ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ

Διπλωματική Εργασία

Ανάπτυξη Συστήματος Αμφίδρομης Μετατροπής

Ήχου - Κειμένου για Ρομποτικές Εφαρμογές.

Λεύκελης Βασίλειος - Στυλιανός

Αρ. Μητρώου: 222017022

Επιβλέπων Καθηγητής:

Παπακίτσος Ευάγγελος

ΑΙΓΑΛΕΩ 2024

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

UNIVERSITY OF WEST ATTICA

SCHOOL OF ENGINEERING DEPARTMENT OF INDUSTRIAL DESIGN
AND PRODUCTION ENGINEERING

Diploma Thesis

Development of a speech-to-text

and text-to-speech system for robotic applications.

Lefkelis Vasileios - Stylianos

Registration Number: 222017022

Supervisor name and surname:

Evangelos Papakitsos

ATHENS 2024

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι

Εξεταστική Επιτροπή:

Α/α	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Ε.Χ. ΠΑΠΑΚΙΤΣΟΣ	ΕΔΙΠ Α΄	Evangelos Papakitsos Digitally signed by Evangelos Papakitsos Date: 2024.08.04 09:25:06 +03'00'
2	Ν. ΛΑΣΚΑΡΗΣ	Επίκουρος Καθηγητής	Nikolaos Laskaris Digitally signed by Nikolaos Laskaris Date: 2024.08.05 18:52:39 +03'00'
3	Χ. ΔΡΟΣΟΣ	Επίκουρος Καθηγητής	Digitally signed by CHRISTOS DROSOS Date: 2024.08.06 15:45:18 +03'00'

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Λεύκελης Βασίλειος Στυλιανός. του ΔΗΜΗΤΡΙΟΥ, με αριθμό μητρώου 222017022 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ & ΠΑΡΑΓΩΓΗΣ δηλώνω υπεύθυνα ότι:

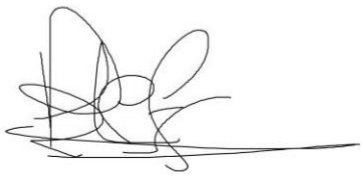
«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία.

Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο.

Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



Λεύκελης
Βασίλειος - Στυλιανός

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της διπλωματικής μου εργασίας κ. Ευάγγελο Παπακίτσο και τον κ Γιάχο Ιωάννη, για την ανάθεση του θέματος της διπλωματικής εργασίας, καθώς και για την καθοδήγηση και τις συμβουλές τους κατά τη συγγραφή της παρούσας εργασίας.

Επιπλέον ευχαριστώ τους κυρίους Ν. Λάσκαρη και Χ. Δρόσο που συμμετείχαν στην τριμελή εξεταστική επιτροπή της εργασίας.

Τέλος, ευχαριστώ την οικογένειά μου για την κατανόηση και στήριξή τους.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Περίληψη

Η παρούσα εργασία επικεντρώνεται στην ανάπτυξη ενός συστήματος που επιτρέπει τη μετατροπή του ήχου σε κείμενο και αντίστροφα, με στόχο την εφαρμογή του σε ρομποτικά συστήματα. Η αμφίδρομη μετατροπή είναι απαραίτητη για τη βελτίωση της επικοινωνίας μεταξύ ανθρώπων και ρομπότ, προσφέροντας φυσικό και αποτελεσματικό μέσο αλληλεπίδρασης. Ο κύριος σκοπός της εργασίας είναι η δημιουργία ενός ολοκληρωμένου συστήματος που θα μπορεί να μετατρέπει την ομιλία σε κείμενο (Speech-to-Text, STT), μετατρέπει το κείμενο σε φυσικό ήχο ομιλίας (Text-to-Speech, TTS). Για την υλοποίηση του συστήματος χρησιμοποιήθηκαν σύγχρονες τεχνολογίες και αλγόριθμοι επεξεργασίας φυσικής γλώσσας (NLP) και μηχανικής μάθησης (Machine Learning). Οι βασικές μέθοδοι περιλαμβάνουν Αναγνώριση Ομιλίας (Speech Recognition), χρήση νευρωνικών δικτύων και αλγορίθμων βαθιάς μάθησης για την ανάλυση και αναγνώριση της ομιλίας και ενσωμάτωση στην προϋπάρχουσα πλατφόρμα αναγνώρισης, όπως το Google Speech to Text API. Σύνθεση Ομιλίας (Speech Synthesis) με την χρήση τεχνικών TTS με αλγόριθμους μετατροπής κειμένου σε ήχο, όπως οι Tacotron και Wavenet. ενσωμάτωση εργαλείων όπως το Google Text to Speech API για την παραγωγή φυσικού ήχου.

Το σύστημα που αναπτύχθηκε δοκιμάστηκε σε διάφορα σενάρια χρήσης σε ρομποτικές εφαρμογές, όπως εντοπισμός και απάντηση σε φωνητικές εντολές από χρήστες, παροχή φωνητικών οδηγιών και πληροφοριών από το ρομπότ προς τους χρήστες. Τα αποτελέσματα έδειξαν υψηλή ακρίβεια στην αναγνώριση ομιλίας και ποιότητα στη σύνθεση φωνής, καθιστώντας το σύστημα χρήσιμο για ποικίλες ρομποτικές εφαρμογές. Η ανάπτυξη του συστήματος αμφίδρομης μετατροπής ήχου-κειμένου προσφέρει σημαντικά πλεονεκτήματα στην αλληλεπίδραση ανθρώπων και ρομπότ. Με τη συνεχή βελτίωση των αλγορίθμων και τη χρήση εξελιγμένων τεχνικών μηχανικής μάθησης, το σύστημα αυτό μπορεί να συμβάλλει σημαντικά στην εξέλιξη των ρομποτικών τεχνολογιών και στην ενίσχυση της επικοινωνίας σε ποικίλα πεδία εφαρμογών: Βελτίωση της αναγνώρισης ομιλίας σε θορυβώδη περιβάλλοντα, προσαρμογή του συστήματος για υποστήριξη σε πολλές γλώσσες και ενσωμάτωση της συναισθηματικής αναγνώρισης στην ανάλυση ομιλίας για πιο φυσική αλληλεπίδραση. Η παρούσα εργασία αποτελεί μια βάση για περαιτέρω έρευνα και ανάπτυξη στον τομέα των ρομποτικών συστημάτων και της επεξεργασίας φυσικής γλώσσας, προάγοντας τη συνεργασία ανθρώπων και μηχανών.

Λέξεις κλειδιά: Μετατροπή ήχου σε κείμενο, Μετατροπή κειμένου σε ήχο, Ρομποτικά συστήματα, Αναγνώριση Ομιλίας, Google Speech to Text API, Φωνητικές εντολές, Φωνητικές οδηγίες.

Abstract

This work focuses on the development of a system that allows the conversion of sound to text and vice versa, with the aim of applying it to robotic systems.

This two-way conversion is necessary to improve the communication between humans and robots, offering a natural and effective means of interaction.

The main purpose of the work is to create an integrated system that will be able to convert speech into text (Speech-to-Text, STT) and text to natural speech sound (Text-to-Speech, TTS). Modern technologies and algorithms of natural language processing (NLP) and machine learning (Machine Learning) were used to implement the system. Key methods include Speech Recognition, by using neural networks and deep learning algorithms for speech analysis and recognition. Integration with pre-existing recognition platform such as Google Speech to Text API, Speech Synthesis Using TTS techniques with text-to-sound algorithms, such as Tacotron and WaveNet, integrating tools like the Google Text to Speech API, to produce natural sound. The developed system was tested in various usage scenarios in robotic applications, such as: Detecting and responding to voice commands from users, provide voice instructions and information from the robot to users. The results showed high accuracy in speech recognition and quality in voice synthesis, making the system useful for a variety of robotic applications. The development of two-way audio-to-text conversion system offers significant advantages in human-robot interaction. With the continuous improvement of algorithms and the use of advanced machine learning techniques, this system can significantly contribute to the evolution of robotic technologies and to the enhancement of communication in a variety of application fields, improving speech recognition in noisy environments, adapting the system to support multiple languages, and embedding emotional recognition in speech analysis for more natural interaction. This work forms a basis for further research and development in the field of robotic systems and natural language processing, promoting human-machine collaboration.

Keywords: Speech-to-Text (STT), Text-to-Speech (TTS), Robotic systems, Speech Recognition, Google Speech to Text API, Voice commands**,** Voice instructions.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Περιεχόμενα

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

ΕΥΧΑΡΙΣΤΙΕΣ

ΠΕΡΙΛΗΨΗ

ABSTRACT

1. **Εισαγωγή**
    - Ιστορική Αναδρομή
    - Σκοπός και στόχοι της εργασίας
    - Αναγκαιότητα και εφαρμογές του συστήματος
    - Δομή της εργασίας
2. **Θεωρητικό Υπόβαθρο**
    - Αναγνώριση φωνής: Βασικές αρχές και τεχνολογίες
    - Τελευταίες Εξελίξεις και προκλήσεις
    - Πως λειτουργεί το ASR
    - Μετατροπή κειμένου σε ομιλία. Αρχές και τεχνολογίες
    - Σύνθεση φωνής
    - Ρομποτικές εφαρμογές: Χρήσεις και απαιτήσεις
3. **Σχεδιασμός Συστήματος**
    - Απαιτήσεις και προδιαγραφές του συστήματος
    - Αρχιτεκτονική του συστήματος
    - Επιλογή εργαλείων και τεχνολογιών
4. **Υλοποίηση Συστήματος**
    - Αναγνώριση φωνής
    - Μετατροπή κειμένου σε ομιλία
    - Ενσωμάτωση και προσαρμογή για ρομποτικές εφαρμογές
    - Συνδυασμός και ενσωμάτωση των δύο υποσυστημάτων
5. **Προγραμματισμός Συστήματος**
    - Διαδικασία προγραμματισμού
    - Προετοιμασία Κώδικα
6. **Συμπεράσματα και Προοπτικές**
    - Συμπεράσματα από την υλοποίηση και την αξιολόγηση
    - Προτάσεις για μελλοντική έρευνα και βελτιώσεις

Πίνακας εικόνων

Εικόνα 1 Αναφορά στο σύστημα text to speech της Apple 'Siri' ……..18
Εικόνα 2 Δημιουργημένη αποκλειστικά από εργαλείο AI.....19
Εικόνα 3 20
Εικόνα 4 – Θεωριτικό κομμάτι 22
Εικόνα 5 – Διάγραμμα λειτουργίας του συνολικού Project ……24
Εικόνα 6 ……25
Εικόνα 7……30
Εικόνα 8 – APIs ……33
Εικόνα 9 – Διάγραμμα λειτουργίας της μονάδας Text to Speech 36
Εικόνα 10 – Σύμβολο Python ……37
Εικόνα 11 ……39
Εικόνα 12……40
Εικόνα 13 ……41
Εικόνα 14 ……43
Εικόνα 15 ……43
Εικόνα 16 ……44
Εικόνα 17 ……44
Εικόνα 18 ……45
Εικόνα 19 ……51

Κεφάλαιο 1

Εισαγωγή

- **Ιστορική Αναδρομή**

Η ανάπτυξη συστημάτων αμφίδρομης μετατροπής ήχου - κειμένου (speech- to-text και text-to-speech) για ρομποτικές εφαρμογές είναι ένα πεδίο που έχει γνωρίσει σημαντική εξέλιξη τις τελευταίες δεκαετίες. Ακολουθεί μια ανασκόπηση των ιστορικών στοιχείων αυτής της τεχνολογίας.

Έρευνες και Ανάπτυξη 1950

Η αρχή της έρευνας στη φωνητική αναγνώριση και σύνθεση έγινε στα τέλη της δεκαετίας του 1950. Μία από τις πρώτες σημαντικές προσπάθειες ήταν του Bell Labs, που ανέπτυξε το σύστημα Audrey το 1952, το οποίο μπορούσε να αναγνωρίσει δέκα αριθμούς που εκφωνούνταν από μία μόνο φωνή.

Πρόοδος στη Δεκαετία του 1970

Στη δεκαετία του 1970, έγιναν σημαντικά βήματα προς την κατεύθυνση της βελτίωσης της ακρίβειας και της ταχύτητας των συστημάτων αναγνώρισης φωνής. Ένα αξιοσημείωτο παράδειγμα είναι το "Harpy" του Carnegie Mellon University, που μπορούσε να αναγνωρίσει περίπου 1.000 λέξεις.

Ανάπτυξη στην Δεκαετία του 1980

Η δεκαετία του 1980 είδε την εισαγωγή πιο εξελιγμένων τεχνικών, όπως τα κρυφά μοντέλα Markov (Hidden Markov Models - HMMs), που επέτρεψαν την ανάλυση της φωνής με μεγαλύτερη ακρίβεια και ευελιξία. Αυτές οι τεχνικές αποτελούν τη βάση πολλών σύγχρονων συστημάτων αναγνώρισης φωνής.

Ανάδυση Εμπορικών Εφαρμογών στη Δεκαετία του 1990

Η δεκαετία του 1990 σημείωσε την εμπορική χρήση των συστημάτων φωνητικής αναγνώρισης, με εταιρείες όπως η Dragon Systems να κυκλοφορούν προϊόντα για προσωπικούς υπολογιστές. Τα συστήματα αυτά άρχισαν να χρησιμοποιούνται σε εφαρμογές όπως η μεταγραφή κειμένων και οι τηλεφωνικές υπηρεσίες.

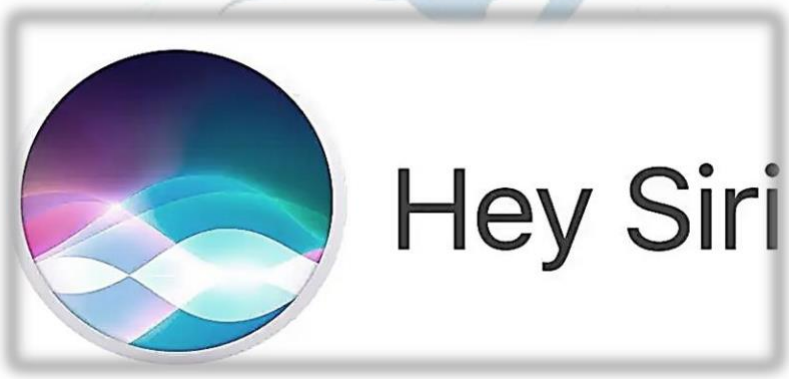
Έκρηξη των Δεδομένων και η Δεκαετία του 2000

Με την έκρηξη της χρήσης των διαδικτυακών δεδομένων, η δεκαετία του 2000 έφερε την εισαγωγή τεχνικών μηχανικής μάθησης και βαθιάς μάθησης. Η Google, για παράδειγμα, χρησιμοποίησε τεράστια σύνολα δεδομένων και ισχυρούς υπολογιστές για να βελτιώσει δραματικά την ακρίβεια των συστημάτων αναγνώρισης φωνής.

Σήμερα, τα συστήματα αναγνώρισης φωνής και σύνθεσης κειμένου έχουν ενσωματωθεί σε πολλές ρομποτικές εφαρμογές. Ρομπότ όπως το Pepper και το NAO χρησιμοποιούν αυτά τα συστήματα για να επικοινωνούν με ανθρώπους σε φυσική γλώσσα. Τεχνολογίες όπως η Siri της Apple, η Alexa της Amazon, και ο Βοηθός της Google έχουν κάνει την αναγνώριση φωνής και την παραγωγή κειμένου αναπόσπαστο μέρος της καθημερινότητας μας.

Η αμφίδρομη μετατροπή ήχου - κειμένου παραμένει ένα πεδίο με συνεχή ανάπτυξη, επηρεάζοντας όχι μόνο τις ρομποτικές εφαρμογές αλλά και μια ευρεία γκάμα τεχνολογικών τομέων.

Η έρευνα συνεχίζεται με στόχο την ανάπτυξη ακόμα πιο έξυπνων και προσαρμοστικών συστημάτων. Οι μελλοντικές προοπτικές περιλαμβάνουν την ενσωμάτωση της τεχνητής νοημοσύνης για την καλύτερη κατανόηση της ανθρώπινης γλώσσας και συναισθημάτων, καθιστώντας τα ρομπότ ικανότερα να αλληλοεπιδρούν με ανθρώπους με πιο φυσικό και ουσιαστικό τρόπο.



ΕΙΚΟΝΑ 01 [12]

- **Σκοπός και Στόχοι της Εργασίας**

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός συστήματος αμφίδρομης μετατροπής ήχου-κειμένου για ρομποτικές εφαρμογές. Στόχος είναι να δημιουργηθεί ένα σύστημα που να μπορεί να αναγνωρίζει φωνή και να την μετατρέπει σε κείμενο καθώς και να μετατρέπει κείμενο σε ομιλία με τρόπο που να είναι λειτουργικός και αποδοτικός για χρήση σε ρομποτικά συστήματα.

ΕΙΚΟΝΑ 02

(πρόθεση της φωτογραφίας ειναι η ένδειξη της χρησιμότητας των νέων τεχνολογιών)

- **Αναγκαιότητα και Εφαρμογές του Συστήματος**

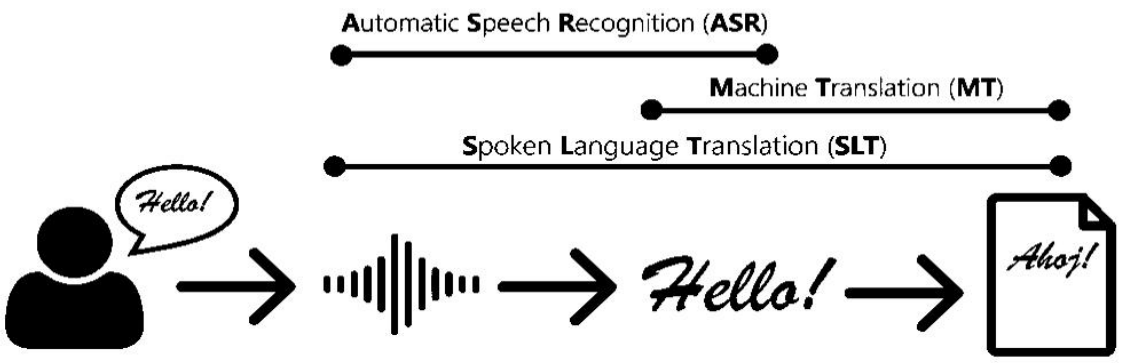
Η επικοινωνία μεταξύ ανθρώπων και ρομποτικών συστημάτων είναι κρίσιμης σημασίας για την ανάπτυξη ευφυών ρομπότ. Η αμφίδρομη μετατροπή ήχου-κειμένου παρέχει έναν φυσικό τρόπο αλληλεπίδρασης που μπορεί να βελτιώσει σημαντικά την ευχρηστία και την αποτελεσματικότητα των ρομποτικών εφαρμογών. Στην πράξη, η δυνατότητα των ρομπότ να κατανοούν και να παράγουν φυσικό ανθρώπινο λόγο επιτρέπει την πιο ομαλή και αποτελεσματική ενσωμάτωσή τους στην καθημερινή ζωή και τις βιομηχανικές διαδικασίες. Οι χρήστες μπορούν να δίνουν εντολές, να κάνουν ερωτήσεις και να λαμβάνουν απαντήσεις με τρόπο που είναι πιο διαισθητικός και άμεσος σε σύγκριση με τις παραδοσιακές μεθόδους εισαγωγής δεδομένων, όπως τα πληκτρολόγια και τα κουμπιά.

Η αμφίδρομη μετατροπή ήχου-κειμένου περιλαμβάνει δύο κύριες τεχνολογίες, την αναγνώριση φωνής (speech recognition) και τη σύνθεση φωνής (speech synthesis). Η αναγνώριση φωνής επιτρέπει στα ρομποτικά συστήματα να κατανοούν τις φωνητικές εντολές των χρηστών και να τις μετατρέπουν σε κείμενο, το οποίο μπορεί να επεξεργαστεί το σύστημα. Η σύνθεση φωνής, από την άλλη, επιτρέπει στα ρομποτικά συστήματα να παράγουν φυσικό ήχο από κείμενο, δίνοντας απαντήσεις και πληροφορίες με τρόπο που μοιάζει με ανθρώπινη ομιλία. Η επιτυχής ενσωμάτωση αυτών των τεχνολογιών μπορεί να βελτιώσει πολλές πτυχές των ρομποτικών εφαρμογών Η φωνητική αλληλεπίδραση είναι

πιο φυσική και μπορεί να μειώσει την ανάγκη για εκπαίδευση των χρηστών. Η γρήγορη και άμεση επικοινωνία μπορεί να αυξήσει την ταχύτητα και την ακρίβεια των λειτουργιών του ρομπότ. Οι τεχνολογίες αυτές μπορούν να καταστήσουν τα ρομποτικά συστήματα πιο προσιτά σε άτομα με αναπηρίες ή περιορισμένη κινητικότητα. Η φωνητική αλληλεπίδραση μπορεί να βελτιώσει την συνολική εμπειρία του χρήστη, κάνοντάς την πιο ευχάριστη και αλληλεπιδραστική. Η ανάπτυξη ευφυών ρομπότ με δυνατότητα αμφίδρομης μετατροπής ήχου-κειμένου απαιτεί συνδυασμό προχωρημένων τεχνολογιών μηχανικής μάθησης, επεξεργασίας φυσικής γλώσσας και αναγνώρισης προτύπων. Καθώς οι τεχνολογίες αυτές εξελίσσονται, οι ρομποτικές εφαρμογές θα γίνονται ολοένα και πιο αποτελεσματικές και προσαρμοστικές στις ανάγκες των χρηστών.

Δομή της Εργασίας

Η εργασία αυτή δομείται σε έξι κύρια κεφάλαια. Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή στο θέμα και παρουσιάζονται οι στόχοι. Στο δεύτερο κεφάλαιο αναλύονται οι θεωρητικές βάσεις των τεχνολογιών που χρησιμοποιούνται. Στο τρίτο κεφάλαιο παρουσιάζεται ο σχεδιασμός του συστήματος, ενώ στο τέταρτο η υλοποίηση. Στο πέμπτο κεφάλαιο γίνεται η αξιολόγηση του συστήματος και στο έκτο παρουσιάζονται τα συμπεράσματα και οι προοπτικές για μελλοντική έρευνα.



ΕΙΚΟΝΑ 03 [13]

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

ΕΙΚΟΝΑ 04 [14]

- **Αναγνώριση Φωνής : Βασικές Αρχές και Τεχνολογίες**

Η αναγνώριση φωνής αποτελείται από μια σειρά από πολύπλοκα στάδια και διαδικασίες που συνεργάζονται για την επιτυχή μετατροπή της ομιλίας σε κείμενο. Ας εξετάσουμε πιο αναλυτικά μερικές από τις τεχνικές και τις τεχνολογίες που χρησιμοποιούνται σε αυτή τη διαδικασία:

Προ-επεξεργασία Ηχητικού Σήματος: Πριν από οποιαδήποτε ανάλυση, το ηχητικό σήμα πρέπει να προ-επεξεργαστεί. Αυτό περιλαμβάνει την απομάκρυνση του θορύβου και τη βελτίωση της ποιότητας του σήματος. Χρησιμοποιούνται τεχνικές όπως το φιλτράρισμα συχνοτήτων και η εξομάλυνση του ήχου για να εξασφαλιστεί ότι το σήμα είναι όσο το δυνατόν καθαρότερο και έτοιμο για ανάλυση. Η εξαγωγή χαρακτηριστικών (feature extraction) είναι ένα κρίσιμο στάδιο όπου το ηχητικό σήμα αναλύεται και μετατρέπεται σε μια σειρά από χαρακτηριστικά που μπορούν να επεξεργαστούν από τα γλωσσικά μοντέλα και τους αλγόριθμους μηχανικής μάθησης. Κοινά χαρακτηριστικά περιλαμβάνουν τα Mel Frequency Cepstral Coefficients (MFCCs), τα οποία αναπαριστούν την φασματική περιεκτικότητα του ήχου με τρόπο που είναι κοντά στην ανθρώπινη ακουστική αντίληψη. Τα ακουστικά μοντέλα χρησιμοποιούν τα χαρακτηριστικά του ήχου για να αναγνωρίσουν τα φωνήματα, τις μικρότερες μονάδες ήχου της ομιλίας. Αυτά τα μοντέλα εκπαιδεύονται με τη χρήση μεγάλων συνόλων δεδομένων από ηχητικά δείγματα και χρησιμοποιούν τεχνικές μηχανικής μάθησης, όπως τα κρυφά μοντέλα (Hidden Markov Models - HMMs) και τα νευρωνικά δίκτυα.

Η αναγνώριση φωνής απαιτεί την αποτελεσματική ενσωμάτωση των γλωσσικών και ακουστικών μοντέλων. Τα γλωσσικά μοντέλα χρησιμοποιούνται για την πρόβλεψη της ακολουθίας λέξεων με βάση τα ακουστικά σήματα, λαμβάνοντας υπόψη τη συντακτική και σημασιολογική πληροφορία της γλώσσας. Αυτή η συνεργασία επιτρέπει στο σύστημα να διορθώνει πιθανά λάθη και να βελτιώνει την ακρίβεια της αναγνώρισης.

- **Τελευταίες Εξελίξεις και Προκλήσεις**

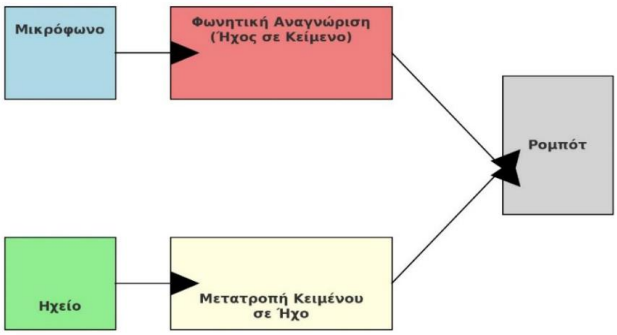
Τα τελευταία χρόνια, η βαθιά μάθηση και τα νευρωνικά δίκτυα έχουν φέρει επανάσταση στον τομέα της αναγνώρισης φωνής. Τα συστήματα αυτά, όπως τα Convolutional Neural Networks (CNNs) και τα Recurrent Neural Networks (RNNs), μπορούν να επεξεργάζονται μεγάλες ποσότητες δεδομένων και να μαθαίνουν πολύπλοκα μοτίβα με υψηλή ακρίβεια. Τα Transformers, μια πιο πρόσφατη τεχνολογία, έχουν επίσης βελτιώσει σημαντικά την απόδοση των συστημάτων αναγνώρισης φωνής, επιτρέποντας την καλύτερη κατανόηση του πλαισίου της ομιλίας. Μια από τις μεγάλες προκλήσεις είναι η προσαρμογή των συστημάτων αναγνώρισης φωνής σε διαφορετικές φωνές, διαλέκτους και προφορές.

Η εξατομίκευση των συστημάτων, δηλαδή η δυνατότητα να προσαρμόζονται στις μοναδικές φωνητικές χαρακτηριστικές του κάθε χρήστη, είναι κρίσιμη για την αύξηση της ακρίβειας και της αποδοχής από το κοινό. Η ευρεία χρήση της αναγνώρισης φωνής φέρνει στο προσκήνιο ζητήματα ασφάλειας. Η καταγραφή και η επεξεργασία των φωνητικών δεδομένων πρέπει να γίνονται με σεβασμό στα προσωπικά δεδομένα των χρηστών. Επιπλέον, πρέπει να εξασφαλιστεί ότι τα συστήματα είναι ασφαλή και προστατευμένα από κακόβουλες επιθέσεις.

Με την τεχνολογία της αναγνώρισης φωνής να συνεχίζει να εξελίσσεται, νέες καινοτομίες και βελτιώσεις αναδύονται συνεχώς. Η συνδυασμένη χρήση της αναγνώρισης φωνής με άλλες τεχνολογίες, όπως η επαυξημένη πραγματικότητα και η τεχνητή νοημοσύνη , ανοίγει νέους δρόμους και εφαρμογές στη βελτίωση της αλληλεπίδρασης ανθρώπου-υπολογιστή.

Η αύξηση της προσβασιμότητας και η ενίσχυση της παραγωγικότητας είναι μόνο μερικά από τα οφέλη που μπορούμε να περιμένουμε στο μέλλον.

Με την περαιτέρω ανάπτυξη της αναγνώρισης φωνής, οι δυνατότητες αυτής της τεχνολογίας θα επεκταθούν ακόμη περισσότερο, προσφέροντας νέες λύσεις και βελτιώνοντας την ποιότητα ζωής σε πολλούς τομείς.



ΕΙΚΟΝΑ 05

Πως Λειτουργεί το ASR

Η αναγνώριση φωνής (ASR - Automatic Speech Recognition) είναι η διαδικασία κατά την οποία ένα σύστημα μετατρέπει την ομιλία σε κείμενο. Η τεχνολογία αυτή χρησιμοποιείται σε διάφορες εφαρμογές, όπως οι ψηφιακοί βοηθοί, οι τηλεφωνικές υπηρεσίες και οι συσκευές IoT. Η λειτουργία της ASR περιλαμβάνει αρκετά στάδια και τεχνικές που συνεργάζονται για να επιτύχουν αυτόν τον σκοπό. Ας δούμε τα βασικά στάδια της αναγνώρισης φωνής. Η διαδικασία ξεκινά με την καταγραφή του ηχητικού σήματος από ένα μικρόφωνο. Το ηχητικό σήμα μπορεί να περιέχει θόρυβο από το περιβάλλον, οπότε είναι σημαντικό να καθαριστεί και να βελτιωθεί η ποιότητά του. Η φωνή αναλύεται για να εξάγει φασματικά χαρακτηριστικά, τα οποία είναι οι πληροφορίες του σήματος που σχετίζονται με τις συχνότητες και την ένταση του ήχου. Αυτή η διαδικασία περιλαμβάνει τη μετατροπή του σήματος από τον χρόνο στην συχνότητα και χώρο, χρησιμοποιώντας τεχνικές όπως η Γρήγορη Μετατροπή Fourier (FFT), στη συνέχεια, εξάγονται τα βασικά χαρακτηριστικά του ήχου, όπως τα Mel-Frequency Cepstral Coefficients (MFCCs) ή τα Perceptual Linear Predictive coefficients (PLPs). Αυτά τα χαρακτηριστικά χρησιμοποιούνται για να εκπροσωπήσουν την ακουστική πληροφορία του ήχου με πιο κατανοητό και επεξεργάσιμο τρόπο για τα μοντέλα μηχανικής μάθησης. Τα ακουστικά μοντέλα χρησιμοποιούν τα εξαγόμενα χαρακτηριστικά για να αναγνωρίσουν βασικά φωνήματα (phonemes) ή άλλες μονάδες ομιλίας. Αυτά τα μοντέλα συνήθως βασίζονται σε τεχνικές όπως τα Κρυφά Μαρκοβιανά Μοντέλα (HMMs) ή τα Νευρωνικά Δίκτυα (Neural Networks). Τα γλωσσικά μοντέλα χρησιμοποιούνται για να προβλέψουν την πιθανότητα μιας ακολουθίας λέξεων και να βοηθήσουν στη διάκριση μεταξύ παρόμοιων ήχων βάσει του πλαισίου. Τα γλωσσικά μοντέλα μπορούν να βασίζονται σε ν-γράμματα (n-grams).

Μετατροπή Κειμένου σε Ομιλία: Αρχές και Τεχνολογίες

Η μετατροπή κειμένου σε ομιλία (Text-to-Speech, TTS) είναι η τεχνολογία που επιτρέπει την παραγωγή ομιλίας από γραπτό κείμενο. Η διαδικασία αυτή περιλαμβάνει διάφορα στάδια και τεχνολογίες για να επιτευχθεί η επιθυμητή ποιότητα και φυσικότητα της ομιλίας.
Ανάλυση Κειμένου και Προ-επεξεργασία: σε αυτό το στάδιο, το γραπτό κείμενο αναλύεται και προετοιμάζεται για τη σύνθεση, περιλαμβάνει την αναγνώριση και την εκκαθάριση ειδικών χαρακτήρων, τη μετατροπή αριθμών σε λέξεις, καθώς και την ανάλυση της δομής των προτάσεων και των γραμματικών τους στοιχείων. Το κείμενο μετατρέπεται σε φωνητική μορφή, χρησιμοποιώντας φωνητική ανάλυση. Αυτό περιλαμβάνει τη χρήση φωνητικών αλφαβήτων, όπως το Διεθνές Φωνητικό Αλφάβητο (International Phonetic Alphabet, IPA) για να αποδοθούν οι ήχοι των λέξεων.

ΕΙΚΟΝΑ 06 [15]

Σύνθεση Φωνής

Προκαθορισμένα Μοντέλα (Concatenative Synthesis): Χρησιμοποιούνται προεγγεγραμμένα τμήματα ομιλίας (φωνητικές μονάδες) από μια βάση δεδομένων, τα οποία συνδέονται για να παράγουν την τελική ομιλία. Αυτός ο τρόπος σύνθεσης μπορεί να αποδώσει πολύ φυσική φωνή, αλλά περιορίζεται από την ποιότητα και την ποικιλία της βάσης δεδομένων.
Σύνθεση Μέσω Παραμετρικών Μοντέλων (Parametric Synthesis): Χρησιμοποιούνται μαθηματικά μοντέλα για να παράγουν φωνή. Τα παραμετρικά μοντέλα (όπως τα μοντέλα βασισμένα σε DNN - Deep Neural Networks) μπορούν να προσφέρουν μεγαλύτερη ευελιξία και προσαρμοστικότητα, αλλά μπορεί να μην είναι τόσο φυσικά όσο

τα προκαθορισμένα μοντέλα.

Προσαρμογή Παραμέτρων Ομιλίας Για την παραγωγή φυσικής και κατανοητής ομιλίας

Οι παράμετροι της φωνής (όπως τονικότητα, ταχύτητα, έμφαση, και ρυθμός) προσαρμόζονται ανάλογα με το περιεχόμενο και το πλαίσιο του κειμένου. ***Ανατροφοδότηση και Βελτίωση:*** Η διαδικασία περιλαμβάνει συνεχείς βελτιώσεις και ανατροφοδότηση για να βελτιωθεί η ποιότητα της σύνθεσης ομιλίας. Οι σύγχρονες TTS τεχνολογίες χρησιμοποιούν τεχνικές μηχανικής μάθησης για να εκπαιδεύσουν τα μοντέλα και να βελτιώσουν την απόδοσή τους με την πάροδο του χρόνου. Η TTS τεχνολογία βρίσκει εφαρμογές σε πολλούς τομείς, όπως οι βοηθοί φωνής, οι συσκευές για άτομα με προβλήματα όρασης, τα συστήματα πλοήγησης, και οι πλατφόρμες εκμάθησης γλωσσών. Οι εξελίξεις στην τεχνητή νοημοσύνη και στη μηχανική μάθηση συνεχίζουν να βελτιώνουν την ποιότητα και τις δυνατότητες των TTS συστημάτων, καθιστώντας τα πιο προσιτά και αποτελεσματικά.

Ρομποτικές Εφαρμογές: Χρήσεις και Απαιτήσεις

Οι ρομποτικές εφαρμογές που αξιοποιούν την αμφίδρομη μετατροπή ήχου-κειμένου (Speech-to-Text και Text-to-Speech) παίζουν σημαντικό ρόλο σε πολλούς τομείς όπου η φυσική επικοινωνία είναι κρίσιμη. Αυτές οι τεχνολογίες συνδυάζουν την αναγνώριση ομιλίας και τη σύνθεση φωνής για να επιτρέψουν στους χρήστες να αλληλοεπιδρούν με συστήματα και ρομπότ με φυσικό και ανθρώπινο τρόπο. Οι βασικές απαιτήσεις για τέτοια συστήματα περιλαμβάνουν την ακρίβεια, την ταχύτητα απόκρισης, και την προσαρμοστικότητα σε διαφορετικές γλώσσες και διαλέκτους.

Οι ρομποτικές εφαρμογές στην εξυπηρέτηση πελατών χρησιμοποιούν τεχνολογίες αμφίδρομης μετατροπής ήχου κειμένου για να παρέχουν άμεση και αποτελεσματική υποστήριξη στους πελάτες. Αυτά τα συστήματα μπορούν να απαντούν σε συχνές ερωτήσεις χρησιμοποιώντας προγραμματισμένα σενάρια και φυσική γλώσσα, μπορούν να αναγνωρίσουν τις ερωτήσεις των πελατών και να παρέχουν ακριβείς απαντήσεις.

Αναγνωρίζοντας το αίτημα του πελάτη, μπορεί να δρομολογεί την κλήση στον κατάλληλο ανθρώπινο εκπρόσωπο χρησιμοποιώντας δεδομένα πελατών, ενώ τα ρομπότ μπορούν να προσαρμόσουν τις απαντήσεις τους για να καλύψουν τις συγκεκριμένες ανάγκες του πελάτη.

Στην εκπαίδευση, τα συστήματα αυτά μπορούν να βοηθήσουν μαθητές και εκπαιδευτικούς, μπορούν να εξασκούν τις γλωσσικές τους δεξιότητες συνομιλώντας με τα ρομπότ, τα οποία παρέχουν άμεση ανατροφοδότηση και διορθώσεις.

Ρομποτικοί βοηθοί μπορούν να διεξάγουν διαδραστικά μαθήματα, απαντώντας σε ερωτήσεις των μαθητών και προσαρμόζοντας το υλικό διδασκαλίας ανάλογα με τις ανάγκες τους.

Υποστήριξη μαθητών με ειδικές ανάγκες: μπορούν να παρέχουν υποστήριξη σε μαθητές με μαθησιακές δυσκολίες ή αναπηρίες, προσφέροντας εξατομικευμένες εκπαιδευτικές λύσεις.

Στην υγειονομική περίθαλψη, η χρήση των τεχνολογιών αυτών μπορεί να βελτιώσει την ποιότητα και την αποτελεσματικότητα των υπηρεσιών ρομπότ που μπορούν να αναγνωρίσουν και να απαντήσουν σε ερωτήσεις των ασθενών, να δίνουν συμβουλές για την υγεία και να υπενθυμίζουν τα φάρμακα.

Χρησιμοποιώντας αναγνώριση ομιλίας και σύνθεση φωνής, οι γιατροί μπορούν να αλληλοεπιδρούν με ασθενείς απομακρυσμένα, διατηρώντας υψηλή ποιότητα επικοινωνίας.

Τα συστήματα μπορούν να καταγράφουν τις συνομιλίες και τις παρατηρήσεις των γιατρών σε πραγματικό χρόνο, βελτιώνοντας την ακρίβεια των ιατρικών αρχείων.

Χρήση ρομποτικών συστημάτων σε γραφεία για την εκτέλεση διοικητικών εργασιών και την απάντηση σε τηλεφωνικές κλήσεις.

Παροχή βοήθειας σε άτομα με κινητικές ή οπτικές αναπηρίες μέσω φωνητικών εντολών, ρομπότ σε ψυχαγωγικά πάρκα ή εκδηλώσεις, που αλληλοεπιδρούν με τους επισκέπτες προσφέροντας πληροφορίες και καθοδήγηση.

Οι απαιτήσεις για την επιτυχή λειτουργία αυτών των συστημάτων περιλαμβάνουν την ικανότητα των συστημάτων να αναγνωρίζουν σωστά τις λέξεις και τις προθέσεις του ομιλητή, τη γρήγορη απόκριση στις εντολές και τις ερωτήσεις για να διατηρείται η ροή της επικοινωνίας, να κατανοούν και να μιλούν διαφορετικές γλώσσες και διαλέκτους για να είναι χρήσιμα σε παγκόσμιο επίπεδο.

Η συνεχής πρόοδος στις τεχνολογίες τεχνητής νοημοσύνης και μηχανικής μάθησης αναμένεται να βελτιώσει ακόμη περισσότερο την ποιότητα και την απόδοση αυτών των συστημάτων, καθιστώντας τα απαραίτητα εργαλεία σε πολλούς τομείς της καθημερινής ζωής.

Κεφάλαιο 3

Σχεδιασμός Συστήματος

- **Απαιτήσεις και Προδιαγραφές του Συστήματος**

Η ακρίβεια στην αναγνώριση φωνής αποτελεί μία από τις πιο κρίσιμες απαιτήσεις για το σύστημα. Για να επιτευχθεί αυτό, θα χρησιμοποιηθούν προηγμένοι αλγόριθμοι αναγνώρισης φωνής που μπορούν να ανιχνεύσουν και να αναλύσουν φωνητικά δεδομένα με υψηλή πιστότητα. Οι τεχνολογίες μηχανικής μάθησης και τα βαθιά νευρωνικά δίκτυα θα ενσωματωθούν για να βελτιωθεί η ακρίβεια, ακόμα και σε περιβάλλοντα με θόρυβο ή όταν οι ομιλητές έχουν διαφορετικές προφορές και τονισμούς, η παραγωγή ομιλίας πρέπει να είναι φυσική και κατανοητή για να εξασφαλίζεται η αποτελεσματική επικοινωνία με τους χρήστες. Η συνθετική ομιλία θα πρέπει να ακούγεται ανθρώπινη, με κατάλληλες διακυμάνσεις στον τόνο, το ρυθμό και την έμφαση, η τεχνολογία Text-to-Speech (TTS) θα προσαρμοστεί για να δημιουργεί φωνές που είναι όχι μόνο ευχάριστες στην ακρόαση αλλά και ικανές να μεταφέρουν συναισθήματα και εννοιολογικές αποχρώσεις. Η ταχύτητα απόκρισης είναι κρίσιμη για την επιτυχία του συστήματος, ειδικά όταν αυτό χρησιμοποιείται σε δυναμικά περιβάλλοντα ή σε εφαρμογές όπου η άμεση ανταπόκριση είναι απαραίτητη. Οι χρόνοι επεξεργασίας θα βελτιστοποιηθούν μέσω της χρήσης αποδοτικών αλγορίθμων και της αξιοποίησης της επεξεργαστικής ισχύος των σύγχρονων υπολογιστικών συστημάτων. Η αρχιτεκτονική του συστήματος θα σχεδιαστεί ώστε να μειώσει την καθυστέρηση από τη στιγμή της λήψης της φωνητικής εντολής μέχρι την απόκριση. Για να είναι ευέλικτο και χρήσιμο σε ποικιλία εφαρμογών, το σύστημα πρέπει να είναι συμβατό με διάφορες ρομποτικές πλατφόρμες.

Αυτό σημαίνει ότι η ανάπτυξη θα βασιστεί σε τυποποιημένα πρωτόκολλα επικοινωνίας και προγραμματισμού εφαρμογών (APIs) που μπορούν να ενσωματωθούν εύκολα σε διαφορετικά ρομποτικά συστήματα. Η διαλειτουργικότητα θα εξασφαλιστεί μέσω της υποστήριξης πολλαπλών λειτουργικών συστημάτων και του σχεδιασμού για ευέλικτη ενσωμάτωση υλικού και λογισμικού. Η ανάπτυξη αυτών των απαιτήσεων θα οδηγήσει σε ένα σύστημα φωνητικής αλληλεπίδρασης που θα είναι ακριβές, φυσικό, γρήγορο και ευέλικτο, καλύπτοντας τις ανάγκες ενός ευρέος φάσματος εφαρμογών και χρηστών. Η αρχιτεκτονική του συστήματος περιλαμβάνει δύο κύρια υποσυστήματα: το υποσύστημα αναγνώρισης φωνής και το υποσύστημα μετατροπής κειμένου σε ομιλία. Τα δύο υποσυστήματα συνδέονται μέσω ενός κεντρικού διαχειριστή που διαχειρίζεται την ροή των δεδομένων.

ΕΙΚΟΝΑ 07 [16]

- **Αρχιτεκτονική Συστήματος**

Η ανάπτυξη ενός ρομποτικού συστήματος με δυνατότητες αναγνώρισης ομιλίας και επεξεργασίας φυσικής γλώσσας απαιτεί μια καλά σχεδιασμένη αρχιτεκτονική. Αυτή η αρχιτεκτονική πρέπει να υποστηρίζει την ενσωμάτωση πολλαπλών τεχνολογιών και εργαλείων, διασφαλίζοντας ταυτόχρονα την αποδοτικότητα και την αξιοπιστία του συστήματος. Σε αυτό το κεφάλαιο, παρουσιάζεται η αρχιτεκτονική ενός τέτοιου συστήματος, καλύπτοντας τα κύρια συστατικά και τις διεπαφές τους.

Η γενική αρχιτεκτονική ενός ρομποτικού συστήματος με δυνατότητες ομιλίας και επεξεργασίας φυσικής γλώσσας μπορεί να περιλαμβάνει τα ακόλουθα βασικά συστατικά: Μικρόφωνα και Αισθητήρες Συλλέγουν ηχητικά δεδομένα από το περιβάλλον.

Μονάδα Προεπεξεργασίας Ήχου

Φιλτράρει τον ήχο και βελτιστοποιεί τα σήματα πριν την αναγνώριση. Μετατρέπει τον προ-επεξεργασμένο ήχο σε κείμενο, Automatic Speech Recognition. ***Μονάδα Κατανόησης Φυσικής Γλώσσας***: Αναλύει το κείμενο και εξάγει τις σχετικές πληροφορίες. Διαχειρίζεται την αλληλεπίδραση με τον χρήστη μέσω διαλόγου. ***Μονάδα Λήψης Αποφάσεων***: Αξιολογεί τις πληροφορίες και λαμβάνει αποφάσεις για τις ενέργειες του ρομπότ, Ελέγχει τις φυσικές ενέργειες και κινήσεις του ρομπότ. ***Cloud Services***: Παρέχουν επιπλέον υπολογιστική ισχύ και αποθήκευση δεδομένων. Σε δικτυακή Επικοινωνία, επιτρέπει την επικοινωνία μεταξύ των διάφορων μονάδων του συστήματος.

Μικρόφωνα και Αισθητήρες

Χρησιμοποιούνται για τη συλλογή ηχητικών δεδομένων. Τα μικρόφωνα μπορεί να είναι κατευθυντικά ή πολυκατευθυντικά, ανάλογα με τις ανάγκες του συστήματος. ***Μονάδα Επεξεργασίας Ήχου***: Εφαρμόζει φίλτρα θορύβου και άλλες τεχνικές βελτίωσης σήματος για να διασφαλίσει την καθαρότητα των ηχητικών δεδομένων πριν αυτά εισαχθούν στη μονάδα ASR. Εδώ χρησιμοποιούνται μοντέρνοι αλγόριθμοι και μοντέλα μηχανικής μάθησης για τη μετατροπή του ήχου σε κείμενο. Δημοφιλείς πλατφόρμες περιλαμβάνουν το Google Cloud Speech-to-Text, το Microsoft Azure Speech Service, και το IBM Watson Speech to Text: Χρησιμοποιεί εργαλεία όπως το spaCy, το NLTK, ή το Hugging Face Transformers για την ανάλυση και κατανόηση του κειμένου που έχει προκύψει από τη μονάδα ASR. Διαχειρίζεται τις αλληλεπιδράσεις με τον χρήστη, κατανοώντας τα συμφραζόμενα και διατηρώντας τη συνοχή του διαλόγου. Μπορεί να χρησιμοποιήσει πλαίσια όπως το Rasa ή το Dialogflow. Βασισμένη στις πληροφορίες που παρέχει η μονάδα κατανόησης φυσικής γλώσσας, λαμβάνει αποφάσεις για τις ενέργειες που θα πραγματοποιήσει το ρομπότ.

Η επικοινωνία μεταξύ των διαφορετικών μονάδων του συστήματος γίνεται μέσω πρωτοκόλλων δικτύου, όπως το ROS (Robot Operating System), που επιτρέπει την εύκολη ενσωμάτωση και επικοινωνία των διαφόρων συστατικών.

Σενάριο 1: Αναγνώριση Φωνητικών Εντολών

1. Ο χρήστης δίνει φωνητική εντολή στο ρομπότ.

2. Τα μικρόφωνα συλλέγουν τον ήχο και η μονάδα προεπεξεργασίας ήχου τον καθαρίζει.

3. Ο ήχος μετατρέπεται σε κείμενο από την ASR engine.

4. Το κείμενο αναλύεται για να κατανοηθεί η εντολή.

5. Η μονάδα λήψης αποφάσεων επεξεργάζεται την εντολή και επιλέγει την κατάλληλη ενέργεια.

6. Η μονάδα ελέγχου κινητήρα και ενέργειας εκτελεί την ενέργεια.

Σενάριο 2: Διαλογική Επικοινωνία

1. Ο χρήστης ξεκινά έναν διάλογο με το ρομπότ.

2. Κάθε φράση του χρήστη περνά από τη διαδικασία ASR.

3. Η μονάδα διαχείρισης διαλόγου διατηρεί τη συνοχή του διαλόγου και απαντά καταλλήλως.

4. Το ρομπότ εκτελεί ενέργειες ή παρέχει πληροφορίες βάσει των απαντήσεων που έχουν προδηλωθεί στη μονάδα .

Η αρχιτεκτονική ενός ρομποτικού συστήματος με δυνατότητες αναγνώρισης ομιλίας και επεξεργασίας φυσικής γλώσσας πρέπει να είναι καλά σχεδιασμένη για να διασφαλίζει την αποδοτικότητα και την αξιοπιστία. Η ενσωμάτωση σύγχρονων τεχνολογιών ASR, η χρήση υπηρεσιών cloud, και η διαχείριση της επικοινωνίας μεταξύ των μονάδων του συστήματος είναι κρίσιμα στοιχεία για την επιτυχία ενός τέτοιου συστήματος.

- ## Επιλογή Εργαλείων και Τεχνολογιών

Για την υλοποίηση του συστήματος, θα χρησιμοποιηθούν τα εξής εργαλεία και τεχνολογίες: ***Google Speech Recognition API***. Αυτό το API θα χρησιμοποιηθεί για την αναγνώριση φωνής. Πρόκειται για μια υπηρεσία που επιτρέπει την μετατροπή φωνής σε κείμενο, παρέχοντας ακριβή και γρήγορη αναγνώριση φωνής.

Αυτή η υπηρεσία θα χρησιμοποιηθεί για τη μετατροπή κειμένου σε ομιλία. Το API προσφέρει φυσικές και ευχάριστες φωνές για την ανάγνωση του κειμένου. Θα χρησιμοποιηθεί ως προγραμματιστικό περιβάλλον για την ενσωμάτωση των υποσυστημάτων.

Η Python είναι μια ισχυρή γλώσσα προγραμματισμού που παρέχει πληθώρα βιβλιοθηκών και εργαλείων για την ανάπτυξη και την διαχείριση σύνθετων συστημάτων, είναι γνωστή για τη λιτή και κατανοητή σύνταξή της, γεγονός που την καθιστά ιδανική για γρήγορη ανάπτυξη εφαρμογών, υπάρχουν πολλές βιβλιοθήκες που μπορούν να χρησιμοποιηθούν για διάφορες λειτουργίες, όπως η επεξεργασία ήχου, η αναγνώριση φωνής, η μετατροπή κειμένου σε ομιλία και πολλά άλλα. Έχει μια μεγάλη κοινότητα χρηστών και προγραμματιστών, προσφέροντας εκτενή τεκμηρίωση και υποστήριξη.

Χρησιμοποιώντας pip, θα εγκατασταθούν οι βιβλιοθήκες google cloud speech και google cloud text to speech που απαιτούνται για την επικοινωνία με τα APIs της Google.

Θα δημιουργηθούν και θα ρυθμιστούν τα κλειδιά API για πρόσβαση στις υπηρεσίες της Google.

Χρήση του Google Speech Recognition API για την καταγραφή και αναγνώριση της φωνής του χρήστη.

Χρήση του Google Text to Speech API για τη μετατροπή των αποκληθέντων κειμένων σε ομιλία.

Ανάπτυξη ενός ενοποιημένου συστήματος που θα ενσωματώνει και τις δύο λειτουργίες, επιτρέποντας την επικοινωνία του χρήστη με το σύστημα μέσω φωνής.

ΕΙΚΟΝΑ 08 [17]

Για την αναγνώριση φωνής, χρησιμοποιείται το Google Speech Recognition API, το οποίο παρέχει υψηλή ακρίβεια και ταχύτητα απόκρισης. Η ενσωμάτωση του API γίνεται μέσω Python και η προσαρμογή του για ρομποτικές εφαρμογές περιλαμβάνει την επεξεργασία και την ερμηνεία των αποτελεσμάτων της αναγνώρισης.

Κεφάλαιο 4

Υλοποίηση Συστήματος

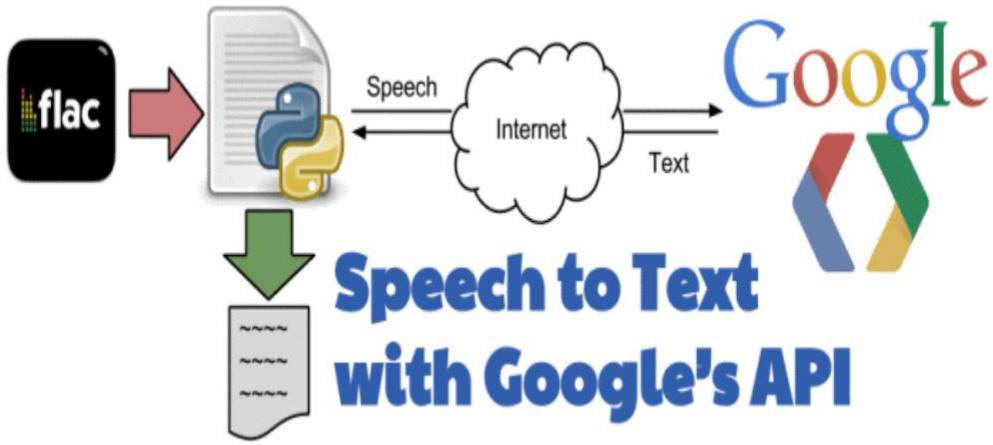
- **Αναγνώριση Φωνής**

Η ενσωμάτωση της αναγνώρισης φωνής και της μετατροπής κειμένου σε ομιλία σε ρομποτικές εφαρμογές μπορεί να βελτιώσει σημαντικά την αλληλεπίδραση ανθρώπου-ρομπότ, καθιστώντας τα ρομπότ πιο ευέλικτα και φιλικά προς το χρήστη. Για την εφαρμογή αυτής της τεχνολογίας μπορούμε να χρησιμοποιήσουμε διάφορα API όπως το Google Speech Recognition, ένα ισχυρό API που υποστηρίζει πολλές γλώσσες και προσφέρει υψηλή ακρίβεια αναγνώρισης φωνής, εύκολα προσαρμόσιμο σε ρομποτικές εφαρμογές μέσω των REST ful υπηρεσιών, του CMU Sphinx ένα ανοικτού κώδικα σύστημα αναγνώρισης φωνής που μπορεί να προσαρμοστεί και να ενσωματωθεί εύκολα σε διάφορες πλατφόρμες, προσφέροντας την ευελιξία προσαρμογής στα ιδιαίτερα χαρακτηριστικά του περιβάλλοντος χρήσης.
Η επιλογή API γίνεται ανάλογα με τις απαιτήσεις της εφαρμογής και τους πόρους του συστήματος.

Για την δημιουργία και διαμόρφωση λογαριασμού για υπηρεσίες όπως το Google Speech Recognition, αποκτούμε τα απαραίτητα κλειδιά πρόσβασης, ενσωμάτωση στον Κώδικα Μέσω κατάλληλων βιβλιοθηκών (π.χ. speech recognition για Python), ενσωματώνουμε την αναγνώριση φωνής στον κώδικα του ρομπότ. Βελτιστοποιούμε την αναγνώριση φωνής για συγκεκριμένες εντολές ή συνθήκες λειτουργίας (π.χ. θορυβώδη περιβάλλοντα).

Το πιο διαδεδομένο API είναι το Google Text to Speech, το οποίο προσφέρει υψηλής ποιότητας συνθετική ομιλία με πολλές φωνές και γλώσσες διαθέσιμες.
Microsoft Azure TTS: Παρέχει επίσης ποιοτική συνθετική ομιλία με επιπλέον δυνατότητες προσαρμογής της έκφρασης και του τόνου της φωνής. Επιλέγουμε το κατάλληλο API σύμφωνα με τις απαιτήσεις της εφαρμογής, δημιουργούμε λογαριασμό και αποκτούμε κλειδιά πρόσβασης, μέσω κατάλληλων βιβλιοθηκών (π.χ. gTTS για Google Text to Speech), ενσωματώνουμε την τεχνολογία TTS στον κώδικα του ρομπότ, ανάλογα με την εφαρμογή, προσαρμόζουμε τον τόνο, την ταχύτητα και την ένταση της φωνής.

Ο συνδυασμός της αναγνώρισης φωνής και της μετατροπής κειμένου σε ομιλία επιτρέπει τη δημιουργία ενός ολοκληρωμένου συστήματος αλληλεπίδρασης φωνής. Ορίζουμε τη ροή των λειτουργιών, π.χ. από την αναγνώριση φωνής στην ανάλυση κειμένου και από εκεί στη μετατροπή κειμένου σε ομιλία, διασφαλίζουμε ότι τα API επικοινωνούν ομαλά μεταξύ τους, ενσωματώνοντάς τα σε μια κοινή πλατφόρμα ή εφαρμογή, αναπτύσσουμε μηχανισμούς διαχείρισης σφαλμάτων και βελτιώνουμε την απόδοση του συστήματος για πραγματικές συνθήκες χρήσης.
Εκτελούμε εκτενείς δοκιμές για να διασφαλίσουμε ότι το σύστημα ανταποκρίνεται στις απαιτήσεις και λειτουργεί αποτελεσματικά στο περιβάλλον χρήσης.
Η υλοποίηση αυτών των βημάτων δημιουργεί ένα αξιόπιστο ρομποτικό σύστημα που μπορεί να αναγνωρίζει φωνητικές εντολές και να απαντά μέσω συνθετικής ομιλίας, προσφέροντας μια βελτιωμένη και φυσικότερη εμπειρία αλληλεπίδρασης.



ΕΙΚΟΝΑ 09 [18]

- **Μετατροπή Κειμένου σε Ομιλία**

Για τη μετατροπή κειμένου σε ομιλία, χρησιμοποιείται το Google Text to Speech API το οποίο παρέχει φυσική και κατανοητή παραγωγή ομιλίας. Η ενσωμάτωση του API γίνεται μέσω Python και η προσαρμογή του περιλαμβάνει τη ρύθμιση της ταχύτητας και της έντασης της ομιλίας.

ΕΙΚΟΝΑ 10 [19]

- **Ενσωμάτωση και Προσαρμογή για Ρομποτικές Εφαρμογές**

Automatic Speech Recognition (ASR) είναι η τεχνολογία που επιτρέπει στους υπολογιστές να κατανοούν και να επεξεργάζονται την ανθρώπινη ομιλία. Ορισμένα από τα πιο δημοφιλή εργαλεία και πλατφόρμες είναι:

- Google Cloud Speech-to-Text: Παρέχει υψηλής ακρίβειας αναγνώριση ομιλίας με υποστήριξη για πολλές γλώσσες και δυνατότητες real-time αναγνώρισης.
- Microsoft Azure Speech Service: Προσφέρει ευέλικτες επιλογές για αναγνώριση ομιλίας, με δυνατότητες προσαρμογής στα συγκεκριμένα δεδομένα του χρήστη.
- IBM Watson Speech to Text: Εστιάζει στην παροχή ακρίβειας μέσω της συνεχούς μάθησης και προσαρμογής στις ανάγκες των χρηστών.

Υπάρχουν επίσης ανοικτού κώδικα λύσεις που επιτρέπουν μεγαλύτερη ευελιξία και προσαρμογή, μια ισχυρή εργαλειοθήκη για αναγνώριση ομιλίας που προσφέρει μεγάλες

δυνατότητες προσαρμογής και βελτιστοποίησης, Kaldi .Ένα παλιό αλλά ευέλικτο εργαλείο που εξακολουθεί να χρησιμοποιείται ευρέως για εφαρμογές αναγνώρισης ομιλίας, CMU Sphinx. Η επεξεργασία των αποτελεσμάτων της αναγνώρισης ομιλίας συχνά περιλαμβάνει την ανάλυση και κατανόηση της φυσικής γλώσσας. Ορισμένα βασικά εργαλεία περιλαμβάνουν:

- spaCy: Μια γρήγορη και αποδοτική βιβλιοθήκη για NLP, που υποστηρίζει ανάλυση κειμένου, εξαγωγή οντοτήτων, και πολλά άλλα.
- NLTK (Natural Language Toolkit): Μια δημοφιλής βιβλιοθήκη που προσφέρει εργαλεία για την ανάλυση κειμένου, τη συντακτική ανάλυση και την εξαγωγή δεδομένων.
- Hugging Face Transformers: Παρέχει σύγχρονες λύσεις για μοντέλα γλώσσας όπως BERT, GPT-3, και άλλα, που μπορούν να χρησιμοποιηθούν για προηγμένες εφαρμογές NLP.

Το ROS είναι ένα ευρέως χρησιμοποιούμενο πλαίσιο για την ανάπτυξη ρομποτικών εφαρμογών που υποστηρίζει την ενσωμάτωση διαφόρων τεχνολογιών και εργαλείων:

Ένα πακέτο του ROS που επιτρέπει τη χρήση εργαλείων ASR σε ρομποτικές εφαρμογές, ros_speech_recognition. Υποστηρίζει την ενσωμάτωση εργαλείων NLP για την επεξεργασία και ανάλυση των αποτελεσμάτων της αναγνώρισης ομιλίας. ros_nlp: Η χρήση των υπηρεσιών αναγνώρισης ομιλίας και NLP από το cloud μπορεί να βελτιώσει την απόδοση και να μειώσει την ανάγκη για τοπική επεξεργαστική ισχύ.

Cloud Robotics: Η ενσωμάτωση cloud υπηρεσιών σε ρομποτικά συστήματα επιτρέπει τη χρήση προχωρημένων εργαλείων και υπηρεσιών χωρίς την ανάγκη για ισχυρό hardware.

Η αξιολόγηση της απόδοσης των συστημάτων αναγνώρισης ομιλίας γίνεται μέσω διαφόρων μετρικών. Ένα κοινό μέτρο για την αξιολόγηση της ακρίβειας της αναγνώρισης ομιλίας είναι το Word Error Rate (WER). Χρησιμοποιούνται για την αξιολόγηση της ακρίβειας στην εξαγωγή οντοτήτων και άλλων στοιχείων από την αναγνώριση ομιλίας. Precision and Recall: Η συνεχής ανατροφοδότηση από τους χρήστες και η προσαρμογή των συστημάτων βάσει των αναγκών τους είναι κρίσιμη για τη βελτίωση της απόδοσης.

Διεξαγωγή ερευνών και μελετών χρηστών για την κατανόηση των αναγκών και των προκλήσεων

Η επιλογή των κατάλληλων εργαλείων και τεχνολογιών για την αναγνώριση ομιλίας και την επεξεργασία των αποτελεσμάτων σε ρομποτικές εφαρμογές είναι κρίσιμη για την επιτυχία αυτών των συστημάτων. Η ενσωμάτωση προηγμένων εργαλείων ASR και NLP, η χρήση του ROS, καθώς και η αξιοποίηση cloud υπηρεσιών, μπορούν να βελτιώσουν σημαντικά την απόδοση και την αξιοπιστία των ρομποτικών συστημάτων.

ΕΙΚΟΝΑ 11 [20]

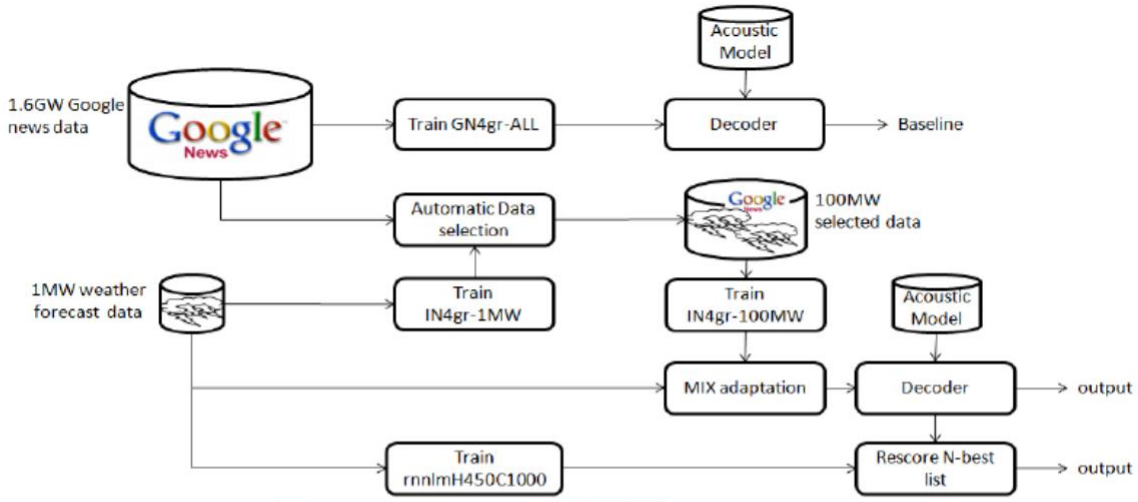
- **Συνδυασμός και Ενσωμάτωση των Δύο Υποσυστημάτων**

Ο συνδυασμός των δύο υποσυστημάτων γίνεται μέσω ενός κεντρικού διαχειριστή που διαχειρίζεται την ροή των δεδομένων από την αναγνώριση φωνής στην παραγωγή ομιλίας και αντίστροφα. Η ενσωμάτωση περιλαμβάνει τη δημιουργία ενός ενιαίου interface για την αλληλεπίδραση του χρήστη με το σύστημα.

Η ενσωμάτωση των δύο υποσυστημάτων, της αναγνώρισης φωνής και της παραγωγής ομιλίας, είναι ιδιαίτερα κρίσιμη διαδικασία και απαιτεί έναν κεντρικό διαχειριστή για τη διαχείριση της ροής των δεδομένων. Ο κεντρικός διαχειριστής (central manager) λειτουργεί ως ο εγκέφαλος του συστήματος, συντονίζοντας την επικοινωνία μεταξύ των υποσυστημάτων και εξασφαλίζοντας την ομαλή και συνεχή ροή των πληροφοριών. Η αναγνώριση φωνής (speech recognition) είναι το υποσύστημα που λαμβάνει και αναλύει τις ηχητικές εισροές του χρήστη, χρησιμοποιεί αλγόριθμους μηχανικής μάθησης για να μετατρέψει την ομιλία σε κείμενο. Αυτή η διαδικασία περιλαμβάνει τα διάφορα προαναφερθέντα στάδια, όπως προ-επεξεργασία του ήχου, εξαγωγή χαρακτηριστικών και τελικά, αναγνώριση λέξεων και φράσεων.

Η παραγωγή ομιλίας (speech synthesis) είναι το υποσύστημα που μετατρέπει το κείμενο σε φυσικό ήχο ομιλίας. Χρησιμοποιεί τεχνολογίες όπως η TTS (Text-to-Speech) για να δημιουργήσει μια ρεαλιστική και κατανοητή φωνή.

Ο κεντρικός διαχειριστής διασφαλίζει ότι η ροή των δεδομένων από την αναγνώριση φωνής στην παραγωγή ομιλίας και αντίστροφα είναι αδιάλειπτη και αποδοτική. Για να επιτευχθεί αυτό, απαιτείται η εφαρμογή ενός συνόλου πρωτοκόλλων και αλγορίθμων που επιτρέπουν τη συνεργασία των υποσυστημάτων σε πραγματικό χρόνο. Ο κεντρικός διαχειριστής πρέπει να διαχειρίζεται / συντονίζει την ανταλλαγή δεδομένων μεταξύ των υποσυστημάτων.



ΕΙΚΟΝΑ 12 [21]

Εξασφαλίζει ότι η αναγνώριση και η παραγωγή γίνονται σε συγχρονισμό.

Διαχειρίζεται την ανθεκτικότητα του συστήματος σε σφάλματα και την αποκατάσταση από πιθανές διακοπές.

Η δημιουργία ενός ενιαίου interface (διεπαφής) είναι κρίσιμη για την αλληλεπίδραση του χρήστη με το σύστημα. Αυτό το interface πρέπει να είναι φιλικό προς το χρήστη και να επιτρέπει την εύκολη και αποτελεσματική χρήση των δυνατοτήτων του συστήματος. Κύρια χαρακτηριστικά ενός τέτοιου interface περιλαμβάνουν:

- Ευκολία στη χρήση, με καθαρό και κατανοητό σχεδιασμό.
- Ανταπόκριση σε πραγματικό χρόνο στις ενέργειες του χρήστη.
- Δυνατότητα προσαρμογής στις ανάγκες και τις προτιμήσεις του χρήστη.
- Παροχή σαφών και χρήσιμων πληροφοριών στον χρήστη κατά την αλληλεπίδραση.

EIKONA 13

Κεφάλαιο 5

Προγραμματισμός Εφαρμογής

- **Διαδικασία Προγραμματισμού**

ΕΙΚΟΝΑ 14

Η εφαρμογή που υλοποιήθηκε στα πλαίσια του project είναι για ένα ρομπότ οικιακό βοηθό όπου αναγνωρίζει τις εντολές που του έχουμε προκαθορίσει και μας απαντάει αναλόγως για την επόμενη κίνηση που θα κάνει .

Στον παρακάτω κώδικα υπάρχουν κάποιες βασικές εντολές για την υλοποίηση.

Ο κώδικας και οι δοκιμές για την λειτουργικότητα έγιναν στο Pycharm και χρησιμοποιήθηκε το μικρόφωνο και το ηχείο ενός φορητού υπολογιστή.

ΕΙΚΟΝΑ 15 [22]

Για την υλοποίηση χρησιμοποιήθηκε το IDE Pycharm (community edition) [1]

Ο interpreter του project ήταν ο default Python 3.10 interpreter του Pycharm.

Τα επιπλέον πακέτα που χρησιμοποιήθηκαν είναι τα εξής:

- Pydub 0.25.1
- Speech Recognition 3.10.4
- gTTS 2.5.1
- sounddevice 0.4.6
- numpy 1.26.4
- difflib 3.10.3

ΕΙΚΟΝΑ 16 [23]

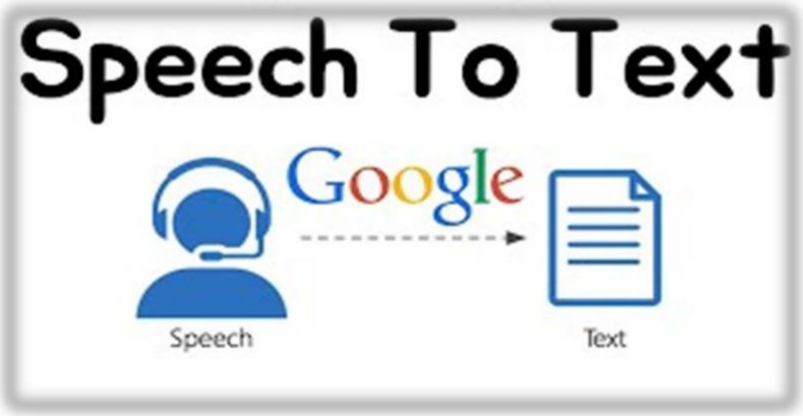
Επιπλέον των παραπάνω, θα πρέπει να κατεβάσουμε το ‘ffmpeg’ και να το δηλώσουμε στο system PATH του υπολογιστή.

Προαπαιτούμενο για την λειτουργία της αναγνώρισης ομιλίας είναι η σύνδεση στο internet.

ΕΙΚΟΝΑ 17 [24]

Για να λειτουργήσει το πρόγραμμα πρεπει να ελέγξουμε οτι υπάρχουν τα εξής πακέτα των interpreters στο Pycharm :

- PyAudio
- SpeechRecognition
- certifi
- cffi
- charset-normalizer
- click
- colorama
- gTTS **(Google Text To Speech , ενσωματώνει το Text to Speech σύστημα της Google)**
- idna
- numpy
- pip
- pycparser
- pudub
- requests
- setuptools
- sounddevice
- typing_extentions
- urlib3

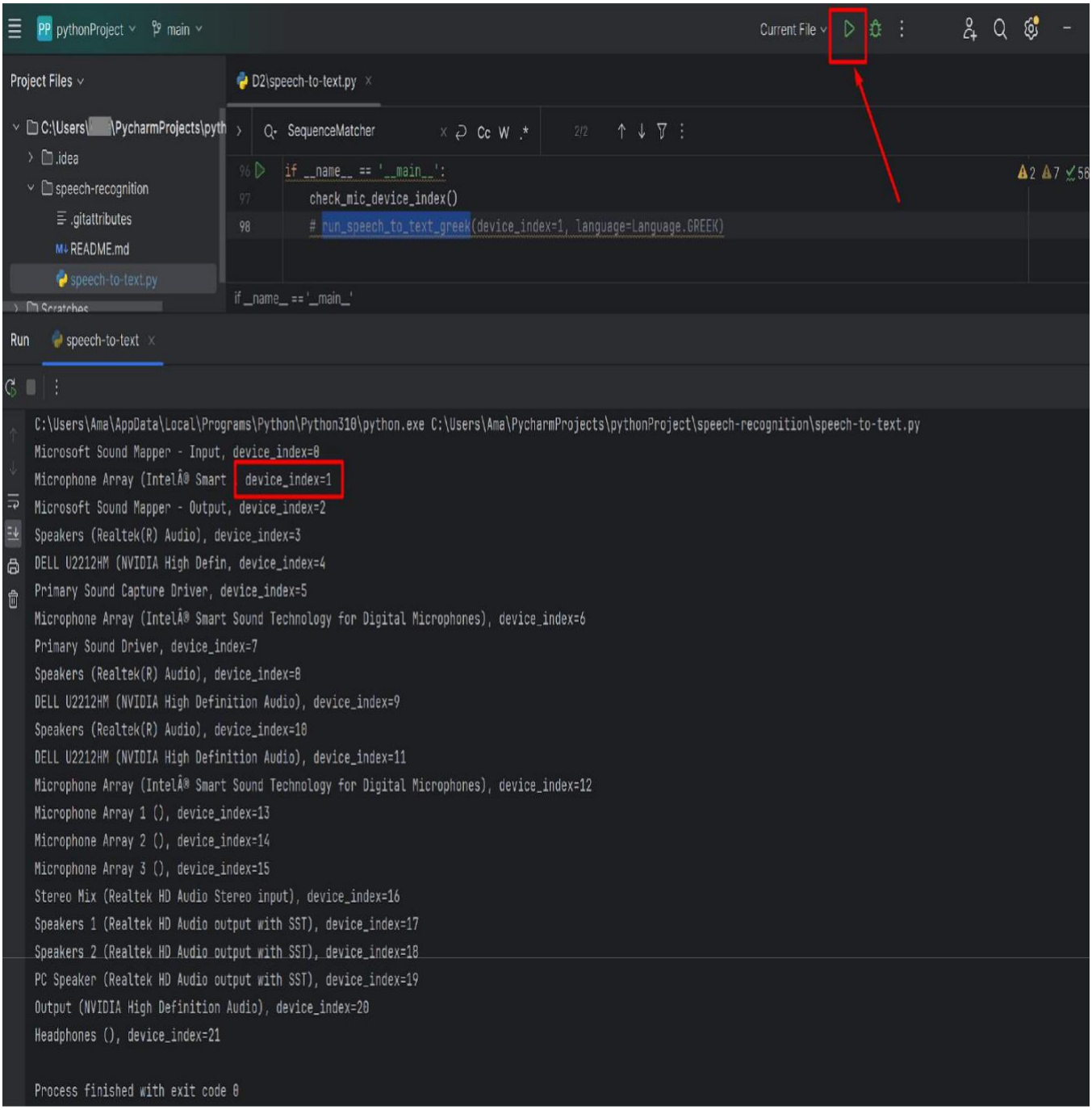


ΕΙΚΟΝΑ 18 [25]

- **Προετοιμασία Κώδικα**

Για να ξέρουμε ποια έξοδο θα χρησιμοποιήσει το πρόγραμμα για να λάβει και να παράγει ήχο θα πρέπει πρώτα να τρέξουμε την εντολή που βρίσκεται στην γραμμή 97, σχολιάζοντας την γραμμή 98 και πατώντας το ‘Run’:

Στο συγκεκριμένο σύστημα επιλέχθηκε η συσκευή με index = 1:



Θα πρέπει να σχολιαστεί και πάλι η γραμμή 97, και να αποσχολιαστεί η γραμμή 98, αφού έχει μπει το device_index που επιλέχθηκε στο προηγούμενο βήμα ως παράμετρος

```
84                  #     robot.move(sit)
85
86                  return True
87              return False
88
89      def check_mic_device_index():
90          # βοηθητική συνάρτηση για να βρούμε την συσκευή μας
91          SpeechToText.print_mic_device_index()
92
93      def run_speech_to_text_greek(device_index, language):
94          SpeechToText.speech_to_text(device_index, language)
95
96      if __name__ == '__main__':
97          # check_mic_device_index()
98          run_speech_to_text_greek(device_index=1, language=Language.GREEK)
```

Αφού επιλεγεί και η συσκευή, πατώντας και πάλι ‘Run’, το πρόγραμμα τρέχει.

Το πρόγραμμα θα περιμένει να ακούσει κάποια από τις εντολές που έχουν δηλωθεί στις γραμμές 17 – 22:

```
15      class SpeechToText:
16          # Εδώ δηλώνονται οι εντολές που περιμένει να ακούσει το ρομπότ
17          commands = [
18              "καθάρισε την κουζίνα",
19              "κάθισε κάτω",
20              "μαγείρεψε μακαρόνια",
21              "κι άλλη εντολή"
22          ]
```

Όταν το πρόγραμμα ξεκινήσει, θα γράψει στην κονσόλα κατάλληλο μήνυμα, θα κάνει έναν ήχο και θα περιμένει να ακούσει την εντολή. Σε περίπτωση που το πρόγραμμα ακούσει κάποια εντολή που δεν υπάρχει στην λίστα του, θα τυπώσει κατάλληλο μήνυμα και θα πει με λόγια ότι δεν αναγνωρίζει την εντολή:

```
Δώσε την εντολή μετά τον ήχο:
You said: είμαι ένα ρομπότ
Δεν αναγνωρίζω την εντολή: είμαι ένα ρομπότ

Process finished with exit code 0
```

Αν η εντολή βρίσκεται στην λίστα του, τότε θα τυπώσει μήνυμα επιτυχίας και θα πει με λόγια για την επόμενη κίνηση του ρομπότ

```
Δώσε την εντολή μετά τον ήχο:
You said: καθάρισε την κουζίνα
Match: καθάρισε την κουζίνα Similarity: 1.0
τέλος
```

Επειδή πολλές φορές ανάλογα με το περιβάλλον και το μικρόφωνο της εισόδου, μπορεί κάποιος να πει σωστά την εντολή, αλλά το σύστημα να την πιάσει λίγο παραλλαγμένη, έχει προστεθεί ένα threshold ομοιότητας της εντολής που άκουσε το σύστημα με τις εντολές που περιμένει. Πάνω από το threshold αυτό, η εντολή θεωρείται επιτυχημένη. Ακολουθεί ένα παράδειγμα:

```
Δώσε την εντολή μετά τον ήχο:
You said: άρεσε την κουζίνα
Match: καθάρισε την κουζίνα Similarity: 0.8648648648648649
τέλος
```

Το threshold αυτό έχει οριστεί σε ποσοστό ομοιότητας 70% (0.7) και μπορεί να αλλάξει από την γραμμή 73

```
def check_command(text):
    # Έλεγχος για το αν ταιριάζει η εντολή που κατάλαβε το ρομπότ με κάποια από τις προκαθορισμένες
    text = text.lower()
    for command in SpeechToText.commands:
        similarity = SequenceMatcher(None, text, command).ratio()
        if similarity >= 0.7:
            print("Match: " + command + " Similarity: " + str(similarity))
            SpeechToText.text_to_speech("Η επόμενη κίνηση που θα κάνει το ρομπότ είναι: " + command, languag

            # εδώ το ρομπότ θα κινηθεί με βάση την παραπάνω εντολή που πήρε (command)
            # για παράδειγμα:
            # if command == SpeechToText.commands[0]:
            #     print("πάω να καθαρίσω την κουζίνα")
```

Στις γραμμές 77 – 84 θα πρέπει να μπει ο κώδικας που θα εκτελείται μετά τον εντοπισμό της εντολής

```
77              # εδώ το ρομπότ θα κινηθεί με βάση την παραπάνω εντολή που πήρε (command)
78              # για παράδειγμα:
79              # if command == SpeechToText.commands[0]:
80              #     print("πάω να καθαρίσω την κουζίνα")
81              #     robot.clean(kitchen)
82              # elif command == SpeechToText.commands[1]:
83              #     print("κάθομαι κάτω")
84              #     robot.move(sit)
```

Παρακάτω παρατείθεται ολόκληρος ο κώδικας σε python που υλοποιήθηκε για την ανάπτυξη της εφαρμογής

```
from pydub import AudioSegment
import speech_recognition as sr
from enum import Enum
from gtts import gTTS
import os
from pydub.playback import play
import sounddevice as sd
import numpy as np
from difflib import SequenceMatcher

class Language(Enum):
ENGLISH = "en-US"
GREEK = "el-GR"

class SpeechToText:
# Εδώ δηλώνονται οι εντολές που περιμένει να ακούσει το
ρομπότ commands = [
"καθάρισε την κουζίνα",
"κάθισε κάτω",
"μαγείρεψε μακαρόνια",
"κι άλλη εντολή"
]

def beep(frequency, duration, volume=0.5, sample_rate=44100):
# η συνάρτηση για να κάνει το αρχικό "μπιπ"
t = np.linspace(0, duration, int(sample_rate * duration),
endpoint=False) wave = volume * np.sin(2 * np.pi * frequency * t)
sd.play(wave,
samplerate=sample_rate) sd.wait()

@staticmethod
def text_to_speech(text,
language='el'): tts = gTTS(text=text,
lang=language) filename = "temp.mp3"
tts.save(filename)
sound = AudioSegment.from_file(filename, format="mp3")
play(sound)
os.remove(filename)

@staticmethod
def print_mic_device_index():
for index, name in enumerate(sr.Microphone.list_microphone_names()):
print("{1}, device_index={0}".format(index, name))

@staticmethod
def speech_to_text(device_index, language=Language.ENGLISH):
r = sr.Recognizer()
with sr.Microphone(device_index=device_index) as source:
print("Δώσε την εντολή μετά τον ήχο: ")
SpeechToText.beep(1000, 0.5)
audio = r.listen(source)
try:
text = r.recognize_google(audio, language=language.value)
print("You said: {}".format(text))
if SpeechToText.check_command(text):
print("τέλος")
else:
print("Δεν αναγνωρίζω την εντολή: " + text)
SpeechToText.text_to_speech("Δεν αναγνωρίζω την εντολή: " + text, language='el')
except sr.UnknownValueError:
print("Speech Recognition could not understand audio")



SpeechToText.text_to_speech("Δεν κατάλαβα αυτό που μου είπες")
except sr.RequestError as e:
```

```
print("Could not request results from Speech Recognition service;
{0}".format(e))
SpeechToText.text_to_speech("Σφάλμα")

@staticmethod
def check_command(text):
# Ελεγχος για το αν ταιριάζει η εντολή που κατάλαβε το ρομπότ με κάποια από
τις προκαθορισμένες
text = text.lower()
for command in SpeechToText.commands:
similarity = SequenceMatcher(None, text,
command).ratio() if similarity >= 0.7:
print("Match: " + command + " Similarity: " + str(similarity))
SpeechToText.text_to_speech("Η επόμενη κίνηση που θα κάνει το ρομπότ είναι: "
+ command, language='el')

# εδώ το ρομπότ θα κινηθεί με βάση την παραπάνω εντολή που πήρε (command)
# για παράδειγμα:
# if command == SpeechToText.commands[0]:
#     print("πάω να καθαρίσω την κουζίνα")
#     robot.clean(kitchen)
# elif command == SpeechToText.commands[1]:
#     print("κάθομαι κάτω")
#     robot.move(sit)

return True
return False

def check_mic_device_index():
# βοηθητική συνάρτηση για να βρούμε την συσκευή μας
SpeechToText.print_mic_device_index()

def run_speech_to_text_greek(device_index, language):
SpeechToText.speech_to_text(device_index, language)

if __name__ == '__main__':
check_mic_device_index()
# run_speech_to_text_greek(device_index=1, language=Language.GREEK)+6359
```

ΕΙΚΟΝΑ 19 [26]

Κεφάλαιο 6

Συμπεράσματα & Προοπτικές

- ## Συμπεράσματα από την υλοποίηση και αξιολόγηση

Για το Project της εφαρμογής του ρομποτικού οικιακού βοηθού που λειτουργεί με το να ανιχνεύει και να αναγνωρίζει εντολές, δείχνοντας μια καλή κατανόηση της φωνητικής ή γραπτής αλληλεπίδρασης με τον χρήστη, είναι προγραμματισμένο να απαντά αναλόγως στις εντολές που λαμβάνει, επικοινωνώντας με τον χρήστη για να τον ενημερώσει για την επόμενη κίνησή του. Η χρήση των λειτουργιών Text-to-Speech / Speech-to-Text είναι απαραίτητη για να είναι αποτελεσματικός. Αυτό δείχνει μια σαφή και εστιασμένη προσέγγιση στην υλοποίηση του Project.

Η εφαρμογή αυτή είναι σχεδιασμένη για χρήση σε οικιακό περιβάλλον, πράγμα που σημαίνει ότι το ρομπότ πρέπει να είναι ικανό να αναγνωρίζει και να προσαρμόζεται σε διάφορες καθημερινές ανάγκες και δραστηριότητες.

Το Project φαίνεται να έχει τη δυνατότητα για μελλοντικές επεκτάσεις, όπως η προσθήκη νέων εντολών και η βελτίωση της αλληλεπίδρασης με τον χρήστη.

Για την υλοποίηση της εφαρμογής χρησιμοποιήθηκαν τεχνολογίες αναγνώρισης φωνής, φυσικής γλώσσας, και αυτοματισμού. Αυτές οι τεχνολογίες είναι ζωτικής σημασίας για τη δημιουργία ενός αποτελεσματικού και αποδοτικού ρομποτικού συστήματος.

Κατά την υλοποίηση και τη δοκιμή του συστήματος, προέκυψαν ορισμένα προβλήματα και περιορισμοί όπως η ασυμβατότητα μεταξύ του Pycharm και του υλικού, το οποίο επιλύνεται με την ενσωμάτωση του interpreter setuptools. Αυτό προκάλεσε και τις μεγαλύτερες καθυστερήσεις, καθώς χρειάστηκε πρόσθετο χρόνο για την επίλυση του.

- ## Προτάσεις για μελλοντική έρευνα και βελτιώσεις

1. Ανάπτυξη μεθόδων για τη βελτίωση της συλλογής, αποθήκευσης και επεξεργασίας δεδομένων, για τη διασφάλιση της ποιότητας και της αξιοπιστίας τους.
2. Αναβάθμιση των μέτρων ασφαλείας για την προστασία των δεδομένων και την διασφάλισης των χρηστών.

Η ενσωμάτωση νέων τεχνολογιών στο καθολικό ρομποτικό σύστημα μπορεί να προσδώσει σημαντικά πλεονεκτήματα στο σύστημα με την αξιοποίηση προηγμένων τεχνικών τεχνητής νοημοσύνης και μηχανικής μάθησης για την περαιτέρω βελτίωση της ανάλυσης και της πρόβλεψης, με την χρήση AR και VR για την παροχή πιο διαδραστικών και εντυπωσιακών εμπειριών στους χρήστες, με την ενσωμάτωση συσκευών IoT για τη συλλογή δεδομένων σε πραγματικό χρόνο και την αποτελεσματικότερη διαχείριση των πόρων.

Η χρήση τεχνολογίας blockchain και η ενσωμάτωση τραπεζικών πληροφοριών για την ασφαλή και διαφανή διαχείριση των δεδομένων και των συναλλαγών που ίσως χρειαστεί να γίνουν κατά την χρήση του οικιακού βοηθού (π.χ. αν προστεθεί η εντολή ‘παράγγειλε φαγητό, προϊόντα σούπερ μάρκετ κ.ά.).

Οι νέες εφαρμογές και σενάρια χρήσης του συστήματος μπορούν να περιλαμβάνουν:

- Επέκταση της χρήσης του συστήματος σε νέους τομείς, όπως η υγεία, η εκπαίδευση, και οι χρηματοοικονομικές υπηρεσίες.
- Προσαρμογή του συστήματος σε διαφορετικά γεωγραφικά και πολιτιστικά περιβάλλοντα για την καλύτερη κάλυψη των τοπικών αναγκών.
- Εξερεύνηση νέων σεναρίων χρήσης, όπως η βελτίωση της εμπειρίας των πελατών, η διαχείριση των πόρων, και η υποστήριξη στη λήψη αποφάσεων.

Με την εφαρμογή αυτών των προτάσεων, το σύστημα μπορεί να γίνει ακόμα πιο αποδοτικό, ασφαλές και προσαρμοστικό στις ανάγκες των χρηστών, προσφέροντας νέα επίπεδα αξίας και καινοτομίας.

Συνολικά, το Project φαίνεται να έχει υλοποιήσει επιτυχώς ένα βασικό σύστημα οικιακού βοηθού, το οποίο είναι ικανό να αναγνωρίζει και να απαντά σε προκαθορισμένες εντολές, υποδεικνύοντας τις επόμενες ενέργειές του και παρέχοντας έτσι μια χρήσιμη και λειτουργική υποστήριξη στο οικιακό περιβάλλον.

Οι υπόλοιπες λειτουργίες του οικιακού βοηθού αναμένεται να δημοσιοποιηθούν από μελλοντική ερευνητική εργασία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. “FFmpeg.” Accessed: Jun. 16, 2024. [Online]. Available: https://ffmpeg.org/download.html
2. “Pycharm IDE.” Accessed: Jun. 16, 2024. [Online]. Available: https://www.jetbrains.com/pycharm/download/?section=windows
3. Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing*. Pearson.
4. Rabiner, L., & Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall.
5. Google's official documentation for Speech Recognition and Text-to-Speech APIs.
6. Siciliano, B., & Khatib, O. (2016). *Springer Handbook of Robotics*. Springer.
7. Murphy, R. R. (2019). *Introduction to AI Robotics*. MIT Press.
8. Google Scholar
9. IEEE Xplore Digital Library
10. Speechify.com
11. Python.com
12. https://www.apple.com/siri/
13. https://elitr.eu/technologies/
14. https://www.linkedin.com/pulse/introduction-openai-text-to-speech-tts-technology- roman- kulibaba-knkef
15. https://www.quora.com/How-can-I-build-AI-voice-models-on-my-computer-rather-than- using-online-services-I-want-to-use-voice-to-voice-not-text-to-voice
16. https://stringeex.com/en/blog/post/Exploring-The-Text-To-Speech-Definition-And-Its-Impact-Across-Industries
17. https://dictate.com.au/blogs/news/voice-recognition-options-for-mac-windows-google-vs-microsoft-vs-apple-vs-dragon
18. https://murf.ai/speech-to-text
19. https://www.egemen.no/how-to-install-python-via-homebrew-on-mac/
20. https://www.pexels.com/photo/bionic-hand-and-human-hand-finger-pointing-6153354/
21. https://www.researchgate.net/figure/Flow-chat-of-Google-speech-API-converting-speech-to-text_fig3_331451704
22. https://www.onmed.gr/ygeia-eidhseis/story/352068/protoporiako-rompot-oikiakos-voithos-gia-ilikiomenoys-me-noitiki-diataraxi
23. https://www.vidnoz.com/ai-solutions/text-to-speech-extension.html
24. https://www.scorchsoft.com/blog/text-to-mic-for-meetings
25. https://cloud.google.com/text-to-speech
26. https://gr.newtechstore.eu