

Ενσωματωμένα Συστήματα  
Τμήμα Μηχανικών Η/Υ και  
Πληροφορικής

---

Personal  
Digital  
Assistant(PDA)

Εργασία Εξαμήνου

---

**Για περισσότερες πληροφορίες παρακαλώ επικοινωνήστε μαζί μας:**

- **Βασιλακοπούλου Χαρίκλεια-Ειρήνη**

AM: 1059714

Φοιτήτρια

Εμμανουήλ Παππά 30

26332 Αχαΐα

Ελλάδα

Τηλέφωνο: +30 6947810539

e-mail: xara160699@gmail.com

---

- **Γκίζας Μαρίνος**

AM: 1054348

Φοιτητής

Μεγάλου Αλεξάνδρου 127

26332 Αχαΐα

Ελλάδα

Τηλέφωνο: +30 6984429072

e-mail: mgizas13@gmail.com

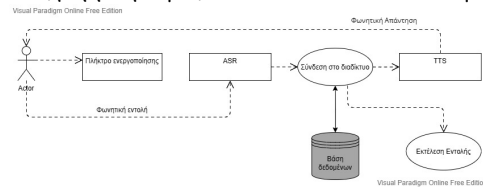
## Περιεχόμενα

1	Στάδιο πρωτο .....	4
1.1	Περιγραφή εφαρμογής, στόχων .....	4
1.2	Απαιτήσεις χρήστη και λειτουργιών .....	4
1.3	Προδιαγραφές χρήστη και λειτουργιών .....	5
1.4	Προσέγγιση αρχιτεκτονικής λογισμικού .....	5
1.5	Προσέγγιση αρχιτεκτονικής υλικού .....	6
2	Στάδιο Β.....	7
2.1	Αρχιτεκτονική Υλικού .....	7
2.2	Αρχιτεκτονική Λογισμικού .....	7
2.3	Σχεδιασμός Συνολικού Συστήματος .....	8
2.4	Σχεδιασμός Υποσυστημάτων .....	13
2.5	Σχεδιασμός Συστατικών Συστήματος/Υποσυστημάτων .....	14
3	Στάδιο Γ.....	16
3.1	Υλοποίηση (ή θεωρητική προσέγγιση) .....	16
3.2	Πειραματικές μετρήσεις (ή εκτίμηση) .....	17
3.3	Διαδικασίες ελέγχου ορθής λειτουργίας .....	17
3.4	Βελτιστοποιήσεις (ή προτάσεις) .....	17
4	Βιβλιογραφία .....	19

## 1 Στάδιο πρωτο

### 1.1 Περιγραφή εφαρμογής, στόχων

Το Personal Digital Assistant (από εδώ και στο εξής θα αναφερόμαστε σε αυτό ως PDA), είναι ένας πολύ μικρός και φορητός υπολογιστής, ο οποίος έχει παραπάνω λειτουργίες από ένα κλασσικό κομπιουτεράκι και η χωρητικότητα της πληροφορίας που μπορεί να αποθηκεύσει είναι ανάλογη με αυτή ενός Personal Computer(PC)[1]. Στόχος του PDA είναι η διευκόλυνση του χρήστη, σε επαγγελματικό αλλά και προσωπικό επίπεδο, μέσω των πολλών λειτουργιών που προσφέρει αφού αυτές γίνονται πιο γρήγορα, πιο εύκολα και πιο αποδοτικά. Θα επικεντρωθούμε σε PDA με φωνητικές εντολές ( ή αλλιώς Echo) όπως οι Siri, Google Home, Alexa. Ένα διάγραμμα ροής για το PDA είναι το εξής:



### 1.2 Απαιτήσεις χρήστη και λειτουργιών

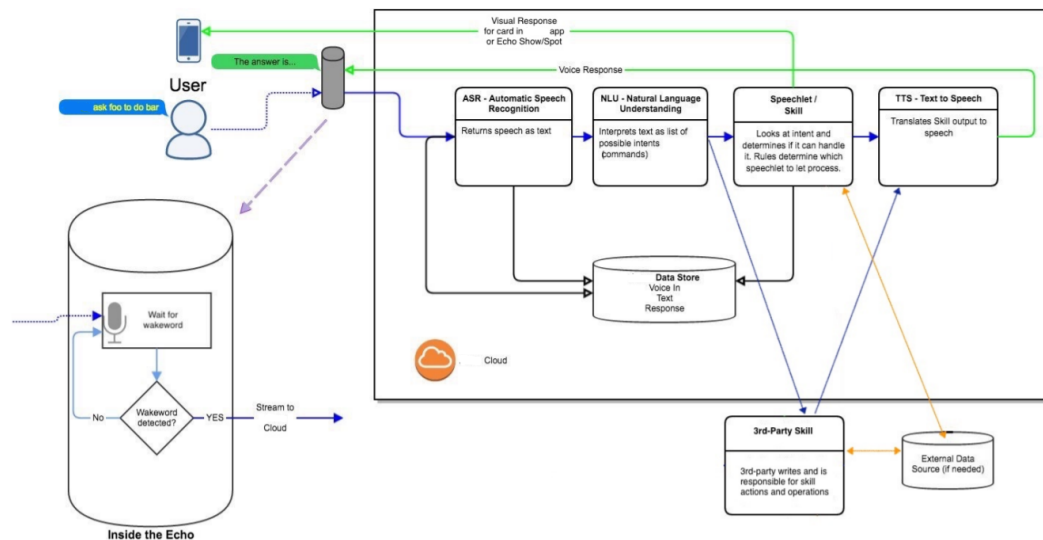
Όνομα	Personal digital assistant με φωνητικές εντολές
Σκοπός	Συσκευή προσωπικού βοηθού με φωνητικές εντολές για καθημερινή χρήση
Είσοδοι	Ένα πλήκτρο ενεργοποίησης και δύο πλήκτρα ρύθμισης έντασης
Έξοδοι	Ηχείο μέχρι 81dBA
Λειτουργίες	Απαντάει στις ερωτήσεις του χρήστη και εκτελεί εντολές
Απόδοση	Οι φωνητικές εντολές έχουν διάρκεια αναμονής μέχρι την απάντηση 8 δευτερόλεπτα
Κόστος κατασκευής	20€
Ισχύς	3W ανά ημέρα
Φυσικό μέγεθος και βάρος	Όχι μεγαλύτερο από 350 gr

### 1.3 Προδιαγραφές χρήστη και λειτουργιών

Το PDA πρέπει να περιλαμβάνει:

1. Μηχανή αναγνώρισης λέξεων και φωνής
2. Λειτουργίες για εκτέλεση των εντολών αυτών
3. Λειτουργίες για απάντηση στον χρήστη
4. Ηχεία για απάντηση

### 1.4 Προσέγγιση αρχιτεκτονικής λογισμικού



[2]

Figure: Overview of Echo and the System



## 2 Στάδιο Β

### 2.1 Αρχιτεκτονική Υλικού

Το PDA αποτελείται από 6 μικρόφωνα, 1 CPU, 1 RAM, 1 Flash Memory, 1 LED, 1 Wifi, 1 Bluetooth, 1 ASR(Automatic Speech Recognition), 1 TTS(Text-to-Speech) και 1 ηχείο. Αρχικά, τα 6 μικρόφωνα χρησιμοποιούνται για να μετατρέψουμε τα αναλογικά σήματα από την φωνή μας σε ψηφιακά σήματα. Έπειτα, η CPU χρησιμοποιείται για διάφορες διεργασίες όπως η αναγνώριση της λέξης αφύπνισης και η συμπίεση/αποσυμπίεση δεδομένων όπως των ηλεκτρικών σημάτων [3]. Με το ASR εντοπίζονται οι ήχοι της φυσικής γλώσσας και αναγνωρίζονται σε λέξεις [4]. Το Wifi χρησιμοποιείται για την σύνδεση του PDA με το διαδίκτυο και την ανάλογη βάση δεδομένων όπου θα στέλνονται οι λέξεις που αναγνωρίζει το ASR και από όπου θα επιστρέφεται η κατάλληλη απάντηση ενώ το Bluetooth για την σύνδεση με κινητό ή και άλλες συσκευές. Το TTS είναι μια τεχνολογία η οποία διαβάζει ένα κείμενο από έναν υπολογιστή (στην περίπτωσή μας από την βάση δεδομένων) και επιστρέφει στον χρήστη ως φωνή σε φυσική γλώσσα[5]. Τέλος, η μνήμη RAM χρησιμοποιείται για να κάνει buffering τα δεδομένα απο την βάση δεδομένων και από την CPU, η Flash Memory για αποθήκευση για μεταφορά δεδομένων, ενώ το ηχείο χρησιμοποιείται για την μετατροπή των ψηφιακών σημάτων σε αναλογικά. Το LED χρησιμοποιείται για την οπτική επικοινωνία με τον χρήστη.

### 2.2 Αρχιτεκτονική Λογισμικού

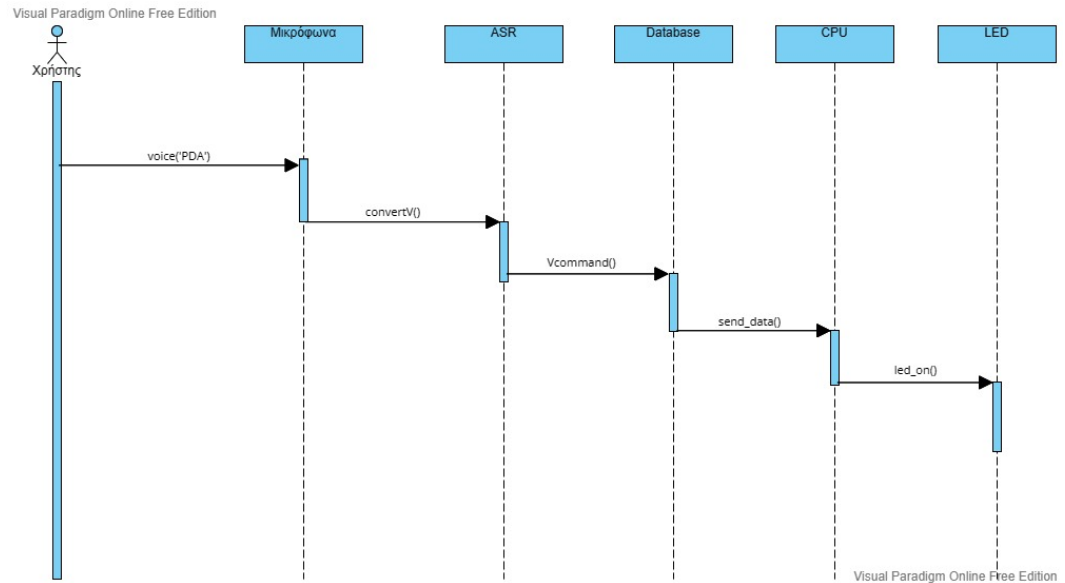
Η αρχιτεκτονική λογισμικού του PDA με ηχητικές εντολές θα βασιστεί στα παρακάτω use cases.

1. Η αφύπνισή του θα γίνεται μέσω της λέξης "PDA" όπου θα ανάβει το LED.
2. Το PDA λαμβάνει μια ερώτηση από τον χρήστη. Στην συνέχεια, επεξεργάζεται την ερώτηση και κάνει αναζήτηση της απάντησης στην βάση δεδομένων της. Τέλος, επιστρέφει στον χρήστη την απάντηση που πρέπει.
3. Το PDA με την εντολή "Set alarm at ..." θα θέτει το ξυπνητήρι στην ώρα που του ζητήθηκε. Στην συνέχεια, θα ρωτάει "For which day?" και ανάλογα την απάντησή μας θα προγραμματίζει και το ξυπνητήρι. Τέλος, θα επιστρέφει μήνυμα επιβεβαίωσης για την εντολή που εκτελέστηκε.
4. Το PDA με την εντολή "Cancel the alarm at ... on ..." θα ακυρώνει το ξυπνητήρι που έχει οριστεί για την ώρα και την ημέρα που του ζητήθηκε. Στην συνέχεια, θα επιστρέφει μήνυμα επιβεβαίωσης για την εντολή που εκτελέστηκε.
5. Το PDA με την εντολή "Add ... to my notes" θα προσθέτει στις σημειώσεις αυτό που του ζητήθηκε από τον χρήστη. Στην συνέχεια, θα επιστρέφει μήνυμα επιβεβαίωσης για την εντολή που εκτελέστηκε.
6. Το PDA με την εντολή "Remove the ... from my notes" θα αφαιρεί την σημείωση που ζητήθηκε από τον χρήστη. Στην συνέχεια, θα επιστρέφει μήνυμα επιβεβαίωσης για την εντολή που εκτελέστηκε.
7. Το PDA με την εντολή "Play the song ... from my library" θα κάνει αναπαραγωγή του τραγουδιού που του ζητήθηκε από την βιβλιοθήκη του χρήστη. Αν ο χρήστης θέλει να διακόψει την ακρόαση του τραγουδιού, μπορεί να χρησιμοποιήσει την εντολή "Stop".

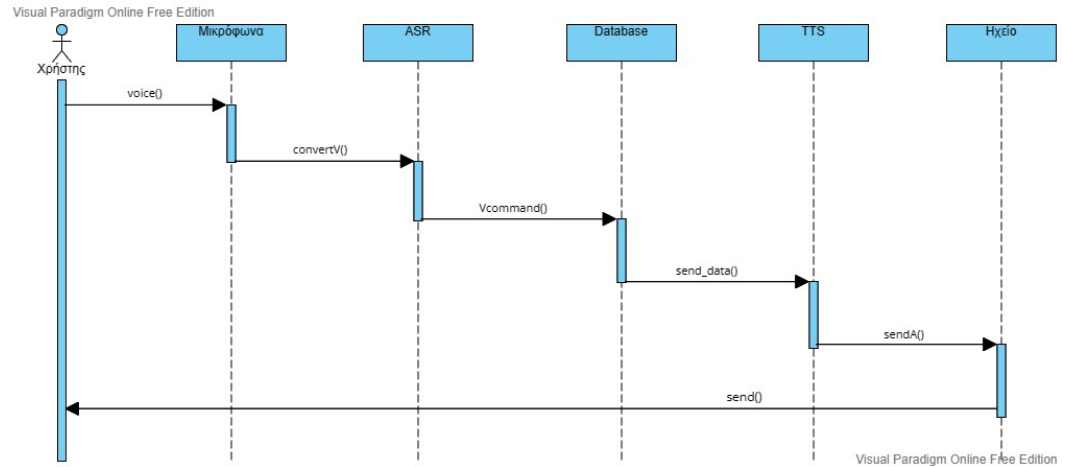
8. Το PDA με την εντολή "Go to sleep" θα μπαίνει σε κατάσταση αδράνειας, σβήνοντας το LED.
9. Ο χρήστης πατάει παρατεταμένα το κουμπί ενεργοποίησης του PDA, όπου ενεργοποιεί το Bluetooth συνδέεται με την επιθυμητή συσκευή

### 2.3 Σχεδιασμός Συνολικού Συστήματος

#### 1. Sequence diagram για use case 1

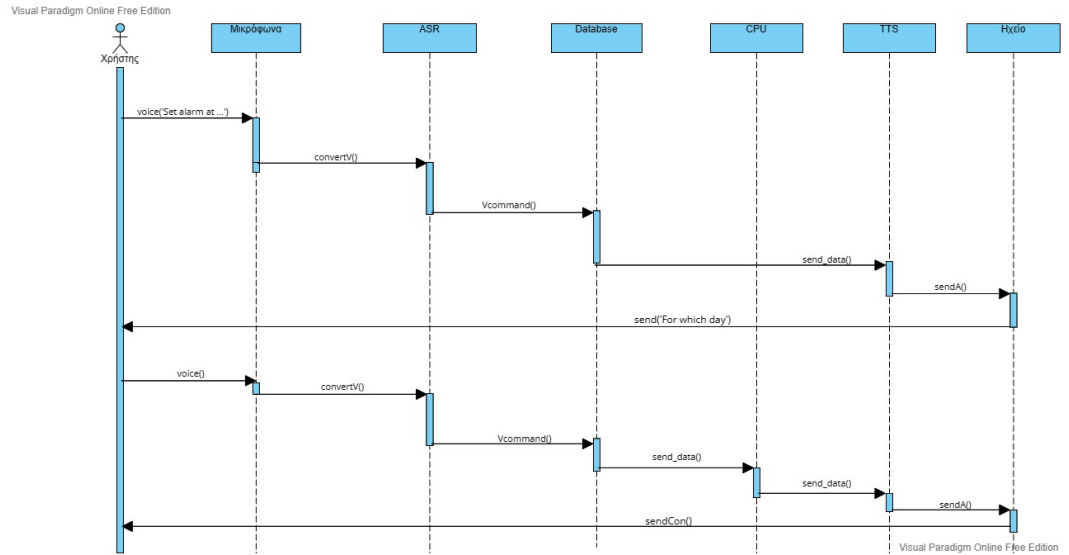


#### 2. Sequence diagram για use case 2

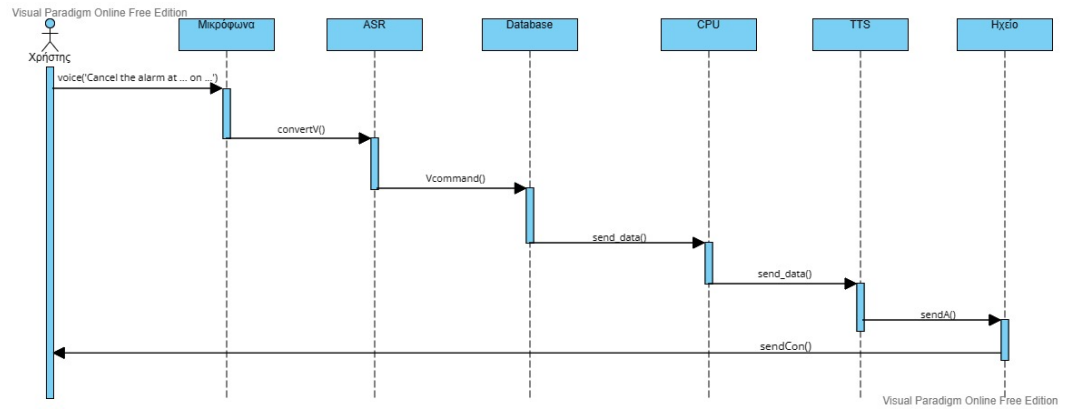




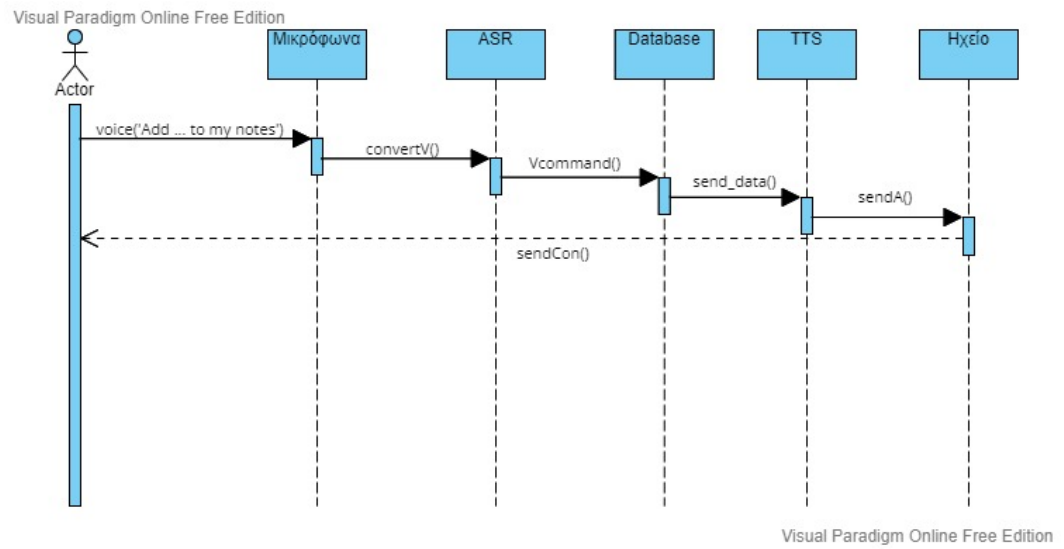
### 3. Sequence diagram για use case 3



### 4. Sequence diagram για use case 4

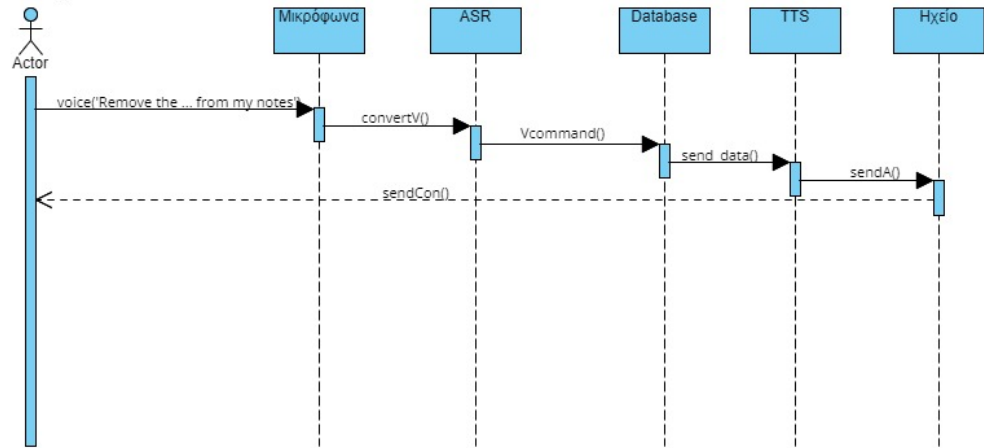


## 5. Sequence diagram για use case 5



6. Sequence diagram για use case 6

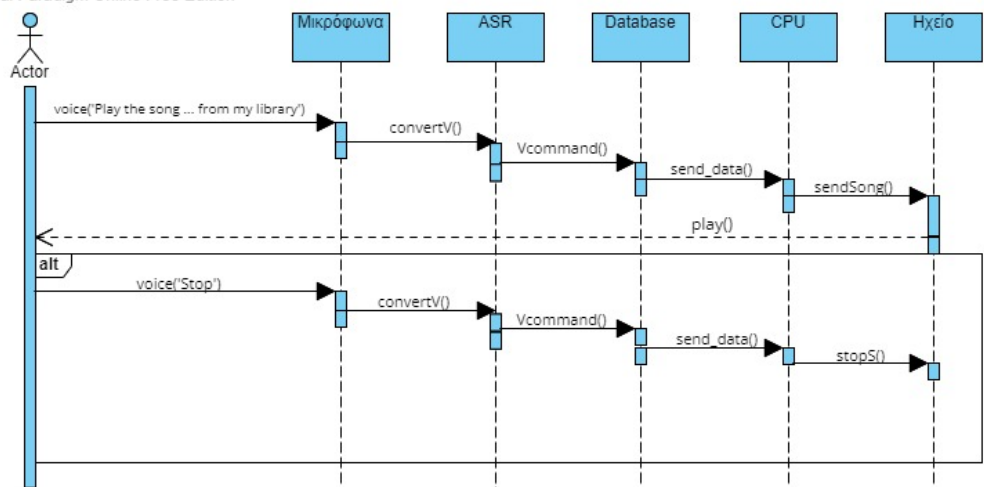
Visual Paradigm Online Free Edition



Visual Paradigm Online Free Edition

7. Sequence diagram για use case 7

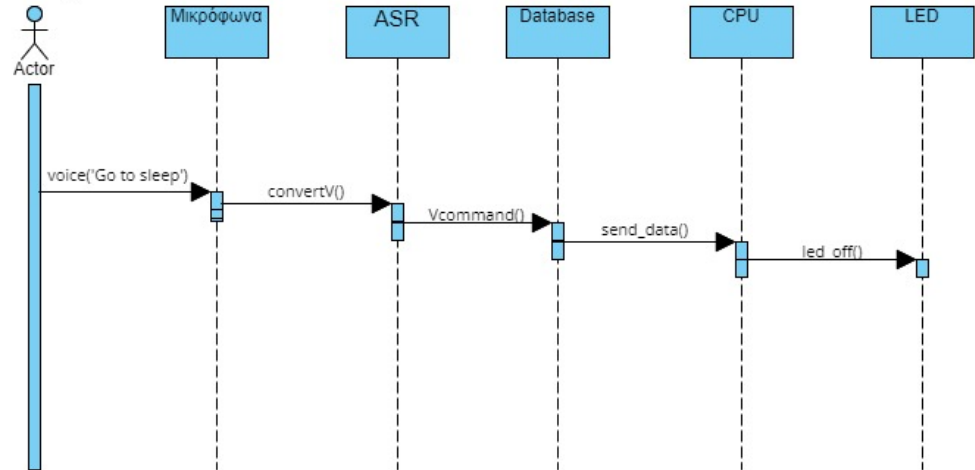
Visual Paradigm Online Free Edition



Visual Paradigm Online Free Edition

## 8. Sequence diagram για use case 8

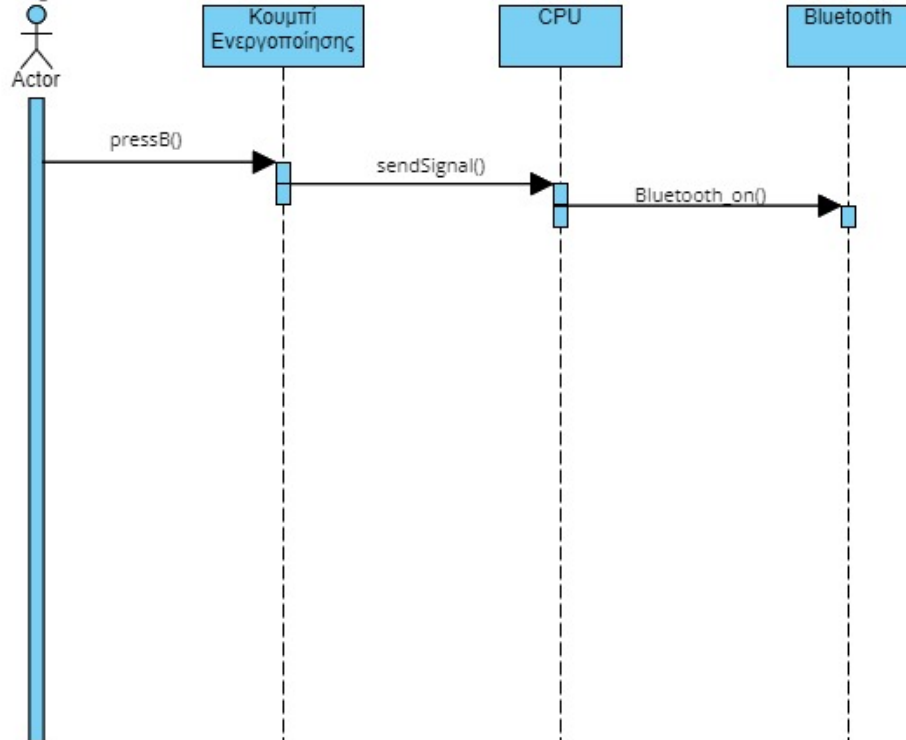
Visual Paradigm Online Free Edition



Visual Paradigm Online Free Edition

## 9. Sequence diagram για use case 9

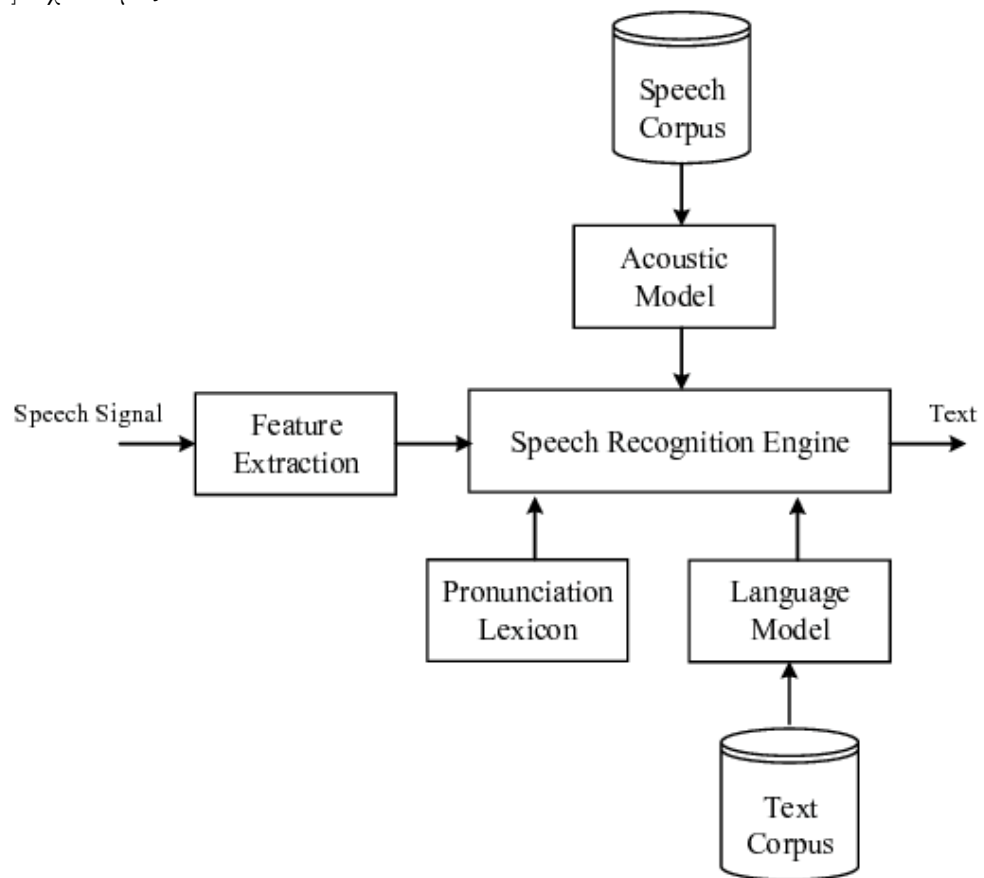
Visual Paradigm Online Free Edition



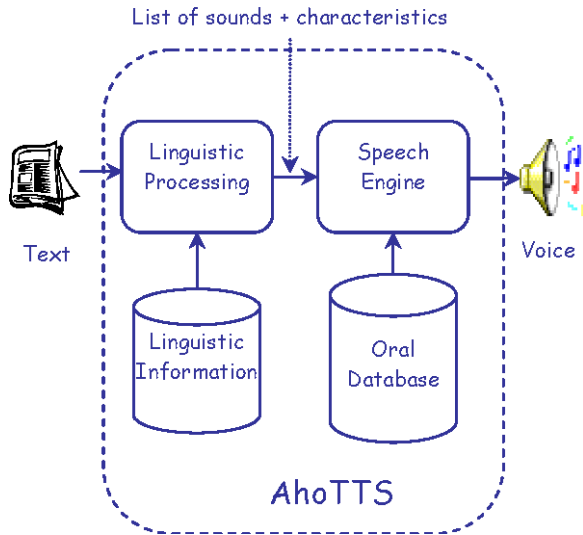
Visual Paradigm Online Free Edition

## 2.4 Σχεδιασμός Υποσυστημάτων

[6] Σχεδιασμός ASR



## [7] Σχεδιασμός TTS

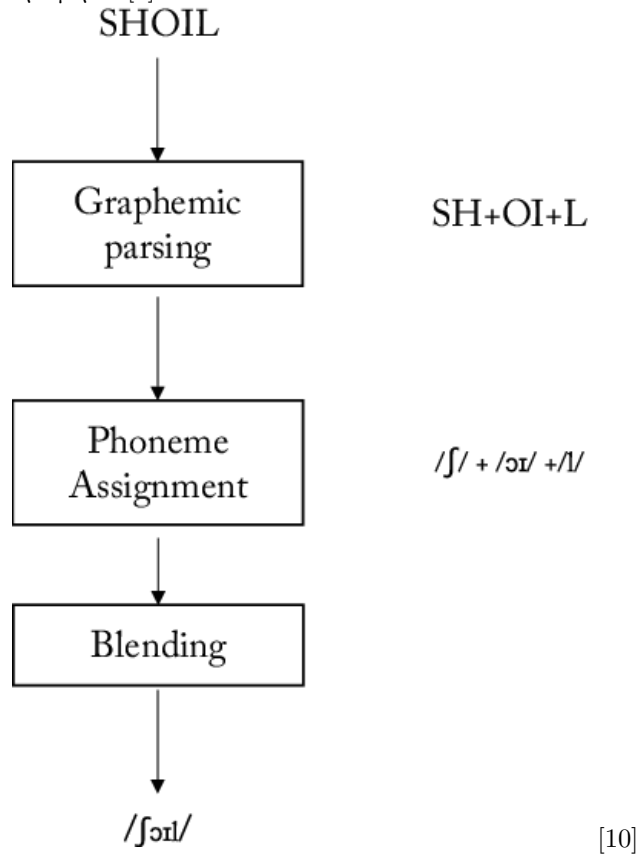


## 2.5 Σχεδιασμός Συστατικών Συστήματος/Υποσυστημάτων

Το υποσύστημα ASR αποτελείται από 5 κύρια συστατικά, όπως φαίνεται και στο παραπάνω σχήμα:

1. Η **Εξαγωγή Χαρακτηριστικών** είναι το πρώτο βήμα του ASR, το οποίο μετατρέπει την κυματομορφή του σήματος της ομιλίας, σε μία ομάδα από χαρακτηριστικά διανύσματα. Ο κύριος στόχος είναι να κάνει αυτά τα διανύσματα να έχουν μεγάλη διαφοροποίηση μεταξύ των διανυσμάτων. [8]
2. Η **Μοντελοποίηση γλώσσας** έχει ως κύριο στόχο την εκτίμηση της κατανομής, με όση μεγαλύτερη ακρίβεια γίνεται, της φυσικής γλώσσας. Η βασική ιδέα είναι ότι η n-οστή λέξη μπορεί να εκτιμηθεί από έναν συνδυασμό των προηγούμενων n-1 λέξεων. Η μοντελοποίηση αυτή μπορεί να περιγραφεί μαθηματικά από το  $P(W)$  όπου  $W=w_1, w_2, \dots, w_k$  είναι μια ακολουθία από λέξεις. [6]
3. Η **Ακουστική μοντελοποίηση** είναι ένα από τα βασικά συστατικά του ASR συστήματος αφού οι περισσότεροι υπολογισμοί και η φόρτωση της απόδοσης γίνονται σε αυτό το μοντέλο. Ακόμα, σε αυτό το μοντέλο ονομάζουμε φώνημα την ευθυγράμμιση μεταξύ του ακουστικού χαρακτηριστικού διανύσματος και των ηχητικών μονάδων. [6] Ουσιαστικά, έρχεται αντιμέτωπο με ηχητικά κύματα της ανθρώπινης γλώσσας και κάνει την αντιστοίχιση κάθε ηχητικού κύματος σε ένα φώνημα. [9]
4. Το **Λεξικό προφοράς** χρησιμοποιεί το Myanmar λεξικό για την προφορά των λέξεων. Η προφορά των νέων γλωσσών παράγονται από το G2P (Grapheme to phoneme) σύστημα, όπου η σχηματική απεικόνισή του φαίνεται παρακάτω, χρησιμοποιώντας το εργαλείο Phonetisaurus. Το λεξιλόγιο αυτού του λεξικού προφοράς περιέχει πάνω από 39.000 λέξεις και χρησιμοποιούνται σε αυτές

108 διαφορετικές φωνητικές μονάδες(a.,k,kh,a-) για την αναπαράσταση της προφοράς.[6]



5. Η **Μηχανή αποκωδικοποίησης** αποτελείται από έναν κωδικοποιητή-αποκωδικοποιητή, ο οποίος αντιστοιχίζει τις ακολουθίες εισόδου που είναι ηχητικές με ακολουθίες χαρακτήρων.[11]

Το υποσύστημα TTS αποτελείται από 2 κύρια συστατικά, όπως φαίνεται και στο παραπάνω σχήμα:

1. Στην **Γλωσσική Επεξεργασία** εισάγεται κείμενο και μετατρέπεται σε αναπαράσταση χαρακτήρων από τους οποίους μπορεί να προχωρήσει η πραγματική σύνθεση του ήχου. [12]
2. Η **Μηχανή ομιλίας** έχει τα ίδια συστατικά με την μηχανή αποκωδικοποίησης του ASR, αλλά,σε αντίθεση με το ASR, μετατρέπει τις ακολουθίες χαρακτήρων σε ηχητικές ακολουθίες.

### 3 Στάδιο Γ

#### 3.1 Υλοποίηση (ή θεωρητική προσέγγιση)

- **CPU**[13]: ARM Cortex-A8 processor  
Frequency: from 600 MHz to 1 GHz  
Architecture: Superscalar dual-issue microarchitecture  
Pipeline: 13-stage integer pipeline and 10-stage NEON pipeline  
Cache: Integrated level 2 Cache (0–4 MiB)
- **RAM** [14]: 256MB of LPDDR2  
Memory array clock (MHz): 200  
I/O bus clock frequency (MHz): 400  
Data transfer rate (DDR) (MT/s): 800  
Supply voltage: 1.2v  
Command/Address bus: 10 bits, DDR
- **Flash Memory** [15]: SanDisk SDIN7DP2-4G 4 GB iNAND Ultra  
Memory Size: 4 GB  
Configuration: SLC  
Sustained Read: 90 MB/s  
Sustained Write: 12 MB/s  
Operating Supply Voltage: 2.7 V to 3.6 V  
Minimum Operating Temperature: - 25 C  
Maximum Operating Temperature: + 85 C
- **WiFi** [16]: Wi-Fi 802.11a/b/g/n  
Max speed/radio: 54 Mbps  
Frequency (GHz): 5  
Bandwidth: 5 MHz
- **Bluetooth** [17]: Bluetooth Advanced Audio Distribution Profile (A2DP)  
Το συγκεκριμένο bluetooth μας δίνει την δυνατότητα να συνδεθούμε με άλλες συσκευές και ταυτόχρονα με το Advanced Audio Distribution Profile μας επιτρέπει να αναπαράξουμε τον ήχο μέσω ηχείων όπως αυτά στο PDA.
- **Microphone**[18]: MSM261D4030H1AP  
Supply Voltage: −0.3V to 4.0V  
Sound Pressure Level: 140 dB SPL  
Temperature Range: −40°C to 100°C  
Electrostatic discharge protection: 2 (HBM) kV  
Fall-asleep Time: 30 μs  
Wake-up Time: 200μs
- **Speaker**[19]: JBL Link 300  
Output power (W) : 2 x 25



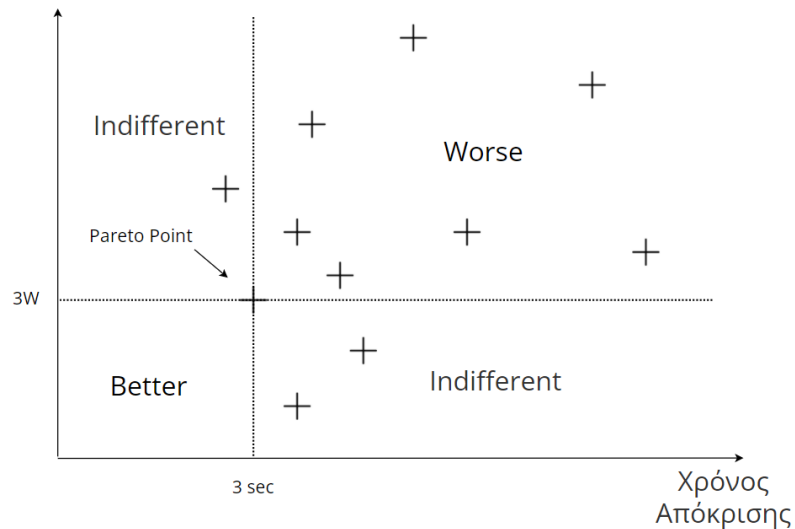
### 3.2 Πειραματικές μετρήσεις (ή εκτίμηση)

Με βάση το υλικό που χρησιμοποιήθηκε μπορούμε να κάνουμε κάποιες εκτιμήσεις για την απόδοση του συστήματός μας. Αρχικά, ο χρόνος ανταπόκρισης μιας ερώτησης/εντολής εκτιμάται ότι θα είναι περίπου 3-5 δευτερόλεπτα. Ο χρόνος μέχρι να υπάρξει η αφύπνιση του μικροφώνου για να δεχθεί την εντολή είναι 200-400μs, ενώ ο χρόνος για να μπει σε λειτουργία ύπνου είναι 30-100μs. Επιπρόσθετα, ο χρόνος αναμονής για της σύνδεσης της συσκευής με μία άλλη μέσω Bluetooth είναι 2-5sec. Τέλος, η εκτιμώμενη κατανάλωση ενέργειας του συστήματος είναι 3-4W ανά ημέρα.

### 3.3 Διαδικασίες ελέγχου ορθής λειτουργίας

- Για να ελέγξουμε αν το σύστημά μας λειτουργεί ορθά ή όχι, θα κάνουμε κάποια ερώτηση ή θα δώσουμε κάποια εντολή στο PDA και θα μετρήσουμε τον χρόνο ανταπόκρισης, καθώς και την ορθότητα της απάντησης που μας έδωσε. Αν για κάποιες ερωτήσεις/εντολές, το σύστημα μάς δώσει τις ίδιες απαντήσεις, τότε θα συμπεράνουμε ότι υπάρχει κάποιο σφάλμα τύπου stuck-at και χρειάζεται να εξετάσουμε από πού προέκυψε το σφάλμα.
- Σχηματικό διάγραμμα: Pareto Point

Κατανάλωση ενέργειας



### 3.4 Βελτιστοποιήσεις (ή προτάσεις)

Κάποιες προτάσεις για βελτιστοποίηση του PDA είναι οι εξής:

1. Γρηγορότερη ανταπόκριση της συσκευής με χρήση ταχύτερης CPU και ταχύτερης μνήμης (RAM και Flash Memory).

2. Ελαχιστοποίηση της καταλάνωσης ενέργειας με την επιλογή του κατάλληλου υλικού.
3. Προσθήκη νέων λειτουργιών, όπως να καλεί επαφές από το κινητό μας ή να μπορεί να συνδέεται και με άλλες έξυπνες συσκευές (όπως ψυγεία, τηλεοράσεις, ήχεια, συσκευές φωτισμού και καθαριότητας) και να τις χειρίζεται, ώστε να επιτευχθεί μεγαλύτερο φάσμα χρήσης.
4. Μεγαλύτερο εύρος αναγνώρισης λέξεων, για να μπορέσει το σύστημα να δώσει την ακριβέστερη απάντηση στον χρήστη.
5. Κατάλληλη τροποποίηση του συστήματος, ώστε να μπορεί να "μαθαίνει" τις προτιμήσεις του κάθε χρήστη και να προσαρμόζεται ανάλογα (Artificial Intelligence).

## 4 Βιβλιογραφία

### References

1. A. Lindquist, P. Johansson, G. Petersson, B.-I. Saveman, and G. Nilsson, "The use of the personal digital assistant (pda) among personnel and students in health care: a review," *Journal of medical Internet research*, vol. 10, no. 4, p. e1038, 2008.
2. J. Leblang, "White Paper - Alexa Privacy and Data Handling Overview." <https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf>. Created: 21/7/18.
3. B. Mars, "Sizing up cpu, memory, and storage for your alexa built-in device." <https://developer.amazon.com/blogs/alexa/post/2a32d792-d471-4136-8262-79962a2b4d72/cpu-memory-and-storage-for-alexa-built-in-devices>. Created: 13/9/18.
4. Amazon, "Alexa by amazon." <https://developer.amazon.com/en-US/alexa/alexa-skills-kit/asr>.
5. U. Team, "What is text-to-speech technology (tts)?" <https://www.understood.org/articles/en/text-to-speech-technology-what-it-is-and-how-it-works>.
6. H. M. S. Naing and W. P. Pa, "Automatic speech recognition on spontaneous interview speech," in *Sixteenth International Conferences on Computer Applications (ICCA 2018)*, Yangon, Myanmar, pp. 203–208, 2018.
7. J. Sanchez, I. Luengo, E. Navas, and I. Hernaez, "Adaptation of the ahotts text to speech system to pda platforms," *Proceedings of the SPECOM 2006*, pp. 292–296, 2006.
8. W. Wang, "Feature extraction for asr: Intro." <http://wantee.github.io/2015/03/14/feature-extraction-for-asr-intro/>, 2015.
9. R. Team, "What is an acoustic model in speech recognition?" <https://www.rev.com/blog/resources/what-is-an-acoustic-model-in-speech-recognition>.
10. M. Coltheart and A. Ulicheva, "Why is nonword reading so variable in adult skilled readers?," *PeerJ*, vol. 6, p. e4879, 2018.
11. X. Zhou, G. Lee, E. Yilmaz, Y. Long, J. Liang, and H. Li, "Self-and-mixed attention decoder with deep acoustic structure for transformer-based lvcsr," *arXiv preprint arXiv:2006.10407*, 2020.
12. R. Sproat, "Linguistic processing for speech synthesis," *Springer Handbook of Speech Processing*, pp. 457–470, 2008.
13. "Arm cortex-a8." [https://en.wikipedia.org/wiki/ARM\\_Cortex-A8](https://en.wikipedia.org/wiki/ARM_Cortex-A8).
14. "Lpddr." <https://en.wikipedia.org/wiki/LPDDR>.
15. M. Elecetronics, "Sdin7dp2-4g?qs=EgF7oUuTQmpAPyQtQI49Xw%3D%3D." <https://eu.mouser.com/ProductDetail/SanDisk/SDIN7DP2-4G?qs=EgF7oUuTQmpAPyQtQI49Xw%3D%3D>.
16. "Ieee 802.11." [https://en.wikipedia.org/wiki/IEEE\\_802.11](https://en.wikipedia.org/wiki/IEEE_802.11).
17. "List of bluetooth profiles." [https://en.wikipedia.org/wiki/List\\_of\\_Bluetooth\\_profiles#Advanced\\_Audio\\_Distribution\\_Profile\\_\(A2DP\)](https://en.wikipedia.org/wiki/List_of_Bluetooth_profiles#Advanced_Audio_Distribution_Profile_(A2DP)).
18. Dell, "Microsoft word - msm261d4030hlap v1.3-eng." <https://www.memsensing.com/webupfile/1510023505867767780.3-ENG.pdf>.
19. JBL, "Jbl link 300." <https://www.jbl.com.sg/home-audio/JBL+LINK+300.htmlf>.