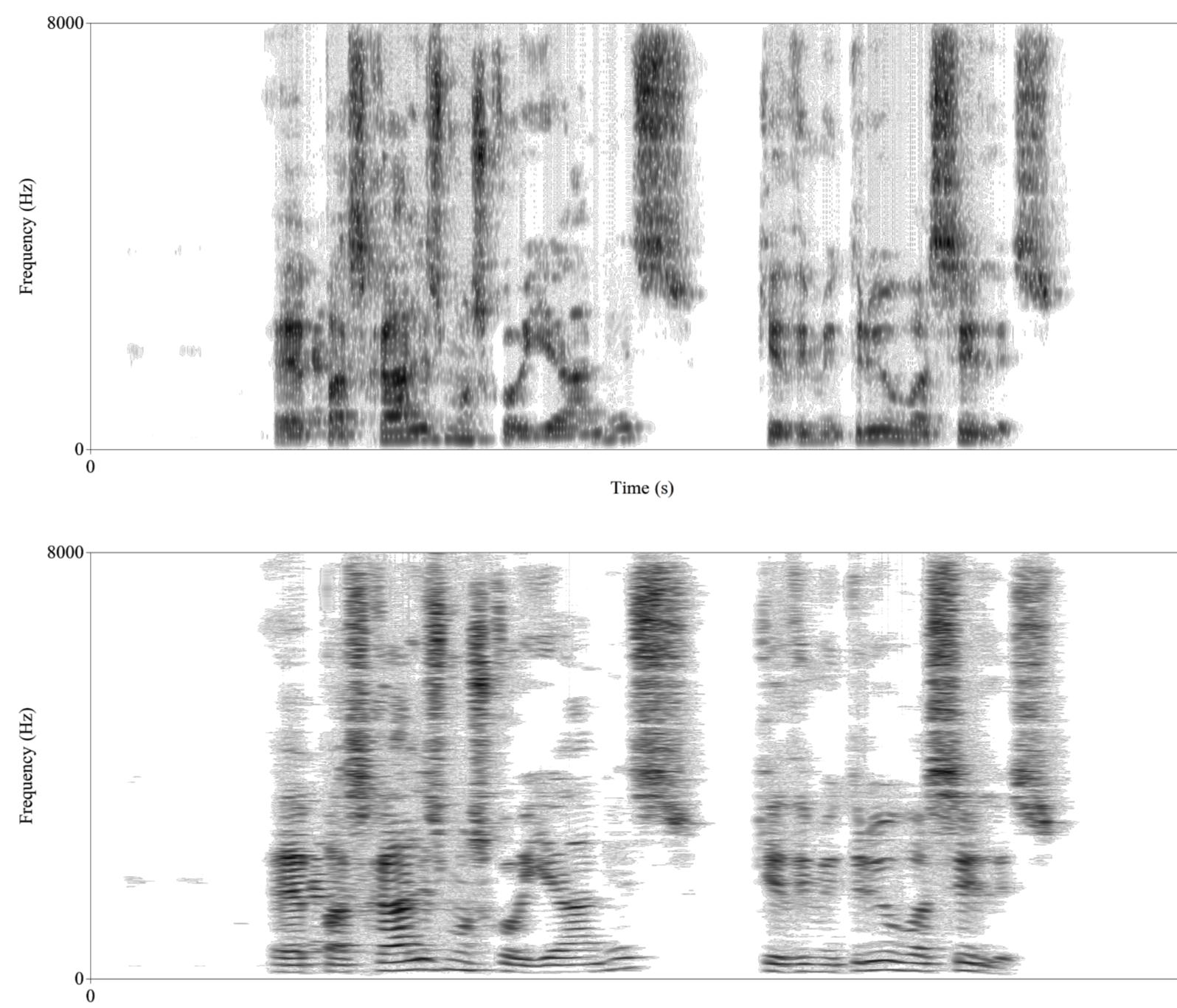
Εικόνα 1. Αποτύπωση της κυματομορφής του ηχητικού σήματος της φράσης "Μοσχογιάννης Πασχάλης και Δημητρίου Βασίλης" στο Praat [1].



Εικόνα 2. Αποτύπωση του ηχητικού σήματος του ενός ονοματεπώνυμου "Μοσχογιάννης Πασχάλης" στο Praat [1].

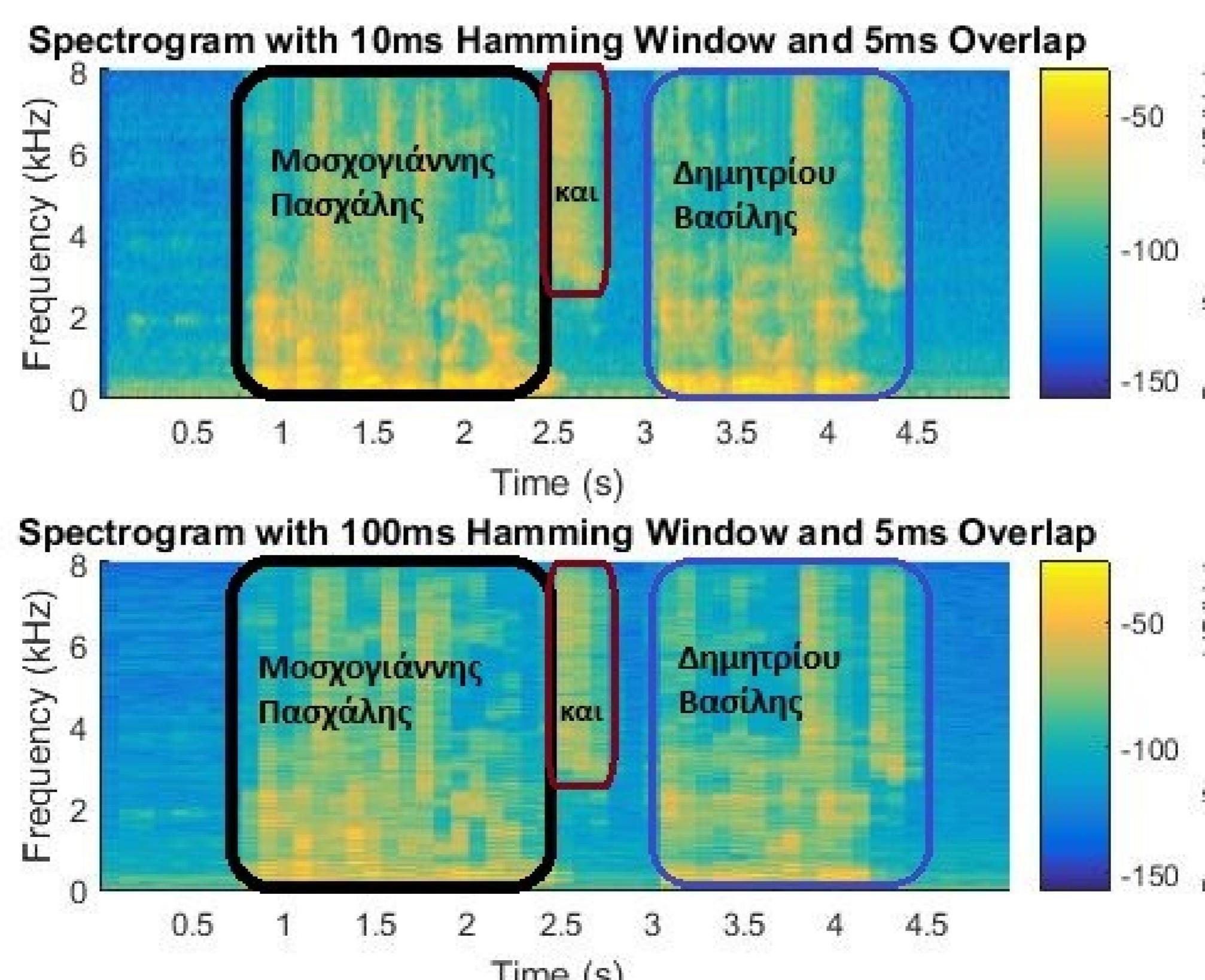
Φασματόγραμμα

Η δημιουργία φασματογραμμάτων μέσω της χρήσης μετασχηματισμού **Fourier** σε ηχητικά σήματα επιτρέπει την ανάλυση της συχνοτικής περιεκτικότητας σε σχέση με τον χρόνο, παρέχοντας μια οπτική απεικόνιση των αλλαγών συχνοτήτων καθ' όλη τη διάρκεια της εγγραφής. Η χρήση διαφορετικών μεγεθών παραθύρων **Hamming** επηρεάζει την ανάλυση, με τα μικρότερα παράθυρα να βελτιώνουν τη χρονική λεπτομέρεια και τα μεγαλύτερα τη συχνοτική λεπτομέρεια.



Εικόνα 3. Φασματογράμματα για 10msec και 100msec κατά αντιστοιχία στο Praat [1] για τη φράση "Μοσχογιάννης Πασχάλης και Δημητρίου Βασίλης".

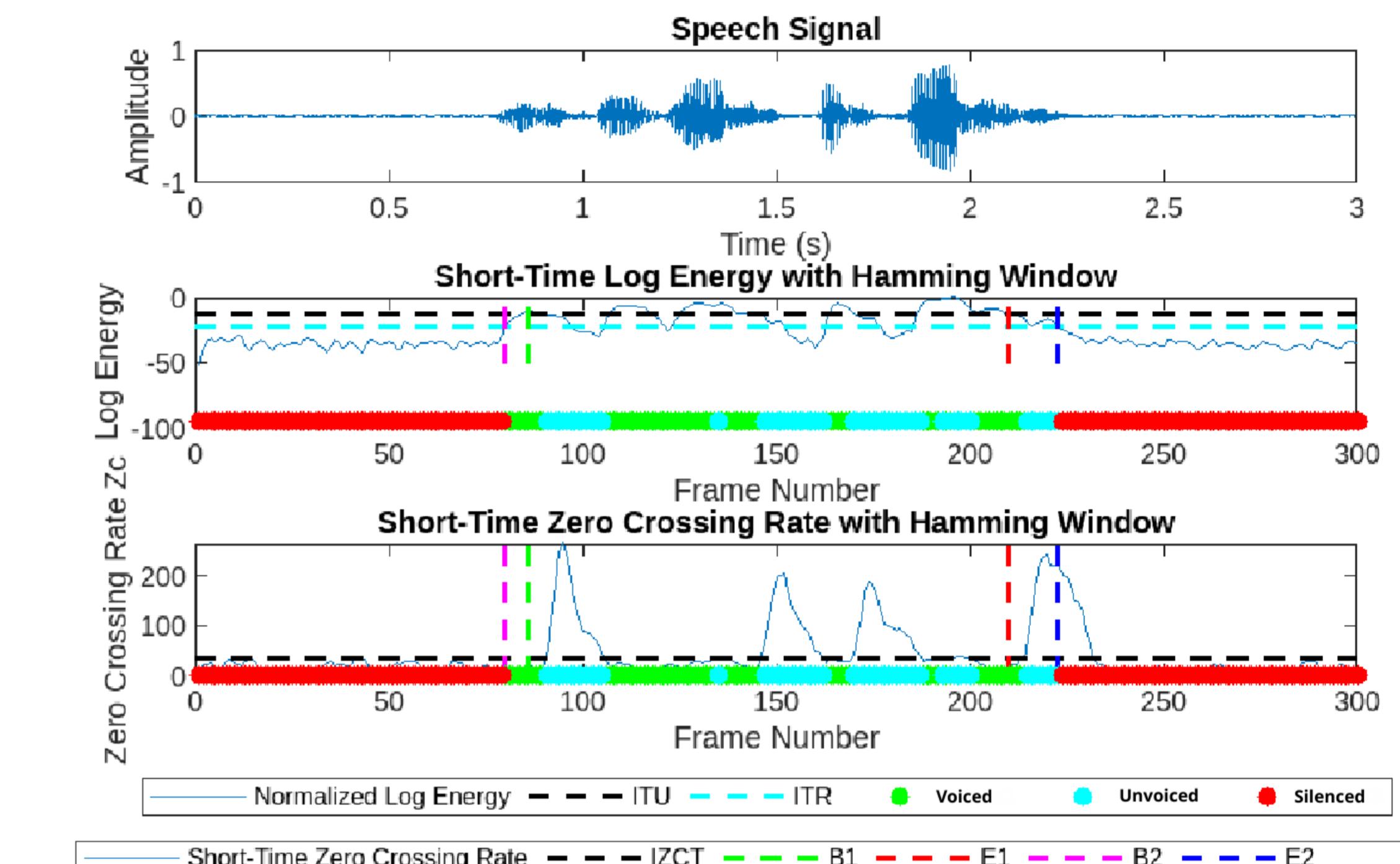
Στο περιβάλλον του MATLAB στην Εικόνα 4 διακρίνονται τα ίδια φασματογράμματα με τον άξονα X να αντιπροσωπεύει τον χρόνο και τον άξονα Y τις συχνότητες. Η ένταση κάθε συχνότητας σε κάθε χρονική στιγμή απεικονίζεται με διαφορετικές αποχρώσεις ή χρώματα.



Εικόνα 4. Υπόδειξη του περιεχομένου (λέξεις) των φασματογραμμάτων.

Διαχωρισμός Φωνητικών, Άφωνων και Σιωπηλών Ήχων

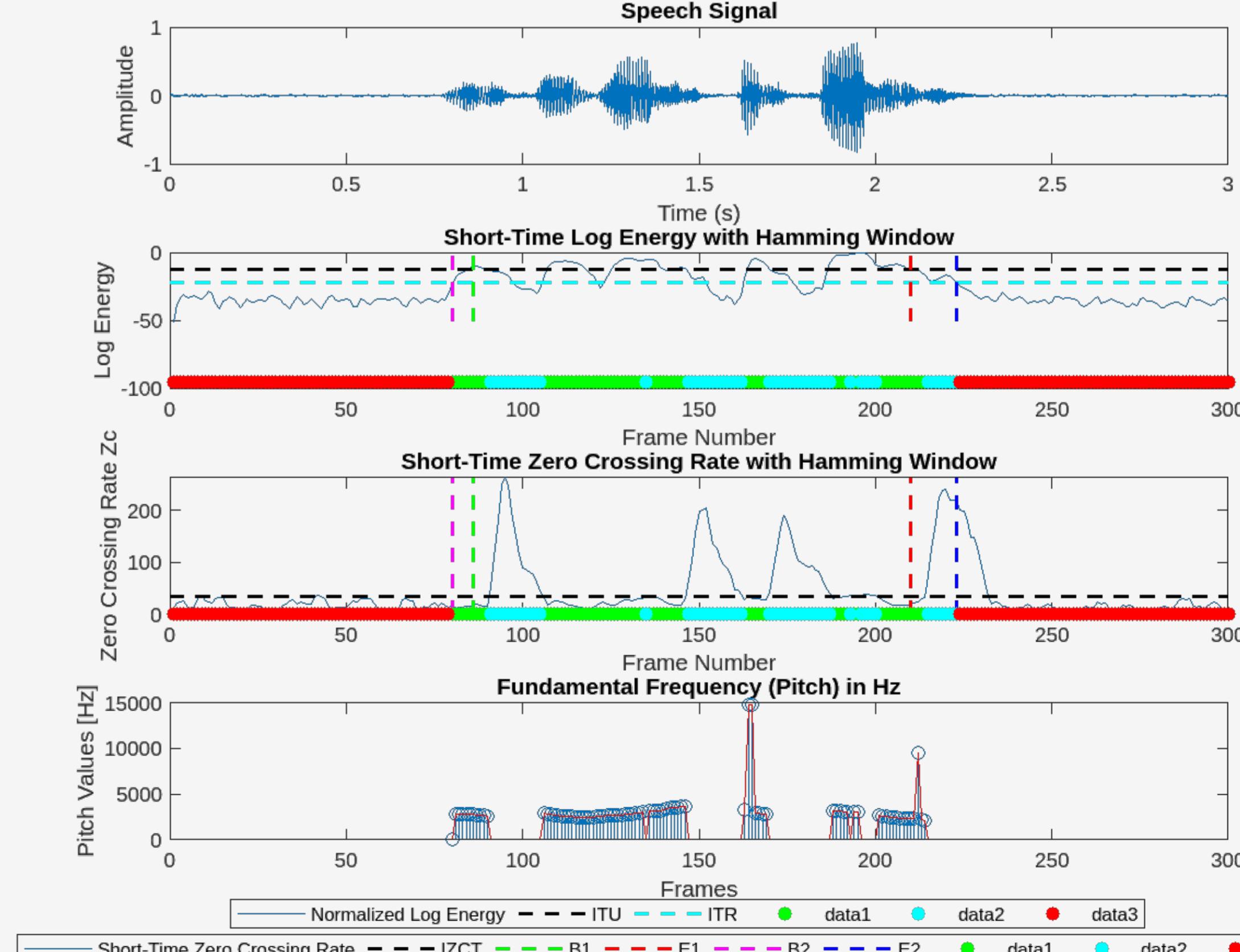
Στην περύπτωση του ηχητικού σήματος του ενός ονοματεπώνυμου αναπτύχθηκε αλγόριθμος για τον διαχωρισμό των φωνητικών, άφωνων και σιωπηλών τμημάτων ενός ηχητικού σήματος, χρησιμοποιώντας την ανάλυση της ενέργειας και του ρυθμού διαβάσεων μηδενικής τιμής (**Zero-Crossing Rate, ZCR**), βελτιώνοντας την ακρίβεια στην επεξεργασία και ανάλυση ομιλίας.



Εικόνα 5. Ανάλυση της ενέργειας και του ρυθμού διαβάσεων μηδενικής τιμής (Zero-Crossing Rate, ZCR).

Ανάλυση Pitch με Αυτοσυσχέτιση και Hamming

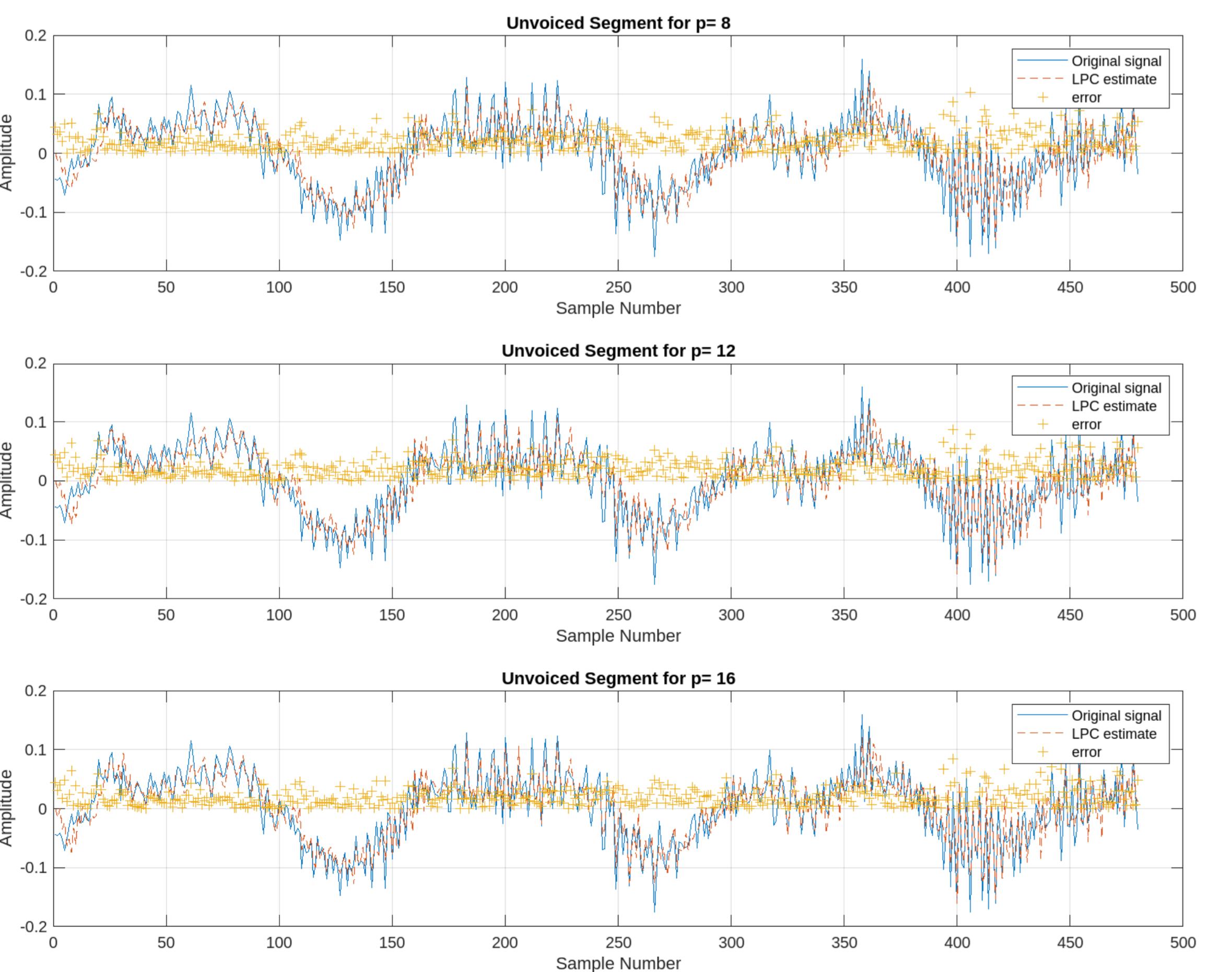
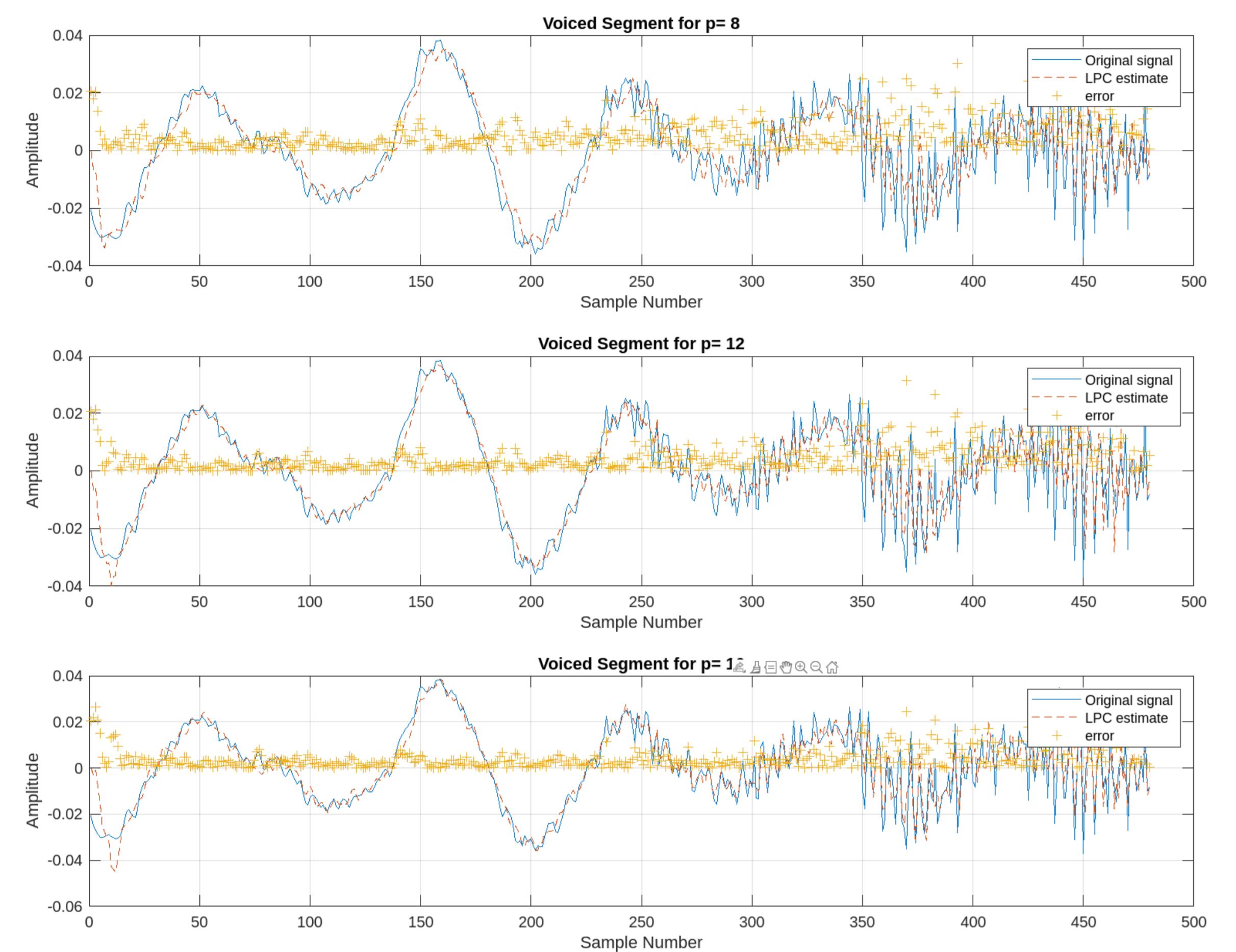
Υλοποιήθηκε μια μέθοδος για την εύρεση της θεμελιώδους συχνότητας έμφωνων ήχων στο ηχογραφημένο σήμα, χρησιμοποιώντας αυτοσυσχέτιση και εφαρμόζοντας παράθυρο Hamming. Τα αποτελέσματα παρουσιάζονται οπτικά, διακρίνοντας τις θεμελιώδεις συχνότητες στα έμφωνα τμήματα και θεωρώντας μηδενική την τιμή στα άφωνα και σιωπηλά τμήματα.





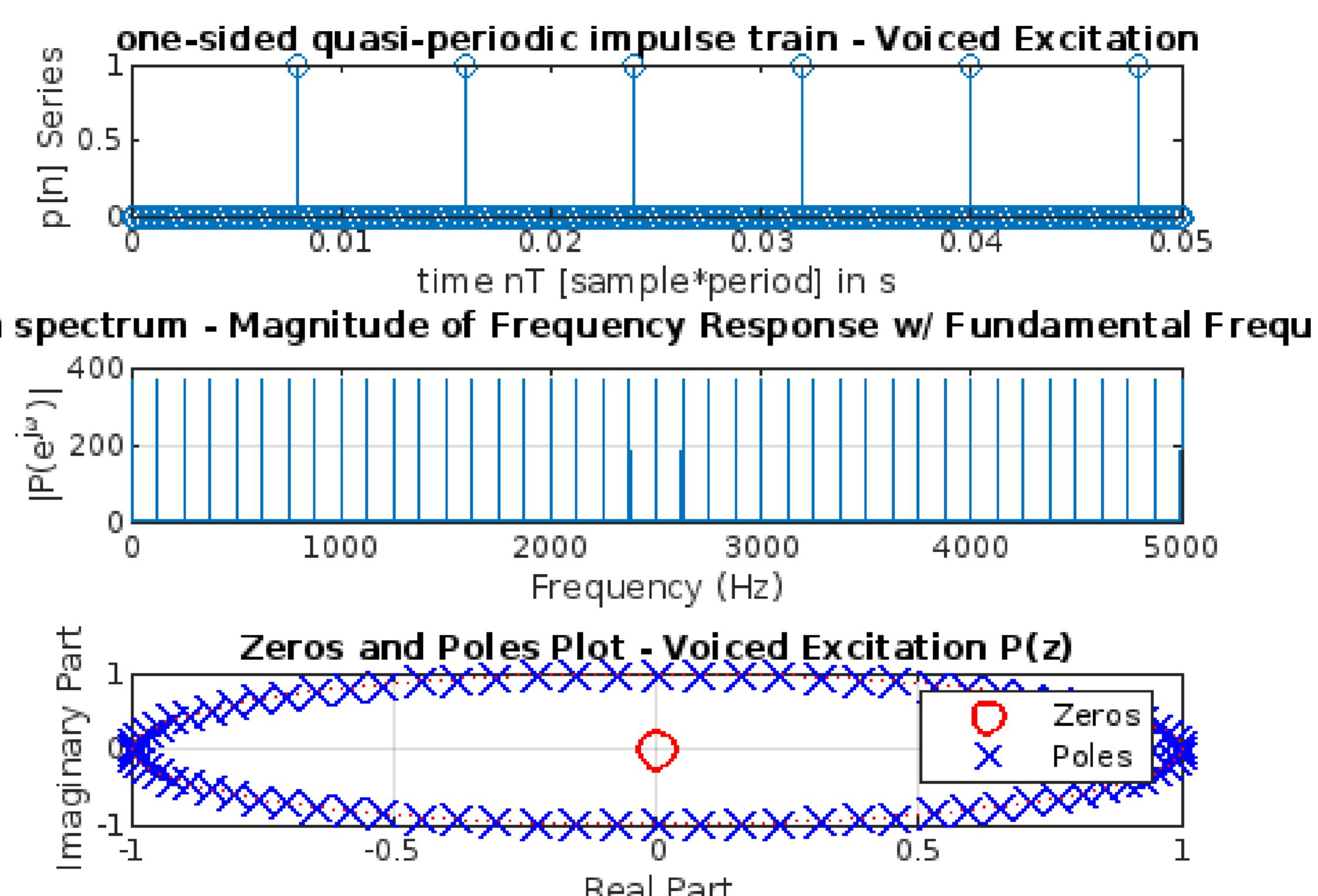
Ανάλυση Ηχητικών Σημάτων με Μοντέλο LPC

Χρησιμοποιώντας το μοντέλο Linear Predictive Coding (LPC), αναλύονται ηχητικά σήματα διάρκειας 30 msec σε έμφωνα και άφωνα τμήματα, εκτιμώντας τα χαρακτηριστικά LPC και εφαρμογή φίλτρου σε ένα έμφωνο και ένα άφωνο frame.



Συνθετική Παραγωγή Φωνηντών με Μοντέλο Διέγερσης Pitch

Δημιουργήστε έναν συνθετικό ήχο 3 δευτερολέπτων με συχνότητα δειγματοληψίας 10 kHz, περιλαμβάνοντας έξι φωνήντα κάθε ένα διάρκειας 0.5 sec, χρησιμοποιώντας περίοδο pitch 8 msec και διέγερση από την εξίσωση $p[n] = \sum_{k=0}^{\infty} \beta^k \delta[n - kNp]$. Σχεδιάστε το $p[n]$, το φάσμα $P(e^{j\omega})$ και το διάγραμμα πόλων-μηδενικών της $P(z)$. $P[z] = \sum_{k=0}^{\infty} \beta^k z^{kNp} = \frac{1}{1 - bz^{-Np}}$

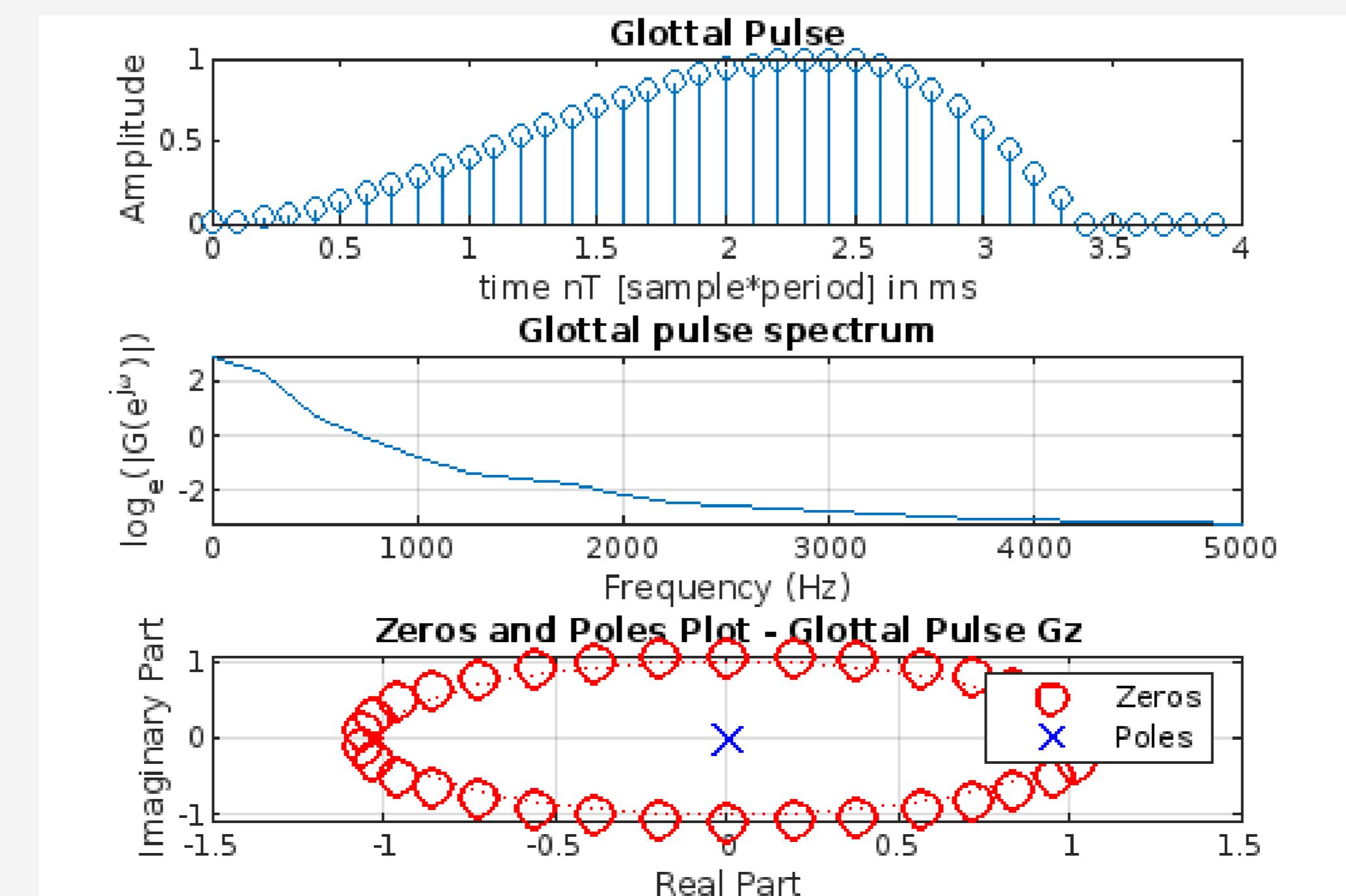


Ανάλυση Γλωττιδιακού Παλμού: Χρονική Σειρά, Φάσμα και Πόλοι-Μηδενικά

Σχεδιάστε το μοντέλο του γλωττιδιακού παλμού $g[n]$ βάσει της εξίσωσης

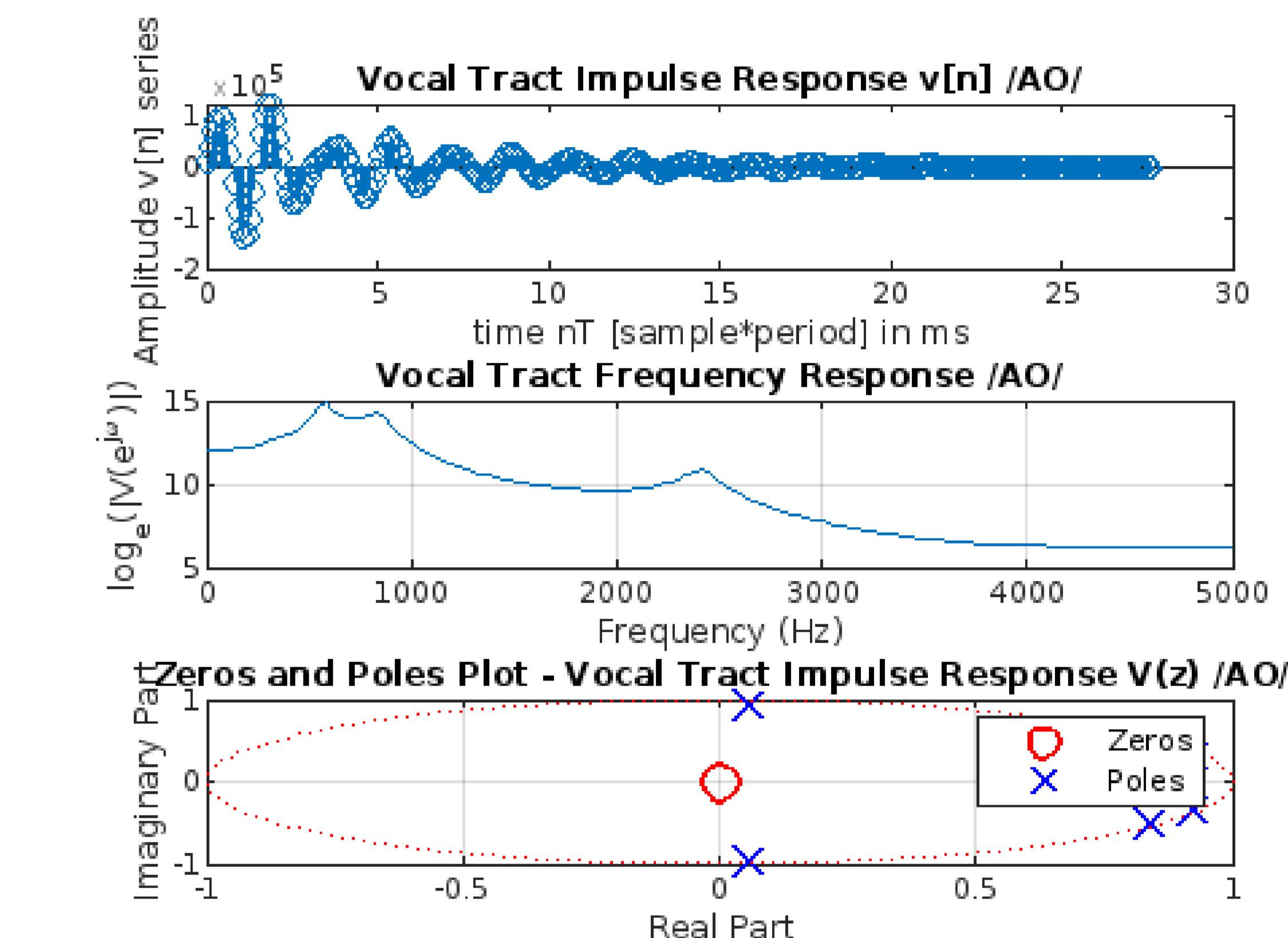
$$g[n] = \begin{cases} 0.5(1 - \cos(\pi(n+1)/25)), & \text{για } 0 \leq n \leq 24 \\ \cos(0.5\pi(n-24)/10), & \text{για } 25 \leq n \leq 33 \\ 0, & \text{αλλού.} \end{cases}$$

, το φάσμα $G(e^{j\omega})$, και το διάγραμμα πόλων-μηδενικών της $G(z)$.



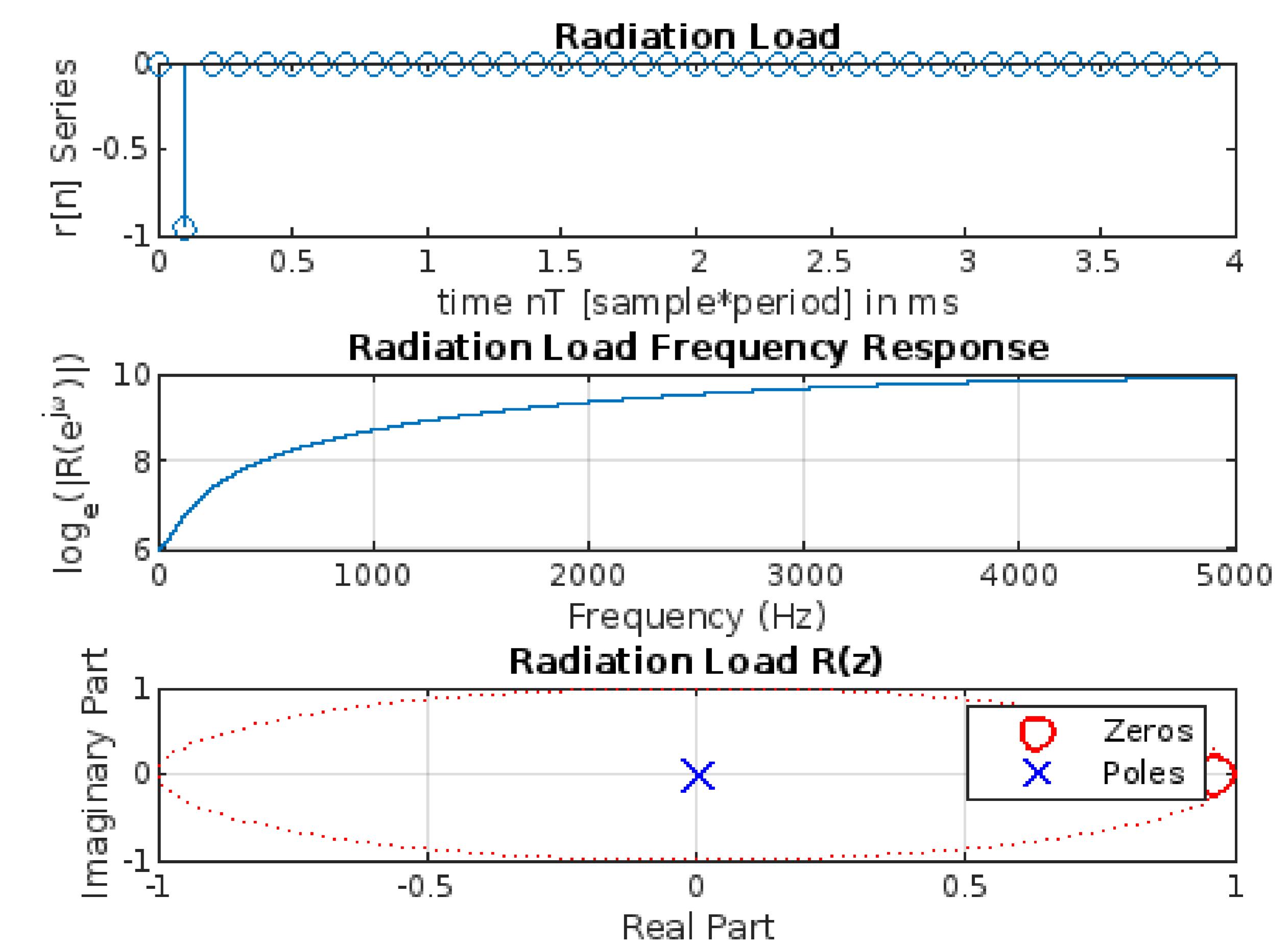
Ανάλυση Φορτίου Ακτινοβολίας: Χρόνος, Φάσμα και Πόλοι-Μηδενικά

Αναλύστε το σύστημα φωνητικής οδού ως φίλτρο $V(z)$, χρησιμοποιώντας τρία formants για το φωνήν /AO/ με συχνότητες $F1=570Hz$, $F2=840Hz$, $F3=2410Hz$ και εύρη bandwidths αντίστοιχα $60Hz$, $100Hz$, $120Hz$. Σχεδιάστε το φάσμα $V(e^{j\omega})$ και το διάγραμμα πόλων-μηδενικών της $V(z)$.



Ανάλυση Φορτίου Ακτινοβολίας: Χρόνος, Φάσμα και Πόλοι-Μηδενικά

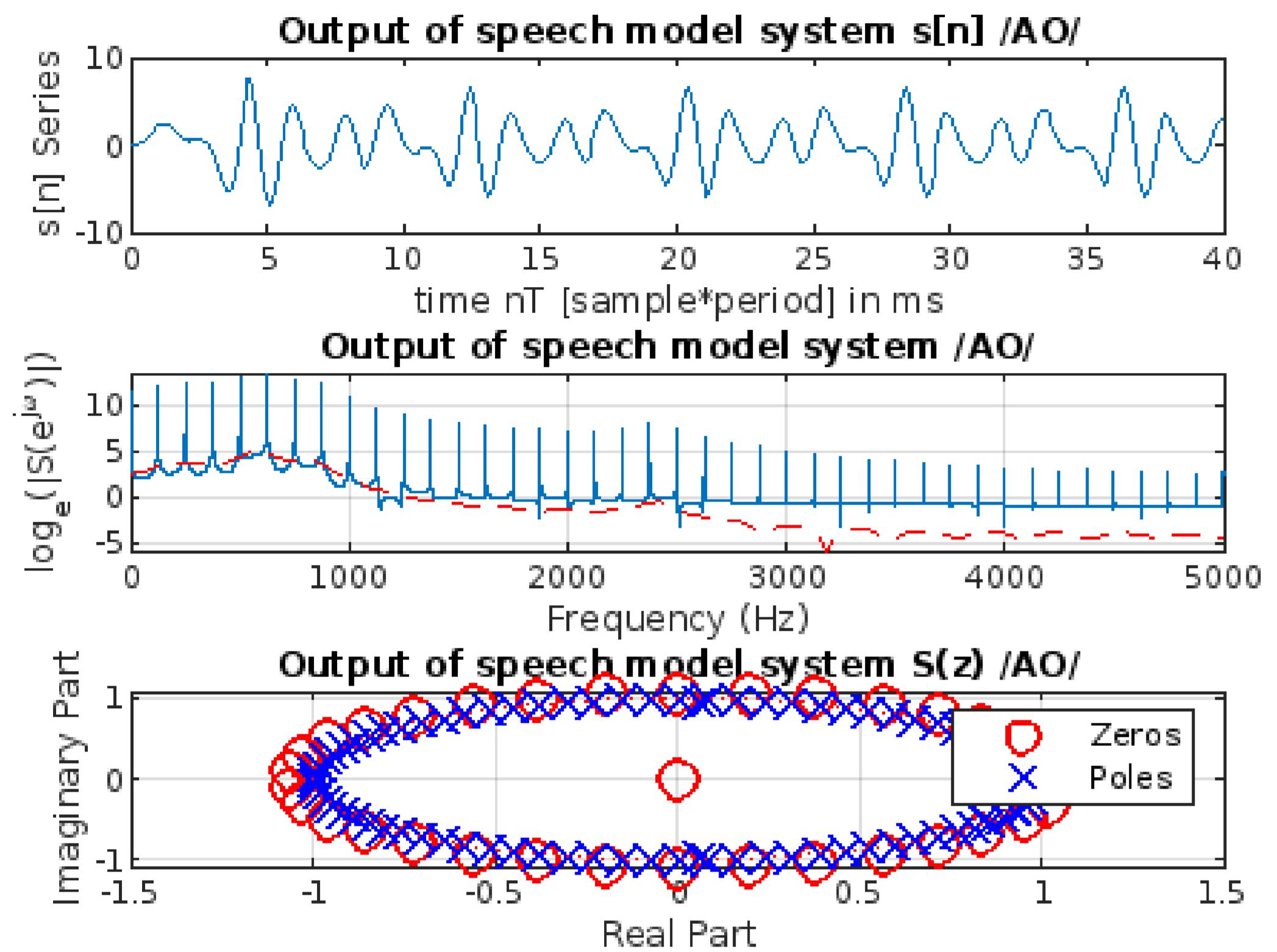
Σχεδιάστε το $r[n]$, το φάσμα $R(e^{j\omega})$ και το διάγραμμα πόλων-μηδενικών της $R(z)$ για το φορτίο ακτινοβολίας $r[n] = \delta[n] - 0.96\delta[n-1]$.





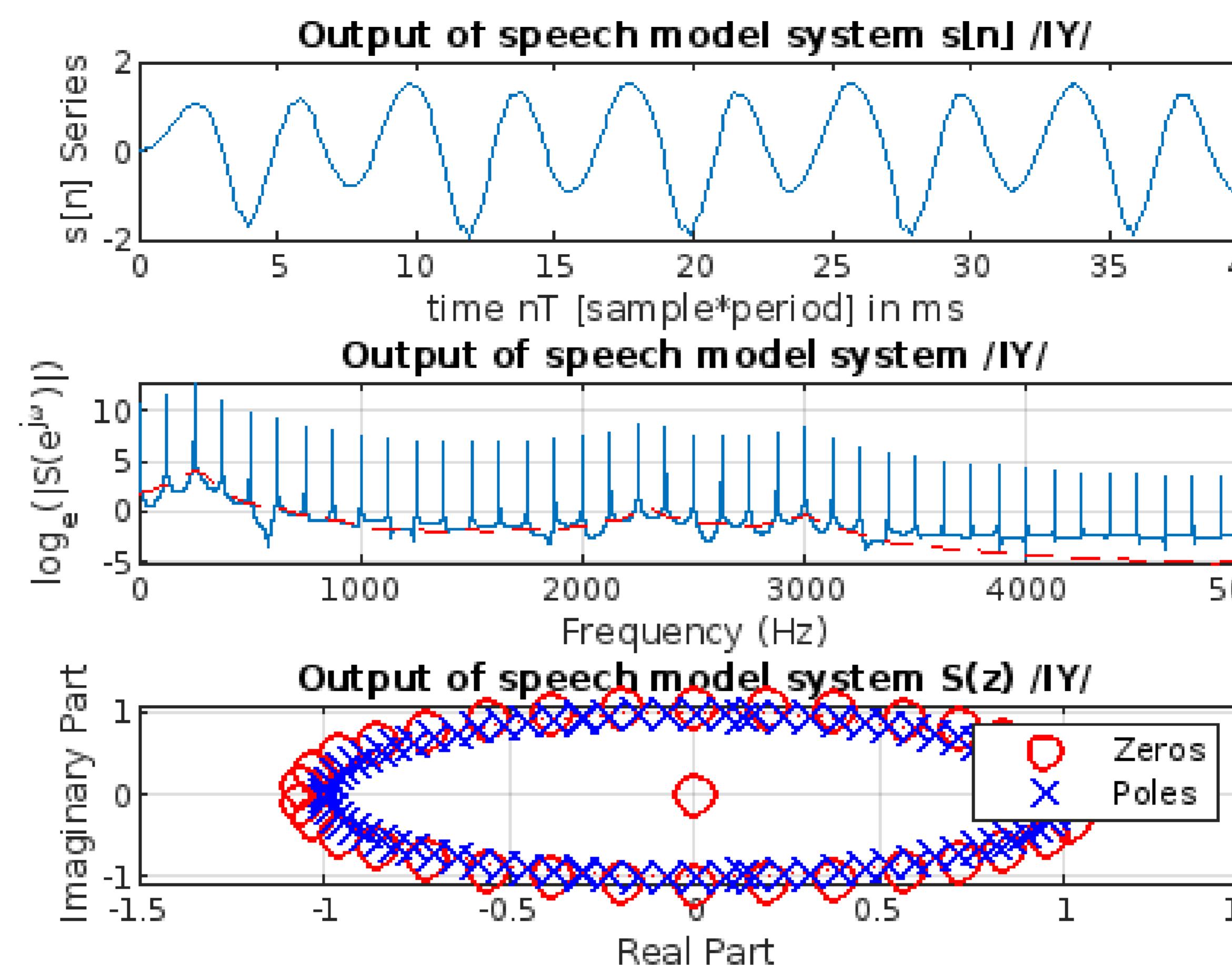
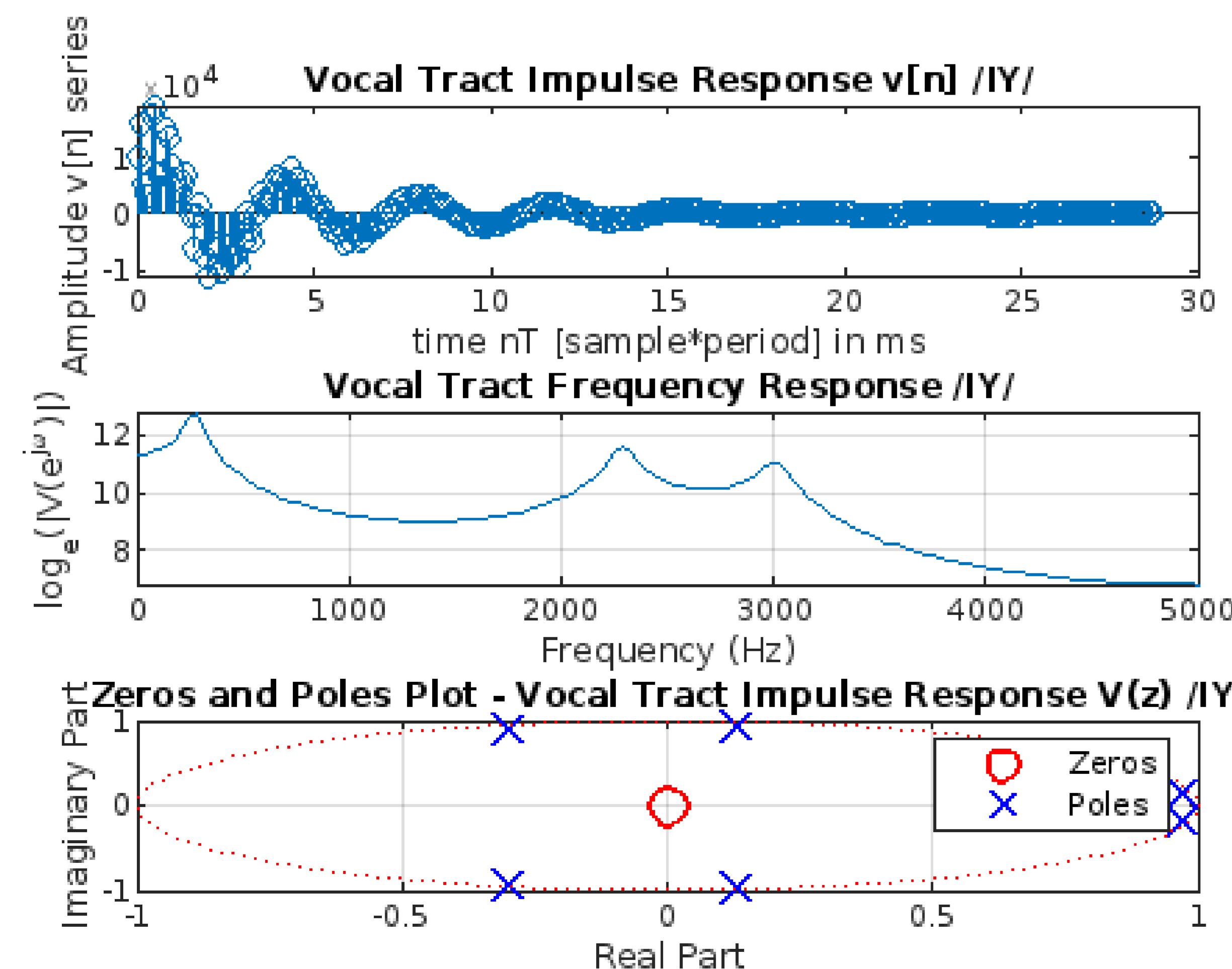
Φασματική Ανάλυση Σήματος Φωνής

Σχεδιάστε το φάσμα $S(e^{j\omega})$ και το διαγράμμα πόλων-μηδενικών της $S(z)$ για το συνολικό σήμα φωνής $s[n]$, που προκύπτει από τη συνέλιξη των $p[n]$, $g[n]$, $v[n]$, και $r[n]$ με κέρδος $A = 5000$.

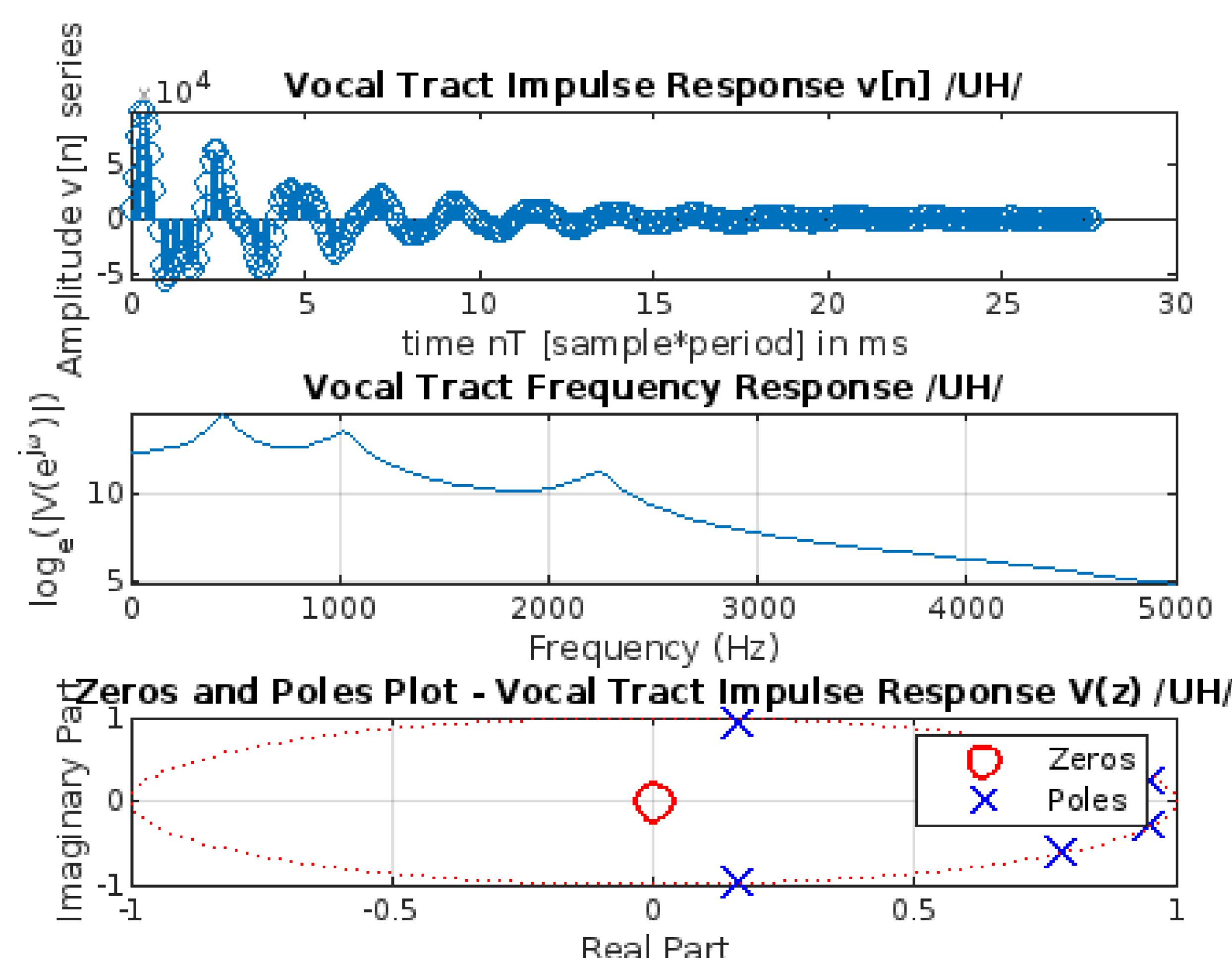


Σύνθεση και Ανάλυση Φωνηέντων: Φάσματα και Πόλοι-Μηδενικά

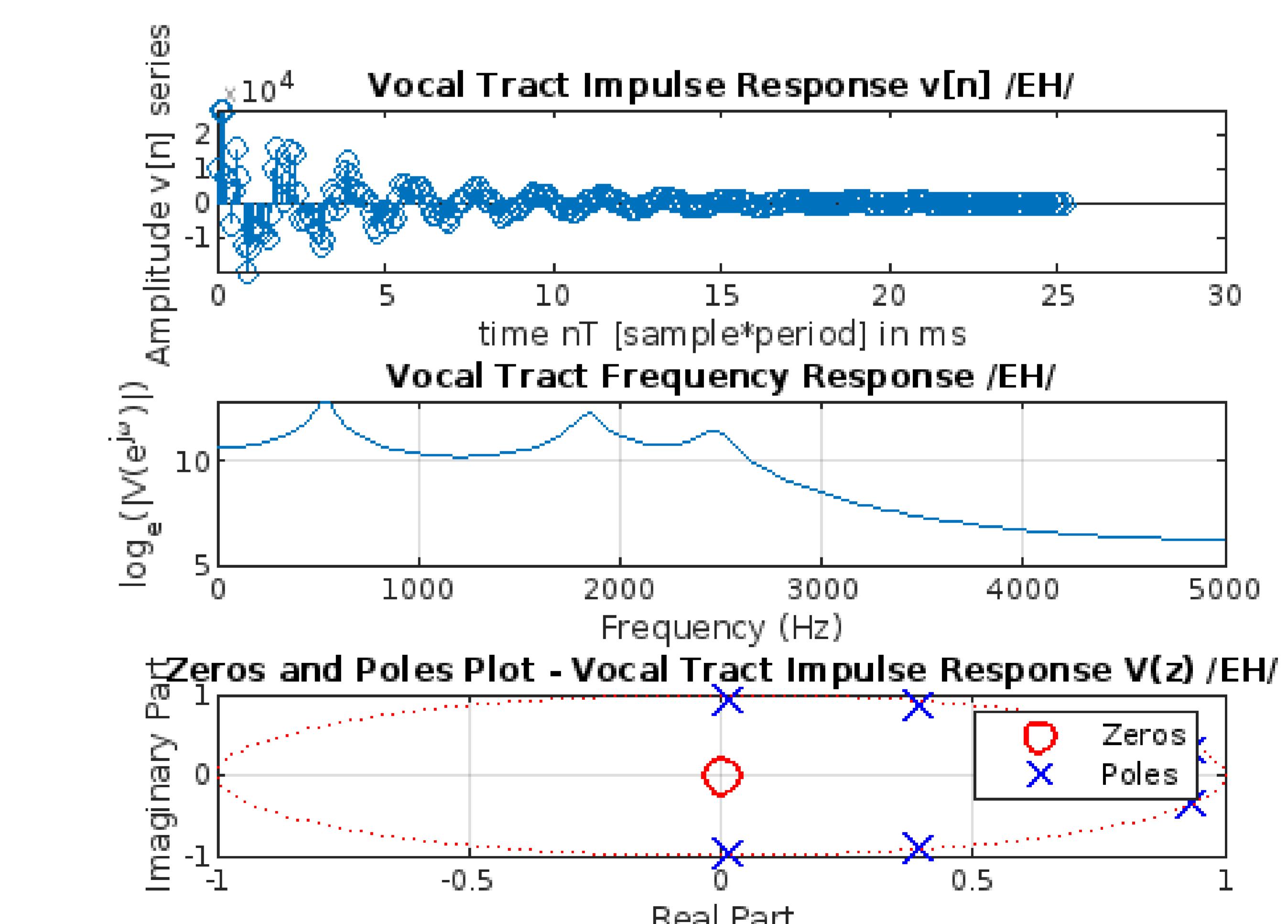
Συνθέστε σήματα φωνής 0.5 sec για τα φωνήντα /IY/, /UH/, /EH/, /AH/, /IH/, ρυθμίζοντας τις συχνότητες F1, F2, F3 ανάλογα. Σχεδιάστε τα φάσματα $V(e^{j\omega})$ και $S(e^{j\omega})$ και τα διαγράμματα πόλων-μηδενικών $V(z)$ και $S(z)$ για κάθε φωνήν.

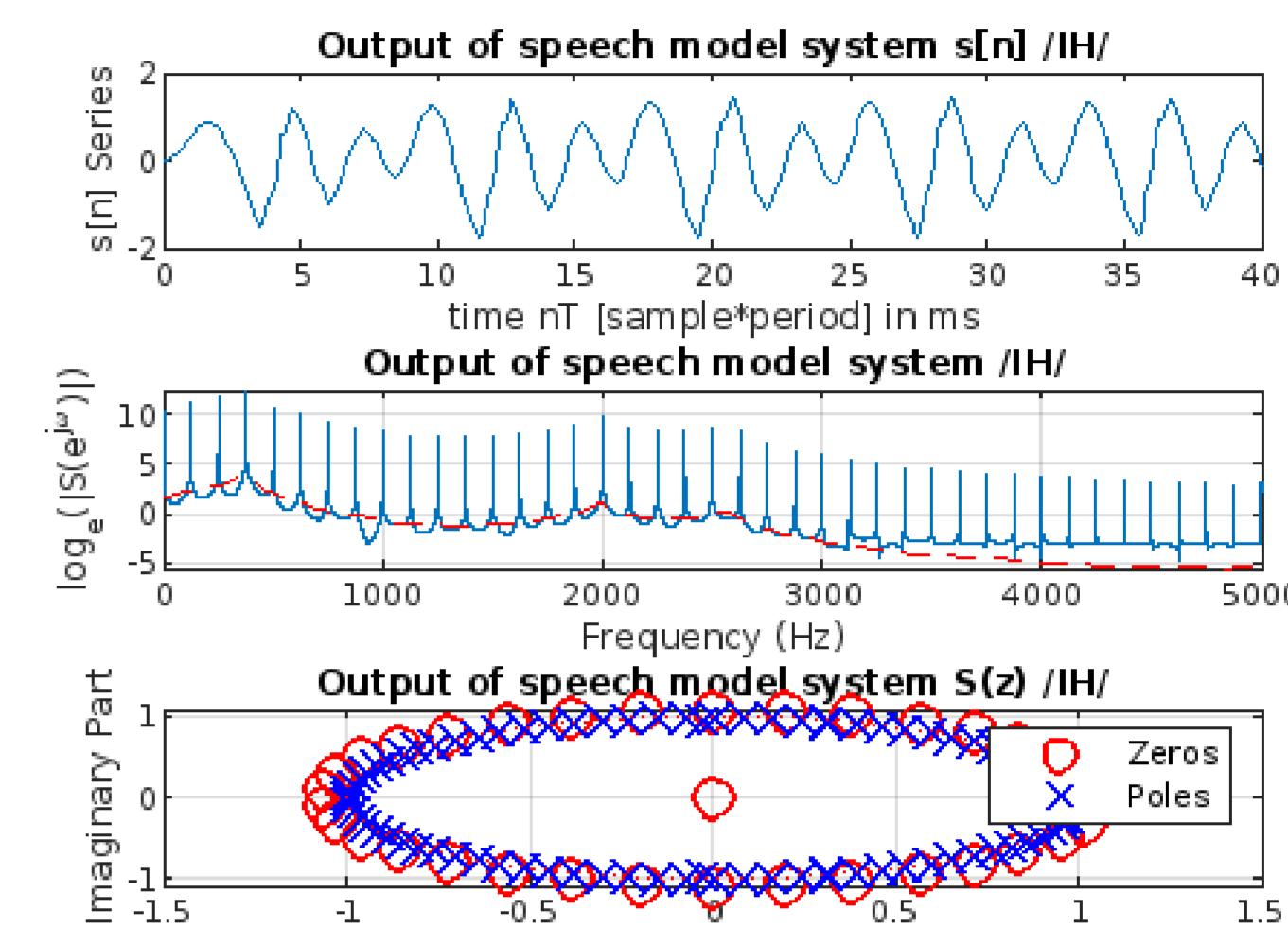
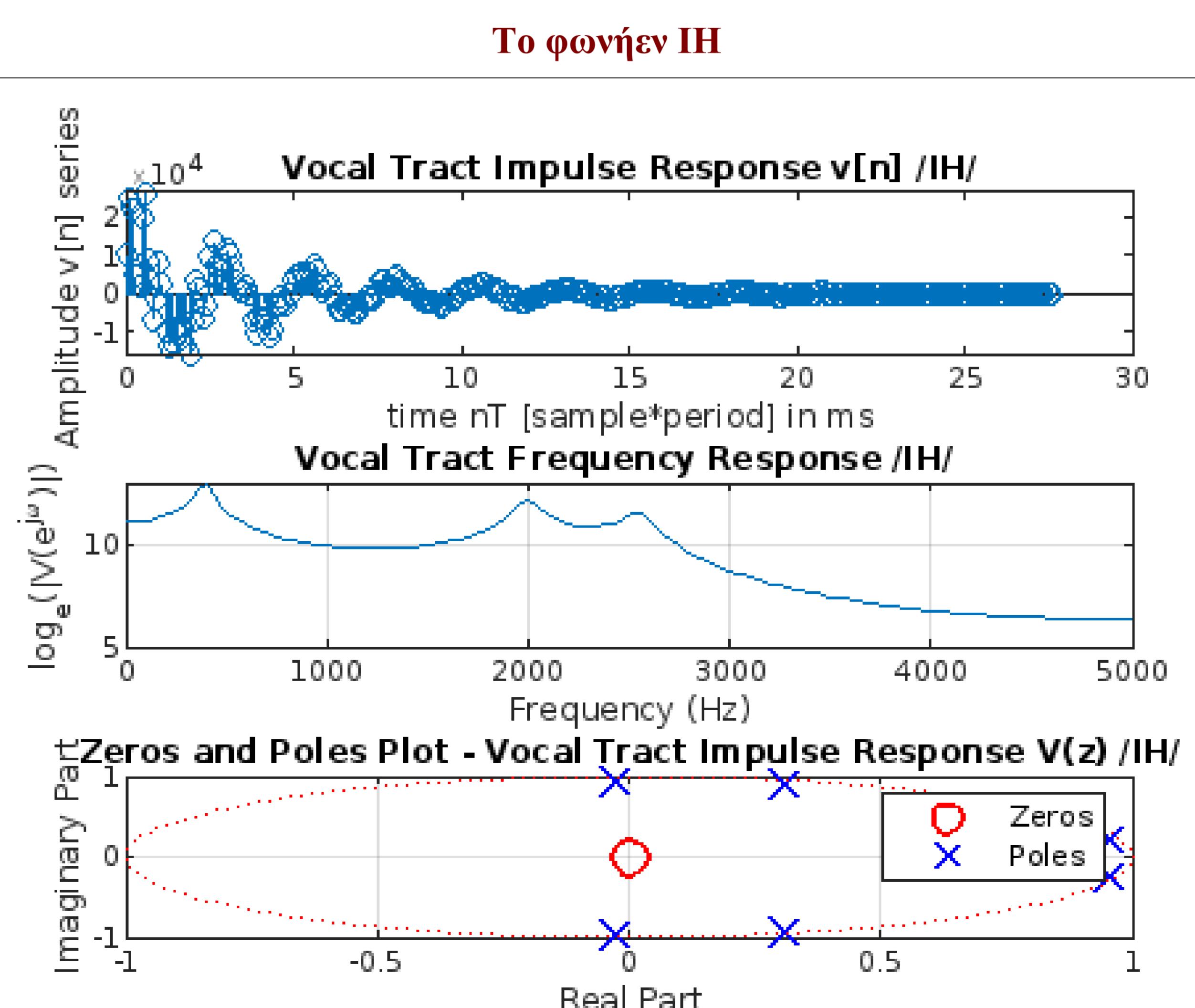
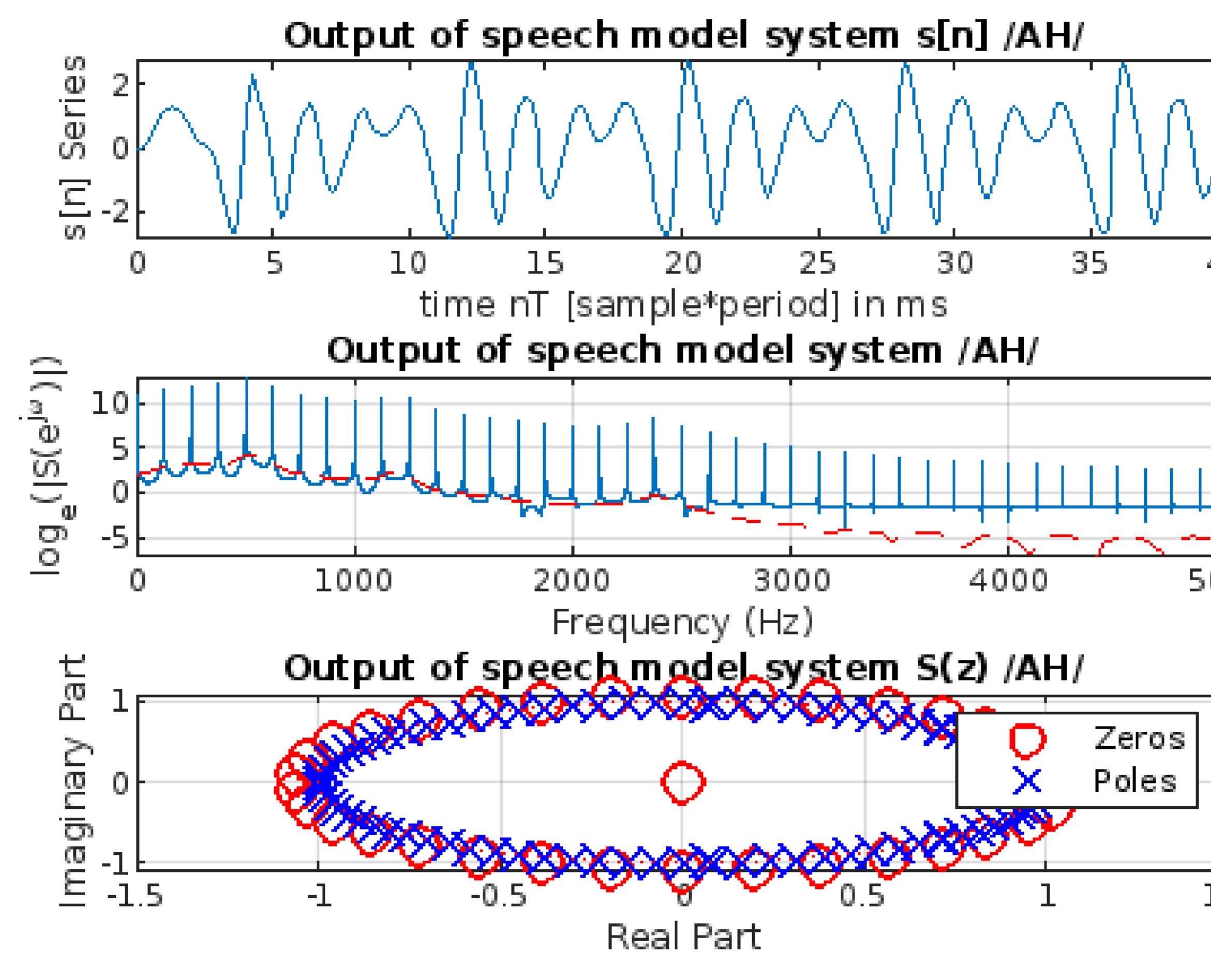
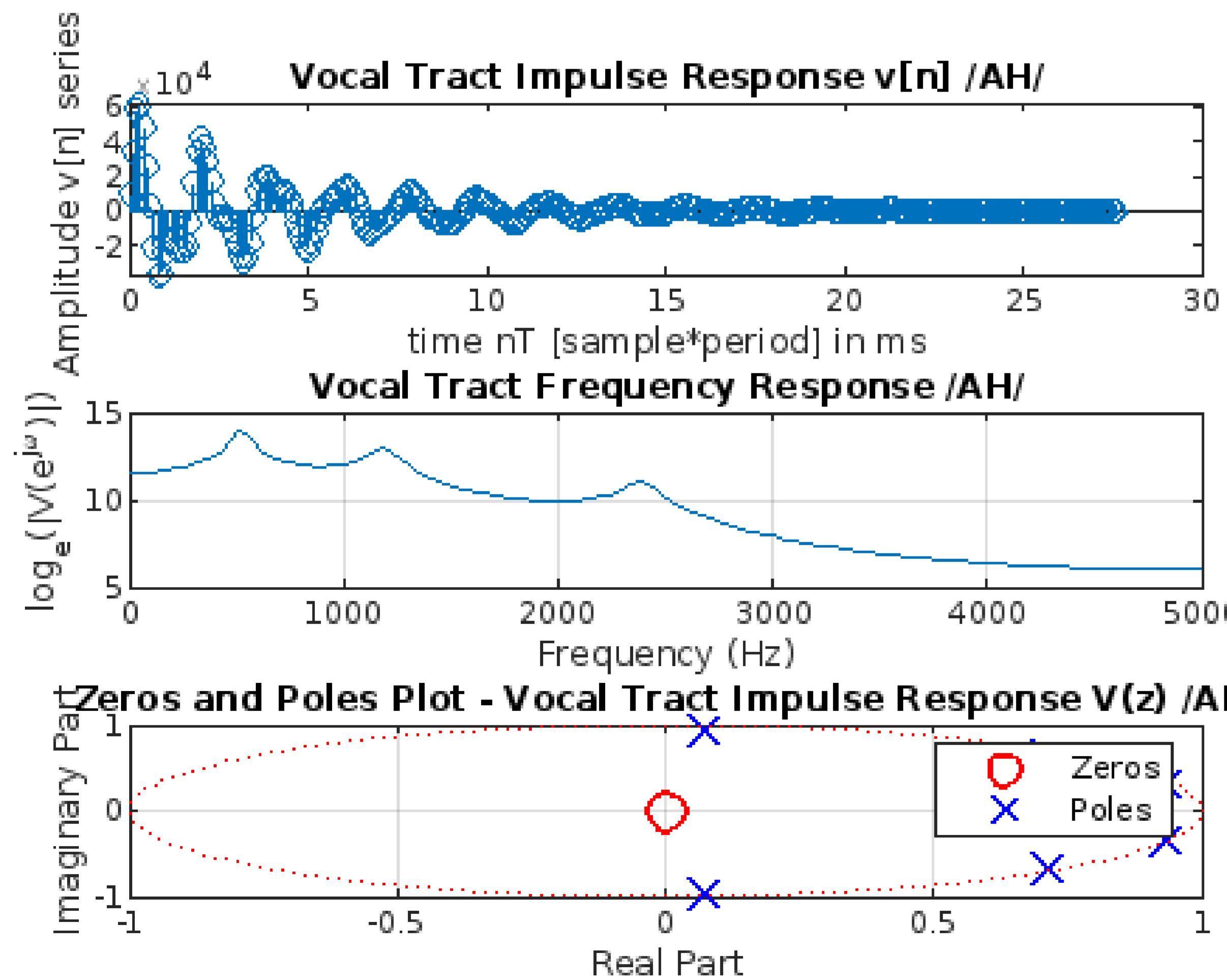
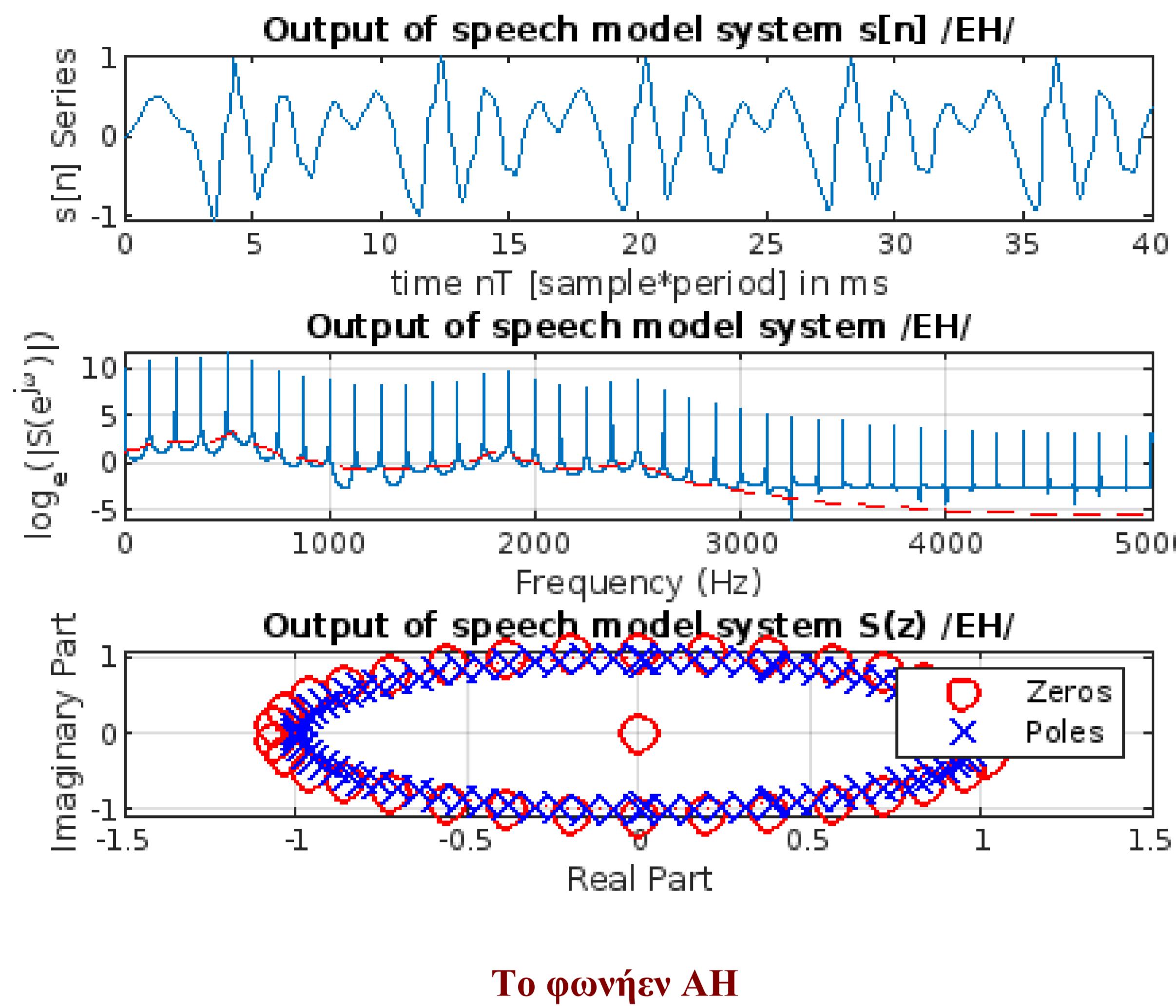


Το φωνήν UH



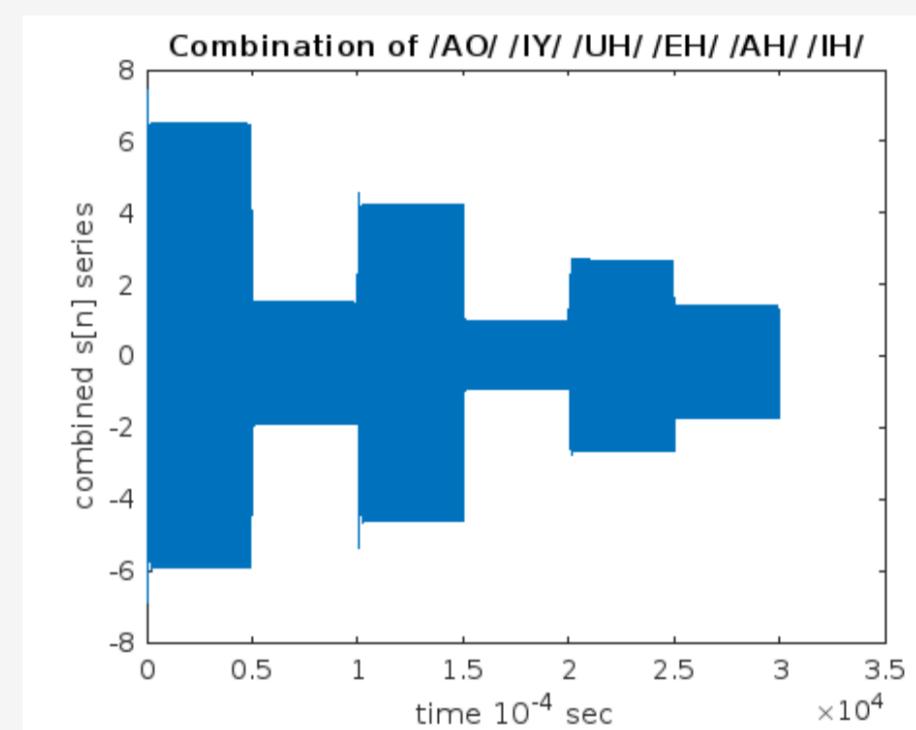
Το φωνήν EH





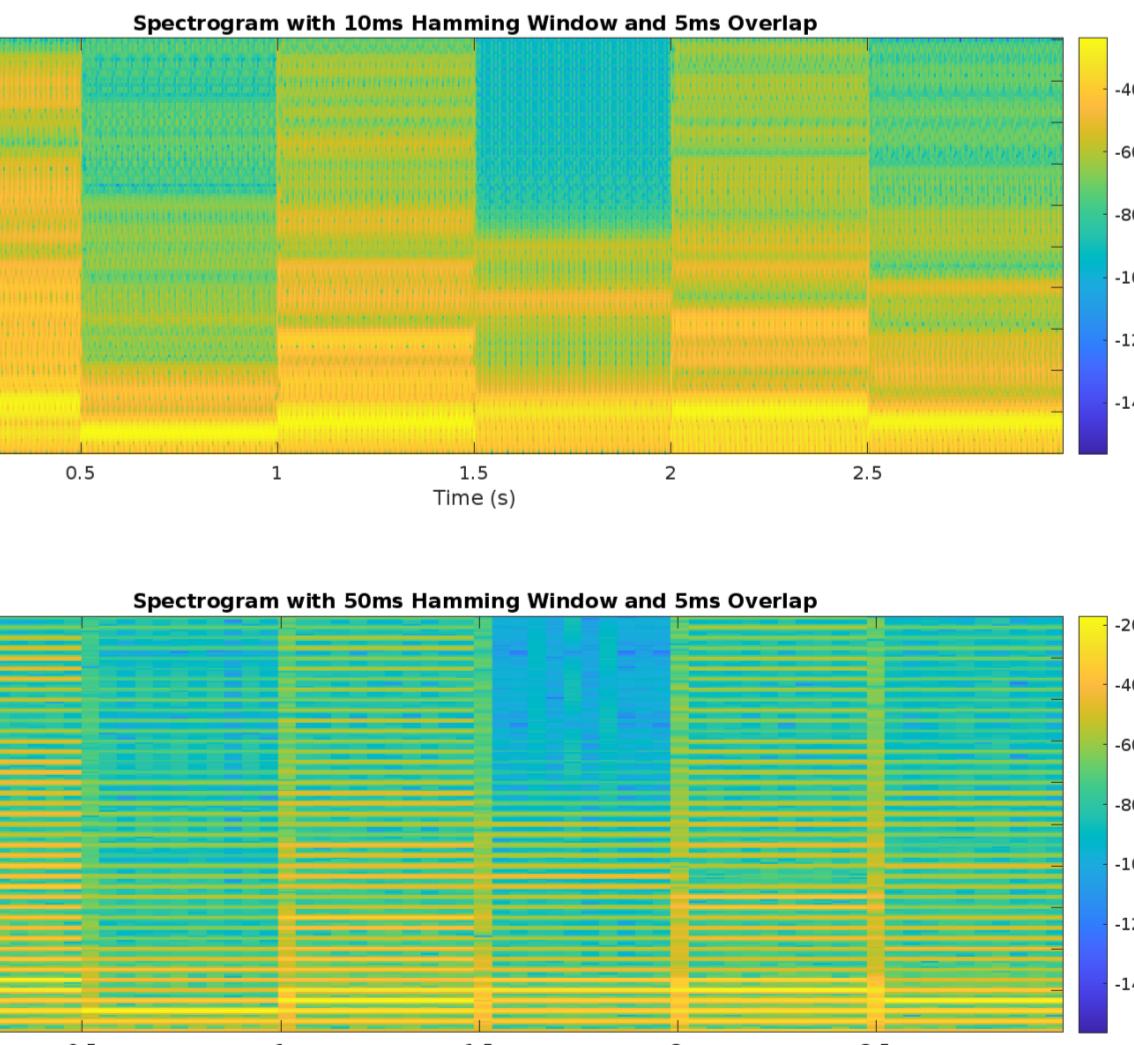
Σύγκριση Συνθετικών Σημάτων Φωνηέντων

Σχεδιάστε και σχολιάστε τα έξι συνθετικά σήματα $s[n]$ (για 1000 δείγματα) που αντιστοιχούν στα φωνήεντα, επισημαίνοντας διαφορές στη μορφή τους.



Φασματογράμματα Συνθετικής vs Φυσικής Φωνής

Σχεδιάστε το φασματόγραμμα του συνθετικού σήματος φωνής, ηχογραφήστε και αναλύστε το φυσικό σήμα φωνής με τα ίδια φωνήεντα και συγκρίνετε τα φασματογράμματα.



Σχετικές Βιβλιογραφικές Πηγές

- [1] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." [Computer program], 2024. Retrieved 6 January 2024 from <http://www.praat.org/>.
- [2] L. R. Rabiner and R. W. Schafer, *Introduction to Digital Speech Processing*. 2007.
- [3] The SoX Project. Available online at <http://sox.sourceforge.net>, SoX: Sound eXchange, 2024.