ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΦΥΣΙΚΗΣ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΟΝ ΗΛΕΚΤΡΟΝΙΚΟ ΑΥΤΟΜΑΤΙΣΜΟ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Συμβολή στην Ελληνικοποίηση της πλατφόρμας μετατροπής κειμένου σε ομιλία OpenMary

Γεώργιος Μ. Σανιόγλου

Επιβλέποντες: Γεώργιος Κουρουπέτρογλου, Αναπληρωτής Καθηγητής
Δημήτριος Τσώνος, Δρ Πληροφορικής

ΑΘΗΝΑ

ΙΑΝΟΥΑΡΙΟΣ 2015

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Συμβολή στην Ελληνικοποίηση της πλατφόρμας μετατροπής κειμένου σε ομιλία OpenMary

Γεώργιος Μ. Σανιόγλου

Α.Μ.: 2012528

ΕΠΙΒΛΕΠΟΝΤΕΣ: Γεώργιος Κουρουπέτρογλου, Αναπληρωτής Καθηγητής
Δημήτριος Τσώνος, Δρ Πληροφορικής

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:
Γεώργιος Κουρουπέτρογλου, Αναπληρωτής Καθηγητής
Αγγελική Αραπογιάννη, Καθηγήτρια
Ευστάθιος Χατζηευθυμιάδης, Αναπληρωτής Καθηγητής

Ιανουάριος 2015

ΠΕΡΙΛΗΨΗ

Αντικείμενο της παρούσας διπλωματικής διατριβής ήταν η συμβολή στην Ελληνικοποίηση της πλατφόρμας μετατροπής κειμένου σε ομιλία OpenMary. Η πλατφόρμα OpenMary (Open Modular Architecture for research on speech Synthesis, επίσης γνωστή και ως MARYTTS) είναι μία ανοιχτού κώδικα (open source) πολυγλωσσική πλατφόρμα Κείμενο-Σε-Ομιλία.

Σχεδιάστηκε και υλοποιήθηκε η υποστήριξη για την Ελληνική γλώσσα, με σκοπό την αναγνώριση των μερών του λόγου των ελληνικών προτάσεων και την βέλτιστη ακουστική απόδοσή τους ανάλογα με το είδος της πρότασης. Με την ολοκλήρωση του συνθέτη ομιλίας τα είδη των προτάσεων που αναγνωρίζονται είναι οι καταφατικές, οι ερωτηματικές, οι επιφωνηματικές και οι αρνητικές προτάσεις. Επιπλέον, γίνεται αντιστοίχηση των ερωτηματικών και των αρνητικών προτάσεων σε κατάλληλο προσωδιακό μοντέλο ομιλίας.

Το πρόβλημα στον παρόν συνθέτη ομιλίας ήταν ότι δεν μπορούσε να κατανοήσει τις ερωτηματικές προτάσεις, με αποτέλεσμα να μην μπορεί να κατανοήσει επιπρόσθετα ούτε το είδος της πρότασης. Επιπλέον, δεν μπορούσε να αποδώσει το σωστό μοντέλο προσωδιακής ομιλίας. Ο λόγος που δεν μπορούσε να γίνει αυτή η κατανόηση ήταν γιατί δεν αντιλαμβανότανε τα γραμματικά μέρη του λόγου που υπήρχαν μέσα στην πρόταση. Χωρίς την γραμματική αναγνώριση δεν μπορούσε να αποδώσει σε πρώτο χρόνο το είδος της πρότασης και στην συνέχεια τον σωστό τόνο επιτονισμού. Στην παρούσα εργασία παρουσιάζουμε τα βήματα που γίνανε για τον εμπλουτισμό του.

Σε αυτή την εργασία θα παρουσιάσουμε τα βήματα που γίνανε για την αναγνώριση του είδους των προτάσεων αλλά και για την απόδοση του προσωδιακού μοντέλου. Με τη χρήση του κατάλληλου αλγορίθμου Επεξεργασίας Φυσικής Γλώσσας (NLP) επιτυγχάνεται αρχικά η γραμματική αναγνώριση των λέξεων της πρότασης και στην συνέχεια το είδος της πρότασης. Έπειτα γίνεται η αντιστοίχιση και διόρθωση του επιτονισμού των λέξεων της πρότασης. Για τα καινούργια είδη των προτάσεων που εισήχθησαν στο σύστημα, δημιουργήσαμε επιπλέον κανόνες για τον επιτονισμό τους. Τέλος, έχοντας τους κανόνες επιτονισμού πραγματοποιείται η μετατροπή Κειμένου-σε-Ομιλία χρησιμοποιώντας το αντίστοιχο προσωδιακό μοντέλο.

Η πλατφόρμα είναι σε θέση να αναγνωρίζει και να ξεχωρίζει, εκτός από το είδος της πρότασης και τον τύπο της ερώτησης, δηλαδή αν είναι ερώτηση ολικής άγνοιας (ερωτηματικές προτάσεις Ναι-Όχι), ερώτηση μερικής άγνοιας (ερωτηματικές προτάσεις Ποιος-Ποια) ή αρνητική ερώτηση. Κάνοντας αυτόν τον διαχωρισμό αποδίδεται διαφορετικό προσωδιακό μοντέλο σε κάθε είδος.

Ο συνθέτης ομιλίας, οι βιβλιοθήκες και οι συναρτήσεις, είναι υλοποιημένα στη γλώσσα προγραμματισμού Java. Η παρούσα υλοποίηση αξιολογήθηκε μέσα από μία πειραματική διαδικασία. Στην πειραματική διαδικασία ζητήθηκε από 37 ακροατές να αξιολογήσουν ερωτήσεις που εκφωνήθηκαν με συνθετική ομιλία. Οι ερωτήσεις χωρίζονταν σε δύο κατηγορίες. Η πρώτη κατηγορία ήταν με το βασικό προσωδιακό μοντέλο και η δεύτερη κατηγορία ήταν με το νέο στοχευμένο προσωδιακό μοντέλο. Τα αποτελέσματα έδειξαν ότι οι χρήστες αναγνωρίζουν τις διαφοροποιήσεις στο προσωδιακό μοντέλο και επιπλέον ότι είχαν καλύτερη αναγνώριση στις προτάσεις ολικής άγνοιας από τις προτάσεις μερικής άγνοιας.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Τεχνολογίες φωνής – Μετατροπή Κειμένου-σε-Ομιλία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: OpenMary, Κείμενο-σε-Ομιλία, Τεχνολογίες φωνής, Προσωδιακό μοντέλο, Ελληνικοποίηση του MaryTTS

ABSTRACT

The object of this thesis was to contribute to the Greek versions of the text-to-speech platform OpenMary. The platform OpenMary (Open Modular Architecture for research on speech Synthesis, also known as MARYTTS) is an open source multilingual Text-To-Speech platform.

We designed and implemented the support for the Greek language, in order to identify the different sentence types in Greek and define the optimal prosody specification based on the sentence type. On completion of the speech synthesizer the sentence types that are recognized are declarative, interrogative, exclamatory and negative sentences. In addition interrogative and negative sentences were mapped to an appropriate prosodic specification.

The old speech synthesizer was unable to identify interrogative sentences, so it cannot be understood in addition neither the nature of the proposal. Moreover it could not assign the correct intonation specification, partly due to the fact that there was no means for identifying the grammatical parts of speech of the words in the sentence. Without this information we could not identify the sentence type and subsequently the appropriate prosody specification. In this paper we present the steps that were made for the enrichment of the relevant modules.

In this paper we present the steps that were made to identify the sentence type and for assigning the correct prosody specification. By using the appropriate Natural Language Processing algorithm we initially achieved identification of the parts of speech and consequently the corresponding sentence type. Following we assigned and corrected the intonation of the words in the sentence.

For the new sentence types which were introduced to the system, we created additional rules for their intonation. Finally, having the intonation rules in place we proceed with the conversion of Text-to-Speech using the corresponding prosodic model. The platform is able to recognize and distinguish between the different types of questions, namely whether it is a Yes-No question, a Wh-question or negative question. Based on this distinction a different prosodic model is assigned to each type.

The speech synthesizer, libraries and functions are implemented in the Java programming language. The present implementation was evaluated through an experimental process. In the experimental procedure 37 listeners were asked to rate questions which were produced with synthetic speech. The questions were divided into two categories. The first category was produced with the baseline prosodic model and the second category was produced with the new prosodic model. The results showed that users recognize the differences in the prosodic model and in addition, they had higher recognition rates for polar questions compared to Wh-questions.

SUBJECT AREA: Voice Technologies – Transformation text to speech

KEYWORDS: OpenMary, Text to Speech, Voice technologies, Prosodic model, Greek version of OpenMary

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ευχαριστίες μου στον επιβλέποντα καθηγητή, τον κ Κουρουπέτρογλου Γεώργιο, και στους βοηθούς του Τσώνο Δημήτριο και Σταυροπούλου Πέπη, του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών για την καθοδήγηση, τον πολύτιμο χρόνο και την πολύτιμη βοήθεια που μου διέθεσαν.

Επίσης θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την ανεκτίμητη στήριξη σε όλη τη διάρκεια των σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΠΡΟΛΟΓΟΣ

Ο επιστημονικός τομέας της Αλληλεπίδρασης Ανθρώπου-Υπολογιστή, και πιο συγκεκριμένα τα συστήματα μετατροπής Κειμένου-σε-Ομιλία (ΚσΟ), συγκαταλέγεται στους πλέον υποσχόμενους τομείς της Σύγχρονης Τεχνολογικής Έρευνας. Οι εφαρμογές των συστημάτων ΚσΟ μπορούν να χρησιμοποιηθούν σε ποικίλες εφαρμογές, όπως ακουστική πρόσβαση στην ηλεκτρονική αλληλογραφία και διάφορα είδη βάσεων δεδομένων, έως ανάγνωση ηλεκτρονικών εγγράφων για τυφλούς.

Στόχος της διπλωματικής διατριβής είναι η βελτιστοποίηση της υποστήριξης της Ελληνικής γλώσσας από το σύστημα OpenMary [1]. Σε μία πρώτη προσπάθεια υποστήριξης των ελληνικών από την πλατφόρμα OpenMary [2], έγινε υλοποίηση του βασικού μοντέλου ομιλίας με τις ελάχιστες απαιτούμενες συναρτήσεις βασικής υποστήριξης του υποσυστήματος επισημείωσης Μερών-του-Λόγου (minimal part of speech tagger) και η δημιουργία μίας βασικής Ελληνικής φωνής.

Στην παρούσα εργασία εμπλουτίστηκε το σύστημα με την εισαγωγή του προσωδιακού μοντέλου για κάθε είδος πρότασης της Ελληνικής γλώσσας. Επιτεύχθηκε δηλαδή η εισαγωγή και η εναλλαγή του προσωδιακού μοντέλου ομιλίας ανάλογα με το είδος και τον τύπο της πρότασης. Από το σύστημα αναγνωρίζονται οι καταφατικές, ερωτηματικές και επιφωνηματικές προτάσεις.

Οι ερωτηματικές προτάσεις μπορούν να χωριστούν σε ερωτήσεις μερικής άγνοιας (ποιος-ποια ερώτηση), ερωτήσεις ολικής άγνοιας (ναι-όχι ερώτηση) και μετά επιπλέον σε καταφατικές και αποφατικές(αρνητικές). Οι ερωτήσεις μερικής άγνοιας είναι ερωτήσεις που εισάγονται με κάποια ερωτηματική αντωνυμία ή επίρρημα.

Οι ερωτήσεις ολικής άγνοιας απατώνται μέσω του Ναι-Όχι. Οι αποφατικές ερωτήσεις έχουν το Δεν/Δε στην αρχή της ερωτηματικής πρότασης ή στο μέσο της. Οι δηλωτικές προτάσεις χωρίζονται σε καταφατικές και σε αποφατικές δηλωτικές αν υπάρχει το Δεν/Δε στην αρχή της δηλωτικής πρότασης ή στο μέσο της. Τελευταία κατηγορία προτάσεων είναι οι επιφωνηματικές οι οποίες πάλι χωρίζονται σε καταφατικές και αποφατικές αν υπάρχει το Δεν/Δε στην αρχή της επιφωνηματικές πρότασης ή στο μέσο της.

Για να υλοποιηθεί η υποστήριξη της αναγνώρισης του είδους των προτάσεων χρειάζεται να αναγνωρίζονται τα μέρη του λόγου των λέξεων που απαρτίζουν την πρόταση. Για την αναγνώριση του είδους των προτάσεων, που μπορεί να συναντήσει κάποιος στην Ελληνική γλώσσα, χρησιμοποιήθηκε το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger που έχει αναπτυχθεί στην εργασία «Ένας νέος ελληνικός επισημειωτής μερών του λόγου, βασισμένος σε ταξινομητή μεγίστης εντροπίας» [3].

Έγινε αρχικά αντικατάσταση του minimal PoS tagger με τον AUEB_POS_tagger. Όμως το συγκεκριμένο σύστημα γραμματικής αναγνώρισης υποστήριζε μόνο την βασική γραμματική αναγνώριση των μερών του λόγου. Δεν είναι πλήρης, σύμφωνα με τις αρχικές μας απαιτήσεις. Γι' αυτό έπρεπε να εμπλουτιστεί με επιπλέον κανόνες. Επίσης, προέκυπταν σφάλματα κατά την διάρκεια της γραμματικής αναγνώρισης. Αυτά τα προβλήματα διορθώθηκαν με την επιπλέον υποστήριξη κανόνων Επεξεργασίας Φυσικής Γλώσσας (NLP). Αναπτύχθηκε ένας μηχανισμός αναγνώρισης των ελληνικών προτάσεων και διαχωρισμού τους.

Στην συνέχεια έχοντας ολοκληρώσει την γραμματική αναγνώριση των λέξεων και την ανάλυση του είδους της πρότασης, υλοποιήθηκε ένας αλγόριθμος αντιστοίχισης των προσωδιακών κανόνων. Δημιουργήθηκαν κανόνες επιτονισμού για το είδος των προτάσεων που εισήχθησαν στο σύστημα. Με την χρήση του αλγορίθμου διορθώθηκαν

οι κανόνες επιτονισμού των προτάσεων. Δημιουργήθηκε στην πλατφόρμα μια καινούργια Ελληνική φωνή με σκοπό την βελτιστοποίηση της συνθετικής ομιλίας για την Ελληνική γλώσσα. Ανάλογα το είδος της πρότασης και την απόδοση του σωστού προσωδιακού μοντέλου από το σύστημα μας η καινούρια φωνή που εκπαιδεύτηκε είναι σε θέση μέσω του συστήματος επισημείωσης να δώσει την προσωδία, το ύφος και τον τόνο του γραπτού προφωρικού που θέλει να εκφράσει ο χρήστης.

Τέλος, πραγματοποιήθηκε μια πειραματική διαδικασία όπου αξιολογήθηκε η παρούσα υλοποίηση.

Η εργασία αποτελείται από τα εξής κεφάλαια:

- Στο κεφάλαιο 2, περιγράφουμε κάποιες βασικές έννοιες της ομιλίας και της Σύνθεσης Ομιλίας από Υπολογιστή (Text-To-Speech).
- Στο κεφάλαιο 3, αρχικά περιγράφουμε τις μεθόδους σύνθεσης ομιλίας. Στην συνέχεια κάνουμε μια επισκόπηση στα ελεύθερα συστήματα σύνθεσης ομιλίας και γίνεται μια πρώτη εισαγωγή στο σύστημα μετατροπής ΚσΟ, που χρησιμοποιήθηκε για την ολοκλήρωση της διπλωματικής διατριβής, OpenMary.
- Στο κεφάλαιο 4, περιγράφουμε το σύστημα μετατροπής ΚσΟ OpenMary και αναλύουμε την βασική αρχιτεκτονική του συστήματος.
- Στο κεφάλαιο 5, περιγράφουμε το σύνολο των δεδομένων αλλά και των αλγορίθμων που χρησιμοποιήσαμε για την ανάπτυξη του υποσυστήματος της συνάρτησης UOA_POS_tagger του προσωδιακού μοντέλου ομιλίας και την ελληνικοποίηση της πλατφόρμας μετατροπής κειμένου σε ομιλία Open Mary.
- Στο κεφάλαιο 6, παρουσιάζουμε την πειραματική διαδικασία και τον τρόπο με τον οποίο χρειάστηκε να γίνει η αξιολόγηση των δύο συστημάτων, του συστήματος με το βασικό προσωδιακό μοντέλο και του συστήματος που αναπτύξαμε με τον NLP και το πλήρες προσωδιακό μοντέλο. Τέλος, αναλύουμε και εξετάζουμε τα δεδομένα και τις γραφικές παραστάσεις της πειραματικής διαδικασίας.
- Στο κεφάλαιο 7, αναλύουμε τα συμπεράσματα της διπλωματικής διατριβής και τέλος προτείνουμε μελλοντικές επεκτάσεις για την βελτίωση του προσωδιακού μοντέλου και την καλύτερη διάκριση των προτάσεων.

1. ΕΙΣΑΓΩΓΗ

Η παρούσα διπλωματική διατριβή αποσκοπεί στην δημιουργία ενός αποτελεσματικού μοντέλου παραγωγής του κατάλληλου επιτονισμού κατά την εκφώνηση των ελληνικών προτάσεων.

Μέσω της πλατφόρμας μετατροπής ΚσΟ OpenMary το προσωδιακό μοντέλο που δημιουργήθηκε χωρίζει και αποδίδει διαφορετική προσωδία σε δηλωτικές, ερωτηματικές, επιφωνηματικές και αποφτικές προτάσεις.

Το OpenMary κατά κύριο λόγο υποστηρίζει έναν απλό διαχωρισμό μεταξύ λειτουργικών λέξεων και λέξεων περιεχομένου. Έπειτα μέσω της γραμματικής ανάλυσης έχει κάποιους βασικούς κανόνες για την απόδοση ενός απλού προσωδιακού μοντέλου.

Επιτεύχθηκε η δημιουργία του προσωδιακού μοντέλου ομιλίας της Ελληνικής γλώσσας του υπό ανάπτυξη συστήματος εισάγοντας και δημιουργώντας ένα σύστημα επεξεργασίας φυσικής γλώσσας (NLP) το οποίο επεξεργάζεται τα δεδομένα που δίνει ο χρήστης στην πλατφόρμα. Μέσω της επεξεργασίας της φυσικής γλώσσας επιτυγχάνετε η γραμματική αναγνώριση των λέξεων. Επόμενο βήμα ήταν η αναδιαμόρφωση της εξόδου του υποσυστήματος NLP, ώστε να είναι συμβατά (υποσύστημα NLP και OpenMary).

Έπειτα έγινε προσθήκη αλγορίθμων για την γραμματική αναγνώριση των ερωτηματικών αντωνυμιών, προθέσεων και των αρνητικών μορίων του λόγου. Έγινε ανάπτυξη του υποσυστήματος για να μπορεί να αναγνωρίζει η πλατφόρμα το είδος της πρότασης που έχει εισάγει ο χρήστης. Χρησιμοποιώντας τις συναρτήσεις του συστήματος, τροποποιώντας τις κατάλληλα και δημιουργώντας επιπλέον συναρτήσεις έγινε εφικτή η αναγνώριση των αποφατικών δηλωτικών προτάσεων (negative declarative sentence), των ερωτήσεων μερικής άγνοιας (Who-questions), των ερωτήσεων ολικής άγνοιας (Yes/No-questions), των αποφατικών ερωτήσεων (Negative questions), κα των αποφατικών επιφωνηματικών προτάσεων (Negative exclamatory sentence).

Το επόμενο στάδιο ήταν η δήλωση κανόνων επιτονισμού για κάθε είδος της πρότασης που δημιουργήσαμε και εισάγαμε στο σύστημα. Μετά την εισαγωγή των κανόνων επιτονισμού δημιουργήθηκαν οι κατάλληλοι αλγόριθμοι για την απόδοση του επιτονισμού στις ερωτηματικές και αρνητικές προτάσεις. Έπρεπε να γίνει διόρθωση και επαναδημιουργία στους υπάρχοντες κανόνες της πλατφόρμας MARYTTS γιατί δεν κάλυπταν την Ελληνική γλώσσα.

2. Ομιλία και σύνθεση ομιλίας

Οι ραγδαίες εξελίξεις στην μικροηλεκτρονική και την τεχνολογία των υπολογιστών έχουν σαν αποτέλεσμα την απότομη αύξηση της χρήσης των υπολογιστών για την επεξεργασία της πληροφορίας. Η πληροφορία προέρχεται από κάποιον άνθρωπο και πρόκειται να χρησιμοποιηθεί επίσης από κάποιον άνθρωπο. Αυτό γεννάει την ανάγκη για αποτελεσματικούς τρόπους μεταφοράς της πληροφορίας. Ένας βολικός τρόπος για αυτήν την ανταλλαγή πληροφοριών είναι η ομιλία, καθόσον αυτή αποτελεί τον συνηθισμένο τρόπο επικοινωνίας μεταξύ των ανθρώπων. Οι άνθρωποι συνήθως μαθαίνουν να ομιλούν πολύ πριν μάθουν να γράφουν και να διαβάζουν. Ο γραπτός και ο προφορικός λόγος όμως παρουσιάζουν πολλές διαφορές. Η ομιλία έχει τη δυνατότητα να αποδίδει λεπτές αποχρώσεις του νοήματος που είναι δύσκολο να εκφραστούν σε ένα κείμενο. Η ομιλία μεταφέρει διάφορα είδη πληροφοριών αποτελούμενες από γλωσσολογικές πληροφορίες που δείχνουν το νόημα που ο ομιλητής επιθυμεί να μεταδώσει.

Η ένταση ή ο δυναμικός τονισμός (intensity or stress), η μουσικότητα ή μουσικός τονισμός (pitch) και ο χρονισμός (timing) παίζουν σημαντικό ρόλο στην επικοινωνία του ανθρώπου με ομιλία. Ο μουσικός τονισμός, ο δυναμικός τονισμός και η διάρκεια ορίζονται όλα μαζί σαν προσωδιακές ιδιότητες της ομιλίας ή προσωδία (prosody). Η προσωδία παρέχει σημαντική πληροφορία για το τι λέγεται. Τα παιδιά μαθαίνουν να διαβάζουν προτάσεις, τις οποίες δεν καταλαβαίνουν, και όπως είναι αδύνατον να τοποθετήσουν την έμφαση σωστά, χωρίς να καταλαβαίνουν πλήρως το νόημα, έχουν την συνήθεια είτε να διαβάζουν μονότονα, είτε στην προσπάθεια τους να ξεχωρίσουν μια λέξη από τις υπόλοιπες, να τοποθετούν την έμφαση τυχαία αλλοιώνοντας έτσι την έννοια αυτού που διαβάζουν [4].

Ο όρος σύνθεση ομιλίας αναφέρεται στην τεχνητή παραγωγή της ανθρώπινης ομιλίας. Η σύνθεση ομιλίας είναι μια διαδικασία η οποία παράγει τεχνητά ομιλία για διάφορες εφαρμογές. Οι μέθοδοι σύνθεσης ομιλίας δίνουν την δυνατότητα σε μία μηχανή να μεταφέρει οδηγίες ή πληροφορίες στο χρήστη «μιλώντας». Η συνθετική ομιλία μπορεί να δημιουργηθεί από την συνένωση τεμαχισμένων ηχογραφημένων ομιλιών τα οποία είναι αποθηκευμένα σε μία βάση δεδομένων.

Ένα σύστημα υπολογιστή που χρησιμοποιείται για το σκοπό αυτό καλείται συνθέτης ομιλίας, και μπορεί να χρησιμοποιηθεί σαν προϊόν λογισμικού (software) ή υλικού (hardware). Ένας συνθέτης ομιλίας είναι σε θέση να συνδυάσει και να ενσωματώσει ένα μοντέλο φωνητικού συστήματος και ένα μοντέλο με χαρακτηρίστηκα ανθρώπινης φωνής για να δημιουργηθεί ένα μοντέλο «συνθετικής» φωνής. Τα συστήματα Σύνθεσης Ομιλίας από Υπολογιστή (Text-to-Speech) συναντώνται σε ένα πολύ μεγάλο εύρος εφαρμογών. Οι πρώτες τους εφαρμογές ήταν για την επικοινωνία ανθρώπου υπολογιστή, για άτομα με ειδικές ανάγκες, και πιο συγκεκριμένα για άτομα με προβλήματα μερικής όρασης ή ολικής όρασης [5]. Το γεγονός αυτό έδωσε το κίνητρο για την ευρεία εξάπλωση των TTS εφαρμογών. Η ποιότητα ενός συνθέτη ομιλίας κρίνεται από την ομοιότητα του με την ανθρώπινη φωνή και από την ικανότητα να μπορεί να γίνει αντιληπτή, κατανοητή και με σαφήνεια.

3. ΕΠΙΣΚΟΠΗΣΗ ΤΩΝ ΜΕΘΟΔΩΝ ΚΑΙ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΣΥΝΘΕΣΗΣ ΟΜΙΛΙΑΣ

3.1 Μέθοδοι σύνθεσης ομιλίας

Στο πλαίσιο της σύνθεσης ομιλίας είναι αδύνατο να καταγραφούν και να αποθηκευτούν όλες οι λέξεις μιας γλώσσας. Ακόμα και ο άνθρωπος δεν είναι δυνητικά ικανός να προφέρει σωστά μια άγνωστη φράση ή λέξη πριν την διδαχτεί [6]. Οι μέθοδοι σύνθεσης ομιλίας έχουν πλεονεκτήματα και μειονεκτήματα. Άλλες υστερούν σε φυσικότητα αλλά έχουν καλύτερη γλωσσολογική απόδοση, ενώ άλλες έχουν καλύτερη φυσικότητα αλλά όχι τόσο καλή απόδοση.

Οι μέθοδοι σύνθεσης ομιλίας μπορούν να διαιρεθούν στους παρακάτω τύπους:

- Σύνθεση με βάση την κωδικοποίηση της κυματομορφής ή συνένωση μονάδων, όπου τα κύματα ομιλίας της ηχογραφημένης ανθρώπινης φωνής η οποία αποθηκεύεται ύστερα από κωδικοποίηση κυματομορφής ή αμέσως μετά την ηχογράφηση χρησιμοποιούνται ώστε να αναπαράγουν επιθυμητά μηνύματα.
- Σύνθεση ομιλίας με μοντελοποίηση άρθρωσης.
- Σύνθεση με κανόνα, όπου η ομιλία παράγεται με βάση τους φωνητικούς και γλωσσολογικούς κανόνες από ακολουθίες γραμμάτων ή ακολουθίες φωνητικών συμβόλων και προσωδιακά χαρακτηριστικά.
- Σύνθεση με χρήση Κρυφών Μαρκοβιανών Μοντέλων, η λειτουργία του στηρίζεται στην ανάλυση και την παραμετρική αναπαράσταση της φωνής, η οποία οδηγεί στην δυνατότητα στατιστικής μοντελοποίησης, με αποτέλεσμα να την καθιστά διαχειρίσιμη μέσω των Κρυφών Μαρκοβιανών Μοντέλων [7].
- Υβριδικές τεχνικές σύνθεσης ομιλίας.

3.1.1 Σύνθεση με βάση την συνένωση μονάδων ή κωδικοποίηση κυματομορφής

Η σύνθεση με βάση την συνένωση μονάδων ομιλίας – unit selection (ή κωδικοποίηση κυματομορφής) είναι η μέθοδος που λέξεις ή φράσεις της ανθρώπινης φωνής αποθηκεύονται και η επιθυμητή πρόταση ομιλίας συντίθεται διαβάζοντας και συνδέοντας τις κατάλληλες ενότητες. Η σύνθεση με συνένωση μονάδων χρησιμοποιεί μεγάλες βάσεις δεδομένων με ηχογραφημένη πραγματική ομιλία [8]. Περιλαμβάνει την κατάτμηση προ-ηχογραφημένης πραγματικής ομιλίας και τη μετέπειτα συγκόλληση των κατάλληλων λεκτικών τμημάτων, για την παραγωγή ενός συνθετικού εκφωνήματος. Κατά τη δημιουργία της βάσης δεδομένων, κάθε ηχογραφημένη έκφραση κατακερματίζεται σε μερικά ή όλα από τα ακόλουθα: ξεχωριστά φωνήματα (individual phones), δίφωνα (diphones), ημί-φωνα (half-phones), συλλαβές (syllables), λέξεις (words), φράσεις (phrases), και προτάσεις (sentences) [8]. Σε αυτή τη μέθοδο η ποιότητα της πρότασης ομιλίας που προέρχεται από σύνθεση επηρεάζεται γενικά από την ποιότητα της συνέχειας των ακουστικών γνωρισμάτων στις συνδέσεις ανάμεσα στις ενότητες. Τα ακουστικά γνωρίσματα περιλαμβάνουν την φασματική περιβάλλουσα, το πλάτος κύματος, την θεμελιώδη συχνότητα και τον ρυθμό ομιλίας. Όταν μικρές μονάδες όπως συλλαβές ή φωνήματα χρησιμοποιούνται, μπορεί να συντεθεί μία μεγάλη κλίμακα λέξεων και προτάσεων αλλά η ποιότητα ομιλίας εκφυλίζεται κατά πολύ. Ενώ όταν αποθηκεύονται και χρησιμοποιούνται μεγάλες ενότητες όπως φράσεις ή προτάσεις, η ποιότητα της ομιλίας που προέρχεται από σύνθεση είναι καλύτερη [4].

Τα πιο διαδεδομένα δομικά στοιχεία είναι τα δίφωνα (diphones), που αποτελούν μονάδες που αρχίζουν από το κέντρο της σταθερής κατάστασης ενός φωνήματος και τελειώνουν στο αντίστοιχο κέντρο του επόμενου. Σύμφωνα με τη θεωρία, αυτές οι

μονάδες είναι πιο εύκολο να συρραφούν απ' ότι χωριστά φωνήματα λόγω της σταθερής κατάστασης στα δύο άκρα [9].

Τα φωνήματα είναι ίσως η πιο συχνά χρησιμοποιημένη μονάδα στη σύνθεση ομιλίας γιατί υπάγονται στην φυσιολογική γλωσσική παρουσίαση της ομιλίας. Ο όρος συνένωση μονάδων (unit selection) χρησιμοποιείται συχνά για να περιγράψει αυτόν τον τύπο σύνθεσης. Οι προσεγγίσεις επιλογής μονάδας προσφέρουν τα πιο φυσικά ακουστικά αποτελέσματα, επειδή ελαχιστοποιούν την επεξεργασία του σήματος ομιλίας τόσο κατά τη δημιουργία του αποθέματος των δειγμάτων όσο και κατά τη σύνθεση.

3.1.2 Σύνθεση ομιλίας με μοντελοποίηση άρθρωσης (Articulatory synthesis)

Η αρθρωτική σύνθεση προσπαθεί να μοντελοποιήσει τα ανθρώπινα φωνητικά όργανα όσο τον δυνατόν καλύτερα γίνεται, έτσι ώστε να είναι η πιο ικανοποιητική μέθοδος για την παραγωγή υψηλής ποιότητας συνθετικής ομιλίας. Αποτελεί μία πολύπλοκη διεργασία, αφενός λόγω της δυσκολίας να μετρηθεί η πραγματική διαδικασία της άρθρωσης καθώς παράγεται η φυσική ομιλία, και αφετέρου λόγω της μαθηματικής και υπολογιστικής πολυπλοκότητας που απαιτείται για αυτά τα μοντέλα. Έτσι παρότι από πλευράς πιστότητας είναι η πιο αποτελεσματική μέθοδος παραγωγής ομιλίας, είναι η λιγότερο ανεπτυγμένη τεχνική και έχει λάβει την λιγότερη προσοχή από τις άλλες μεθόδους σύνθεσης. Αυτό έχει σαν αποτέλεσμα να μην έχει επιτευχθεί το ίδιο επίπεδο επιτυχίας [10] [11].

Η σύνθεση ομιλίας με μοντελοποίηση άρθρωσης τυπικά περιλαμβάνει μοντέλα των ανθρώπινων αρθρωτών και των φωνητικών χορδών. Οι αρθρωτές συνήθως μοντελοποιούνται από ένα σύνολο συναρτήσεων μεταξύ της περιοχής της γλωττίδας και του στόματος.

Το πρώτο μοντέλο αρθρωτικής ομιλίας βασίστηκε σε ένα πρότυπο της φωνητικής οδού από τον λάρυγγα στα χείλη για κάθε φωνητικό τμήμα. Όταν μιλάμε, οι μύες της φωνητικής οδού προκαλούν τους αρθρωτές να αλλάξουν με αποτέλεσμα να αλλάζει το σχήμα της φωνητικής οδού και να προκαλούνται διαφορετικοί ήχοι. Τα δεδομένα για το μοντέλο των αρθρωτών συνήθως προέρχονται από την ανάλυση ακτίνων x του φυσικού λόγου.

Υπάρχουν μερικά συστήματα που έχουν επιδείξει κάποια ενθαρρυντικά αποτελέσματα [10] [11], με πιο πρόσφατο το HLSyn [12]. Πρόοδος επίσης επιτελείται στις μετρήσεις της διαδικασίας άρθρωσης με διάφορες τεχνικές όπως το ηλεκτροπαλατογράφημα (electropalletography), οι μικροδέσμες ακτίνων-Χ (x-ray microbeam) και το ηλεκτρομαγνητικό αρθρογράφημα (ElectroMagnetic Articulograph).

Όσο εξελίσσεται η τεχνολογία και βελτιώνεται η ικανότητα μας να μοντελοποιήσουμε τέτοιες διαδικασίες, η σύνθεση με μοντελοποίηση άρθρωσης θα καταστεί πιο διαδεδομένη. Προς το παρόν όμως στερείται πρακτικότητας στη χρήση [9].

3.1.3 Σύστημα σύνθεσης ομιλίας βασισμένη σε κανόνες (Formant synthesis)

Η πιο ευρέως χρησιμοποιημένη μέθοδος σύνθεσης είναι η σύνθεση ομιλίας βασισμένη σε κανόνες. Η σύνθεση με κανόνες είναι μια μέθοδος για να παράγουμε οποιαδήποτε λέξη ή πρόταση η οποία βασίζεται σε ακολουθίες φωνητικών / συλλαβικών συμβόλων ή δειγμάτων. Σε αυτή τη μέθοδο, οι κυριότερες παράμετροι για τις θεμελιώδεις μικρές μονάδες τις ομιλίας όπως συλλαβές, φωνήματα ή ομιλία περιόδου θεμελιώδους συχνότητας αποθηκεύονται και συνδέονται με κανόνες [4]. Υπάρχουν δύο βασικές δομές, η παράλληλη και η σε σειρά, αλλά την μέγιστη απόδοση μπορούμε να την πετύχουμε με τον συνδυασμό αυτών των δυο τεχνικών [9]. Η σύνθεση ομιλίας βασισμένη σε κανόνες παρέχει απεριόριστο αριθμό ήχων κάτι που την καθιστά πιο

ευέλικτη από ότι την σύνθεση ομιλίας βασισμένη στην συνένωση μονάδων. Όμως η σύνθεση ομιλίας βασισμένη σε κανόνες είναι μικρότερης καταληπτότητας από αυτή των συστημάτων συνένωσης μονάδων. Το γεγονός αυτό οφείλεται αφενός στην έλλειψη ακόμα επαρκούς γνώσης για την λειτουργία του ανθρώπινου μηχανισμού παραγωγής ομιλίας και αφετέρου στην δυσκολία εξαγωγής των παραμέτρων του μοντέλου αυτού μελετώντας μόνο το σήμα ομιλίας [13].

Η σύνθεση ομιλίας βασισμένη σε κανόνες βασίζει την λειτουργία της σε ένα θεωρητικό μοντέλο εξομοίωσης του ανθρώπινου μηχανισμού παραγωγής ομιλίας. Βάση αυτού του μοντέλου η στοματική κοιλότητα μπορεί να αναπαρασταθεί με ένα σύστημα χρονικά μεταβαλλόμενων ψηφιακών φίλτρων, σύμφωνα με κανόνες, τα οποία διεγείρονται από ένα σήμα που αντιστοιχεί στη μεταβολή της πίεσης του αέρα που ρέει μέσα σε αυτή.

Η αρκετά καλή ποιότητα ομιλίας με σχετικά μικρές απαιτήσεις σε αποθηκευτικό χώρο και η μεγάλη ευχέρεια στην μετατροπή του κωδικοποιημένου σήματος ομιλίας (αλλαγή χροιάς, ταχύτητα ομιλίας) είναι από τα βασικά πλεονεκτήματα αυτής της μεθόδου. Όμως το κύριο μειονέκτημα αυτής της τεχνικής είναι το γεγονός ότι η παραγωγή συνθετικής ομιλίας υψηλής ποιότητας είναι μια χρονοβόρα διαδικασία, κάνοντας την δημιουργία μεγάλων βάσεων ομιλίας δύσκολη διαδικασία.

3.1.4 Μέθοδος σύνθεσης ομιλίας με Hidden Markov Model (HMM)

Η μέθοδος αναγνώρισης λέξεων όπου κάθε λέξη αναπαρίσταται με ένα κρυφό μοντέλο Markov έχει ευρέως ερευνηθεί και αργότερα χρησιμοποιηθεί από το 1975 και μετά. Το HMM είναι ένα πανίσχυρο στατιστικό εργαλείο για την μοντελοποίηση των παραγωγικών ακολουθιών που μπορούν να χαρακτηριστούν από μία ελλοχεύουσα διαδικασία που παράγει μία αισθητή ακολουθία [14]. Σε αυτή τη μέθοδο, κάθε λέξη μοντελοποιείται με ένα δίκτυο μετάβασης το οποίο έχει ένα μικρό αριθμό καταστάσεων N όπως ακριβώς ο αριθμός των φωνημάτων. Κάθε κατάσταση αντιστοιχεί σε ένα σύνολο χρονικών γεγονότων στην ομιλούμενη λέξη. Μολονότι μπορούμε να παρατηρήσουμε ένα γεγονός σε κάθε κατάσταση, η ίδια κατάσταση δεν είναι δυνατόν να παρατηρηθεί. Έτσι αυτό το μοντέλο αναφέρεται ως «κρυφό» («hidden») [4].

Χρησιμοποιώντας την πλήρως επισημειωμένη βάση δεδομένων, το στάδιο της εκπαίδευσης είναι υπεύθυνο για την παραμετροποίηση του σήματος φωνής και την εξαγωγή της κατάλληλης πληροφορίας τόσο σε ακουστικό όσο και σε γλωσσολογικό επίπεδο. Κατόπιν, πραγματοποιείται η εκτίμηση των παραμέτρων των μοντέλων HMM, σύμφωνα με το κριτήριο μέγιστης πιθανοφάνειας (ML - maximum likelihood criterion) [15].

3.1.5 Υβριδικές τεχνικές σύνθεσης ομιλίας

Οι υβριδικές τεχνικές αναφέρονται σε προσπάθειες αποδοτικού συνδυασμού των υπαρχόντων προσεγγίσεων με στόχο την εκμετάλλευση των πλεονεκτημάτων που προσφέρει η καθεμία. Οι γνωστότερες υβριδικές τεχνικές αφορούν προσπάθειες ενοποίησης: α) της σύνθεσης βασισμένη σε κανόνες με την βοήθεια HMM [16], της σύνθεσης βασισμένης σε κανόνες και της σύνθεσης με συνένωση μονάδων [17] [18], γ) της σύνθεσης με μοντελοποίηση άρθρωσης με την βοήθεια HMM [19] και δ) της σύνθεσης με HMM και της σύνθεσης με συνένωσης μονάδων [20].

3.2 Συστήματα μετατροπής Κειμένου-σε-Ομιλία ανοιχτού κώδικα

Στον χώρο της σύνθεσης ομιλίας υπάρχει μία μεγάλη ποικιλία μη εμπορικών συστημάτων σύνθεσης φωνής χάρη στο μεγάλο εύρος εφαρμογών και στην συστηματική έρευνα που γίνεται στον χώρο τα τελευταία χρόνια από ακαδημαϊκά ιδρύματα αλλά και από ιδιωτικούς ερευνητικούς φορείς. Αυτά τα συστήματα ως επί τον πλείστον προέρχονται από ερευνητικά προγράμματα στο τομέα της τεχνολογίας ομιλίας. Παρακάτω θα παρουσιαστούν και θα γίνει μία πρώτη εισαγωγή σε συστήματα σύνθεσης ομιλίας. Η επιλογή των συστημάτων έγινε με γνώμονα την διαθεσιμότητά τους, που σημαίνει ότι προτιμώνται συστήματα ελεύθερα (ανοιχτού κώδικα), αλλά και σύμφωνα με την αποτελεσματικότητα, την διαθεσιμότητα τους σε διάφορες γλώσσες και την επεκτασιμότητα τους.

3.2.1 eSpeak

Το σύστημα eSpeak είναι ένα λογισμικό μετατροπής κειμένου σε ομιλία ανοιχτού κώδικα που υποστηρίζει πολλές γλώσσες και μπορεί να δουλέψει είτε σε πλατφόρμα linux είτε σε πλατφόρμα windows. Το σύστημα αυτό περιλαμβάνει μια παραμετρική σύνθεση ομιλίας με κανόνες. Ο έλεγχος των παραμέτρων της σύνθεσης γίνεται βάση κανόνων που αναπτύσσονται από ειδικούς. Αυτό του επιτρέπει την ύπαρξη πολλών γλωσσών σε μικρό μέγεθος. Μερικές από τις 50 γλώσσες που υποστηρίζει είναι τα αλβανικά, τα δανέζικα, τα γερμανικά και τα ελληνικά. Η ομιλία η οποία παράγεται είναι με περιορισμένη φυσικότητα αλλά υπάρχει η δυνατότητα να επιτευχθεί υψηλή ευκρίνεια, ιδιαίτερα για τους χρήστες που χρησιμοποιούν το σύστημα συχνά. Συνήθως χρησιμοποιείται σε συνδυασμό με προγράμματα ανάγνωσης οθόνης για να παρέχετε πρόσβαση σε χρήστες με προβλήματα όρασης σε υπολογιστές. Το συγκεκριμένο σύστημα επιπλέον έχει το πλεονέκτημα ότι δεν απαιτείται πολύ μνήμη και μπορεί να μεταφερθεί εύκολα σε μια νέα πλατφόρμα. Ωστόσο, η χρήση του δεν είναι συνήθως αποδεκτή γιατί η φωνή που περιμένει η πλειοψηφία των χρηστών να ακούσει θέλει να μοιάζει με ανθρώπινη και όχι με ρομποτική.

3.2.2 Festival

Το σύστημα Festival αναπτύχθηκε με στόχο να είναι μια πλατφόρμα ανάπτυξης για την αξιολόγηση των διάφορων ενοτήτων που συνθέτουν ένα σύστημα μετατροπής ΚσΟ. Είναι ανοιχτού κώδικα και η διάδοσή του κατά κύριο λόγο οφείλεται στο γεγονός ότι παρέχει μία εργαλειοθήκη για την ανάπτυξη συνθετικών φωνών. Στο σύνολο του προσφέρει, δυνατότητες σύνθεσης ομιλίας μέσω ενός πλήθους API: από επιπέδου κελύφους (shell level), μέσω ενός Scheme command interpreter, μέσω μίας βιβλιοθήκης στην C++, στην Java αλλά και μίας διεπαφής Emacs. Υπάρχουν διαθέσιμες διάφορες εκδόσεις του συστήματος με διαφορετικές τεχνολογίες και για διαφορετικές γλώσσες. Το Festival είναι ένα πολυγλωσσικό σύστημα το οποίο υποστηρίζει Αγγλικά και Ισπανικά, με την αγγλική γλώσσα να είναι η πιο ανεπτυγμένη. Άλλες ερευνητικές ομάδες αναπτύσσουν νέες γλώσσες για το σύστημα. Στην βασική του έκδοση είναι ένα σύστημα που βασίζεται στην σύνθεση με βάση την κωδικοποίηση κυματομορφής και έχει την δυνατότητα να χρησιμοποιεί διάφορες φωνές. Ωστόσο, επειδή είναι ένα ερευνητικό σύστημα, οι φωνές είναι ως επί τον πλείστον από μη επαγγελματίες ομιλητές και η τμηματοποίηση των ηχογραφήσεων δεν έχει αρκετή ποιότητα σε σύγκριση με τα εμπορικά συστήματα. Ένα μεγάλο μειονέκτημα του συστήματος είναι ότι ναι μεν αναπτύχθηκε με ελάχιστες απαιτήσεις συστήματος, αλλά κάνει χρήση υπερβολικά μεγάλο μέρους της μνήμης και θέλει πολύ χρόνο για την επεξεργασία της κάθε πρότασης. για τον λόγο αυτό αναπτύχθηκε μια αποτελεσματικότερη έκδοση η FLite (Festival Lite). Το Flite είναι ανεπτυγμένο αποκλειστικά στην γλώσσα προγραμματισμού C και έχει σχεδιαστεί με στόχο να είναι μεταφέρσιμο σε κάθε πλατφόρμα. Αυτή η

έκδοση είναι πολύ πιο αποτελεσματική, αλλά λιγότερο ευέλικτη, και εξακολουθεί να περιορίζεται από την ποιότητα των φωνών. Η πλατφόρμα FLite αρχικά σχεδιάστηκε για μικρά ενσωματωμένα συστήματα ή για μεγάλους διακομιστές.

Τέλος το Festival έχει την δυνατότητα να συνθέτει ομιλία με τρείς διαφορετικούς τρόπους. Με την μέθοδο συνένωσης μονάδων (Unit Selection), με την χρήση κρυφών Μαρκοβιανών Μοντέλων (HMM-based method) και με την χρήση δίφωνων (Diphone). Η τελευταία μέθοδος βέβαια θεωρείται παλιάς τεχνολογίας και τείνει να εγκαταλείπεται. Εργαλεία και οδηγίες για την κατασκευή νέων φωνών και την ανάπτυξη νέων γλωσσών παρέχονται από την βιβλιοθήκη Edinburgh Speech Tools και είναι διαθέσιμα μέσω του Carnegie Mellon's FestVox project.

3.2.3 FreeTTS

Το σύστημα FreeTTS είναι ένα μεταφερόμενο κομμάτι της πλατφόρμας FLite, γραμμένο στην γλώσσα προγραμματισμού java με σκοπό την αύξηση της ευελιξίας του, την μεταφερσιμότητα του αλλά και την ανεξαρτησία του από την πλατφόρμα. Μία από τις δυσκολίες του FLite είναι η διασύνδεση με την συσκευή ήχου που πρέπει να προσαρμοστεί για κάθε διαφορετική πλατφόρμα. Το περιβάλλον της java προσφέρει μια αφαίρεση για την συσκευή ήχου, καθιστώντας την μεταφερσιμότητα ασήμαντη. Το σύστημα FreeTTS μπορεί να δουλέψει σε όλες τις πλατφόρμες λογισμικού, όπως είναι τα Mac, το Linux και τα windows. Η μόνη προϋπόθεση είναι η ύπαρξη εγκατεστημένου λογισμικού της java. Το FreeTTS χρησιμοποιεί τις ίδιες φωνές με το Festival με αποτέλεσμα να μοιράζονται τους ίδιους περιορισμούς ποιότητας.

3.2.4 MARYTTS

Το σύστημα μετατροπής κειμένου σε ομιλία OpenMary (Modular Architecture for research on speech Synthesis - Σπονδυλωτή Αρχιτεκτονική για την Έρευνα στην Σύνθεση Ομιλίας.) δημιουργήθηκε μετά από την συνεργασία από το Γερμανικό ερευνητικό κέντρο DKFI και από τα πανεπιστήμια του Saarbrucken και του Saarland. Αυτή την στιγμή το πρόγραμμα συντηρείται και αναβαθμίζεται από το ερευνητικό κέντρο του DFKI, το οποίο παράγει νέες εκδόσεις του συστήματος. Το πρόγραμμα παρέχει υψηλή ποιότητα συνθετικής ομιλίας, όπως για παράδειγμα για την γερμανική και την αγγλική γλώσσα [21]. Αυτή η πλατφόρμα χρησιμοποιήθηκε στην παρούσα διπλωματική διατριβή. Είναι μία ανοιχτού κώδικα πολυγλωσσική πλατφόρμα, γραμμένη σε JAVA. Το Mary περιέχει μία εργαλειοθήκη για την «εισαγωγή νέας γλώσσας» που προσφέρει ένα σύνολο εργαλείων για την εισαγωγή του σώματος κειμένων της νέας γλώσσας σε μία βάση δεδομένων, την επιλογή και ηχογράφηση των κατάλληλων προτάσεων από το σώμα, και την κατασκευή κάποιων κανόνων επεξεργασίας φυσικής γλώσσας με την χρήση της τεχνικής «γράμματος-προς-ήχο» (letter-to-sound) αναπαράστασης. Τέλος για την σύνθεση φωνής χρησιμοποιεί τεχνικές βασισμένες σε Κρυφά Μαρκοβιανά Μοντέλα (HMM), στην τεχνική συνένωσης μονάδων (Unit Selection) αλλά και στην συνένωση δίφωνων με τη χρήση του αλγόριθμου MBROLA [22]. Επιμέρους ανάλυση του συστήματος γίνεται στο επόμενο κεφάλαιο.

4. ΣΥΣΤΗΜΑ ΜΕΤΑΤΡΟΠΗΣ ΚΕΙΜΕΝΟΥ ΣΕ ΟΜΙΛΙΑ OPENMARY

4.1 Βασικά μέρη της πλατφόρμας OpenMary

Μπορούν να διακριθούν τέσσερα μέρη του συστήματος μετατροπής κειμένου σε ομιλία

- Η προεπεξεργασία ή κανονικοποιήση κειμένου
- Η επεξεργασία φυσικής γλώσσας, κάνοντας γλωσσική ανάλυση και σήμανση
- Ο υπολογισμός των ακουστικών παραμέτρων, στον οποίο μεταφράζεται η γλωσσολογική δομή της συμβολικής σήμανσης σε έναν πίνακα που περιέχει μόνο τις φυσικές παραμέτρους
- Η σύνθεση, στην οποία μετατρέπεται ο πίνακας παραμέτρων σε αρχεία ήχου

4.1.1 Προ-επεξεργασία κειμένου (ή κανονικοποίηση κειμένου)

Η προ-επεξεργασία κειμένου ή κανονικοποίηση κειμένου περιλαμβάνει τον κατακερματισμό, την επέκταση των συντομογραφιών και την επέκταση των αριθμών. Παράλληλα μία υποτυπώδης εσωτερική δομή XML χτίζεται γύρω από το κείμενο εισόδου και σταδιακά μεταφράζει κάθε προκαθορισμένη μορφή τονισμού που μπορεί να δοθεί στο κείμενο εισόδου.

4.1.2 Επεξεργασία φυσικής γλώσσας (Natural language processing – NLP)

Η επεξεργασία φυσικής γλώσσας είναι υπεύθυνη για τον υπολογισμό των λεκτικών δεδομένων από το γραπτό κείμενο εισόδου, δηλαδή τα φωνήματα και τις ετικέτες τονισμού. Στο πρώτο βήμα του NLP εκτελείται μέρος της επισήμανσης των μερών του λόγου και της συντακτικής κατάτμησης. Στην συνέχεια πραγματοποιείται μία αναζήτηση στο λεξικό προφοράς. Άγνωστες λέξεις αποσυντίθεται μορφολογικά και βάση των κανόνων που χρησιμοποιούνται για την μορφολογική ανάλυση γίνεται φωνολογική επεξεργασία. Ανεξάρτητα από την αναζήτηση στο λεξικό, τα σύμβολα για τον τονισμό και την δομή της φράσης αποδίδονται βάση της χρήσης κανόνων, χρησιμοποιώντας σημεία στίξης, μέρος των πληροφοριών των μερών του λόγου και την τοπική συντακτική πληροφορία.

Η επεξεργασία φυσικής γλώσσας οργανώνεται σε έναν σπονδυλωτό τρόπο και περιέχει τα ακόλουθα εργαλεία:

- Το υποσύστημα επισημείωσης μερών του λόγου (part of speech tagger)
- Τον κατατμητή (chunker), που κάνει μερική συντακτική ανάλυση
- Μετατροπή γραφημάτων σε φωνήματα χρησιμοποιώντας:
    - Ένα λεξικό για τα γνωστά στοιχεία
    - Κανόνες που χρησιμοποιούν μορφολογική ανάλυση για τα άγνωστα στοιχεία
    - Κανόνες συλλαβισμού φωνολογικής και έντασης της λέξης (word stress)
- Επισημείωση επιτόνων (pitch accent) και τόνων ορίου (boundary tone) με βάση το σύστημα επισημείωσης ToBI.
- Φωνολογικοί κανόνες (postlexical phonological rules)

4.1.3 Υπολογισμός των ακουστικών παραμέτρων

Αυτή η πλούσια σε δεδομένα είσοδος που προκύπτει από τον NLP μεταφράζεται σε ένα αρχείο ακουστικών παραμέτρων, εφαρμόζοντας ένα μοντέλο για την διάρκεια και τον τονισμό.

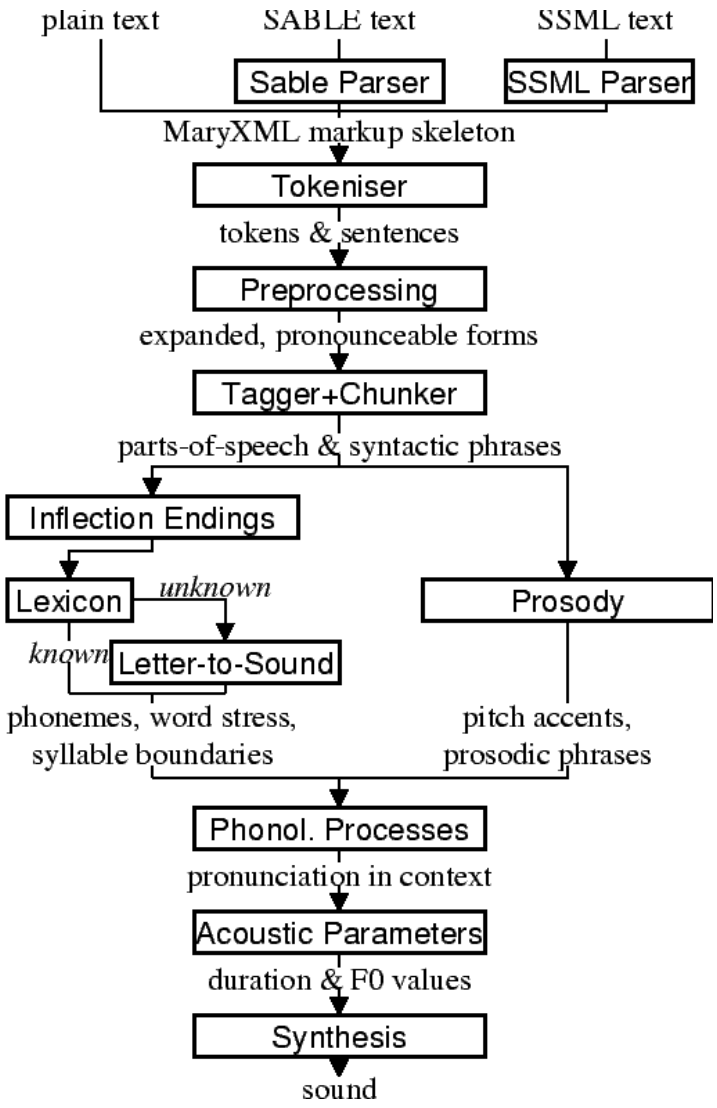
Η έξοδος που προκύπτει είναι ένα αρχείο παραμέτρων που χρησιμοποιείται από πολλά συστήματα σύνθεσης ομιλίας. Σαν έναν τύπο κυματοειδούς μορφής συνθέτη, χρησιμοποιείται το σύστημα σύνθεσης MBROLA, οπότε η μορφή του αρχείου παραμέτρων που παράγεται είναι συμβατή με αυτή του MBROLA. Κάθε φωνητικό σύμβολο αποδίδεται σε διάρκεια χιλιοστών δευτερολέπτου, και σε μερικά αποδίδεται άλλο ένα ζεύγος τιμής (χρόνος, συχνότητα), όπου ο χρόνος είναι το ποσοστό της διάρκειας της φωνητικής μονάδας και συχνότητα είναι μετρημένη σε HZ.

4.1.4 Ο συνθέτης

Τέλος, ο συνθέτης ήχου παράγει ένα αρχείο ήχου. Γίνεται χρήση του MBROLA για την σύνθεση των δίφωνων καθώς και του αλγορίθμου συνένωσης μονάδων ομιλίας – unit selection ο οποίος προέρχεται από το σύστημα FreeTTS. Επιπλέον μπορεί να γίνει χρήση και του αλγορίθμου HMM. Μπορούν να παραχθούν πολλές μορφές του ηχητικού αρχείου συμπεριλαμβανομένων και 16bit wav, aiff , au και mp3.

4.2 Αρχιτεκτονική συστήματος

Στο παρακάτω διάγραμμα φαίνεται η αρχιτεκτονική διάταξη του MARY και των διεργασιών που επιτελούνται προκειμένου να παραχθεί ομιλία βάσει του γερμανικού συστήματος. Η αρχιτεκτονική επεξεργασίας για τις άλλες γλώσσες είναι παρόμοια. Παρακάτω γίνεται μία μικρή επεξήγηση της βασικής αρχιτεκτονικής επεξεργασίας μιας γλώσσας [23].



Εικόνα 1: Η Αρχιτεκτονική της πλατφόρμας OpenMary.

4.2.1 Είσοδος κειμένου

Το απλό κείμενο είναι πιο βασική μορφή εισόδου. Τίποτα δεν είναι γνωστό ούτε για την δομή ή για το νόημα του κειμένου. Το κείμενο ενσωματώνεται σε ένα αρχείο της μορφής MaryXML για την περαιτέρω επεξεργασία του. Άλλες μορφές εισαγωγής κειμένου είναι η SABLE και η SSML. Η SABLE και SSML είναι μία γλώσσα επισημείωσης (markup language) για τον σχολιασμό κειμένων της σύνθεσης ομιλίας. Οι γλώσσες επισημείωσης για την σύνθεση ομιλίας είναι χρήσιμες για την παροχή πληροφοριών σχετικές με την δομή ενός κειμένου, την σημασία των αριθμών, ή την σημαντικότητα των λέξεων, έτσι ώστε αυτή η πληροφορία να μπορεί να εκφραστεί κατάλληλα στην ομιλία (όπως οι παύσεις στα κατάλληλα σημεία, η ανάγνωση τηλεφωνικών αριθμών, ή η έμφαση στις λέξεις που χρειάζεται).

4.2.2 Γλώσσα επισημείωσης MaryXML

Το MaryXML είναι μια εσωτερική, σχετικά χαμηλού επιπέδου γλώσσα σήμανσης που αντανακλά τις δυνατότητες διαμόρφωσης αυτού του συγκεκριμένου ΚσΟ συστήματος. Αυτή η βασική δομή πριν τον κατακερματισμό (tokenization), δεν απαιτείται να συμμορφωθεί με το MaryXML σχήμα, το οποίο υποθέτει ότι τα δεδομένα του κειμένου έχουν κατακερματιστεί. Όλα τα επόμενα ενδιάμεσα αποτελέσματα (αποτελέσματα εξόδου) συμμορφώνονται με το σχήμα MaryXML. Μέσω των επόμενων ενοτήτων επεξεργασίας, η δομή του MaryXML εμπλουτίζεται. Εάν μία ενότητα βρει τον τύπο της πληροφορίας που πρέπει να προσθέσει δίνεται προτεραιότητα σε αυτήν. Αυτό σημαίνει ότι οι ενδείξεις που εκφράζονται στη σήμανση εισαγωγής (π.χ. SABLE, SSML) θεωρούνται ως συμπληρώματα στη ΚσΟ ανάλυση των ενοτήτων της εισαγωγής [23].

4.2.3 Υποσύστημα κατακερματισμού (Tokenizer)

Το υποσύστημα κατακερματισμού (Tokenizer) τεμαχίζει το κείμενο σε tokens, για παράδειγμα λέξεις και σημεία στίξης. Χρησιμοποιεί ένα σύνολο κανόνων που καθορίζονται μέσω της ανάλυσης του πυρήνα του σώματος των κειμένων της γλώσσας (corpus) για να επισημάνουν την έννοια των κατακερματισμένων λέξεων που βασίζεται στο περιβάλλον πλαίσιο. Κάθε κατακερματισμένη λέξη περικλείεται από μία ετικέτα (tag) MaryXML του τύπου <t>...</t> . Όλες οι τοπικές πληροφορίες για ένα μέρος του λόγου που καθορίζονται από τα επόμενα βήματα επεξεργασίας προστίθενται στην ετικέτα του (<t>) σαν ζεύγος ιδιότητας/αξίας. Επιπλέον, η στίξη χρησιμοποιείται για να καθορίσει την έναρξη και το τέλος των προτάσεων που είναι χαρακτηρισμένες χρησιμοποιώντας την ετικέτα MaryXML <s>...< /s> η οποία εσωκλείει μια πρόταση [23].

Παράδειγμα κατακερματισμού μίας πρότασης στην Ελληνική γλώσσα:

```
<p>
    <s>
        <t g2p_method="lexicon" ph="k a - ' l o s" pos="RB">
            Καλώς
        </t>
        <t g2p_method="rules" ph="' i - r T a - t e" pos="VB">
            ήρθατε
        </t>
        <t g2p_method="lexicon" ph="s t o" pos="PDT">
            στο
        </t>
        <t g2p_method="lexicon" ph="' s i - s t i - m a" pos="NN">
            σύστημα
        </t>
        <t g2p_method="rules" ph="m e - t a - t R o - ' p i s" pos="NN">
            μετατροπής
        </t>
        <t g2p_method="rules" ph="c i - ' m e - n u" pos="NN">
            κειμένου
        </t>
        <t g2p_method="lexicon" ph="s e" pos="IN">
            σε
        </t>
        <t g2p_method="lexicon" ph="o - m i - ' l i - a" pos="NN">
            ομιλία
        </t>
        <t pos="$PUNCT">
            .
        </t>
    </s>
</p>
```

4.2.4 Κανονικοποίηση Κειμένου (Text normalization)

Στην ενότητα προ-επεξεργασίας, εκείνα τα tokens στα οποία η προφορική μορφή δεν αντιστοιχεί επακριβώς στη γραπτή μορφή, αντικαθίστανται από ένα token το οποίο έχει μία μορφή η οποία μπορεί να προφερθεί πιο εύκολα [23].

Η προφορά των αριθμών για παράδειγμα εξαρτάται σε μεγάλο βαθμό από τη σημασία τους. Διαφορετικοί τύποι αριθμών, όπως οι βασικοί και οι τακτικοί αριθμοί, τα ποσά νομίσματος, ή οι τηλεφωνικοί αριθμοί, πρέπει να προσδιοριστούν υπό αυτήν τη μορφή, είτε από τη σήμανση εισαγωγής είτε από τα συμφραζόμενα του κειμένου, και να αντικατασταθούν από τις κατάλληλες συμβολικές σειρές. Ενώ η επέκταση των βασικών αριθμών είναι απλή διαδικασία, η επέκταση των τακτικών αριθμών δημιουργεί ενδιαφέροντα προβλήματα στα ελληνικά αλλά και σε άλλες γλώσσες, λόγω των κλίσεων τους. Από την μία, η επέκταση ενός τακτικού αριθμού εξαρτάται από το μέρος του λόγου (επίρρημα ή επίθετο), αλλά από την άλλη, όταν πρόκειται για επίθετα, η κατάληξη τους εξαρτάται από το γένος και τον αριθμό (ενικός, πληθυντικός) της λέξης που προσδιορίζει και στην περίπτωση της ονοματικής φράσης που ο τακτικός αριθμός ανήκει. Στο στάδιο της προ-επεξεργασίας, καμία τέτοιου είδους πληροφορία δεν είναι διαθέσιμη, έτσι ο τακτικός αριθμός απλά χαρακτηρίζεται υπό αυτήν τη μορφή και του δίνεται μία τυπική επέκταση. Για παράδειγμα, ο αριθμός «1», θα γινόταν «πρώτος» στην επιρρηματική θέση και «ένας» όταν θα συμμετείχε στην πρόταση ως επίθετο. Αυτή η ενότητα προσθέτει αυτήν την πληροφορία κατάληξης στην ετικέτα (tag) ενός αριθμού. Με βάση αυτήν την σήμανση, η κατάλληλη κατάληξη θα επιλεχτεί κατά τη διάρκεια της φωνητικοποίησης (phonemisation) [23].

Επιπλέον πρόβλημα είναι η προφορά των συντμήσεων. Υπάρχουν δύο κύριες κατηγορίες συντμήσεων και διαχωρίζονται: Σε αυτές που προφέρονται έτσι όπως είναι , όπως η «ΗΠΑ», και σε αυτές που χρειάζονται επέκταση. Η πρώτη κατηγορία προφέρεται σωστά με την χρήση κανόνων συλλαβισμού.

Η δεύτερη κατηγορία διαβάζεται χρησιμοποιώντας έναν πίνακα επέκτασης, που περιέχει μία γραφηματική και προαιρετικά μία φωνητική επέκταση. Το τελευταίο είναι ιδιαίτερα χρήσιμο για τις ξένες συντμήσεις, όπως η λέξη «FBI» που προφέρεται στα γερμανικά με την αγγλική προφορά «Ef-bi-‘ai» [23].

4.2.5 Επισημείωση και Κατάτμηση των μερών του λόγου (Part-of-speech tagger and chunker)

Ένας αναλυτής κατάτμησης χρησιμοποιείται για να καθορίσει τα όρια των ονοματικών φράσεων ουσιαστικού, των εμπρόθετων φράσεων και των φράσεων επιθέτου. Η πληροφορία του Μέρους-Του-Λόγου και της κατάτμησης προστίθεται στην ετικέτα <t> για κάθε token. Για την πληροφορία κατάτμησης, αυτό δεν είναι πραγματικά μια πολύ ικανοποιητική λύση, αφού η τοπική συντακτική δομή δύσκολα μπορεί να θεωρηθεί ιδιοκτησία ενός μεμονωμένου token. Παρ’ όλα αυτά, η πιο λογική αντιπροσώπευση της συντακτικής δομής ως δομή δέντρων XML ενδεχομένως θα συγκρουόταν με την προσωδιακή δομή, εξαιτίας του γεγονότος ότι η συντακτική και προσωδιακή δομή δεν είναι βέβαιο ότι θα συμπέσουν σε όλες τις περιπτώσεις. Αφού η XML επιτρέπει μόνο μια κατάλληλη δομή δέντρων, χωρίς το πέρασμα των ακραίων περιπτώσεων, η μόνη εναλλακτική λύση φαίνεται να είναι να σταματήσει η αντιπροσώπευση XML στην παρούσα περίπτωση. Για παράδειγμα μία αναπαράσταση διαγράμματος θα επέτρεπε μεγαλύτερη ευελιξία. Εντούτοις, η προς το παρόν χρησιμοποιημένη κωδικοποίηση με τη δομή XML που αντιπροσωπεύει την προσωδιακή δομή και τη συντακτική δομή να «ενθυλακώνεται» στις συμβολικές ετικέτες φαίνεται να είναι μια βιώσιμη λύση [23].

4.2.6 Μετατροπή Γράμμα-σε-Ήχο (Letter-to-sound conversion)

Οι άγνωστες λέξεις που δεν μπορούν να ακουστικοποιηθούν με τη βοήθεια του λεξικού αναλύονται από έναν αλγόριθμο «μετατροπής γράμματος-σε-ήχο» ("letter-to-sound conversion" algorithm). Οι κανόνες γράμματος-σε-ήχο είναι στατιστικά εκπαιδευμένοι στο λεξικό του OpenMary.

Ο συλλαβισμός (syllabification) των μεταγραφόμενων λέξεων είναι βασισμένο στις τυποποιημένες φωνολογικές αρχές όπως η ιεραρχία ηχηρότητας των φωνημάτων, της αρχής maximal onset, της υποχρεωτικής αρχής coda και των φωνητικών περιορισμών για τη γερμανική γλώσσα.

Στο τέλος, ένας αλγόριθμος ανάθεσης τόνου λέξης (word stress assignment algorithm) αποφασίζει ποια συλλαβή παίρνει τον αρχικό λεξικό τόνο. Καμία βασισμένη σε κανόνες δευτερογενής ανάθεση τόνου δεν προστίθεται προς το παρόν [23].

4.2.7 Έξοδος Φωνημάτων (Phonemisation output)

Η έξοδος των στοιχείων φωνημάτων περιλαμβάνει την φωνητική μετατροπή (με την χρήση του προτύπου SAMPA) για κάθε token [23].

4.2.8 Υποσύστημα Προσωδίας

Η προσωδία μοντελοποιείται χρησιμοποιώντας τον αλγόριθμο ToBI, ("Tones and Break Indices"). Ο ToBI περιγράφει τον τονισμό από την σκοπιά των σημείων θεμελιώδους συχνότητας (F0), που διακρίνουν μεταξύ των επιτόνων που συνδέονται με τις προεξέχουσες λέξεις και των τόνων ορίου που συνδέονται με το τέλος μιας φράσης. Το μέγεθος διακοπής στην φράση κωδικοποιείται με δείκτες διακοπής. Στο OpenMary, οι δείκτες διακοπής που χρησιμοποιούνται είναι οι ακόλουθοι: "δείκτης διακοπής 2" είναι μία πιθανή θέση ορίου που δηλώνει το όριο προσωδιακής λέξης, "δείκτης διακοπής 3" όριο ενδιάμεσης φράσης, "δείκτης διακοπής 4" όριο επιτονικής φράσης, και οι δείκτες "δείκτης διακοπής 5" και "δείκτης διακοπής 6" αντιπροσωπεύουν τα όρια του τέλους της πρότασης και του τέλους της παραγράφου αντίστοιχα. Οι κανόνες της ενότητας της προσωδίας βασίζονται σε συμβολισμούς του ToBI, σε επόμενο βήμα, αυτοί μεταφράζονται σε διακριτές F0 συχνότητες και διάρκειες παύσεων [23].

Κάποια μέρη του λόγου, όπως τα ουσιαστικά και τα επίθετα, πάντα δέχονται έναν επίτονο. Υπάρχουν όμως και μέρη του λόγου τα οποία συχνά δεν δέχονται επίτονο. Τα μέρη του λόγου ταξινομούνται βάση ιεραρχίας (κατά προσέγγιση: πλήρη ρήματα > ρήματα > επιρρήματα) , σύμφωνα με την έμφαση που τους δίνεται. Αυτή η ταξινόμηση χρησιμοποιείται όπου οι υποχρεωτικοί κανόνες ανάθεσης δεν τοποθετούν κάποιο επίτονο μέσα σε κάποια ενδιάμεση φράση. Σύμφωνα με την αρχή ToBI, κάθε ενδιάμεση φράση πρέπει να περιέχει τουλάχιστον έναν επίτονο (pitch accent). Σε αυτή την περίπτωση, το σημείο σε εκείνη την ενδιάμεση φράση με την υψηλότερη ταξινόμηση σύμφωνα με το μέρος του λόγου που ανήκει φέρει αυτόν τον επίτονο [23].

Μετά τον καθορισμό της θέσης των προσωδιακών ορίων και των επιτόνων (pitch accents), τα είδη των επιτόνων ορίζονται σύμφωνα με τον τύπο της πρότασης (καταφατική, ερωτηματική ολικής άγνοιας, ερωτηματική μερικής άγνοιας και επιφωνηματική). Για κάθε τύπο πρότασης ορίζονται τα είδη των επιτόνων (για κάθε προπυρηνική και πυρηνική θέση), οι ενδιάμεσοι τόνοι φράσης και τόνοι ορίου. Ο τελευταίος τόνος επιτονισμού φράσης και προσωδίας σε μια πρόταση είναι συνήθως διαφορετικός από τους υπόλοιπους [23].

4.2.9 Φωνολογικοί κανόνες (Post lexical phonological rules module)

Αφού μεταγραφούν οι λέξεις σε μια τυποποιημένη φθογγική σειρά συμπεριλαμβανομένων των ορίων των συλλαβών και του τόνου από τη μία, και τις ετικέτες προσωδίας για τους επίτονους και τα προσωδιακά όρια φράσης από την άλλη, η προκύπτουσα φωνολογική αναπαράσταση μπορεί να αναδομηθεί από ένα σύνολο φωνολογικών κανόνων. Αυτοί οι κανόνες λειτουργούν βάσει του γενικότερου πλαισίου των φωνολογικών πληροφοριών όπως ο επίτονος (pitch accent), η φραστική περιοχή ή, προαιρετικά, η ζητούμενη ακρίβεια της άρθρωσης. Οι κανόνες αυτοί μπορούν να εφαρμοστούν στις καταλήξεις. Όμως στην ομιλία με δίφωνα τέτοιες μειώσεις περιορίζουν την σαφήνεια κι έτσι απενεργοποιούνται εξορισμού [23].

4.2.10 Υπολογισμός των ακουστικών παραμέτρων (Calculation of acoustic parameters)

Αυτή η μονάδα εκτελεί την μετατροπή από το συμβολικό στο φυσικό επίπεδο. Η δομή MaryXML ερμηνεύεται από τους κανόνες διάρκειας και τους κανόνες ToBI. Η αναπαράσταση των τόνων χρησιμοποιεί ένα σύνολο σημείων για κάθε σύμβολο τόνου. Αυτά τα σημεία τοποθετούνται, στο χρονικό άξονα, σε συνάρτηση με τον πυρήνα της συλλαβής που σχετίζονται, και στον άξονα συχνότητας, τοποθετούνται σε σχέση με ένα φθίνον ζευγάρι τιμών (topline, baseline) που αντιπροσωπεύουν την υψηλότερη και χαμηλότερη πιθανή συχνότητα σε μια δεδομένη στιγμή. Το γεγονός είναι ότι αυτές οι γραμμές έχουν φθίνουσα κλίση (declination), καθώς το συνολικό F0 επίπεδο είναι υψηλότερο στην αρχή μίας φράσης από, τι κοντά στο τέλος. Προφανώς, οι πραγματικές τιμές συχνότητας του topline και της βασικής baseline πρέπει να τεθούν κατάλληλα για τη φωνή που χρησιμοποιείται κατά τη διάρκεια της σύνθεσης, και ιδιαίτερα σύμφωνα με το φύλο του ομιλητή [23].

4.2.11 Μονάδα Σύνθεσης (Synthesis module)

Μεταξύ άλλων, ο MBROLA χρησιμοποιείται για τη σύνθεση της έκφρασης με βάση την έξοδο της προηγούμενης ενότητας. Μπορούν να χρησιμοποιηθούν διάφορα σύνολα δίφωνων για ανδρικές και γυναικείες φωνές. Το MARY επίσης περιέχει τον κώδικα για την παραγωγή φωνής βασισμένη στον αλγόριθμο επιλογής μονάδων (unit selection) και HMM, από το σύστημα ανοιχτού κώδικα FreeTTS [23].

5. ΥΠΟΣΥΣΤΗΜΑ ΑΝΑΓΝΩΡΙΣΗΣ ΚΑΙ ΕΠΙΣΗΜΕΙΩΣΗΣ ΤΩΝ ΜΕΡΩΝ ΤΟΥ ΛΟΓΟΥ ΓΙΑ ΤΗΝ ΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ

5.1 Σύστημα αναγνώρισης και επισημείωσης μερών του λόγου

Ένα σύστημα αναγνώρισης και επισημείωσης των μερών του λόγου μίας πρότασης δημιουργεί και ταξινομεί τις λέξεις σε κλάσεις ανάλογα με το μέρος του λόγου που είναι. Κάποιος ίσως σκεφτεί ότι το πρόβλημα αυτό θα μπορούσε να λυθεί με τη χρήση ενός λεξικού. Η προσέγγιση, όμως, αυτή δεν είναι επαρκής. Τα συστήματα επισημείωσης Μέρος-Του-Λόγου (Part-Of-Speech taggers) έχουν να αντιμετωπίσουν άγνωστες λέξεις και λέξεις με διφορούμενη ετικέτα, όπως ουσιαστικά, επίθετα και ρήματα που λόγω της όμοιας δομής τους μέσα στην πρόταση μπορούν να έχουν δύο διαφορετικές συντακτικές ερμηνείες [24]. Για να μπορέσει να επιτευχθεί το καλύτερο αποτέλεσμα σε έναν συνθέτη ομιλίας χρειάζεται να υπάρχει ένα σύστημα επισημείωσης μερών του λόγου.

5.1.1 Περιγραφή του υποσυστήματος επισημείωσης μερών του λόγου

Το πρόβλημα της παρούσης πλατφόρμας που είχε δημιουργηθεί ήταν ότι έκανε χρήση της βασικής υποστήριξης της Ελληνικής γλώσσας. Οπότε σαν πρώτος στόχος σε αυτή την εργασία είναι η ενσωμάτωση του συστήματος επισημείωσης Μερών-του-Λόγου (Part-Of-Speech Tagger) για την Ελληνική γλώσσα.

Ήταν πολύ σημαντικό να μπορεί το σύστημα να αναγνωρίσει και να κατατάξει τις προτάσεις που δέχεται σαν είσοδο κειμένου. Μέσω της αναγνώρισης των μερών του λόγου θα μπορούσε στην συνέχεια να επιτευχθεί η γραμματική αναγνώριση της πρότασης και να ταξινομηθεί το είδος της πρότασης (καταφατική, ερωτηματική μερικής άγνοιας, ερωτηματική ολικής άγνοιας, επιφωνηματική ή αρνητική πρόταση).

Για τον σκοπό αυτό, έγινε χρήση του υποσυστήματος Μερών-του-Λόγου AUEB_POS_tagger [3]. Στον συγκεκριμένο POS tagger κάθε λέξη του δοθέντος κειμένου κατατάσσεται σε μια κατηγορία, όπου η κατηγορία αντιστοιχεί στα μέρη του λόγου (π.χ. ρήμα, επίθετο, άρθρο). Ο αλγόριθμος εκμάθησης που χρησιμοποιήθηκε για την δημιουργία του POS tagger ήταν ένας ταξινομητής Μέγιστης εντροπίας (Maximum Entropy classifier) του Πανεπιστημίου Stanford [3].

Το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger μας επέτρεπε μόνο την γραμματική αναγνώριση συγκεκριμένων ετικετών μερών του λόγου. Ήταν όμως ελλιπής η γραμματική αναγνώριση που μας παρείχε. Δεν ήταν σε θέση να αναγνωρίσει τα Wh words, τα ρήματα πρώτου και δεύτερου ενικού και τα αρνητικά μόρια (negative particle). Το συγκεκριμένο υποσύστημα επισημείωσης Μερών-του-Λόγου δεν υποστήριζε την ίδια μορφή εξόδου με την μορφή εξόδου της πλατφόρμας μετατροπής ΚσΟ OpenMary. Αυτά τα προβλήματα λύθηκαν στο δικό μας υποσύστημα με την χρήση αλγορίθμων.

Ο POS tagger αποτελείται από 6 συναρτήσεις οι οποίες είναι υπεύθυνες για την γραμματική αναγνώριση των κατακερματισμένων λέξεων της πρότασης. Το σύστημα αναγνώρισης μερών του λόγου δέχεται τις λέξεις της πρότασης μία – μία και τις αναγνωρίζει ως προς την γραμματική τους υπόσταση. Η γραμματική αναγνώριση των λέξεων γίνεται μέσω περίπλοκων αλγορίθμων. Κάθε μία λέξη (για την ακρίβεια εμφάνιση λέξης) σε ένα κείμενο αναπαριστάται ως ένα διάνυσμα ιδιοτήτων που παρέχει πληροφορίες για τη λέξη και τα συμφραζόμενα της. Ως συμφραζόμενα σε αυτό το σύστημα επισημείωσης μερών του λόγου θεωρούνται οι λέξεις («γειτονιά») που βρίσκονται πριν και μετά από την υπό κατάταξη λέξη (πριν και μετά τη λέξη που παριστάνεται με το συγκεκριμένο διάνυσμα) [3].

5.2 Ενσωμάτωση του υποσυστήματος UOA_POS_tagger στην πλατφόρμα OpenMary

Έχοντας αναλύσει το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger που ενσωματώθηκε περνάμε στο επόμενο βήμα το οποίο είναι η ενσωμάτωση του υποσυστήματος στην πλατφόρμα OpenMary TtS για την Ελληνική γλώσσα.

Σκοπός της ενσωμάτωσης είναι να διαχειρίζεται μέσω του διακομιστή η είσοδος κειμένου που πληκτρολόγησε ο χρήστης. Επομένως, για να επιτευχθεί αυτή η διαδικασία έπρεπε η είσοδος κειμένου που δίνεται στον διακομιστή του OpenMary και από αυτόν να αποστέλλεται στο υποσύστημα επισημείωσης μέρων του λόγου.

Για να μπορέσει να γίνει αυτή η διαδικασία δημιουργήθηκε η βασική συνάρτηση του προγράμματος, η UOA_POS_tagger, η οποία δέχεται σαν είσοδο μέσω του διακομιστή την προς επεξεργασία πρόταση. Μέσα σε αυτή την συνάρτηση προσαρμόστηκε και αναδιαμορφώθηκε το σύστημα επισημείωσης μερών του λόγου. Η πρόταση που έχει δοθεί από τον χρήστη σαν είσοδος πληροφορίας κατακερματίζεται μέσα στο υποσύστημα μας και δρομολογείται προς το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger. Εκεί γίνεται η γραμματική αναγνώριση της πρότασης. Πριν σταλθεί ξανά η πληροφορία των μερών του λόγου στο υποσύστημα μας αλλάζουμε τις ετικέτες επισημείωσης των κατακερματισμένων λέξεων με σκοπό να είναι διαχειρίσιμες από την πλατφόρμα OpenMary.

Η αναπροσαρμογή της εξόδου των μερών του λόγου γίνεται μέσα στο υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger. Στον πίνακα 1 στην στήλη «Κατηγορίες του μέρους του λόγου στο OpenMary» μπορείτε να δείτε την αναπροσαρμογή που γίνεται στις ετικέτες των μερών του λόγου του υποσυστήματος μας. Φαίνεται η αντιστοίχηση των ετικετών επισημείωσης ώστε να είναι συμβατή με την πλατφόρμα OpenMary. Συνοψίζοντας, αρχικά δημιουργήσαμε το υποσύστημα επισημείωσης μερών του λόγου UOA_POS_tagger. Εκεί πηγαίνει η είσοδος κειμένου που έχει δώσει ο χρήστης. Αφού κατακερματιστεί η πρόταση γίνεται η γραμματική αναγνώριση στο υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger.

Πριν γίνει η επιστροφή του κατακερματισμένου κειμένου εισόδου στο υποσύστημα επισημείωσης μερών του λόγου UOA_POS_tagger γίνεται αναπροσαρμογή της επισημείωσης των μερών του λόγου με σκοπό τη διόρθωση των ετικετών του μέρους του λόγου. Με την επιστροφή της εξόδου και κάνοντας τους απαραίτητους ελέγχους για την διόρθωση των μερών του λόγου αναπροσαρμόζονται οι κατακερματισμένες λέξεις και η γραμματική ανάλυση των λέξεων, οι οποίες χρησιμοποιούνται στην συνέχεια για την απόδοση του επιτονισμού της πρότασης.

Πίνακας 1: Αντιστοίχηση των ετικετών επισημείωσης

Κατηγορίες του υποσυστήματος Μερών-του-Λόγου AUEB_POS_tagger	Κατηγορίες μερών του λόγου στα αγγλικά	Κατηγορίες του μέρους του λόγου στο OpenMary
Ρήμα	verb	VB
Ουσιαστικό	noun	NN
Επίθετο	adjective	JJ
Επίρρημα	adverb	RB
Άρθρο	particle	RP
Αντωνυμία	pronoun	PRP
Αριθμητικό	numeral	CD
Πρόθεση	preposition	IN
Μόριο	particle	RP
Σύνδεσμος	conjunction	CC
Σημείο στίξης	punctuation	$PUNCT
Λέξη μερικής άγνοιας	Wh word	WP
Αρνητικό μόριο	negative particle	NRP
Άλλο	other	OTHER

- **Έλεγχος και ορισμός των Wh–Words**

Επειδή το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger δεν αναγνώριζε και δεν υποστήριζε τις ερωτηματικές λέξεις (Wh-words), έπρεπε να υλοποιηθεί στο υποσύστημα μας ένας αλγόριθμος ο οποίος θα μπορούσε να τα διακρίνει. Ήταν μέγιστης σημασίας να μπορέσουν να οριστούν τα Wh–Words. Με τον ορισμό τους και την εισαγωγή τους θα μπορούσε να γίνει η διάκριση και η αναγνώριση του είδους των ερωτηματικών προτάσεων, σε ερωτηματικές προτάσεις ολικής άγνοιας (ναι-όχι ερώτηση) και μερικής άγνοιας (ποιος-ποια ερώτηση) (π.χ. “Ποια είσαι εσύ;”, “Από που θέλετε να αναχωρήσετε;”). Επιπλέον εκτός των ερωτηματικών προτάσεων, θα μπορούσαμε να ελέγξουμε το πότε έχουμε μια λέξη η οποία μπορεί να αναγνωριστεί γραμματικά σαν μέρος του λόγου Wh–Word (WP). Η προσθήκη του αλγορίθμου και των κανόνων έγινε μέσα στην συνάρτηση UOA_POS_tagger. Για την επίτευξη της αναγνώρισης των Wh–word, αν μέσα στην πρόταση υπάρχει Wh–Word, γίνεται έλεγχος της κατακερματισμένης πρότασης εισόδου μέσα στο υποσύστημα μας. Με την ολοκλήρωση δηλαδή του υποσυστήματος Μερών-του-Λόγου AUEB_POS_tagger, την διόρθωση και αναπροσαρμογή της εξόδου που κάνουμε πριν σταλθούν οι αναγνωρισμένες λέξεις ξανά στο υποσύστημα μας, πραγματοποιείται μετά μέσα στο υποσύστημα μας ένας δεύτερος γραμματικός έλεγχος στην πρόταση. Κάθε λέξη της πρότασης ελέγχεται μέσω ενός αρχείου κειμένου, με σκοπό τον χαρακτηρισμό της λέξης ως Wh–word. Αυτός ο έλεγχος αφορά λέξεις που βρίσκονται όχι μόνο στην αρχή της πρότασης, αλλά σε οποιοδήποτε μέρος της πρότασης.

- **Έλεγχος και ορισμός των Προθέσεων – Preposition (IN)**

Επόμενο πολύ σημαντικό βήμα ήταν η σωστή αναγνώριση των προθέσεων. Το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger δεν αναγνώριζε με επιτυχία όλες τις προθέσεις. Υπήρχε αστοχία στην αναγνώριση των προθέσεων. Ήταν μεγάλης σημασίας η ορθή αναγνώριση των προθέσεων και η σωστή γραμματική απόδοση τιμής τους. Εμείς θέλαμε να επιτυγχάνεται σωστή αναγνώριση των προθέσεων για να μπορεί να επιτευχθεί ο αρχικός μας στόχος, δηλαδή η αποτελεσματικότερη αναγνώριση των ερωτήσεων μερικής άγνοιας και των προθέσεων κατά την διάρκεια της γραμματικής ανάλυσης των προτάσεων. Η αναγνώριση των προθέσεων θα βοηθούσε την καλύτερη απόδοση προσωδίας σε όλα τα είδη των προτάσεων. Γι αυτό τον λόγο μέσα στο υποσύστημα UOA_POS_tagger δημιουργήθηκε και προστέθηκε ένας αλγόριθμος ελέγχου και ορισμού των προθέσεων. Θα αναλυθεί διεξοδικά παρακάτω αλλά σαν μία πρώτη εισαγωγή θα πρέπει να αναφερθεί ότι μία πρόταση χαρακτηρίζεται ως ερώτηση μερικής άγνοιας όταν η πρώτη λέξη της ερωτηματικής πρότασης είναι Wh-word (WP) ή όταν μέσα σε μία ερωτηματική πρόταση, τις περισσότερες φορές, έχουμε ένα Wh–word να ακολουθείται από μία πρόθεση. Με την αναγνώριση των προθέσεων από το σύστημα επισημείωσης μερών του λόγου, τον περαιτέρω έλεγχο και την διόρθωση τους μπορούσε να επιτευχθεί με επιτυχία η αναγνώριση των ερωτήσεων μερικής άγνοιας. Η αναγνώριση των προθέσεων γίνεται με πανομοιότυπο τρόπο με την αναγνώριση των Wh–Word, γίνεται δηλαδή έλεγχος κάθε λέξης την κατακερματισμένης πρότασης εισόδου μέσω ενός αρχείου κειμένου στο οποίο έχουν οριστεί οι προθέσεις της Ελληνικής γλώσσας.

- **Έλεγχος και ορισμός των αρνητικών μορίων (negative particle)**

Το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger δεν υποστήριζε την αναγνώριση των αρνητικών μορίων (Δεν/Δε). Για την συμπλήρωση και έχοντας εξασφαλίσει την αναγνώριση των ερωτήσεων μερικής και ολικής άγνοιας, πραγματοποιήθηκε η εισαγωγή αλγορίθμου για την αναγνώριση και τον ορισμό της έννοιας των αρνητικών μορίων στο υποσύστημα μας με στόχο την αναγνώριση των αρνητικών ερωτήσεων. Αυτό έγινε όχι μόνο για την πληρότητα των ερωτηματικών προτάσεων στην Ελληνική γλώσσα, αλλά και για την αναγνώριση των αποφατικών προτάσεων, δηλωτικών και επιφωνηματικών. Για τον ορισμό του αρνητικού μορίου, υλοποιήθηκε επιπλέον άλλη μια συνάρτηση μέσα στο υποσύστημα UOA_POS_tagger. Προστέθηκε ένας αλγόριθμος για την γραμματική αναγνώριση των λέξεων που μπορούν να χαρακτηριστούν ως αρνητικά μόρια (negative particle) (NRP). Η σωστή αναγνώριση των μορίων δίνει την δυνατότητα της αναγνώρισης των αποφατικών προτάσεων, ερωτηματικών και μη, και την απόδοση της κατάλληλης τιμής στις αρνητικές λέξεις κατά την γραμματική ανάλυση των προτάσεων. Η αναγνώριση των αρνητικών μορίων (NRP) γίνεται με πανομοιότυπο τρόπο με την αναγνώριση των Wh word και των προθέσεων, δηλαδή γίνεται έλεγχος κάθε λέξης την κατακερματισμένης πρότασης εισόδου μέσω ενός αρχείου κειμένου στο οποίο έχουν δηλωθεί τα αρνητικά μόρια Δεν/Δε. Ένα αρνητικό μόριο μπορεί να βρίσκεται είτε στο μέσο είτε στην αρχή της πρότασης.

5.2.1 Η συνάρτηση UOA_POS_tagger

Για να μπορούν να δουλεύουν ενιαία όλες οι λειτουργικότητες που αναλύθηκαν και περιγράφτηκαν παραπάνω, δημιουργήθηκε μια συνάρτηση μέσα στην οποία ενσωματώθηκαν όλα αυτά. Η συνάρτηση που δημιουργήθηκε ονομάστηκε UOA_POS_tagger και περιέχει μέσα τους κατάλληλους αλγορίθμους για τον κατακερματισμό της πρότασης εισόδου και την γραμματική αναγνώριση και διόρθωση των λέξεων της πρότασης. Ο διακομιστής στέλνει την πρόταση μέσα στην συνάρτηση UOA_POS_tagger. Εκεί κατακερματίζεται η πρόταση και γίνεται χρήση των συναρτήσεων της γραμματικής αναγνώρισης του AUEB_POS_tagger.

Μόλις ολοκληρωθεί ο γραμματικός έλεγχος των λέξεων της πρότασης αναπροσαρμόζεται η πληροφορία εξόδου και επιστρέφει με την κατάλληλη μορφή στην συνάρτηση του υλοποιημένου συστήματος μας, Στην συνέχεια πραγματοποιείται ο έλεγχος των Wh-words, των προθέσεων και των αρνητικών μορίων. Αυτή η διαδικασία του ελέγχου είναι μια διαδικασία η οποία μπορεί να χαρακτηριστεί χρονοβόρα όταν υπάρξουν μεγάλες προτάσεις ή ολόκληροι παράγραφοι κειμένου. Επιπλέον ο έλεγχος κάθε κερματισμένης λέξης σε μία μεγάλη παράγραφο καταναλώνει τους πόρους μνήμης του συστήματος. Όμως, εφόσον το υποσύστημα Μερών-του-Λόγου AUEB_POS_tagger δεν αναγνώριζε αυτά τα τρία γραμματικά είδη η παραπάνω διαδικασία έπρεπε να υλοποιηθεί και να ενσωματωθεί μέσα στην βασική συνάρτηση του συστήματος μας.

Μέσω της γραμματικής αναγνώρισης κάθε λέξης της πρότασης και την ενσωμάτωση της αναγνώρισης των Wh-words, των προθέσεων και των αρνητικών μορίων μας δόθηκε η δυνατότητα της αναγνώρισης των προτάσεων και της ταξινόμησης τους σε δηλωτικές, ερωτηματικές και επιφωνηματικές.

5.3 Αναγνώριση του είδους της πρότασης και επιτονισμός

Η αναγνώριση του είδους της πρότασης και ο επιτονισμός των προτάσεων γίνεται μέσω της συνάρτησης Prosody. Η συνάρτηση αυτή υπάρχει στην πλατφόρμα OpenMary για τις γλώσσες που έχουν υλοποιηθεί, αλλά δεν υπήρχε για την Ελληνική γλώσσα.

Έχοντας ολοκληρώσει την γραμματική ανάλυση των λέξεων της πρότασης σκοπός ήταν η δημιουργία μιας συνάρτησης με στόχο την αναγνώριση αρχικά του είδους της πρότασης και στην συνέχεια με την χρήση κανόνων tobi να μπορεί να περιγραφεί ο επιτονισμός της πρότασης με την κατάλληλη επισημείωση έτσι ώστε να μπορεί να χρησιμοποιηθεί από την πλατφόρμα OpenMary.

Η συνάρτηση Prosody χρησιμοποιείται για να εισαχθούν οι κανόνες επιτονισμού της γλώσσας που έχουν δημιουργηθεί. Η συγκεκριμένη συνάρτηση δέχεται σαν παραμέτρους 4 μεταβλητές. Οι 2 μεταβλητές είναι αρχεία κανόνων και οι άλλες 2 μεταβλητές είναι μεταβλητές συστήματος. Η πρώτη παράμετρος που δέχεται η συνάρτηση αναγνώρισης του είδους της πρότασης και του επιτονισμού είναι οι κανόνες επιτονισμού από το αρχείου tobipredparams, το οποίο έχουμε δημιουργήσει και αναπροσαρμόσει βάσει των απαιτήσεων της Ελληνικής γλώσσας.

Το αρχείο tobipredparams που χρησιμοποιεί το σύστημα σύνθεσης ομιλίας OpenMary είναι πολύ σημαντικό για την σωστή λειτουργία του συνθέτη ομιλίας. Εμπεριέχονται μέσα σε αυτό οι κανόνες για το ποια μέρη του λόγου χρειάζονται κανόνες πρόβλεψης προσωδίας, για το ποια μέρη του λόγου δέχονται επίτονο, κανόνες για την απόδοση της σωστής προσωδίας ανάλογα την θέση του μέρους του λόγου μέσα στην πρόταση και κανόνες για τον επιτονισμό των προτάσεων ανάλογα με το είδος της πρότασης.

Ποιο αναλυτικά στο αρχείο tobipredparams αρχικά ορίζουμε τα μέρη του λόγου που μπορεί να αναγνωρίσει η πλατφόρμα μας. Αυτά τα μέρη του λόγου χωρίζονται σε δυο κατηγορίες. Στα μέρη του λόγου που επιδέχονται κανόνες για τον επιτονισμό τους με

την προσωδιακή ομιλία και στα μέρη του λόγου που δεν δέχονται κανόνες για τον επιτονισμό τους (functional λέξεις). Μέσω της αναγνώρισης του μέρους του λόγου έγινε η εισαγωγή των κανόνων απόδοσης και μη απόδοσης προσωδιακής ομιλίας.

Στην συνέχεια ορίζονται οι λίστες για το αν ένα μέρος του λόγου δέχεται επίτονο. Αν ένα μέρος του λόγου δέχεται επίτονο τότε αρχικά του δίνεται η ετικέτα “tone”, ενώ αν δεν δέχεται επίτονο ή αν ο επίτονος είναι ορισμένος στο null τότε δεν παίρνει κάποια τιμή. Επόμενο βήμα μέσα στο αρχείο TOBI είναι ο ορισμός του τύπου του επιτόνου για κάθε λέξη που έχει δεχτεί την ετικέτα “tone”. Οι κανόνες αυτοί χωρίζονται σε κανόνες για κάθε είδος πρότασης. Ακόμα χωρίζονται ανάλογα με την θέση της λέξης που δέχεται επίτονο μέσα στην πρόταση. Οι κανόνες για τις ερωτήσεις μερικής άγνοιας, ερωτήσεις ολικής άγνοιας και αρνητικών προτάσεων δημιουργήθηκαν κατά την εκπόνηση της διπλωματικής διατριβής. Έγινε προσθήκη των κανόνων επιτονισμού μέσα στο αρχείο tobipredparams. Οι κανόνες δημιουργήθηκαν με στόχο την απόδοση του σωστού προπυρηνικού και πυρηνικού επίτονου στις λέξεις της πρότασης εισόδου.

Παράδειγμα κανόνων που δημιουργήθηκαν για τις ερωτήσεις μερικής άγνοιας:

```
<rule> <!-- prenuclear accent in Wh-question -->
   <sentence type="interrogW"/>
      <prosodicPosition type="prenuclear"/>
            <attributes accent="tone"/>
                  <action accent="L*+H"/>
 </rule>

 <rule> <!-- nuclear accent in Wh-question, not at end of paragraph -->
  <sentence type="interrogW"/>
      <prosodicPosition type="nuclearNonParagraphFinal"/>
            <attributes accent="tone"/>
                  <action accent="L*+H"/>
 </rule>

 <rule> <!-- nuclear accent in Wh-question, if prosodicPosition type="null" -->
  <sentence type="interrogW"/>
       <prosodicPosition type="null"/>
             <attributes accent="tone"/>
             <action accent=""/>
 </rule>
```

Τέλος υπάρχουν κανόνες που καθορίζουν τα όρια της πρότασης. Αν πρέπει να εισαχθεί κάποια παύση ή κάποιο όριο τόνου στην αρχή, ενδιάμεσα ή στο τέλος της πρότασης.

Επόμενες μεταβλητές που δέχεται είναι η syllableaccents και η paragraphdeclination. Η syllableaccents συνδέει τον επίτονο (pitch accent) μιας λέξης με την τονισμένη συλλαβή αυτής της λέξης (οι επίτονοι γενικά ευθυγραμμίζονται/συνδέονται με τονισμένες συλλαβές, και όχι με ολόκληρες λέξεις) ενώ η syllableaccents υλοποιεί το φαινόμενο κατά το οποίο όσο μιλάμε η θεμελιώδης συχνότητα (F0) χαμηλώνει (η μελωδική καμπύλη ακολουθεί μια επικλινή πτωτική πορεία (declination)). Το δεύτερο αρχείο κειμένου, και τέταρτη μεταβλητή που δέχεται αυτή η συνάρτηση, δεν χρειάζεται να αναφερθεί αφού στην Ελληνική γλώσσα δεν έγινε χρήση του. Στα ελληνικά τυπικά όλες οι λέξεις παίρνουν επίτονο εκτός από τις κάποιες λειτουργικές (functional λέξεις). Αυτό επιτυγχανόταν ούτος ή άλλως και με το βασικό μοντέλο οπότε δεν χρειαζόταν για την Ελληνική γλώσσα να προστεθεί.

Η συνάρτηση Prosody καλεί την συνάρτηση ELProsodyGeneric, η οποία χρησιμοποιεί τους παραπάνω κανόνες. Μέσω της συνάρτησης ELProsodyGeneric γίνεται περιγραφή του επιτονισμού της πρότασης σύμφωνα με τους κανόνες που έχουμε ορίσει. Μέσω της δημιουργία της συνάρτησης ELProsodyGeneric και της σύνδεσης της με την συνάρτηση Prosody μας επιτρέπεται πρώτα να αναγνωρίζουμε το είδος των προτάσεων, αν δηλαδή μία πρόταση χαρακτηρίζεται ως δηλωτική, ερωτηματική ολικής άγνοιας, ερωτηματική μερικής άγνοιας, επιφωνηματική ή αρνητική, και δεύτερον να προσθέτουμε και να επεξεργαζόμαστε τον επίτονο κάθε λέξης και τον τόνο ορίου κάθε πρότασης χωριστά.

5.3.1 Αναγνώριση του είδους της πρότασης

Η συνάρτηση ELProsodyGeneric δημιουργήθηκε γιατί έχει ως σκοπό να μπορεί να επιτευχθεί αρχικά η αναγνώριση του είδους της πρότασης και στην συνέχεια να μπορούν να δοθούν τα σωστά επιτονικά γεγονότα στην πρόταση αλλά και στην λέξη που δέχεται και πρέπει να επιτονιστεί. Δεν δέχονται επίτονο όλες οι λέξεις μέσα σε μία πρόταση. Παρακάτω θα αναλυθούν τα βήματα που γίνονται για τον διαχωρισμό των προτάσεων αλλά και τον τρόπο με τον οποίο δίνονται από το σύστημα τα επιτονικά γεγονότα στις προτάσεις. Τέλος, θα αναλύσουμε τον τρόπο διόρθωσης του επιτονισμού στις ερωτηματικές προτάσεις, ολικής και μερικής άγνοιας, και των αρνητικών δηλωτικών προτάσεων.

Η συνάρτηση ELProsodyGeneric καλεί την βασική υπορουτίνα getSentenceType. Μέσω της getSentenceType γίνεται έλεγχος στο είδος της πρότασης. Η συγκεκριμένη υπορουτίνα κάνει έλεγχο στο σημείο στίξης (punctuation mark) και κατατάσσει την πρόταση αρχικά σε δηλωτική, ερωτηματική ή επιφωνηματική. Στην συνέχεια μέσω συγκεκριμένων κανόνων κατατάσσει την πρόταση στο τελικό της είδος.

Αναλυτικά η υπορουτίνα λειτουργεί ως εξής: εφόσον έχει γίνει έλεγχος στο σημείο στίξης και έχει δοθεί το αρχικό είδος της πρότασης, γίνεται έλεγχος στην πρώτη λέξη της πρότασης. Αν η πρώτη λέξη της πρότασης καλύπτει τους κανόνες που έχουμε δημιουργήσει και εισάγει τότε δίνεται στην πρόταση το νέο είδος της.

Αν η πρώτη λέξη είναι WP ή NRP ή πληροί τους κανόνες των ερωτήσεων ολικής άγνοιας τότε κατευθείαν δίνεται το νέο είδος της πρότασης. Αν η πρώτη λέξη δεν καλύψει κάποιον κανόνα τότε γίνεται έλεγχος στο μέσο της πρότασης. Αυτός ο έλεγχος γίνεται για να ελεγχθούν οι επιπλέον κανόνες που έχουμε εισάγει. Μια πρόταση μπορεί επιπλέον να χαρακτηριστεί ως ερώτηση μερικής άγνοιας και όταν η ξεκινάει με πρόθεση και στην συνέχεια υπάρχει Wh–word.

Αντίστοιχα σε μία αρνητική πρόταση, μπορεί το αρνητικό μόριο να βρίσκεται στο μέσο της πρότασης και όχι στην πρώτη λέξη της πρότασης. Μέσω της ολοκλήρωσης των κανόνων δίνεται το τελικό είδος της πρότασης.

Επειδή οι ερωτηματικές προτάσεις είναι πιο περίπλοκες, γίνεται διαδοχικός έλεγχος για την απόδοση του είδους της ερωτηματικής πρότασης. Αρχικά γίνεται έλεγχος για τις ερωτήσεις ολικής άγνοιας, στην συνέχεια ακολουθούν οι κανόνες για τις αρνητικές ερωτηματικές προτάσεις και στο τέλος οι κανόνες για τις ερωτήσεις μερικής άγνοιας.

Παράδειγμα ερωτήσεων ολικής άγνοιας:

«Θέλεις να έρθεις στο μάθημα μαζί μας;»

Η πρόταση αυτή ξεκινάει με ρήμα οπότε κατατάσσεται στις ερωτήσεις ολικής άγνοιας.

Γραμματική ανάλυση:

[Θέλεις, να, έρθεις, στο, μάθημα, μαζί, μας, ;]

[VB, CC, NN, PDT, NN, RB, PRP, $PUNCT]

Παράδειγμα ερωτήσεων μερικής άγνοιας:

«Ποιος μαθητής θέλει να έρθει στο μάθημα;»

Η πρόταση αυτή ξεκινάει με Wh word, οπότε κατατάσσεται στις ερωτήσεις μερικής άγνοιας.

Γραμματική ανάλυση:

[Ποιος, μαθητής, θέλει, να, έρθει, στο, μάθημα, ;] [WP, NN, VB, CC, VB, PDT, NN, $PUNCT]

«Για ποιο θέμα μιλάτε;»

Η πρόταση αυτή ξεκινάει με πρόθεση και μετά υπάρχει Wh word, γι αυτό τον λόγο κατατάσσεται σε ερώτηση μερικής άγνοιας.

Γραμματική ανάλυση:

[Για, ποιό, θέμα, μιλάτε, ;]

[IN, WP, NN, VB, $PUNCT]

Παράδειγμα αποφατικής ερώτησης ολικής άγνοιας:

«Δεν θέλεις να έρθεις στο μάθημα μαζί μας;»

Γραμματική ανάλυση:

[Δεν, θέλεις, να, έρθεις, στο, μάθημα, μαζί, μας, ;]

[NRP, VB, CC, NN, PDT, NN, RB, PRP, $PUNCT]

Παράδειγμα αποφατικής ερώτησης μερικής άγνοιας:

«Γιατί δεν θέλεις να έρθεις στο μάθημα μαζί μας;»

Γραμματική ανάλυση:

[Γιατί, δεν, θέλεις, να, έρθεις, στο, μάθημα, μαζί, μας, ;]

[WP, NRP, VB, CC, NN, PDT, NN, RB, PRP, $PUNCT]

Η πρώτη πρόταση ξεκινάει με το αρνητικό μόριο οπότε κατατάσσεται στις αρνητικές προτάσεις και η δεύτερη πρόταση έχει το αρνητικό μόριο στο μέσο της, οπότε ξανά κατατάσσεται στις αρνητικές προτάσεις.

5.3.2 Ορισμός του επιτονισμού της πρότασης και των λέξεων της πρότασης

Έχοντας ολοκληρώσει την διαδικασία ελέγχου του είδους της πρότασης προχωράμε στην απόδοση επιτονισμού στην πρόταση. Ξανά αναφέρουμε ότι σε μία πρόταση δεν δέχονται επίτονο όλα τα μέρη του λόγου. Τα μέρη του λόγου που δέχονται επίτονο ορίζονται μέσα στο αρχείο tobipredparams. Η πλατφόρμα μετατροπής κειμένου σε ομιλία OpenMary έχει υλοποιημένους αλγορίθμους για την απόδοση επιτονισμού στα μέρη του λόγου μια πρότασης. Όμως οι κανόνες που υπήρχαν, δεν ήταν σε θέση να καλύψουν τον επιτονισμό των ελληνικών αρνητικών καταφατικών και ερωτηματικών προτάσεων. Οι επιπλέον κανόνες που δημιουργήθηκαν μέσα στο αρχείο tobipredparams αφορούν τις ερωτηματικές προτάσεις ολικής και μερικής άγνοιας και τις αρνητικές δηλωτικές προτάσεις (negative declarative).

- **Κανόνες επιτονισμού των ερωτήσεων ολικής άγνοιας**

Οι προτάσεις που χαρακτηρίζονται ως ερωτήσεις ολικής άγνοιας δέχονται τον πυρηνικό επίτονο να πέφτει πάνω στο ρήμα της λέξης/φράσης. Οι λέξεις/φράσεις που είναι πριν το ρήμα δέχονται ως επίτονο τον προπυρηνικό επίτονο και όλες οι λέξεις που ακολουθούν το ρήμα δεν δέχονται καθόλου επίτονο. Για να μπορέσει να επιτευχθεί η θεωρία επιτονισμού των ελληνικών ερωτήσεων ολικής άγνοιας έπρεπε εκτός από την δημιουργία του προπυρηνικού και πυρηνικού επίτονου να δημιουργηθεί και ο κανόνας απόδοσης αυτών των τιμών. Ο προπυρηνικός και πυρηνικός επίτονος δημιουργήθηκε μέσα στο αρχείο tobipredparams και ενσωματώθηκε μέσω της συνάρτησης Prosody στην συνάρτηση ELProsodyGeneric. Εκεί στην υπορουτίνα getAccentShape δημιουργήθηκαν οι κανόνες για την απόδοση του αναφερθέντα επιτονισμού.

Παράδειγμα κανόνων επιτονισμού:

«Θέλεις να έρθεις στο μάθημα μαζί μας;»

Απόδοση επιτονισμού:

[Θέλεις, να, έρθεις, στο, μάθημα, μαζί, μας, ;]

[VB, CC, VB, PDT, NN, RB, PRP, $PUNCT]

Θέλεις ->prenuclear ->L*+H

έρθεις ->nuclearNonParagraphFinal ->L*

μάθημα ->null

Τόνος ορίου -> H-L%

- **Κανόνες επιτονισμού των ερωτήσεων μερικής άγνοιας**

Για τις ερωτηματικές προτάσεις μερικής άγνοιας ισχύει ότι ο πυρηνικός επίτονος του επιτονισμού πέφτει πάνω στην ερωτηματική λέξη/φράση και όλες οι λέξεις που ακολουθούν την ερωτηματική λέξη/φράση δεν παίρνουν καθόλου επίτονο. Ως ερωτηματική λέξη/ φράση μπορεί να υπάρχει σκέτο ένα Wh-word μέσα στην πρόταση ή αλλιώς αν το Wh-word ακολουθείται από ουσιαστικό, τότε ο πυρηνικός επίτονος πέφτει στο ουσιαστικό και το Wh-word δέχεται τον προπυρηνικό επίτονο. Για να μπορέσει να επιτευχθεί η θεωρία επιτονισμού των ελληνικών ερωτήσεων μερικής άγνοιας έπρεπε εκτός από την δημιουργία του προπυρηνικού και πυρηνικού επίτονου να δημιουργηθεί και ο κανόνας απόδοσης αυτών των τιμών. Ο προπυρηνικός και πυρηνικός επίτονος δημιουργήθηκε μέσα στο αρχείο tobipredparams και ενσωματώθηκε μέσω της

συνάρτησης Prosody στην συνάρτηση ELProsodyGeneric. Εκεί στην υπορουτίνα getAccentShape δημιουργήθηκαν οι κανόνες για την απόδοση της προαναφερθείσης επιτονικής καμπύλης.

Παράδειγμα κανόνων επιτονισμού:

«Ποιος μαθητής θέλει να έρθει στο μάθημα;»

Απόδοση επιτονισμού:

[Ποιός, μαθητής, θέλει, να, έρθει, στο, μάθημα, ;]

[WP, NN, VB, CC, VB, PDT, NN, $PUNCT]

Ποιός -> prenuclear -> L*+H

μαθητής -> nuclearNonParagraphFinal -> L*+H

θέλει -> null

έρθει -> null

μάθημα -> null

Τόνος ορίου -> H-L%

- **Κανόνες επιτονισμού των αρνητικών δηλωτικών προτάσεων**

Οι προτάσεις που χαρακτηρίζονται ως αρνητικές δηλωτικές (negative declarative) δέχονται τον πυρηνικό επίτονο να πέφτει πάνω στην αρνητική λέξη/φράση. Οι λέξεις/φράσεις που είναι πριν την αρνητική λέξη δέχονται ως επίτονο τον προπυρηνικό επίτονο και όλες οι λέξεις που ακολουθούν την αρνητική λέξη/φράση δεν δέχονται καθόλου επίτονο. Για να μπορέσει να επιτευχθεί η θεωρία επιτονισμού των ελληνικών αρνητικών καταφατικών προτάσεων έπρεπε εκτός από την δημιουργία του προπυρηνικού και πυρηνικού επίτονου να δημιουργηθεί και ο κανόνας απόδοσης αυτών των τιμών. Ο προπυρηνικός και πυρηνικός επίτονος δημιουργήθηκε μέσα στο αρχείο tobipredparams και ενσωματώθηκε μέσω της συνάρτησης Prosody στην συνάρτηση ELProsodyGeneric. Εκεί στην υπορουτίνα getAccentShape δημιουργήθηκαν οι κανόνες για την απόδοση του αναφερθέντα επιτονισμού.

Παράδειγμα κανόνων επιτονισμού:

«Ο Μανώλης δεν ήρθε στο μάθημα.»

Απόδοση επιτονισμού:

[Ο, Μανώλης, δεν, ήρθε, στο, μάθημα, .]

[PDT, NN, NRP, VB, PDT, NN, $PUNCT]

Μανώλης -> prenuclear -> L*+H

δεν -> nuclearNonParagraphFinal -> L*+H

ήρθε -> null

μάθημα -> null

Τόνος ορίου -> H-L%

5.4 Δημιουργία της Ελληνικής φωνής

Η δημιουργία της Ελληνικής γλώσσας έγινε με τα αυτόματα εργαλεία της πλατφόρμας μετατροπής κειμένου σε ομιλία OpenMary. Με την διαδικασία δηλαδή του Voice Import Tools και με την χρήση των εξ' ορισμού επιλογών για την ανδρική φωνή. Για την εκπαίδευση της φωνής έγινε χρήση του corpus του Ρήτορα του Πανεπιστημίου Αθηνών [25]. Η σημαντική διαφορά με την φωνή της προηγούμενης έκδοσης είναι ότι γίνανε κάποιες επιπλέον ενέργειες που είχαν ως στόχο την βελτιστοποίηση της φωνής. Οι ενέργειες αυτές γίνανε προκειμένου να διορθωθεί η επισημείωση των επιτονικών γεγονότων του ToBI στα αρχεία με το intonation/prosody specification. Τα αρχεία αυτά δημιουργούνται αυτόματα βάσει του prosody specification της Ελληνικής φωνής. Η διόρθωση έγινε ώστε πλέον να συμφωνούν με τη χειροκίνητη επισημείωση στο corpus του Ρήτορα.

6. ΑΞΙΟΛΟΓΗΣΗ ΠΡΟΣΩΔΙΑΚΟΥ ΜΟΝΤΕΛΟΥ

6.1 Εισαγωγή

Υπάρχουν διάφοροι μέθοδοι αξιολόγησης ενός συνθέτη ομιλίας [26]:

Η αξιολόγηση μπορεί να είναι διαγνωστική (diagnostic) ή συγκριτική (comparative), υποκειμενική (subjective) ή αντικειμενική (objective), αρθρωτή (modular) ή συνολική (global), Επίσης υπάρχουν και διάφοροι τρόποι αξιολόγησης, όπως μέσω ιστοσελίδας (web-based) ή ζωντανή (live) αξιολόγηση, με ακουστικά (headphones) ή με ηχεία (loudspeakers), με ειδικό ακροατήριο (specialist) ή με απλό ακροατήριο (naive). Τέλος η αξιολόγηση διαφοροποιείται ανάλογα με το αν τα δείγματα ανθρώπινης ομιλίας περιλαμβάνονται στην ίδια αξιολόγηση με τα δείγματα της συνθετικής ομιλίας.

Σκοπός της παρούσας πειραματικής διαδικασία είναι να αξιολογήσουμε την προτίμηση των ακροατών για το συγκεκριμένο προσωδιακό μοντέλο. Οι χρήστες ακούγοντας συγκεκριμένες προτάσεις ζητούνται να αξιολογήσουν την προτίμηση τους σε αυτές. Η αξιολόγηση έγινε με την χρήση του Mean Opinion Score (MOS).

6.1.1 Συμμετέχοντες

Στην διαδικασία αξιολόγησης συμμετείχαν 21 άνδρες και 16 γυναίκες (συνολικά 27 συμμετέχοντες) – ηλικίας από 20 μέχρι 35 ετών (μέση ηλικία = 27,11 έτη, τυπική απόκλιση = 3,22). Ήταν προπτυχιακοί και μεταπτυχιακοί φοιτητές από διάφορα εκπαιδευτικά ιδρύματα της Ελλάδας. Η μητρική τους γλώσσα είναι η Ελληνική και δεν υπήρχε καμία εμπλοκή σε προηγούμενο πείραμα που να σχετίζεται με τη μελέτη μας. Δεν υπήρχε κάποιο πρόβλημα ακοής από τους συμμετέχοντας στην διαδικασία αξιολόγησης. Τέλος από το μεγαλύτερο μέρος των συμμετεχόντων δεν υπήρχε καμία εξοικείωση με την συνθετική ομιλία.

6.1.2 Ερεθίσματα

Η επιλογή των ερωτηματικών προτάσεων για την αξιολόγηση του προσωδιακού μοντέλου έγινε με πολύ προσοχή, προσπαθώντας να γίνει χρήση προτάσεων που μπορούν να βρεθούν σε πραγματικά συστήματα συνθετικής ομιλίας εξυπηρέτησης πελατών, με προτάσεις διαφορετικού μεγέθους και προτάσεις διαφορετικού τονισμού της ερωτηματικής λέξης.

Τα συνολικά ερωτηματικά ερεθίσματα ήταν 12. Παρακάτω στον πίνακα 2 φαίνονται τα 12 ερωτηματικά ερεθίσματα. Το κάθε ερέθισμα περιείχε δύο διαφορετικές εκδοχές της ίδιας ερώτησης. Στην μία περίπτωση ήταν η ακουστικοποίηση της ερώτησης με το ελάχιστο προσωδιακό μοντέλο ομιλίας και στην δεύτερη περίπτωση με την παρούσα υλοποίηση του προσωδιακού μοντέλου. Οι ερωτήσεις που επιλέχθηκαν στο σύνολο τους είναι 12, από αυτές οι 6 χαρακτηρίζονται ως ερωτήσεις μερικής άγνοιας και οι άλλες 6 ως ερωτήσεις ολικής άγνοιας. Σε κάθε ομάδα από τις 6 ερωτηματικές προτάσεις είχαμε δύο προτάσεις 3 λέξεων(short), 5 λέξεων (medium) και 7 λέξεων (long). Οι προτάσεις των διαφορετικών μοντέλων έχουν ενωθεί μεταξύ τους ανά δύο, βάση του μεγέθους τους, και έχει προστεθεί ανάμεσα τους μία παύση 2 δευτερολέπτων.

Αυτές οι δώδεκα ερωτήσεις ακουστικοποιήθηκαν με την χρήση της πλατφόρμας OpenMary. Μια με την χρήση του ελάχιστου υποσυστήματος επισημείωσης Μερών-του-Λόγου και μία με την χρήση του προτεινόμενου προσωδιακού μοντέλου που αναλύθηκε και αναπτύχθηκε προηγούμενος. Οι δύο διαφορετικές εκδόσεις των ερωτήσεων ενώθηκαν μεταξύ τους με μια παύση των 2 δευτερολέπτων. Χωρίς το προσωδιακό

μοντέλο (Α), με το προσωδιακό μοντέλο (Β). Η σειρά που ακούστηκαν οι ερωτήσεις ήταν τυχαία και μπορούσε να είναι Α-Β ή Β-Α με μία ενδιάμεση παύση 2 δευτερολέπτων. Ήταν τυχαία και η σειρά των ερωτήσεων αλλά και η σειρά του προσωδιακού μοντέλου Α-Β και Β-Α.

Πίνακας 2: Συνολικά ερωτηματικά ερεθίσματα

Αριθμός ερεθίσματος	Είδος ερεθίσματος	Ερέθισμα
	Ερώτηση μερικής άγνοιας	
1		Ποιο δρομολόγιο θέλετε;
2		Ποια ημερομηνία θέλετε;
3		Πως μπορώ να σας εξυπηρετήσω;
4		Ποιο ακριβώς είναι το πρόβλημα;
5		Πείτε μου παρακαλώ, τι πρόβλημα έχετε;
6		Από πού θέλετε να αναχωρήσετε και πότε;
	Ερώτηση ολικής άγνοιας	
7		Συμφωνείται με αυτό;
8		Ρωτήσατε για φραγή;
9		Ενδιαφέρεστε για θέματα σταθερής τηλεφωνίας;
10		Υπάρχει πρόβλημα με τα μηνύματα;
11		Έχετε τεχνικό πρόβλημα με το καινούργιο περιβάλλον;
12		Θέλετε να εντοπίσετε κλήσεις από απόρρητο αριθμό;

6.1.3 Η πειραματική διαδικασία

Οι συμμετέχοντες για την εξοικείωσή τους με την πειραματική διαδικασία, στην αρχή άκουγαν 4 δοκιμαστικά παραδείγματα εκτέλεσης του πειράματος. Στην συνέχεια, άκουγαν με τυχαία σειρά διαφορετικές ερωτήσεις τις οποίες και αξιολόγησαν βάσει της προτίμησής τους. Οι συμμετέχοντες δεν άκουγαν στην ίδια σειρά τις δύο διαφορετικές εκδοχές των προτάσεων. Μπορούσε το πρώτο ερέθισμα να είναι σε σειρά Α-Β ενώ το δεύτερο ερέθισμα σε σειρά Β-Α. Ο κάθε χρήστης μπορούσε να ακούσει το ερέθισμα όσες φορές θεωρούσε ότι απαιτούνταν.

Οι ερωτήσεις που έπρεπε να απαντήσει ο χρήστης ήταν:

α) «*Πιστεύετε ότι ο δεύτερος τρόπος απόδοσης της μελωδίας της ερώτησης είναι καλύτερος από τον πρώτο;*»

β) «*Πιστεύετε ότι η συνολική ποιότητα του δεύτερου τρόπου απόδοσης είναι καλύτερη από του πρώτου;*»

Η διαδικασία, εν συντομία, ήταν η ακόλουθη: Οι συμμετέχοντες άκουγαν τους δύο διαφορετικούς τρόπους εκφώνησης της πρότασης με μία μικρή παύση μεταξύ τους. Η κάθε ερώτηση είχε 5 επιλογές, από τις οποίες μπορούσε να επιλέξει μόνο μία: «καλύτερος», «λίγο καλύτερος», «ίδιος», «λίγο χειρότερος», «χειρότερος». Οι επιλογές για το πολύ καλύτερος ως πολύ χειρότερος αντιστοιχεί από το 2 έως το -2 και η επιλογή για το ίδιος είναι στο 0. Οι ενδιάμεσες απαντήσεις αντιστοιχούν στο 1 και -1 αντίστοιχα. Εφόσον δίνονταν οι απαντήσεις και στις δύο ερωτήσεις οι συμμετέχοντες άκουγαν το επόμενο στην σειρά ερέθισμα.

6.2 Ανάλυση των αποτελεσμάτων

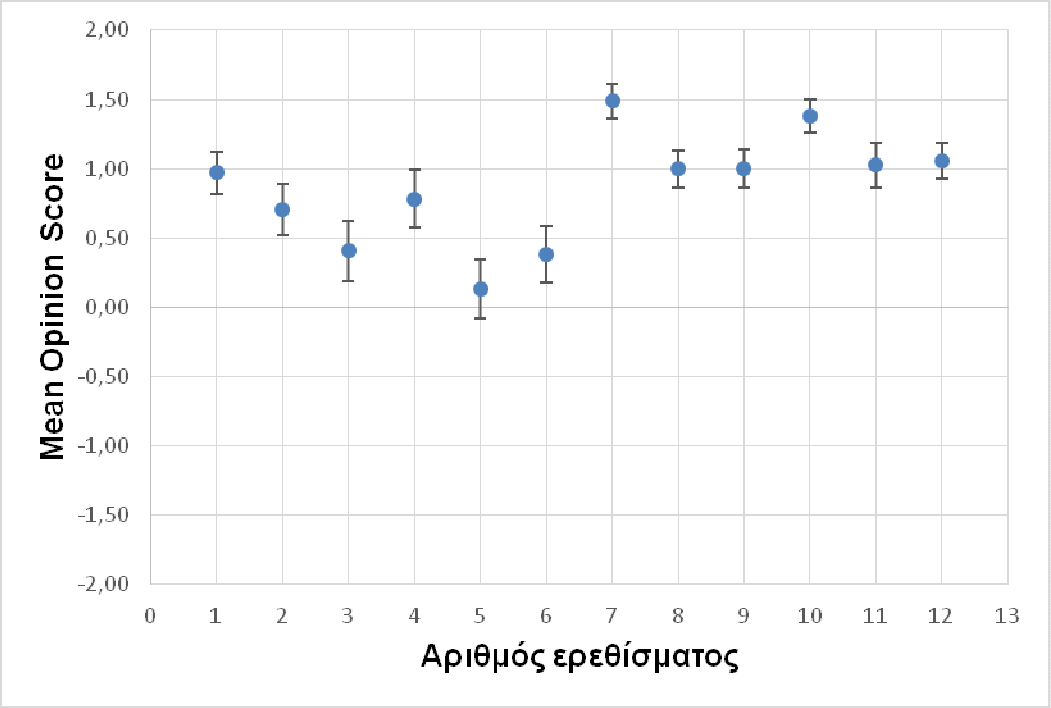
Από τα αποτελέσματα παρατηρήσαμε ότι οι χρήστες ήταν σε θέση να κατανοήσουν και να αναγνωρίσουν τις ερωτήσεις ολικής και μερικής άγνοιας του προσωδιακού μοντέλου με επιτυχία. Προτιμούσαν τον τρόπο εκφώνησης των ερωτηματικών προτάσεων βάσει του νέου προσωδιακού μοντέλου. Βρίσκοντας το Mean Opinion Score και αναλύοντας τα αποτελέσματα βλέπουμε ότι έχουμε θετικό MOS σε κάθε ερέθισμα που αξιολογήθηκε. Το θετικό MOS δείχνει ότι η αξιολόγηση ήταν υπέρ του νέου προσωδιακού μοντέλου και κατά του βασικού προσωδιακού μοντέλου. Επιπλέον παρατηρούμε ότι το MOS για τις ερωτήσεις ολικής άγνοιας είναι υψηλότερο από το αντίστοιχο για τις μερικής άγνοιας. Οπότε μπορούμε να συμπεράνουμε ότι η βελτίωση ήταν μεγαλύτερη για τις ερωτήσεις ολικής άγνοιας και οι χρήστες μπορούσαν να τις κατανοήσουν καλύτερα σε σχέση με τις ερωτήσεις μερικής άγνοιας.

Τα αποτελέσματα παρουσιάζονται αναλυτικά στους παρακάτω πίνακες. Στον πίνακα 2 παρουσιάζονται τα αποτελέσματα από την πρώτη ερώτηση αξιολόγησης και στον πίνακα 3 τα αποτελέσματα από την δεύτερη ερώτηση αξιολόγησης.

Στους δύο αυτούς πίνακες υπάρχουν τα είδη των ερωτήσεων, το μέγεθος της ερώτησης, ένας αύξοντας αριθμός που αντιστοιχεί στην θέση του ερεθίσματος και το MOS μαζί με το αντίστοιχο τυπικό σφάλμα (Standard Error, SE) για κάθε πρόταση. Οι προτάσεις βάση του είδους και του μεγέθους της ερώτησης κατηγοριοποιούνται σε sentence 1 (S1) και sentence 2 (S2). Επιπλέον παρουσιάζονται τα δύο διαγράμματα των αποτελεσμάτων. Στο σχήμα 1 είναι τα αποτελέσματα της πρώτης ερώτησης αξιολόγησης και στον σχήμα δύο τα αποτελέσματα από την δεύτερη ερώτηση αξιολόγησης. Στον κάθετο των δύο σχημάτων έχουμε το MOS και στον οριζόντιο άξονα έχουμε τον αριθμό του ερεθίσματος.

Πίνακας 3: Αποτελέσματα της πρώτης ερώτησης αξιολόγησης

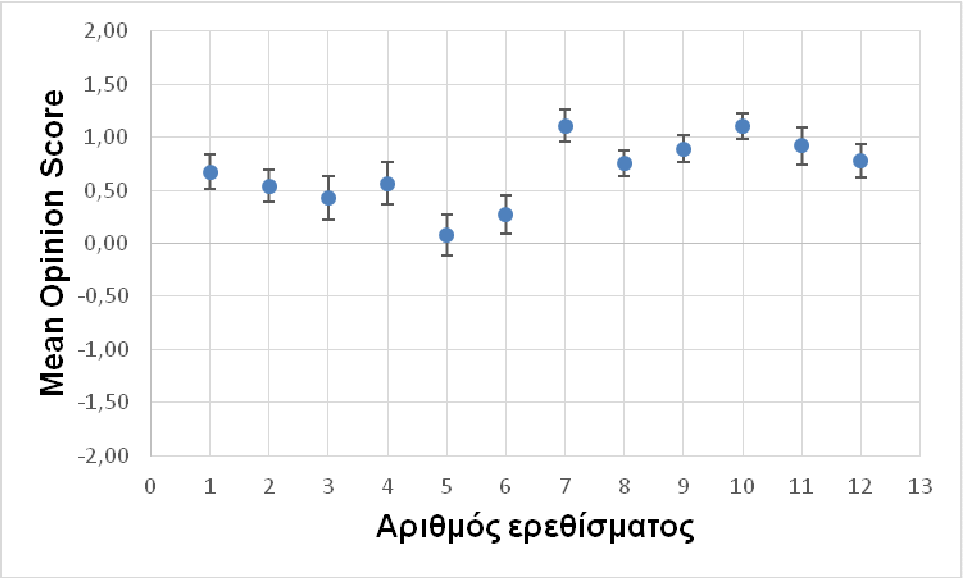
Είδος ερώτησης	Μέγεθος ερώτησης	α/α	Ερωτηματική πρόταση	MOS (SE)
Wh	3 words	1	S1	0,97(0,15)
		2	S2	0,70(0,18)
	5 words	3	S1	0,41(0,21)
		4	S2	0,78(0,20)
	7 words	5	S1	0,14(0,21)
		6	S2	0,38(0,20)
Yes/No	3 words	7	S1	1,49(0,12)
		8	S2	1,00(0,13)
	5 words	9	S1	1,00(0,13)
		10	S2	1,38(0,11)
	7 words	11	S1	1,03(0,16)
		12	S2	1,05(0,12)



Σχήμα 1: Αποτελέσματα της πρώτης ερώτησης αξιολόγησης. Αποτελέσματα MOS με το αντίστοιχο τυπικό σφάλμα.

Πίνακας 4: Αποτελέσματα της δεύτερης ερώτησης αξιολόγησης

Είδος ερώτησης	Μέγεθος ερώτησης	α/α	Ερωτηματική πρόταση	MOS(SE)
Wh	3 words	1	S1	0,68(0,16)
		2	S2	0,54(0,15)
	5 words	3	S1	0,43(0,20)
		4	S2	0,57(0,19)
	7 words	5	S1	0,08(0,19)
		6	S2	0,27(0,18)
Yes/No	3 words	7	S1	1,11(0,14)
		8	S2	0,76(0,11)
	5 words	9	S1	0,89(0,12)
		10	S2	1,11(0,12)
	7 words	11	S1	0,92(0,17)
		12	S2	0,78(0,15)



Σχήμα 2: Αποτελέσματα της δεύτερης ερώτησης αξιολόγησης. Αποτελέσματα MOS με το αντίστοιχο τυπικό σφάλμα.

7. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

7.1 Συμπεράσματα

Σκοπός αυτής της διπλωματικής διατριβής ήταν η υποστήριξη ενός υποσυστήματος συνθέτη ομιλίας που είχε ως στόχο να αναγνωρίζει είδη προτάσεων όπως δηλωτικές, ερωτηματικές, επιφωνηματικές και αρνητικές προτάσεις και να τους αποδίδει κατάλληλη προσωδιακή περιγραφή. Η εργασία αυτή βελτίωσε την προηγούμενη Ελληνική έκδοση της πλατφόρμας μετατροπής κειμένου σε ομιλία OpenMary [2]. Η βελτίωση που είχαμε θέσει ως αρχικό στόχο ήταν η ανάπτυξη και η εισαγωγή ενός προσωδιακού μοντέλου ομιλίας για τις ερωτηματικές προτάσεις της Ελληνικής γλώσσας. Ο στόχος αυτός επιτεύχθηκε και το καινούργιο σύστημα είναι σε θέση μέσω της αναγνώρισης των προτάσεων να αποδώσει μία προσωδία στην ομιλία του αρκετά καλύτερη από του προηγούμενου συστήματος. Το νέο σύστημα χρησιμοποιεί ένα υποσύστημα επισημείωσης των Μερών-του-Λόγου . Χρησιμοποιεί συναρτήσεις για την διόρθωση της γραμματικής αναγνώρισης των λέξεων και αλγόριθμους για την αναγνώριση των προτάσεων. Επιπλέον δημιουργήθηκαν κανόνες για τον επιτονισμό των λέξεων αλλά και για τον αποδοτικότερο επιτονισμό των προτάσεων. Τέλος, εκτός από την προσωδία στις ερωτηματικές προτάσεις, διορθώθηκε η προσωδία στις αρνητικές δηλωτικές προτάσεις και έγινε εισαγωγή των αρνητικών προτάσεων στο σύστημα μας.

7.2 Μελλοντικές επεκτάσεις

θα μπορούσαν να προστεθούν επιπλέον κανόνες για την διόρθωση των λανθασμένων μερών του λόγου. Υπήρχαν προβλήματα με τα ρήματα του πρώτου και δεύτερου ενικού με αποτέλεσμα να χαλάει το προσωδιακό μοντέλο ομιλίας. Τα προβλήματα αυτά δημιουργηθήκαν λόγω της λανθασμένης πρόβλεψης του υποσυστήματος Μερών-του-Λόγου AUEB_POS_tagger. Μία μελλοντική ανάπτυξη θα ήταν η περαιτέρω εκπαίδευση του υποσυστήματος Μερών-του-Λόγου με μεγαλύτερο corpus και επισημειωμένο με προτάσεις που περιέχουν ρήματα στο πρώτο και δεύτερο ενικό. Επιπλέον θα ήταν απαραίτητη η εισαγωγή μεγαλύτερων παύσεων κατά την διάρκεια της ομιλίας, ειδικά στο τέλος μια ερωτηματικής πρότασης. Ακόμα κάτι πολύ σημαντικό είναι η αύξηση της θεμελιώδους συχνότητας (F0). Η αύξηση της θεμελιώδους συχνότητας είναι ένα χαρακτηριστικό της σύνθετης ομιλίας το οποίο θα μπορούσε να αποδώσει με την αύξηση του την καλύτερη απόδοση των ερωτηματικών προτάσεων. Τέλος, για την βελτίωση των ερωτηματικών προτάσεων θα πρέπει να δημιουργηθεί ο κανόνας επιτονισμού L-!H%. Ο συγκεκριμένος κανόνας δεν υπάρχει μέσα στην πλατφόρμα μετατροπής κειμένου σε ομιλία και οι ερωτήσεις μερικής άγνοιας στο 70% των περιπτώσεων πραγματώνονται με τον συγκεκριμένο επίτονο. Τέλος θα ήταν χρήσιμη η περαιτέρω εκπαίδευση της φωνής με μεγαλύτερο corpus ερωτήσεων και καταφάσεων ώστε να γίνει πιο φυσική η ομιλία.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	**Ελληνικός Όρος**
Accent	Τόνος
Article (PDT)	Άρθρο
Verb (VB)	Ρήμα
Punctuation (PUNCT)	Σημείο στίξης
Adjective (JJ)	Επίθετο
Adverb (RB)	Επίρρημα
Conjunction(CC)	Σύνδεσμος
Noun (NN)	Ουσιαστικό
Numeral (CD)	Αριθμητικό
Particle (RP)	Μόριο
Preposition (IN)	Πρόθεση
Pronoun (PRP)	Αντωνυμία

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

ΕΚΠΑ	Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
POS	(Part Of Speech): Μέρος του λόγου
POS Tagger	(Part Of Speech Tagger): Σύστημα επισημείωσης Μερών-του-Λόγου
TtS	(Text to Speech): Κείμενο σε Ομιλία (ΚσΟ)
MARY TtS	(Modular Architecture for research on speech Synthesis Text to Speech): Σπονδυλωτή Αρχιτεκτονική για έρευνα στην σύνθεση ομιλίας μετατροπής κειμένου σε ομιλία
Decl	(declarative): Δηλωτικές
Excl	(exclamatory): Επιφωνηματικές
Interrog	(interrogative): Ερωτηματικές
Neg	(Negative): Αρνητικές
Neg decl	(Negative declarative): Αρνητικές δηλωτικές
Neg ecxl	(Negative exclamatory): Αρνητικές επιφωνηματικές
Neg interrog	(Negative interrogative): Αρνητικές ερωτήσεις
Wh-Questions	(Wh questions): Ερωτήσεις Μερικής Άγνοιας
YN-Questions	(Yes No questions): Ερωτήσεις Ολικής Άγνοιας
Neg Questions	(Negative questions): Αρνητικές ερωτήσεις
UOA	(University of Athens): Πανεπιστήμιο Αθηνών
NLP	(Natural language processing): Επεξεργασία φυσικής γλώσσας
ΕΦΓ	Επεξεργασία φυσικής γλώσσας
HMM	(Hidden Markov models): Κρυφά Μαρκοβιανά Μοντέλα
FFT	(Fast Fourier transform): Γρήγορη μετατροπή Φουριέ

ΑΝΑΦΟΡΕΣ

[1] The MARY Text-to-Speech System (MARYTTS) http://mary.dfki.de/
[2] P. Stavropoulou, D. Tsonos, G. Kouroupetroglou, Language Resources and Evaluation for the Support of the Greek Language in the MARY TtS 2014
[3] Ε. Κολέλη «Ένας νέος ελληνικός επισημειωτής μερών του λόγου, βασισμένος σε ταξινομητή μέγιστης εντροπίας» Πτυχιακή εργασία Τμήμα Πληροφορικής Οικονομικό Πανεπιστήμιο Αθηνών 2011
[4] Γ. Θ. Κουρουπέτρογλου, Μαθήματα Επεξεργασίας Ομιλίας. Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Αθήνα, 1998.
[5] A. Jonathan, M. Sharon, K. Dennis, «*From Text to Speech: The MITalk system. Cambridge University Press»* (1987).
[6] T. Dutoit «A short introduction to text-to-speech synthesis» 1996
[7] A.W Black,., H.Zen, K.Tokuda, «Statistical parametric speech synthesis,» Proc. of IEEE ICASSP 2007, April,2007
[8] W. Alan Black, Perfect synthesis for all of the people all of the time. IEEE TTS Workshop 2002
[9] S.Lemmetty, «Review of Speech Synthesis Technology» Department of Electrical and Communications Engineering Helsinki University of Technolog
[10] J.Rubin, «Study of Cognitive Processes in Second Language Learning» 1981
[11] Browman and Goldstein «Articulatory Phonology: An Overview» 1992
[12] K.Stevens, «Toward formant synthesis with articulatory controls» , Proc. IEEE Workshop on Speech Synthesis, Santa Monica, September 2002, pp. 67-72
[13] Π. Ζέρβας , «Μοντελοποίηση και ψηφιακή επεξεργασία προσωδιακών φαινομένων της ελληνικής γλώσσας με εφαρμογή στην σύνθεση ομιλίας» Διπλωματική διατριβή Ηλεκτρολόγων μηχανικών Μεταπτυχιακό Συστήματα Επεξεργασίας Σημάτων και Εικόνων 2007
[14] P. Blunsom, Hidden Markov Models August 19, 2004.
[15] H Zen, K Tokuda.and A.W. Black, «Statistical parametric speech synthesis», Speech Communication, 2009
[16] A. Acero, «Formant analysis and synthesis using hidden markov models», In Proc. of Eurospeech 1999
[17] S.,R Hertz. «Integration of Rule-based Formant Synthesis and Waveform Concatenation: A Hybrid Approach To Text-to-Speech Synthesis» IEEE 2002 Workshop On Speech Synthesis, 2002
[18] R. Carlson, B. Granström, «Data-driven multimodal synthesis" Speech Communication», Volume 47, Issues 1-2, 2005
[19] Y. Nakamura, T. Toda, Y. Nankaku, K. Tokuda, «On the Use of Phonetic Information for Mapping from Articulatory Movements to Vocal Tract Spectrum» 2006
[20] P. Taylor, «Unifying unit selection and hidden Markov model speech synthesis», in Proc. of Interspeech, Pittsburgh, USA, Sept.2006
[21] T. Black, Caley «The Architecture of the Festival Speech Synthesis System» 1998
[22] S. Pammi, M. Charfuelan, M. Schr¨oder «Multilingual Voice Creation Toolkit for the MARY TTS Platform» 2010
[23] Architecture Walkthrough - MARY Text-to-Speech. Retrieved from http://mary.dfki.de/documentation/module-architecture
[24] A. Trilla, «Natural Language Processing techniques in Text-To-Speech» 2009
[25] D. Spiliotopoulos, G. Petasis, and G. Kouroupetroglou: "A Framework for Language-independent Analysis and Prosodic Feature Annotation of Text Corpora", Lecture Notes in Artificial Intelligence (LNAI), Vol. 5246, 2008, pp 517-524.
[26] L. Dybkjaer, H. Hemsen. W. Minker «Evaluation of Text and Speech Synthesis» Springer – 2007