

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Πτυχιακή Εργασία

**Υποβοήθηση χρηστών με προβλήματα όρασης για πλοήγηση
σε ηλεκτρονικό κατάστημα μέσω φωνητικών εντολών**

Φοιτητής: Χριστόδουλος Καλογιάννης Π2017077

Επιβλέπων: Παναγιώτης Κουρουθανάσης

Θεσσαλονίκη, Σεπτέμβριος 2021

Επιβλέπων

Παναγιώτης Κουρουθανάσης, *Αναπληρωτής Καθηγητής,*

Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Παναγιώτης Κουρουθανάσης, *Αναπληρωτής Καθηγητής,*

Ιόνιο Πανεπιστήμιο

Θεμιστοκλής Έξαρχος, *Επίκουρος Καθηγητής,*

Ιόνιο Πανεπιστήμιο

Βασίλειος Καρυώτης, *Αναπληρωτής Καθηγητής,*

Ιόνιο Πανεπιστήμιο

*Για τον παππού μου,
γιατί μου έμαθε να βλέπω τον κόσμο καθαρά, χωρίς ο ίδιος να μπορεί.*

Περίληψη

Η πρόσβαση στο διαδίκτυο αποτελεί βασική προϋπόθεση για την επιβίωση ενός ανθρώπου στην εποχή της πληροφορίας. Ωστόσο, παρά την τεχνολογική πρόοδο σε διαφορετικούς τομείς της καθημερινότητας, η ίση πρόσβαση σε αυτό αποτελεί μέχρι και σήμερα μία ουτοπία για τα άτομα με προβλήματα όρασης. Η πολύπλοκη δομή των ιστοσελίδων, σε συνδυασμό με τον δύσκολο χειρισμό εργαλείων όπως ο αναγνώστης οθόνης, απομακρύνουν τους χρήστες με χαμηλή ή μηδενική οπτική απόδοση από την αξιοποίηση του περιεχομένου τους. Παράλληλα, φωνητικοί βοηθοί όπως η Alexa της Amazon, δίνουν την δυνατότητα αλληλεπίδρασης με το διαδίκτυο εξ ολοκλήρου μέσω της φωνής, προσθέτοντας στο πεδίο της επικοινωνίας ανθρώπου-υπολογιστή μία νέα μορφή διάδρασης, η οποία μπορεί να υιοθετηθεί πλήρως από ένα άτομο με προβλήματα όρασης. Η παρούσα πτυχιακή μελετά την προοπτική ενσωμάτωσης των τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας που χρησιμοποιεί ένας φωνητικός βοηθός, σε εφαρμογές του διαδικτύου. Για την διαπίστωση της συγκεκριμένης θέσης έγινε η δημιουργία ενός proof of concept ηλεκτρονικού καταστήματος, το οποίο επιτρέπει την πλοήγηση εξ ολοκλήρου μέσω της φωνής. Χωρίς την ανάγκη για συσκευές όπως το πληκτρολόγιο και το ποντίκι, η φωνή αποτελεί το μέσο για την αξιοποίηση των δυνατοτήτων του διαδικτύου από άτομα που αντιμετωπίζουν προβλήματα όρασης, με αποτελεσματικό και εύελικτο τρόπο.

Λέξεις κλειδιά: Αναγνώριση ομιλίας, Σύνθεση ομιλίας, Web Speech API, Ανάπτυξη διαδικτυακών εφαρμογών, Ηλεκτρονική Προσβασιμότητα, Προβλήματα Όρασης, Ηλεκτρονικό Κατάστημα, Επικοινωνία Ανθρώπου-Υπολογιστή

Abstract

The access to the World Wide Web is a basic condition for the survival of a human being in the information age. However, despite the technological progress in different areas of everyday life, an equal access to the Web for people with vision problems is still considered a utopia. The complex structure of websites, combined with the difficult handling of tools such as the screen reader, both distance users who have low or non-existent visual performance from the exploitation of the website's content. In parallel, voice assistants such as Amazon's Alexa enable interaction with the Web entirely through the usage of voice, adding a new form of interaction to the field of human-computer communication, which can be adopted entirely by a person with vision problems. The present dissertation studies the prospect of integrating voice recognition and synthesis technologies that are previously used by a voice assistant, in web applications. For the determination of the specific position, a proof of concept e-shop which allows full voice navigation was created. Without the need for hardware such as the keyboard and the mouse, the voice is the mean for the exploitation of the Web's possibilities in an efficient and flexible way by people who face vision problems.

Keywords: Speech Recognition, Speech Synthesis, Web Speech API, Web Development, Web Accessibility, Vision Problems, E-shop, Human Computer Interaction

Ευχαριστίες

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω τον επιβλέπων καθηγητή της πτυχιακής μου εργασίας κ. Παναγιώτη Κουρουθανάση, ο οποίος μου έδωσε την δυνατότητα να βάλω ένα λιθαράκι στο πεδίο που με ενδιαφέρει πραγματικά. Την οικογένεια μου, γιατί μου παρείχε το εισιτήριο και τα ψυχικά εφόδια για όλο αυτό το ταξίδι. Τέλος, τους φίλους και την κοπέλα μου, που στέκονται ως συνοδοιπόροι μου κάθε μέρα.

Περιεχόμενα

Περίληψη	i
Abstract.....	ii
Ευχαριστίες.....	iii
Εισαγωγή	1
1. Χρήστες με προβλήματα όρασης και διαδίκτυο.....	3
1.1 Υπόθεση έρευνας.....	3
1.2 Σχετικές εργασίες.....	7
2. Web Speech API.....	11
2.1 Εισαγωγή	11
2.2 Πορεία δημιουργίας Web Speech API	12
2.3 Web Speech API specification	14
2.4 Περιγραφή	18
2.5 SpeechRecognition Interface.....	21
2.6 SpeechSynthesis Interface	34
2.7 Συμβατότητα	43
2.8 Συμπεράσματα.....	46
3. Ηλεκτρονικό κατάστημα με δυνατότητες υποβοήθησης ατόμων με προβλήματα όρασης για πλοήγηση μέσω φωνητικών εντολών	47
3.1 Εισαγωγή	47
3.2 Αναλυτική περιγραφή των λειτουργιών του ηλεκτρονικού καταστήματος.....	48
3.2.1 Αρχική σελίδα.....	48
3.2.2 Προϊόντα	55
3.2.3 Σελίδα προϊόντος.....	64
3.2.4 Αναζήτηση προϊόντων	67
3.2.5 Φόρμα επικοινωνίας	70
3.2.6 Καλάθι αγορών	73
3.3 Τεχνική υποδομή	78
3.3.1 Τεχνολογίες.....	78
3.3.2 Μεθοδολογία ενσωμάτωσης φωνητικού βοηθού	82
3.3 Προβλήματα	93
4. Συμπεράσματα και μελλοντική έρευνα	98
Βιβλιογραφία	101

Δικτυογραφία	101
---------------------------	------------

Εισαγωγή

Στην εποχή της πληροφορίας, όλο και περισσότερες πτυχές της καθημερινότητας διαβαίνουν το μονοπάτι του ψηφιακού μετασχηματισμού, συνθέτοντας έναν άυλο, απεριόριστων δυνατοτήτων, κόσμο, ο οποίος γίνεται όλο και πιο αναγκαίος για την συμβίωση ενός ατόμου στην κοινωνία. Στο διαδίκτυο, οι λέξεις τοποθεσία και όριο χάνουν το νόημα τους, δίνοντας σε ένα χρήστη ικανότητες για επικοινωνία, ενημέρωση, αγοραπωλησία, εκπαίδευση και η λίστα αυξάνεται καθημερινά, μέσω της ραγδαίας ανάπτυξης της τεχνολογίας. Καταλαβαίνει κανείς, λοιπόν, την ανάγκη για ίση πρόσβαση στο διαδίκτυο, αφού πλέον αποτελεί αναπόσπαστο κομμάτι του τρόπου με τον οποίο εξελίσσονται οι κοινωνίες. Ωστόσο, παρά τις συνεχής επιτυχίες στον τεχνολογικό τομέα, η ικανοποίηση αυτής της ανάγκης αποτελεί μέχρι και σήμερα μία ουτοπία για τα άτομα με προβλήματα όρασης. Συσκευές όπως το πληκτρολόγιο και το ποντίκι δεν αποτελούν κατάλληλα μέσα για την αλληλεπίδραση ενός τέτοιου χρήστη στο διαδίκτυο. Για τον λόγο αυτό, το πεδίο της υποστηρικτικής τεχνολογίας (assistive technology) προσφέρει στα συγκεκριμένα άτομα εργαλεία όπως ο αναγνώστης οθόνης, με στόχο την πρόσβαση τους στον κόσμο της πληροφορίας. Παρά την πληθώρα δυνατοτήτων πλοήγησης που δίνει ένα τέτοιο εργαλείο, βασίζεται αποκλειστικά στην HTML δομή της εκάστοτε σελίδας που θα επισκεφτεί ο χρήστης, κάτι το οποίο σημαίνει πως εάν αυτή δεν ακολουθεί κανόνες προσβασιμότητας, όπως αυτούς που έχει ορίσει το W3C (World Wide Web Consortium), τότε ο αναγνώστης οθόνης δεν μπορεί να χρησιμοποιηθεί. Παράλληλα, ένα τέτοιο εργαλείο δεν εξαλείφει την χρήση του πληκτρολογίου, αφού η πλοήγηση του χρήστη βασίζεται σε πληθώρα συνδυασμού πλήκτρων, μετατρέποντας την διαδικασία της πλοήγησης σε δύσκολη και σε πολλές περιπτώσεις χρονοβόρα. Την ίδια στιγμή, οι εξελίξεις στο πεδίο της τεχνητής νοημοσύνης έχουν φέρει στην ζωή τεχνολογία η οποία ακούει στο όνομα φωνητικός βοηθός (voice assistant). Εταιρείες κολοσσοί όπως η Amazon, η Apple και η Google, παρέχουν φωνητικούς βοηθούς, οι οποίοι εκμεταλλεύονται τις δυνατότητες του διαδικτύου για την επίτευξη πολυποίκιλων λειτουργιών. Οι χρήστες τους δεν χρειάζεται να έρθουν σε επαφή με ένα γραφικό περιβάλλον, παρά μόνο να αξιοποιήσουν το εργαλείο που τους έχει δοθεί εκ γενετής. Τη φωνή τους. Από την αναπαραγωγή μουσικής, μέχρι την αγορά υπηρεσιών και προϊόντων, ένας χρήστης μπορεί να αλληλεπιδράσει με το διαδίκτυο μέσω του φωνητικού βοηθού, ο οποίος περιλαμβάνει δυνατότητες αναγνώρισης και σύνθεσης ομιλίας. Ένα τέτοιο μοντέλο αλληλεπίδρασης αποφέρει πολλαπλά οφέλη στην επικοινωνία ανθρώπου-υπολογιστή, με ένα από τα κυριότερα, τη δυνατότητα χρήσης του από άτομα με προβλήματα όρασης. Η παρούσα πτυχιακή, βασισμένη στην συγκεκριμένη τοποθέτηση, μελετά την προοπτική ενσωμάτωσης τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας σε εφαρμογές του διαδικτύου, με στόχο την υποβοήθηση ατόμων με προβλήματα όρασης για την αλληλεπίδραση τους με αυτές. Για τη μελέτη της συγκεκριμένης προοπτικής δημιουργήθηκε ως proof of concept ένα ηλεκτρονικό κατάστημα, στο οποίο περιλαμβάνεται ένας φωνητικό βοηθός, στόχος του οποίου

αποτελεί η υποβοήθηση χρηστών με προβλήματα όρασης για την αλληλεπίδραση τους με το κατάστημα εξ ολοκλήρου μέσω της φωνής τους. Αδιαμφισβήτητα, ο κλάδος του ηλεκτρονικού εμπορίου αποτελεί μία από τις μεγαλύτερες πτυχές του διαδικτύου, μέσω της οποίας πραγματοποιούνται αγοραπωλησίες χωρίς γεωγραφικούς περιορισμούς και όρια σε επιλογές προϊόντων και υπηρεσιών. Συνεπώς, η δυνατότητα αλληλεπίδρασης σε ένα τέτοιο περιβάλλον από άτομα με προβλήματα όρασης θα προσδώσει στις επιχειρήσεις μεγαλύτερο κοινό, χαρακτήρα κοινωνικής αλληλεγγύης, προωθώντας παράλληλα την καινοτομία. Παράλληλα, οι χρήστες με προβλήματα όρασης θα μπορούν να ενσωματώσουν στην καθημερινότητα τους τις διαδικτυακές αγορές, κάτι το οποίο αποτελεί δικαίωμα και όχι προνόμιο. Στο πρώτο κεφάλαιο, αναφέρεται η υπόθεση έρευνας αλλά και σχετικές εργασίες. Στο δεύτερο κεφάλαιο, μελετάται το Web Speech API εργαλείο, όπου μέσω της χρήσης Javascript γίνεται εφικτή η ενσωμάτωση δυνατοτήτων αναγνώρισης και σύνθεσης ομιλίας σε εφαρμογές του διαδικτύου. Στο τρίτο κεφάλαιο, περιγράφεται το ηλεκτρονικό κατάστημα που δημιουργήθηκε ως proof of concept, η τεχνική του υποδομή, αλλά και προβλήματα που αντιμετωπίστηκαν. Τέλος, ακολουθούν τα συμπεράσματα από την διεξαγωγή της συγκεκριμένης εργασίας, η μελλοντική έρευνα και η βιβλιογραφία.

1. Χρήστες με προβλήματα όρασης και διαδίκτυο

1.1 Υπόθεση έρευνας

Σύμφωνα με το WHO (World Health Organization), σε παγκόσμιο επίπεδο τουλάχιστον 2,2 δισεκατομμύρια άνθρωποι αντιμετωπίζουν προβλήματα όρασης. Οι συνέπειες της χαμηλής ή μηδενικής οπτικής απόδοσης, μπορεί να έχουν διαφορετικό αντίκτυπο στη ζωή ενός ασθενή, ανάλογα με την παράμετρο της ηλικίας. Σε νέα παιδιά, υπάρχει το ενδεχόμενο να αντιμετωπίσουν δυσκολίες οι οποίες να αφορούν τις κοινωνικές τους δεξιότητες, ενώ πιθανό είναι να βιώσουν και χαμηλότερα επίπεδα εκπαιδευτικών επιτευγμάτων. Αντίστοιχα, ενήλικες συχνά παρουσιάζουν χαμηλότερα ποσοστά παραγωγικότητας και υψηλότερα ποσοστά κατάθλιψης και άγχους, ενώ σε μεγαλύτερες ηλικίες συχνό είναι και το φαινόμενο της κοινωνικής απομόνωσης. Καταλαβαίνει κανείς, λοιπόν, τον καθοριστικό ρόλο που διαδραματίζει σήμερα το διαδίκτυο στη ζωή ενός ατόμου με προβλήματα όρασης, αφού τα παραπάνω φαινόμενα είναι δυνατό να μειωθούν μέσω αυτού, αλλά αντίστοιχα να αυξηθούν χωρίς την ικανότητα αξιοποίησης των δυνατοτήτων του, πόσο μάλλον σε μία κοινωνία, η οποία ενσωματώνει την χρήση του όλο και πιο βαθιά στο πυρήνα της.

Ο εφευρέτης του παγκόσμιου ιστού Tim Berners-Lee, δήλωσε ότι η δύναμη αυτού βρίσκεται στην καθολικότητα του και ότι η πρόσβαση όλων ανεξαρτήτως αναπηρίας αποτελεί μία ουσιώδη πτυχή του. Για την επίτευξη της παραπάνω θέσης, πραγματοποιούνται μεγάλες προσπάθειες για ίση πρόσβαση στο διαδίκτυο, οι οποίες τοποθετούνται κάτω από την ομπρέλα του όρου διαδικτυακή προσβασιμότητα (web accessibility). Μία από αυτές τις προσπάθειες, αφορά ο σχηματισμός του WAI (Web Accessibility Initiative) από το W3C (World Wide Web Consortium), με στόχο την παροχή κατευθυντήριων γραμμών σε προγραμματιστές του διαδικτύου για το πως μπορούν να κάνουν τις σελίδες τους προσβάσιμες σε όλους. Σύμφωνα με την σελίδα του WAI, όλοι οι χρήστες, ανεξαρτήτως αναπηρίας, πρέπει να μπορούν να αντιλαμβάνονται, να κατανοούν, να πλοηγούνται και να αλληλεπιδρούν με τον ιστό, καθώς επίσης να έχουν και την ικανότητα συνεισφοράς σε αυτόν. Έτσι λοιπόν, δημιούργησε τις οδηγίες WCAG (Web Content Accessibility Guidelines), οι οποίες κατευθύνουν έναν προγραμματιστή για τον τρόπο με τον οποίο πρέπει να δομήσει το περιεχόμενο μίας σελίδας, έτσι ώστε να έχουν πρόσβαση σε αυτό όλοι οι χρήστες χωρίς περιορισμούς.

Παρά την ύπαρξη των οδηγιών WCAG, μεγάλος αριθμός σελίδων δεν τις ακολουθεί με αποτέλεσμα άνθρωποι με προβλήματα όρασης να μην έχουν την δυνατότητα αλληλεπίδρασης με το περιεχόμενο τους. Χαρακτηριστικό παράδειγμα αποτελεί η έρευνα του WebAIM, το οποίο αξιολόγησε την προσβασιμότητα των 1.000.000 κορυφαίων ιστοσελίδων για τις χρονιές 2019, 2020, 2021. Αναφορικά με τα αποτελέσματα συμμόρφωσης με το περιεχόμενο του WCAG 2 το 2021, τα ποσοστά έδειξαν ότι στο

97.4% των αρχικών σελίδων (home pages) ανιχνεύθηκαν αποτυχίες, με τις συχνότερες να αφορούν την έλλειψη labels σε φόρμες (missing form input labels 54.4%), την έλλειψη εναλλακτικού κειμένου σε εικόνες (missing alternative text for images 60.6%), άδειους συνδέσμους (empty links 51.3%) κ.α. Τέτοιες ελλείψεις, καθιστούν αδύνατη την πλοήγηση ενός ατόμου με προβλήματα όρασης με εργαλεία όπως είναι ο αναγνώστης οθόνης (screen reader), αφού το συγκεκριμένο εργαλείο βασίζεται στην ύπαρξη τέτοιων στοιχείων. Επιπλέον, μεγάλη εντύπωση προκαλεί και η σύγκριση των ποσοστών αποτυχίας με τις χρονιές 2019 και 2020, όπου τα ποσοστά είναι πολύ κοντά μεταξύ τους, ενώ υπάρχουν και περιπτώσεις που κάθε χρόνο το ποσοστό παρουσιάζει άνοδο όπως αυτή της έλλειψης labels σε φόρμες, όπου το 2019 ήταν 52.8%, το 2020 53.8% και το 2021 54.4%. Τέτοια αποτελέσματα, φανερώνουν την αδυναμία μεγάλου μέρους ιστοσελίδων να παρέχει προσβάσιμο περιεχόμενο, κάτι το οποίο θα έπρεπε να βελτιώνεται με την πάροδο του χρόνου. Βέβαια, αξιοσημείωτη είναι και η εργασία των Power et al (2012), κατά την οποία διαπιστώθηκε πως για προβλήματα που αντιμετωπίζουν τυφλοί χρήστες στο διαδίκτυο και τα οποία καλύπτονταν από το WCAG 2.0, το 16.7% των ιστοσελίδων που χρησιμοποιήθηκαν για την έρευνα είχαν ακολουθήσει τεχνικές που προτείνονται στο WCAG 2.0, αλλά οι τεχνικές αυτές δεν έλυσαν τα προβλήματα. Κάτι τέτοιο, όπως αναφέρουν και οι ίδιοι σημαίνει πως ακόμα και με την συμμόρφωση με το περιεχόμενο WCAG, υπάρχουν ελάχιστες ενδείξεις ότι τα άτομα με αναπηρία θα αντιμετωπίσουν λιγότερα προβλήματα κατά την πλοήγηση τους.

Όταν μία ιστοσελίδα ακολουθεί τις οδηγίες WCAG, στρώνει ένα μονοπάτι, το οποίο έπειτα έχουν την δυνατότητα να διασχίσουν άτομα με προβλήματα όρασης μέσω της χρήσης υποστηρικτικών τεχνολογιών. Η πιο διαδεδομένη τεχνολογία για τον παραπάνω σκοπό είναι ο αναγνώστης οθόνης. Το συγκεκριμένο εργαλείο, έχει την ικανότητα να εκφωνεί την πληροφορία που περιλαμβάνεται σε ένα γραφικό περιβάλλον στον χρήστη. Εκτός από την εκφώνηση της πληροφορίας, υπάρχει και η επιλογή μετατροπής της σε μπράιγ, μέσω της χρήσης μίας επιπλέον συσκευής που ονομάζεται ανανεώσιμη οθόνη μπράιγ (refreshable braille display). Ο χρήστης, αξιοποιεί μία πληθώρα εντολών μέσω διαφορετικών συνδυασμών πλήκτρων, οι οποίες τον βοηθούν να αλληλεπιδράσει με το περιεχόμενο της εκάστοτε ιστοσελίδας. Παρά την αναγκαιότητα του συγκεκριμένου εργαλείου παρουσιάζονται αρκετοί περιορισμοί αναφορικά με την χρήση του. Συγκεκριμένα, η εκάστοτε ιστοσελίδα θα πρέπει να έχει μία δομή βασισμένη στους κανόνες προσβασιμότητας, ενώ παράλληλα ο χρήστης καλείται να απομνημονεύσει μία πληθώρα συνδυασμών πλήκτρων για να πλοηγηθεί, κάτι το οποίο μετατρέπει την διαδικασία χρήσης ενός τέτοιου εργαλείου σε δύσκολη αλλά και χρονοβόρα. Οι Borodin et al (2010), μελέτησαν τις στρατηγικές πλοήγησης που έχουν αναπτύξει άτομα με προβλήματα όρασης για την πλοήγηση τους μέσω ενός αναγνώστη οθόνης. Σύμφωνα με τους ίδιους, *η πρόσβαση στο διαδίκτυο με αναγνώστες οθόνης είναι γενικά αναποτελεσματική λόγω των περιορισμών της σειριακής διεπαφής ήχου*, αφού αυτή η μέθοδος αναγκάζει τον χρήστη να ακούει με την σειρά τις διαθέσιμες επιλογές, χωρίς να έχει την ικανότητα μετάβασης στο σημείο που επιθυμεί

κατευθείαν. Επιπλέον, τονίζουν πως η πρόσβαση σε μια ιστοσελίδα μπορεί να είναι τεχνικά δυνατή με την χρήση ενός αναγνώστη οθόνης, αλλά μπορεί να είναι τόσο δύσκολη, ώστε να καταλήξει στο να μην είναι χρήσιμη.

Από όλα τα παραπάνω, καταλαβαίνει κανείς πως η πλοήγηση ενός ατόμου με προβλήματα όρασης στο διαδίκτυο δεν είναι μία απλή υπόθεση και πως ακόμα και αν υπάρχουν οι προϋποθέσεις, αυτή θα γίνεται με μη αποδοτικό τρόπο. Αποτελεί λοιπόν, καθήκον των προγραμματιστών, να βρουν μία λύση η οποία να προσφέρει μία εύκολη, αλλά ταυτόχρονα αποδοτική εμπειρία πλοήγησης.

Την ίδια στιγμή, οι τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας φέρνουν στο πεδίο της αλληλεπίδρασης ανθρώπου-υπολογιστή ένα νέο μέσο επικοινωνίας, το οποίο ο άνθρωπος χρησιμοποιεί αιώνες τώρα για την αλληλεπίδραση με το περιβάλλον του, πολύ πριν εμφανιστεί ο όρος τεχνολογία. Τη φωνή. Ενώ οι καθιερωμένοι μέθοδοι διάδρασης με συσκευές όπως είναι το πληκτρολόγιο και το ποντίκι, απαιτούν την χρήση των δακτύλων και της όρασης, μέσω της φωνής γίνεται εφικτή η αλληλεπίδραση με την τεχνολογία με τον πιο φυσικό και αβίαστο τρόπο. Δυνατότητες όπως η αναζήτηση στο διαδίκτυο, η υπαγόρευση και ανάγνωση μηνυμάτων ηλεκτρονικού ταχυδρομείου αλλά και η πλοήγηση σε εφαρμογές, αποτελούν κάποιες μόνο από τις ικανότητες που μπορούν να αξιοποιηθούν ίσα από όλους μέσω της αλληλεπίδρασης με τη φωνή. Ωστόσο, ο συνδυασμός των δύο αυτών τεχνολογιών ως μέσο διάδρασης δεν παρατηρείται τόσο στο διαδίκτυο, όσο στην νέα τάση που ακούει στο όνομα φωνητικός βοηθός. Τεχνολογίες όπως η Siri και Alexa, παρέχουν μία προσωποποιημένη δίοδο επικοινωνίας, μέσω της οποίας είναι εφικτή η αλληλεπίδραση με το διαδίκτυο. Μάλιστα, μία έρευνα από το Juniper Research, προέβλεψε πως το 2023 θα χρησιμοποιούνται 8 δις φωνητικοί βοηθοί ενώ στο τέλος του 2018 η εκτίμηση έδειξε ότι ο αριθμός αυτός ήταν 2.5 δις, κάτι το οποίο δείχνει την ολοένα και περισσότερο ενσωμάτωση τους στην καθημερινότητα των ανθρώπων. Παράλληλα, έρευνες όπως των Pradhan, Mehta και Findlater (2018), μελετούν τη χρήση των φωνητικών βοηθών από άτομα με αναπηρίες λόγω του φωνητικού μοντέλου αλληλεπίδρασης που παρέχουν. Αξιοσημείωτη είναι η διαπίστωση τους πως σε αρκετές περιπτώσεις άνθρωποι με προβλήματα όρασης πραγματοποιούν λειτουργίες όπως την αναπαραγωγή μουσικής, αλλά και την ανάκτηση πληροφοριών σχετικά με τον καιρό και τις ειδήσεις, χωρίς να εξαρτώνται από την βοήθεια κάποιου ατόμου.

Γεννάται λοιπόν το ερώτημα του κατά πόσο το μοντέλο αλληλεπίδρασης που χρησιμοποιεί ένας φωνητικός βοηθός θα μπορούσε να ενσωματωθεί σε εφαρμογές του διαδικτύου, με στόχο την υποβοήθηση ατόμων με προβλήματα όρασης. Κάτι τέτοιο, θα εξάλειφε την ανάγκη για χρήση συσκευών όπως το πληκτρολόγιο και το ποντίκι, ενώ παράλληλα θα μπορούσε να συμπληρώσει την σειριακή διάσχιση της πληροφορίας που προσφέρουν οι αναγνώστες οθόνης με μία επιλεκτική ικανότητα. Έτσι, για έναν χρήστη με προβλήματα όρασης, δεν θα απαιτείται να απομνημονεύσει πληθώρα συνδυασμού πλήκτρων, ενώ

παράλληλα θα μπορεί να αλληλεπιδράσει με την πληροφορία που επιθυμεί χωρίς να χρειάζεται να του εκφωνείται σειριακά το ενδιάμεσο περιεχόμενο. Διαδικτυακές εφαρμογές οι οποίες διαθέτουν δυνατότητες διάδρασης με τη φωνή, μπορούν να αυξήσουν την χρηστικότητα τους σε χρήστες που δεν έχουν την ικανότητα χρήσης παραδοσιακών μεθόδων από το πεδίο της αλληλεπίδρασης ανθρώπου-υπολογιστή (Kandhari, Zulkemine, Isah, 2018), όπως είναι το ποντίκι και το πληκτρολόγιο.

Με βάση την παραπάνω θέση, στην παρούσα εργασία δημιουργήθηκε ένα φωνητικό μοντέλο αλληλεπίδρασης για την υποβοήθηση ατόμων με προβλήματα όρασης έτσι ώστε να έχουν την ικανότητα πλοήγησης σε ένα ηλεκτρονικό κατάστημα εξ ολοκλήρου μέσω της φωνής τους. Αδιαμφισβήτητα, το περιβάλλον ενός ηλεκτρονικού καταστήματος αποτελεί ένα από τα δυσκολότερα για την πλοήγηση χρηστών με προβλήματα όρασης και συνεπώς, η εύρεση ενός τρόπου με τον οποίο θα μπορούν να εκμεταλλευτούν τα πλεονεκτήματα του είναι σημαντική. Η φιλοσοφία του μοντέλου που δημιουργήθηκε, συνδυάζει ικανότητες που παρέχονται σήμερα μέσω της χρήσης ενός αναγνώστη οθόνης, με μία περισσότερο ανθρωποκεντρική προσέγγιση, κατά την οποία ένας χρήστης δεν καλείται να αλληλεπιδράσει μόνο με το υπάρχον περιεχόμενο μίας σελίδας και με απρόσωπο τρόπο. Αντίθετα, ανταλλάσσει πληροφορίες μέσω ενός φωνητικού βοηθού, ο οποίος συμβάλει στο να κατανοήσει ο χρήστης τις δυνατότητες που του δίνονται, χωρίς η ικανότητα πλοήγησης να στηρίζεται αποκλειστικά στην δομή της HTML. Παράλληλα, υιοθετήθηκε μία προσέγγιση κατά την οποία δεν χρειάζεται να παρέχεται η δυνατότητα ανάκτησης μόνο του περιεχομένου που υπάρχει στο γραφικό περιβάλλον, αλλά ο χρήστης μπορεί να ανακτήσει πληροφορίες κατευθείαν από την βάση δεδομένων, χωρίς την ανάγκη αναζήτησης τους σε άλλη σελίδα, κάτι το οποίο θα είχε ως αποτέλεσμα την αύξηση χρονικά της διαδικασίας πλοήγησης. Με μία φράση, η ιδέα της συγκεκριμένης εργασίας αφορά στην αξιοποίηση τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας για την υποβοήθηση χρηστών με προβλήματα όρασης, με τελικό στόχο την παροχή μίας εμπειρίας πλοήγησης, η οποία να είναι εύκολη, γρήγορη, κατανοητή και το κυριότερο ανθρωποκεντρική. Με αυτό τον τρόπο, δεν αυξάνεται μόνο η προσβασιμότητα μίας σελίδας αλλά και η χρηστικότητα της. Επιπλέον, η ικανότητα πρόσβασης στο διαδίκτυο από άτομα με προβλήματα όρασης μέσω της φωνής τους φέρει πολλαπλά οφέλη και στην ίδια την επιχείρηση, αφού έχει την ικανότητα να απευθυνθεί σε μεγαλύτερο κοινό, ενώ παράλληλα μπορεί να χρησιμοποιηθεί ως marketing εργαλείο, το οποίο θα αυξήσει την οικονομική της αξία αλλά και τη φήμη της για τις κοινωνικές της ευθύνες (Talib et al, 2009).

Για την δημιουργία του proof of concept αξιοποιήθηκε το Web Speech API σε συνδυασμό με τεχνικές χειρισμού DOM (Document Object Model) και AJAX (Asynchronous JavaScript And XML). Το αποτέλεσμα, αποτελεί ένα φωνητικό μοντέλο αλληλεπίδρασης, το οποίο λειτουργεί δυναμικά και μπορεί να ενσωματωθεί σε ένα κατάστημα με λίγες μόνο προσθήκες. Παράλληλα, με την κατάλληλη

παραμετροποίηση του είναι αξιοποιήσιμο και από εφαρμογές που δεν αφορούν μόνο ηλεκτρονικά καταστήματα, διευκολύνοντας έτσι την πρόσβαση και σε άλλες πτυχές του διαδικτύου.

Η συγκεκριμένη εργασία, εστιάζει στους χρήστες που αντιμετωπίζουν προβλήματα όρασης. Ωστόσο, προφανές είναι πως μπορούν να εξυπηρετηθούν και ομάδες ανθρώπων που αντιμετωπίζουν άλλα προβλήματα όπως RSI (Repetitive strain injury) ή δυσλεξία και γενικότερα άτομα, τα οποία δυσκολεύονται να αλληλεπιδράσουν με καθιερωμένα μέσα, όπως είναι το πληκτρολόγιο και το ποντίκι. Επιπλέον, ακόμα ένα πεδίο εφαρμογής των συγκεκριμένων τεχνολογιών μπορεί να αποτελέσει και η υποβοήθηση ηλικιωμένων, αφού σε αντίθεση με μικρότερες γενιές δεν μεγάλωσαν σε ένα περιβάλλον, το οποίο να χαρακτηρίζεται από έντονη χρήση τεχνολογιών διαδικτύου, με αποτέλεσμα ακόμα και η πραγματοποίηση μίας απλής αναζήτησης να αποτελεί ένα μαύρο κουτί, το οποίο προκαλεί τον αποκλεισμό τους από τις δυνατότητες που δίνει σήμερα ο ψηφιακός κόσμος. Τέλος, αυτονόητο είναι πως οι τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας μπορούν να προσφέρουν στους χρήστες του διαδικτύου έναν επιπλέον τρόπο πλοήγησης, ο οποίος θα είναι πιο ευέλικτος και γρήγορος, ενώ παράλληλα θα είναι ικανός να εξυπηρετήσει σενάρια, όπως η αλληλεπίδραση την ώρα της οδήγησης, κάτι το οποίο μπορεί να έχει τραγικά αποτελέσματα όταν απαιτείται η χρήση των ματιών και των δακτύλων.

1.2 Σχετικές εργασίες

Η συγκεκριμένη εργασία, δεν είναι η πρώτη που μελετά την προοπτική ενσωμάτωσης τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας σε εφαρμογές του διαδικτύου με στόχο την υποβοήθηση χρηστών με προβλήματα όρασης. Στην παρούσα υποενότητα, αναφέρονται σχετικές εργασίες οι οποίες διαπίστωσαν εξίσου, τα πολλαπλά οφέλη της αλληλεπίδρασης με την φωνή και συνεπώς έβαλαν το λιθαράκι τους για ένα διαδίκτυο ίσο προς όλους.

Οι Algarni, Cheung και Lee (2013), αναφέρουν ότι παρά τις καινοτομίες στο πεδίο των τεχνολογιών πληροφορικής, πτυχές του διαδικτύου όπως τα marketplaces, δεν παρέχουν κατάλληλη υποδομή, έτσι ώστε χρήστες με προβλήματα όρασης να αξιοποιήσουν τις δυνατότητες τους. Βασισμένοι σε αυτό, προτείνουν ένα interface το οποίο θα μπορεί να ενσωματωθεί σε τέτοια περιβάλλοντα κάνοντας εφικτή την αλληλεπίδραση των χρηστών μέσω φωνητικών εντολών. Στην εργασία τους, περιγράφεται ένα voice-based eMarketplace κατά το οποίο ένας χρήστης θα μπορεί να αιτείται μία εντολή μέσω της αναγνώρισης ομιλίας. Έπειτα, με την χρήση ενός φωνητικού βοηθού η είσοδος του χρήστη θα μεταβιβάζεται για πιθανή αντιστοίχιση της με το περιεχόμενο της βάσης δεδομένων. Σε περίπτωση που αντιστοιχιστεί κάτι, θα εκφωνηθούν τα αποτελέσματα στον χρήστη και θα συνεχιστεί η αγοραστική διαδικασία φωνητικά. Σε περίπτωση μη αντιστοίχισης, ο φωνητικός βοηθός θα κατευθύνει τον χρήστη σε έναν εκπρόσωπο εξυπηρέτησης πελατών, χρησιμοποιώντας Voice over IP, όπου και θα έχει την δυνατότητα να αναφέρει το προϊόν, το οποίο αναζητεί λαμβάνοντας την αντίστοιχη βοήθεια. Στο τέλος της

εργασίας, αναφέρεται ότι το προτεινόμενο σύστημα μπορεί να δημιουργηθεί ως ένα module το οποίο να ενσωματώνεται σε ένα υπάρχον eMarketplace με δυνατότητα ενεργοποίησης του όποτε απαιτείται.

Οι Kandhari, Zulkemine και Isah (2018), στηριζόμενοι στο ότι τα προβλήματα προσβασιμότητας των websites έχουν ως αποτέλεσμα οι χρήστες με προβλήματα όρασης να στερούνται πληθώρα υπηρεσιών, ενώ παράλληλα επιχειρήσεις ηλεκτρονικού εμπορίου να χάνουν πιθανούς πελάτες, δημιούργησαν ένα πρωτότυπο, το οποίο αφορά μία καθοδηγούμενη από φωνή e-commerce εφαρμογή, με στόχο την προβολή της χρηστικότητας που μπορεί να έχει αναφορικά με τα παραπάνω ζητήματα. Χρησιμοποιώντας τις υπηρεσίες Watson για Speech to Text και Text to Speech, δημιουργήθηκε ένα σενάριο κατά το οποίο ένας χρήστης δίνει μέσω της φωνής του ένα query ως εντολή, το σύστημα βλέπει μέσα στο query για συγκεκριμένες λέξεις κλειδιά και με βάση την πλησιέστερη αντιστοίχιση που θα κάνει, εκτελεί λειτουργίες, όπως η αναζήτηση για συγκεκριμένο προϊόν και η διάσχιση ιστοσελίδων. Σε περίπτωση που δεν βρεθεί ένα προϊόν, χρησιμοποιείται Text to Speech για την ενημέρωση του χρήστη. Στην εργασία τους, σημειώνεται ότι ένα από τα βασικά χαρακτηριστικά της υπηρεσίας Watson Speech to Text αφορά στη δυνατότητα της για αποτελεσματική αναγνώριση λέξεων κλειδιών μέσα σε ένα κείμενο, με αποτέλεσμα την εκτέλεση προκαθορισμένων λειτουργιών ανάλογα με την λέξη κλειδί που θα αναγνωρισθεί. Εκτός των τεχνολογιών για αναγνώριση και σύνθεση ομιλίας, χρησιμοποιήθηκε Javascript και HTML για την πλευρά του πελάτη, Javascript και Python για το κομμάτι του back-end, ενώ για την βάση δεδομένων χρησιμοποιήθηκε MongoDB.

Οι Mohadikar και Nawkhare (2017), στηριζόμενοι στο ότι τα άτομα με προβλήματα όρασης αντιμετωπίζουν δυσκολίες κατά την πλοήγηση τους σε ηλεκτρονικά καταστήματα, δημιούργησαν ένα σύστημα το οποίο τα βοηθάει να ψωνίσουν επιλέγοντας ρούχα διαφορετικών χρωμάτων από διαφορετικές κατηγορίες. Το σύστημα τους, χρησιμοποιεί αναγνώριση ομιλίας μέσω της χρήσης deep learning αλγορίθμου με στόχο την αναζήτηση προϊόντων. Σύμφωνα με την εργασία τους, η χρήση της ομιλίας για την αναζήτηση προϊόντων, αντί του πληκτρολογίου, δίνει πλεονέκτημα στα άτομα με προβλήματα όρασης, επιτρέποντας να χρησιμοποιήσουν την εφαρμογή με πιο ευέλικτο τρόπο. Με την μετατροπή της εισόδου που έδωσε ο χρήστης μέσω της φωνής του σε κείμενο, θα πραγματοποιηθεί ανάλυση και με βάση το αποτέλεσμα θα επιλεγεί το εκάστοτε προϊόν από την βάση δεδομένων όπου και θα τυπωθεί στην οθόνη συνοδευόμενο από audio. Λόγω της ύπαρξης χιλιάδων προϊόντων μαζί με τις κατηγορίες τους, η βάση δεδομένων περιλαμβάνει μεγάλο αριθμό εγγραφών και συνεπώς με στόχο την μεγαλύτερη ευελιξία πραγματοποιήθηκε clustering στα δεδομένα με την χρήση του αλγορίθμου K-mean.

Οι Iyer, Sheth, Shah και Devadkar (2020), διατυπώνουν στην εργασία τους ότι ένας virtual assistant είναι ικανός να παρέχει εύκολη πρόσβαση στο διαδίκτυο για τα άτομα με προβλήματα όρασης. Στηριζόμενοι στο ότι οι κανόνες προσβασιμότητας σε συνδυασμό με τις υποστηρικτικές τεχνολογίες

(assistive technologies) δεν δίνουν σε ικανοποιητικό βαθμό την ικανότητα πρόσβασης στα άτομα με προβλήματα όρασης, δημιούργησαν ένα σύστημα με στόχο την αλληλεπίδραση ενός τέτοιου χρήστη στο διαδίκτυο. Η φιλοσοφία του συστήματος τους αφορά την δημιουργία διαφορετικών modules για κάθε είδος λειτουργικότητας στο διαδίκτυο. Στην εργασία τους, αναλύονται τα τρία modules που έχουν δημιουργήσει, με το πρώτο να αφορά την δυνατότητα χρήσης του Gmail, το δεύτερο το Wikipedia, ενώ το τρίτο την ικανότητα αναζήτησης και αλληλεπίδρασης με τα αποτελέσματα του Google. Όλα τα modules, ενώνονται μέσω ενός μενού, το οποίο μπορεί να χρησιμοποιήσει ο χρήστης για την επιλογή του επιθυμητού module, αλλά και την ρύθμιση παραμέτρων, όπως η ταχύτητα και η προφορά του virtual assistant. Για τις δυνατότητες αναγνώρισης ομιλίας χρησιμοποιείται η βιβλιοθήκη της Google για Python, ενώ αντίστοιχα για την επικοινωνία του συστήματος με τον χρήστη χρησιμοποιείται η text-to-speech βιβλιοθήκη της Python pyttsx3. Κάθε module είναι γραμμένο σε Python, περιλαμβάνει κώδικα αναφορικά με τις δυνατότητες του και χρησιμοποιείται Selenium για την αυτοματοποίηση του. Παράλληλα, γίνεται και χρήση εργαλείων όπως το BeautifulSoup για scraping στο περιεχόμενο μίας σελίδας, αλλά και τεχνικές machine learning για δυνατότητες, όπως απαντήσεις σε ερωτήσεις του χρήστη στην περίπτωση του Wikipedia module.

Οι Rallabhandy και Rodda (2020), αναφέρουν στην εργασία τους ότι η παρουσία του διαδικτύου σε κάθε πτυχή της καθημερινότητας έχει ως αποτέλεσμα την μεγάλη προώθηση των ηλεκτρονικών καταστημάτων. Ωστόσο, τέτοιου είδους ιστότοποι είναι απρόσιτοι σε ανθρώπους με προβλήματα όρασης και βασισμένοι στο γεγονός αυτό χρησιμοποιούν την τεχνητή νοημοσύνη για την υποβοήθηση τους αναφορικά με την εμπειρία των αγορών στο διαδίκτυο. Το σύστημα που έφτιαξαν, κάνει εφικτή την αλληλεπίδραση ενός χρήστη εξ ολοκλήρου μέσω της φωνής του από την φάση του logging μέχρι την προσθήκη ενός προϊόντος στο καλάθι αγορών. Αρχικά, ο χρήστης αναφέρει μέσω της φωνή του το URL του ηλεκτρονικού καταστήματος που επιθυμεί να επισκεφτεί και αυτό θα ανοίξει μέσω του Google Chrome. Για τις διαδικασίες του login και registration, χρησιμοποιείται τεχνολογία αναγνώρισης προσώπου σε συνδυασμό με την αναγνώριση ομιλίας, όπου ο χρήστης για παράδειγμα μπορεί να πει Login και η διαδικασία να ολοκληρωθεί μέσω της αναγνώρισης προσώπου. Το επόμενο κομμάτι του συστήματος, αφορά στην ικανότητα για αναζήτηση προϊόντος και η διαδικασία παρουσιάζεται στο amazon.in. Ο χρήστης, θα εκφωνήσει μέσω της φωνής του το προϊόν που ψάχνει και αφού η είσοδος του μετατραπεί σε κείμενο θα προσκολληθεί στο URL της σελίδας έτσι ώστε να επιστραφούν τα εκάστοτε προϊόντα. Σε εκείνο το σημείο δίνονται δύο επιλογές. Η πρώτη, αφορά είτε την πραγματοποίηση νέας αναζήτησης είτε την επιλογή προϊόντων βάση ενός εύρους τιμών. Η δεύτερη, αφορά την δυνατότητα του συστήματος να επιλέξει τα τρία καλύτερα με βάση τα ratings χρησιμοποιώντας web scraping τεχνικές και στην συνέχεια τα εκφωνεί. Με την επιλογή των τριών καλύτερων προϊόντων, το σύστημα θα τα εκφωνήσει συμπεριλαμβανομένων των περιγραφών τους. Έπειτα, ο χρήστης μπορεί να επιλέξει ένα από αυτά

χρησιμοποιώντας τις λέξεις first/second/third. Ανάλογα με το προϊόν που θα επιλεγεί, το σύστημα θα διαβάσει τις διαθέσιμες παραλλαγές και θα περιμένει από τον χρήστη να επιλέξει μία από αυτές. Τέλος, με την επιλογή της εκάστοτε παραλλαγής το προϊόν θα προστεθεί στο καλάθι με την χρήση του Selenium web driver. Το συγκεκριμένο σύστημα υλοποιήθηκε σε python και τα πακέτα που χρειάστηκαν ήταν τα Gtts, speech_recognition, pyAudio αναφορικά με την δυνατότητα ενσωμάτωσης δυνατοτήτων speech, bs4 για scraping και OpenCV για αναγνώριση προσώπου.

2. Web Speech API

2.1 Εισαγωγή

Πίσω στο 2012, η κοινοπραξία του παγκόσμιου ιστού (World Wide Web Consortium) και συγκεκριμένα το Speech API Community Group παρουσίασε ένα API (Application Programming Interface) το οποίο θα ανοίξει νέους ορίζοντες στο πεδίο της επικοινωνίας ανθρώπου υπολογιστή. Συγκεκριμένα, το Web Speech API προσφέρει την δυνατότητα στους προγραμματιστές του διαδικτύου να πειραματιστούν με την δημιουργία νέων εφαρμογών ή να ενσωματώσουν σε ήδη υπάρχουσες την ικανότητα αλληλεπίδρασης με την ομιλία. Η ικανότητα αυτή, προσδίδει μία νέα διάσταση στον τελικό χρήστη, σύμφωνα με την οποία η είσοδος δεδομένων καθίσταται γρηγορότερη, ενώ ταυτόχρονα δεν απαιτούνται καθόλου τα χέρια και τα μάτια ως μέσο διάδρασης, κάτι το οποίο μπορεί να φανεί εξαιρετικά χρήσιμο σε καταστάσεις όπως η αλληλεπίδραση την ώρα της οδήγησης. Παράλληλα, σημαντική είναι και η συνεισφορά ενός τέτοιου εργαλείου στο κομμάτι της ηλεκτρονικής προσβασιμότητας, αφού για την πλοήγηση του στο διαδίκτυο και την συμπλήρωση για παράδειγμα μιας φόρμας, ένας χρήστης με προβλήματα όρασης δεν χρειάζεται να έρχεται σε επαφή με βοηθητικό εξοπλισμό τόσο σε επίπεδο υλικού, όσο και λογισμικού με την χρήση εργαλείων όπως είναι ο αναγνώστης οθόνης (screen reader) και η ανανεώσιμη οθόνη μπράιγ (refreshable braille display). Η χρήση τέτοιων τεχνολογιών απαιτεί σε πολλές περιπτώσεις την καταβολή υψηλού χρηματικού ποσού για την απόκτηση του εξοπλισμού, καθώς επίσης και την εκπαίδευση του χρήστη με αυτόν, κριτήρια τα οποία μπορούν τελικά να τον απωθήσουν από το να χρησιμοποιήσει το διαδίκτυο, μη αποκομίζοντας έτσι τα πολλαπλά οφέλη του. Επιπλέον, λόγω της μεγάλης τεχνολογικής άνθησης στα πεδία της τεχνητής νοημοσύνης που αφορούν την αναγνώριση ομιλίας (speech recognition), την κατανόηση φυσικής γλώσσας (Natural Language Understanding) και την σύνθεση φωνής από κείμενο (Text to Speech), όλο και περισσότεροι χρήστες έχουν αρχίσει να ενσωματώνουν στην καθημερινότητα τους συσκευές που χρησιμοποιούν τις συγκεκριμένες τεχνικές και ακούν στο όνομα φωνητικοί βοηθοί (voice assistants). Χαρακτηριστικά παραδείγματα τέτοιων τεχνολογιών αποτελούν το Amazon Echo και το Google Home, τα οποία χρησιμοποιούνται για λειτουργίες όπως η αναζήτηση πληροφοριών, η αγορά προϊόντων κ.α. Μάλιστα, στην περίπτωση του τελευταίου αξίζει να σημειωθεί ο όρος Voice Commerce, ο οποίος ακούγεται όλο και περισσότερο και αφορά την πραγματοποίηση ηλεκτρονικών αγορών μέσω της χρήσης ενός φωνητικού βοηθού. Ωστόσο, η αλληλεπίδραση με φωνητικούς βοηθούς και γενικότερα η χρήση φωνητικών εντολών είναι συχνότερη σε κινητά ή συσκευές όπως το Amazon Echo. Έτσι λοιπόν, επιχειρήσεις που θα ενσωματώσουν τέτοιες δυνατότητες στις διαδικτυακές τους εφαρμογές, θα αποκτήσουν προτέρημα έναντι των ανταγωνιστών τους, προσφέροντας μία πρωτόγνωρη εμπειρία στους χρήστες η οποία θα χαρακτηρίζεται από cutting edge τεχνολογίες. Καταλαβαίνει κανείς, λοιπόν, τα πολλαπλά οφέλη που μπορεί να προσφέρει ένα εργαλείο με δυνατότητες

αναγνώρισης και σύνθεσης ομιλίας όπως το Web Speech API. Το παρόν κεφάλαιο, αναλύει τις λειτουργίες του συγκεκριμένου API και μπορεί να αποτελέσει έναν οδηγό για οποιονδήποτε επιθυμεί να ενσωματώσει δυνατότητες ομιλίας στην εφαρμογή του. Παράλληλα, η κατανόηση του τρόπου λειτουργίας του είναι απαραίτητη για όποιον επιθυμεί να κατανοήσει την τεχνική υποδομή του ηλεκτρονικού καταστήματος που περιγράφεται στο επόμενο κεφάλαιο.

2.2 Πορεία δημιουργίας Web Speech API

Με στόχο την έρευνα γύρω από το πώς θα μπορούσαν να ενσωματωθούν δυνατότητες αναγνώρισης και σύνθεσης ομιλίας σε HTML5 σελίδες, το W3C (World Wide Web Consortium) στο πλαίσιο του προγράμματος Incubator Activity σχημάτισε το HTML Speech Incubator Group. Ο ρόλος ενός W3C Incubator Group περιλαμβάνει διαδικασίες όπως η έρευνα πάνω σε νέες ιδέες, η ανάπτυξη σεναρίων χρήσης, καθώς επίσης και ο καθορισμός των απαιτήσεων για μία ιδέα η οποία μελλοντικά είναι πιθανό να μετατραπεί σε ένα W3C Standard. Να σημειωθεί, πως τον Απρίλιο του 2012 το πρόγραμμα Incubator Activity σταμάτησε αριθμώντας 28 Incubator Groups, μεταξύ των οποίων και το HTML Speech Incubator, τα οποία σχηματίστηκαν στα πλαίσια λειτουργίας του, τοποθετώντας καθένα από αυτά το δικό του λιθαράκι για τον σχηματισμό ενός διαδικτύου απεριόριστων δυνατοτήτων. Η κοινοπραξία του παγκόσμιου ιστού ανακοίνωσε πως το πρόγραμμα Incubator Activity αποτέλεσε πηγή πολύτιμης εμπειρίας, βάση της οποίας δημιούργησε ένα ακόμη πιο σταθερό και περιεκτικό πρόγραμμα με όνομα Community and Business Groups. Σύμφωνα με το W3C Incubator Activity, στην περίπτωση του HTML Speech Incubator Group στόχος ήταν η μελέτη σκοπιμότητας σχετικά με την ενσωμάτωση τεχνολογιών ομιλίας σε HTML5 σελίδες, με τρόπο που να αξιοποιούνται οι δυνατότητες και των δύο τεχνολογιών, έτσι ώστε ως τελικό αποτέλεσμα να παρέχεται μία υψηλής ποιότητας εμπειρία αλληλεπίδρασης μέσω της ομιλίας η οποία θα είναι ανεξάρτητη από το πρόγραμμα περιήγησης (browser independent), ενώ παράλληλα σημαντική παράμετρο αποτελεί και η αποφυγή καταστάσεων όπως η περιττή επικάλυψη προτύπων (standards overlap). Στις 6 Δεκεμβρίου του 2011, δημοσιεύτηκε η τελική αναφορά του HTML Speech Incubator Group, η οποία περιέχει περιπτώσεις χρήσης που ανέπτυξαν τα μέλη του Group με στόχο την κάλυψη διαφορετικών σεναρίων αλληλεπίδρασης μέσω της ομιλίας, απαιτήσεις που καθόρισαν, καθώς επίσης και σχεδιαστικές αποφάσεις που έλαβαν. Τέλος, περιλαμβάνονται προτάσεις για HTML στοιχεία (HTML elements), ένα Javascript API αλλά και ένα πρωτόκολλο για την υλοποίησή τους. Η συγκεκριμένη αναφορά αποτελεί τα θεμέλια πάνω στα οποία θα στηριχτεί ένα διαδίκτυο, το οποίο θα δίνει στους χρήστες των εφαρμογών την δυνατότητα της αλληλεπίδρασης μέσω της ομιλίας, ενώ παράλληλα προγραμματιστές από όλο τον κόσμο θα έχουν την ικανότητα να υλοποιήσουν πρωτόγνωρες εμπειρίες διάδρασης.

Όπως αναφέρθηκε παραπάνω, με την λήξη του προγράμματος W3C Incubator Activity, η κοινοπραξία του παγκόσμιου ιστού στηριζόμενη στην εμπειρία που έλαβε στα πλαίσια της λειτουργίας του,

δημιούργησε ένα νέο πρόγραμμα με όνομα Community and Business Groups. Σύμφωνα με την επίσημη σελίδα του προγράμματος αυτού, ένα Community Group στην ουσία αποτελεί ένα φόρουμ όπου οι προγραμματιστές Ιστού (Web developers) και άλλα ενδιαφερόμενα μέρη αναπτύσσουν προδιαγραφές (specifications) και δοκιμαστικές σουίτες (test suites), διεξάγουν συζητήσεις και συνδέονται με την διεθνή κοινότητα ειδικών του W3C σε θέματα ιστού. Επιπλέον, τα Community Groups δίνουν την δυνατότητα σε οποιονδήποτε επιθυμεί να μοιραστεί στο W3C τις ιδέες του σχετικά με το διαδίκτυο με ενδεχόμενο την προτυποποίηση (standardization) τους στο μέλλον. Αντίστοιχα με τα Community Groups, τα Business Groups αφορούν τα άτομα που επιθυμούν να έχουν αντίκτυπο στην ανάπτυξη του διαδικτύου στο βραχυπρόθεσμο μέλλον. Για να επιτευχθεί αυτό, παρέχεται ένα φόρουμ για την συνεργασία αυτών των ατόμων με άλλα ενδιαφερόμενα μέρη που έχουν παρόμοιες ιδέες και μπορεί να θεωρούνται είτε μέλη του W3C είτε όχι.

Σήμερα, ο αριθμός των Community Groups ισούται με 356 μέσα στα οποία περιλαμβάνεται και το Speech API. Το συγκεκριμένο Group, δημιουργήθηκε στις 3 Απριλίου του 2012 και διαθέτει 65 συμμετέχοντες. Ο λόγος της δημιουργίας του αφορούσε την υλοποίηση ενός Javascript Speech API, το οποίο θα μπορεί να υποστηρίξει την πλειοψηφία των περιπτώσεων χρήσης που ορίζονται στην τελική αναφορά του HTML Speech Incubator Group. Σύμφωνα με την σελίδα του συγκεκριμένου Group, το API αυτό αποτελεί ένα απλοποιημένο υποσύνολο των προτάσεων που περιέχονται στην τελική αναφορά. Αυτό, διότι το Speech API Community Group θεωρεί πως η δημιουργία ενός τέτοιου API θα επιταχύνει το κομμάτι της υλοποίησης, το κομμάτι των δοκιμών σχετικά με το αν θα μπορεί να αλληλεπιδρά και να ανταλλάσει δεδομένα με άλλα συστατικά λογισμικού χωρίς να υπάρχουν προβλήματα συμβατότητας (interoperability testing), το κομμάτι της προτυποποίησης (standardization) και τέλος την χρήση του API από τους προγραμματιστές. Με την δημιουργία του συγκεκριμένου API, οι προγραμματιστές θα μπορούν να ενσωματώνουν scripts στις εφαρμογές τους, τα οποία θα χαρακτηρίζονται από λειτουργίες όπως την παραγωγή text to speech ως έξοδο (output), speech recognition τα αποτελέσματα του οποίου θα μπορούν για παράδειγμα να χρησιμοποιούνται ως είσοδος (input) σε φόρμες, καθώς επίσης και δυνατότητα για continuous dictation, το οποίο επιτρέπει την αναγνώριση μεγάλων φράσεων και θα μπορούσε να χρησιμοποιηθεί σε περιπτώσεις όπως είναι η υπαγόρευση ενός μηνύματος ηλεκτρονικού ταχυδρομείου.

Στην επίσημη σελίδα του Speech API Community Group, φαίνεται πως στις 13 Ιουνίου του 2012 δημοσιεύτηκε το πρώτο προσχέδιο (draft) του Speech Javascript API το οποίο αποτελεί τον οδηγό που πρέπει να ακολουθήσουν προγραμματιστές για την υλοποίηση του σε ένα πρόγραμμα περιήγησης. Το συγκεκριμένο specification, δεν αποτελεί μέχρι και σήμερα ένα W3C Standard. Παρά το γεγονός αυτό, προγράμματα περιήγησης όπως το Google Chrome έχουν υλοποιήσει το Web Speech API. Τα περιεχόμενα του specification περιλαμβάνουν την εισαγωγή (Introduction), τα σενάρια χρήσης (Use Cases) που

υποστηρίζει το συγκεκριμένο API από το σύνολο των σεναρίων που περιλαμβάνονται στην τελική αναφορά του HTML Speech Incubator Group, ζητήματα που προκύπτουν σχετικά με την ασφάλεια και ιδιωτικότητα (Security and privacy considerations), περιγραφή του API (API Description) και τέλος κάποια παραδείγματα χρήσης του (Examples).

Στην παρούσα ενότητα, μελετήθηκε η πορεία που έχει ακολουθηθεί από το W3C μέχρι και την δημιουργία του Web Speech API. Συγκεκριμένα, παρουσιάστηκε το πρώτο Incubator Group που σχηματίστηκε στα πλαίσια του προγράμματος Incubator Activity με όνομα HTML Speech Incubator Group, στόχος του οποίου αποτελούσε η μελέτη γύρω από το πως μπορούν να ενσωματωθούν δυνατότητες αναγνώρισης και σύνθεσης ομιλίας σε HTML5 σελίδες. Έπειτα, αναφέρθηκαν τα περιεχόμενα της τελικής αναφοράς του συγκεκριμένου group, η οποία συγκεντρώνει τα αποτελέσματα της έρευνας τους. Στην συνέχεια, παρουσιάστηκε το πρόγραμμα που δημιουργήθηκε από το W3C μετά το Incubator Activity, με όνομα Community and Business Groups. Ακολούθως, παρουσιάστηκε το Speech API Community Group, το οποίο δημιούργησε το Web Speech API ως ένα πλήρως λειτουργικό υποσύνολο της τελικής αναφοράς του HTML Speech Incubator Group. Στην επόμενη ενότητα, περιγράφονται συνοπτικά τα περιεχόμενα του Web Speech API specification.

2.3 Web Speech API specification

Για κάθε τεχνολογία με την οποία αλληλεπιδρούμε σήμερα στο διαδίκτυο και προέρχεται από την κοινοπραξία του παγκόσμιου ιστού, υπάρχει ένα specification, το οποίο δίνει κατευθυντήριες οδηγίες για το πως μπορεί αλλά και το ποιους κανόνες πρέπει να πληρεί αυτή για να υλοποιηθεί σε ένα πρόγραμμα περιήγησης. Οι συγκεκριμένες προδιαγραφές (specifications), δεν αποτελούν ένα εγχειρίδιο χρήσης και συνεπώς δεν απευθύνονται στους προγραμματιστές που θέλουν να χρησιμοποιήσουν την αντίστοιχη τεχνολογία. Αντίθετα, αφορά αυτούς που πρόκειται να υλοποιήσουν μία τεχνολογία, έτσι ώστε να μπορούν έπειτα οι προγραμματιστές διαδικτυακών εφαρμογών να την αξιοποιήσουν. Έτσι λοιπόν, στην παρούσα υποενότητα παρουσιάζεται συνοπτικά το περιεχόμενο του Web Speech API specification, ενώ στις επόμενες αναλύεται ο τρόπος χρήσης του. Αυτό, διότι η αλληλεπίδραση με το περιεχόμενο ενός specification μπορεί να συμβάλει στην καλύτερη κατανόηση των χαρακτηριστικών της τεχνολογίας που πρόκειται να χρησιμοποιηθεί.

Όπως αναφέρθηκε παραπάνω, σε ένα specification περιλαμβάνονται οι κανόνες που πρέπει να πληρούνται όταν πρόκειται να υλοποιηθεί μία τεχνολογία. Αυτό, συμβαίνει διότι δεν μπορεί ο κάθε προγραμματιστής να προβαίνει στην υλοποίηση με δικούς του τρόπους και παραδοχές, αφού κάτι τέτοιο θα οδηγήσει σε ένα διαδίκτυο, του οποίου οι τεχνολογίες λειτουργούν διαφορετικά ανά πρόγραμμα περιήγησης. Για να αποφευχθεί αυτό, μέσα στις προδιαγραφές εμφανίζονται συχνά οι όροι normative και non-normative ή αλλιώς informative. Σύμφωνα με το MDN, στην πρώτη περίπτωση, όταν δηλαδή

εμφανίζεται σε ένα κομμάτι του specification ο όρος normative, τότε το περιεχόμενο αυτού του σημείου πρέπει να υλοποιηθεί ως έχει, αφού αποτελεί έναν κανόνα. Στην δεύτερη περίπτωση, το κομμάτι του specification στο οποίο εμφανίζεται ο όρος non-normative ή informative, έχει ως στόχο του να βοηθήσει τον εκάστοτε προγραμματιστή να κατανοήσει καλύτερα το ίδιο το specification, να δείξει ένα παράδειγμα ή μία καλή πρακτική, αλλά σε κάθε μία από τις τρεις περιπτώσεις δεν θεωρείται ότι πρέπει να ακολουθηθεί ως κανόνας. Έτσι λοιπόν, όπως σε κάθε specification του W3C έτσι και στην περίπτωση του Web Speech API συναντάμε τους αντίστοιχους όρους σε κάθε ενότητα. Στην συνέχεια, ακολουθεί μία περιγραφή των 5 κεφαλαίων η οποία συνοψίζει τόσα τα normative, όσο και τα informative κομμάτια του specification το οποίο αναφέρει το μονοπάτι που πρέπει να ακολουθηθεί σε περίπτωση που προγραμματιστές του διαδικτύου θέλουν να υλοποιήσουν την συγκεκριμένη τεχνολογία σε ένα πρόγραμμα περιήγησης όπως έγινε στην περίπτωση του Google Chrome.

Το πρώτο κεφάλαιο του specification θεωρείται informative και χωρίζεται σε δύο παραγράφους. Στην πρώτη, γίνεται μία πολύ σύντομη παρουσίαση των δυνατοτήτων του συγκεκριμένου API. Στην δεύτερη παράγραφο, αναφέρονται πληροφορίες για το πως αυτό σχετίζεται με την τελική αναφορά του HTML Speech Incubator Group. Ειδικότερα, τονίζεται πως το συγκεκριμένο API αποτελεί ένα υποσύνολο της τελικής αναφοράς, το οποίο είναι πλήρως λειτουργικό αλλά δεν περιλαμβάνει τεχνολογίες που προτάθηκαν από το HTML Speech Incubator Group, όπως για παράδειγμα τα νέα HTML στοιχεία (HTML elements). Ωστόσο, σύμφωνα με το specification αν και υποσύνολο, μπορεί να υποστηρίξει σχεδόν όλα τα σενάρια χρήσης που έχουν οριστεί στην τελική αναφορά.

Αναφέρθηκαν μέχρι τώρα αρκετές φορές τα σενάρια χρήσης, που ορίστηκαν από το HTML Speech Incubator Group, σχετικά με τα είδη αλληλεπίδρασης που θα μπορούσε να έχει ένας χρήστης σε μία εφαρμογή του διαδικτύου, η οποία υποστηρίζει δυνατότητες αναγνώρισης και σύνθεσης ομιλίας. Συγκεκριμένα, τα σενάρια αυτά είναι 17 και στο δεύτερο κεφάλαιο του specification αναγράφεται πως το API μπορεί να υποστηρίξει τα 16 από αυτά, κάτι το οποίο δίνει μεγάλες προοπτικές στους προγραμματιστές του διαδικτύου είτε για νέα project, είτε για προσθήκες στα ήδη υπάρχοντα. Παρακάτω, δίνονται ονοματικά τα 17 σενάρια καθώς έχει μεγάλο ενδιαφέρον να δει κανείς πως είχαν οραματιστεί τα μέλη του HTML Speech Incubator Group ένα διαδίκτυο με δυνατότητες Speech. Να σημειωθεί, πως η περιγραφή κάθε σεναρίου μπορεί να βρεθεί στο 4 κεφάλαιο του HTML Speech Incubator Group Final Report σε αγγλική γλώσσα. Επιπλέον, αυτονόητο είναι πως το κεφάλαιο 2 στο specification που περιγράφεται έχει οριστεί ως informative, αφού δεν αναφέρει κάποιο κανόνα που πρέπει να πληρείται στην υλοποίηση, αλλά αντίστοιχα βοηθάει τον αναγνώστη να κατανοήσει καλύτερα τις δυνατότητες του API.

1. Voice Web Search
2. Speech Command interface

3. Domain Specific Grammars Contingent on Earlier Inputs
4. Continuous Recognition of Open Dialog
5. Domain Specific Grammars Filling Multiple Input Fields
6. Speech UI present when no visible UI need be present
7. Voice Activity Detection
8. Temporal Structure of Synthesis to Provide Visual Feedback
9. Hello World
10. Speech Translation
11. Speech Enabled Email Client
12. Dialog Systems
13. Multimodal Interaction
14. Speech Driving Directions
15. Multimodal Video Game
16. Multimodal Search
17. Rerecognition

Το specification του Web Speech API, αναφέρει ότι από τα 17 σενάρια χρήσης υποστηρίζονται τα 16. Το σενάριο χρήσης που δεν υποστηρίζεται αφορά την δυνατότητα για Rerecognition. Ωστόσο, το specification αναφέρει ότι σε πολλές περιπτώσεις η χρήση του Rerecognition μπορεί να αντικατασταθεί είτε από την χρήση ενός μεγαλύτερου grammar, είτε από τον συνδυασμό πολλαπλών grammars. Να σημειωθεί, πως ο όρος grammar αναφέρεται σε μία από τις δυνατότητες που παρέχει το Web Speech API στο κομμάτι της αναγνώρισης ομιλίας και αφορά ένα σύνολο λέξεων, οι οποίες δίνονται έτσι ώστε το εκάστοτε web service, που πραγματοποιεί την αναγνώριση, να είναι σε θέση να τις καταλάβει. Αυτή η δυνατότητα, περιγράφεται στην επόμενη ενότητα, η οποία αφορά το πως μπορεί να χρησιμοποιηθεί το Web Speech API αναλύοντας διεπαφές (interfaces), μεθόδους (methods), γεγονότα (events), ιδιότητες (properties) κλπ.

Στο τρίτο κεφάλαιο του specification, το Speech API Community Group, αρχικά, αναφέρει τρεις κανόνες, τους οποίους πρέπει να πληροί ένα πρόγραμμα περιήγησης που υλοποιεί το Web Speech API. Σύμφωνα με τον πρώτο κανόνα, ένας χρήστης θα πρέπει να δώσει την άδεια του πριν ξεκινήσει η διαδικασία της αναγνώρισης ομιλίας. Δεδομένου αυτού, προτείνονται τρεις τρόποι, με τους οποίους ο χρήστης θα μπορούσε να δίνει την άδεια του, όπως για παράδειγμα με την εμφάνιση ενός prompt πριν ξεκινήσει η αναγνώριση το οποίο θα ρωτάει τον χρήστη εάν συμφωνεί να αρχίσει η διαδικασία. Με μία τέτοια μέθοδο είναι προφανές, ότι επιχειρείται από την πλευρά των δημιουργών του Web Speech API να αποφευχθεί όσο το δυνατόν περισσότερο γίνεται η χρήση της αναγνώρισης ομιλίας κακόβουλα χωρίς να το έχει αντιληφθεί ο τελικός χρήστης. Ο δεύτερος κανόνας, αναφέρει πως το πρόγραμμα περιήγησης που

θα υλοποιήσει το Web Speech API θα πρέπει να δίνει μία ένδειξη στον χρήστη όταν η αναγνώριση ομιλίας είναι ενεργοποιημένη. Για παράδειγμα, μία τέτοια λειτουργικότητα θα μπορούσε να υλοποιηθεί με μία ηχητική ειδοποίηση, η οποία θα ακούγεται όταν ξεκινάει η αναγνώριση. Ο τρίτος και τελευταίος κανόνας του κεφαλαίου, αναφέρει ότι την πρώτη φορά που θα χρησιμοποιηθεί η αναγνώριση ομιλίας θα πρέπει να δοθεί στον χρήστη μία επεξήγηση για το τι είναι ακριβώς, καθώς επίσης και το πως μπορεί να την σταματήσει σε περίπτωση που δεν επιθυμεί να αλληλεπιδράσει με την εκάστοτε εφαρμογή μέσω της φωνής του. Παρόλο που στο συγκεκριμένο κομμάτι του κεφαλαίου δεν αναγράφεται αν θεωρείται normative ή αντίστοιχα informative, καταλαβαίνουμε ότι ισχύει η πρώτη περίπτωση, αφού λίγο πριν το τέλος της αναφοράς, στην ενότητα με ονομασία Conformance, αναφέρεται πως όλο το κείμενο θεωρείται normative εκτός από τα σημεία στα οποία αναφέρεται ρητά ο όρος non-normative. Εκτός από τους τρεις κανόνες που προαναφέρθηκαν, το τρίτο κεφάλαιο του specification περιλαμβάνει και μία non-normative ενότητα στην οποία αναφέρονται κάποιοι προβληματισμοί σε θέματα ασφάλειας που είναι πιθανό να προκύψουν από την χρήση της αναγνώρισης ομιλίας. Για παράδειγμα, αναφέρεται το σενάριο χρήσης της ως μέσο για την συμπλήρωση ενός πεδίου, το οποίο αφορά κωδικό πρόσβασης. Μια τέτοια περίπτωση είναι προφανές πως μπορεί να εκθέσει τον τελικό χρήστη σε κινδύνους. Ωστόσο, σύμφωνα με το specification η απόφαση βρίσκεται στο χέρι του χρήστη, επιλέγοντας συνειδητά εάν θα δώσει ως είσοδο (input) στην εκάστοτε εφαρμογή τον κωδικό του με χρήση της ομιλίας ως μέσο. Πέρα από τον συγκεκριμένο προβληματισμό, αναφέρονται και κάποιες ακόμα περιπτώσεις κατά τις οποίες η αναγνώριση ομιλίας μπορεί να χρησιμοποιηθεί κακόβουλα, όπως για παράδειγμα το σενάριο κατά το οποίο ο χρήστης μπορεί να θεωρεί πως έχει σταματήσει, ενώ αυτή πραγματοποιείται εν αγνοία του.

Στο τέταρτο κεφάλαιο, περιλαμβάνεται ο πυρήνας του specification. Περιεχόμενο του αποτελεί η δομή του API, παρουσιάζοντας πως πρέπει να υλοποιηθούν τα απαιτούμενα συστατικά του, όπως διεπαφές (interfaces), μέθοδοι (methods), γεγονότα (events) κλπ. Η σύνταξη που χρησιμοποιείται στο συγκεκριμένο κεφάλαιο είναι πιθανό να φανεί πρωτόγνωρη για έναν προγραμματιστή του διαδικτύου που σκοπεύει να αξιοποιήσει το API στην εφαρμογή του. Αυτό, αφού όπως προαναφέρθηκε το συγκεκριμένο specification αφορά προγραμματιστές που επιθυμούν να υλοποιήσουν το Web Speech API σε ένα πρόγραμμα περιήγησης. Αυτονόητο είναι πως το συγκεκριμένο κομμάτι, δεν θα μπορούσε να θεωρηθεί κάτι διαφορετικό από normative.

Στο πέμπτο και τελευταίο κεφάλαιο, παρατίθενται τέσσερα παραδείγματα, από την χρήση του Web Speech API, για αναγνώριση ομιλίας και δύο για σύνθεση. Τα συγκεκριμένα παραδείγματα, αφορούν την χρήση του API από την οπτική των προγραμματιστών που επιθυμούν να το αξιοποιήσουν στην εφαρμογή τους και συνεπώς απεικονίζουν τον τρόπο με τον οποίο μπορούν να χρησιμοποιηθούν οι δυνατότητες αναγνώρισης και σύνθεσης ομιλίας στα προγράμματα περιήγησης που υποστηρίζεται το Web Speech API.

Τέλος, αφού το συγκεκριμένο κεφάλαιο δίνει παραδείγματα χρήσης είναι προφανές πως θεωρείται non-normative.

Στην παρούσα ενότητα, πραγματοποιήθηκε μία συνοπτική περιγραφή των κεφαλαίων του Web Speech API specification, το οποίο έχει γραφεί με στόχο την χρήση του από προγραμματιστές που επιθυμούν να το υλοποιήσουν σε ένα πρόγραμμα περιήγησης. Στις ενότητες που ακολουθούν, αναλύεται το API από την οπτική ενός προγραμματιστή που θέλει να το χρησιμοποιήσει για να ενσωματώσει στην εφαρμογή του ικανότητες αναγνώρισης και σύνθεσης ομιλίας και συνεπώς μελετώνται τα βασικά interfaces του API τα οποία περιλαμβάνουν μεθόδους (methods), γεγονότα (events), ιδιότητες (properties) κ.α.

2.4 Περιγραφή

Όπως έχει προαναφερθεί, η αλληλεπίδραση με την ομιλία είναι μία δυνατότητα που μπορεί να φέρει μία πληθώρα πλεονεκτημάτων. Τι εργαλεία παρέχει όμως το Web Speech API στους προγραμματιστές του διαδικτύου έτσι ώστε αυτοί να δημιουργήσουν τέτοιου είδους εφαρμογές;

Αρχικά, το Web Speech API μπορεί να διακριθεί σε δύο βασικά interfaces. Κάθε ένα από τα αυτά, διαθέτει μία πληθώρα από μεθόδους (methods), γεγονότα (events) και ιδιότητες (properties), τα οποία χρησιμοποιούνται για την ενσωμάτωση αλλά και παραμετροποίηση τόσο των δυνατοτήτων αναγνώρισης όσο και σύνθεσης ομιλίας. Το πρώτο, ακούει στο όνομα SpeechRecognition και δίνει την δυνατότητα σε μία εφαρμογή να δέχεται ως είσοδο (input) από τον χρήστη ομιλία και στην συνέχεια να την μετατρέπει σε κείμενο. Το δεύτερο, ονομάζεται SpeechSynthesis και δίνει την δυνατότητα σε μία εφαρμογή να μετατρέψει οποιοδήποτε κείμενο της δοθεί σε ομιλία. Καταλαβαίνουμε λοιπόν, πως το Web Speech API δίνει στους προγραμματιστές τα εφόδια με τα οποία μπορούν να δημιουργήσουν διαδικτυακές εφαρμογές, οι οποίες να δέχονται δεδομένα εισόδου (input) με ομιλία αλλά και να δίνουν ως έξοδο (output) ομιλία, έχοντας έτσι την δυνατότητα να υλοποιήσουν πλήθος σεναρίων, όπως αυτά της τελικής αναφοράς του HTML Speech Incubator Group. Σε κάθε περίπτωση, είτε μιλάμε δηλαδή για αναγνώριση είτε για σύνθεση ομιλίας, η γλώσσα προγραμματισμού με την οποία υλοποιήθηκαν οι δυνατότητες του Web Speech API είναι εξ ολοκλήρου η Javascript.

Όπως γνωρίζουμε από το πεδίο των τεχνολογιών διαδικτύου, σε μία εφαρμογή, το κομμάτι του κώδικα που εκτελείται στην πλευρά του πελάτη (client side) είναι στο μεγαλύτερο μέρος του event – driven. Η συγκεκριμένη ορολογία, αναφέρεται στο ότι μεγάλο μέρος της αλληλεπίδρασης με μία εφαρμογή στηρίζεται σε μία πληθώρα από γεγονότα (events), τα οποία πυροδοτούνται (fired) με διάφορους τρόπους, όπως για παράδειγμα με το πάτημα ενός κουμπιού από τον τελικό χρήστη, το φόρτωμα μιας σελίδας κ.α. Χτισμένο πάνω σε αυτή την φιλοσοφία, το Web Speech API είναι event – based, κάτι το οποίο του δίνει την δυνατότητα να λειτουργεί ασύγχρονα επιτρέποντας έτσι στην εκάστοτε εφαρμογή να εκτελεί

παράλληλα με την αναγνώριση – σύνθεση ομιλίας οποιαδήποτε άλλη λειτουργικότητα έχουν ορίσει οι δημιουργοί της. Για παράδειγμα, όταν η αναγνώριση ομιλίας είναι ενεργοποιημένη, ο χρήστης μπορεί να αλληλεπιδρά ταυτόχρονα με ένα slider, το οποίο χρησιμοποιεί Javascript για ένα κομμάτι της λειτουργικότητάς του χωρίς να υπάρξει πρόβλημα. Κάτι τέτοιο, δεν θα ήταν εφικτό εάν η φύση του API δεν ήταν ασύγχρονη.

Ένα από τα μεγαλύτερα πλεονεκτήματα του Web Speech API αφορά την ευκολία στην χρήση του. Οποιοσδήποτε προγραμματιστής θέλει να αξιοποιήσει τις δυνατότητες του συγκεκριμένου εργαλείου, δεν έρχεται σε επαφή και ούτε χρειάζεται να παραμετροποιήσει το εκάστοτε web service που χρησιμοποιείται από το API για τις λειτουργίες αναγνώρισης και σύνθεσης ομιλίας. Αντίθετα, μπορεί να ενσωματώσει κατευθείαν δυνατότητες αναγνώρισης και σύνθεσης ομιλίας μέσω της χρήσης πληθώρας interfaces που παρέχει το API, καθένα από τα οποία περιλαμβάνει διαφορετικές ιδιότητες, γεγονότα και μεθόδους. Τα τελευταία, θα αναλάβουν με την σειρά τους λειτουργίες όπως την μεταβίβαση της εισόδου που δίνει ο χρήστης μέσω της φωνής του στο αντίστοιχο web service, για να πραγματοποιηθεί η αναγνώριση και να σταλεί τελικά πίσω ως κείμενο. Κάτι τέτοιο είναι πολύ χρήσιμο, αφού οι προγραμματιστές δεν χρειάζεται να γνωρίζουν τους μηχανισμούς κάτω από το API και μπορούν να αφοσιωθούν στην δημιουργία πρωτόγωνων διεπαφών, οι οποίες θα προσφέρουν στον τελικό χρήστη μία εμπειρία υψηλής ποιότητας. Σημαντικό είναι να σημειωθεί, πως σε έναν user agent, στον οποίο είναι επιθυμητό να ενσωματωθεί το Web Speech API, θα πρέπει να υλοποιηθεί και ο τρόπος επικοινωνίας με το web service που θα επιλεγεί για λειτουργίες, όπως για παράδειγμα την αναγνώριση ομιλίας (speech recognition). Η συγκεκριμένα πληροφορία ωστόσο, δεν αφορά τους προγραμματιστές που θέλουν να χρησιμοποιήσουν το Web Speech API στην εφαρμογή τους. Αντίθετα, αφορά αυτούς που θα υλοποιήσουν το API στο εκάστοτε πρόγραμμα περιήγησης με βάση το περιεχόμενο του specification, που περιγράφηκε στην προηγούμενη ενότητα.

Μέχρι και σήμερα, το Web Speech API θεωρείται πειραματικό (experimental). Μία τεχνολογία που έχει οριστεί με αυτή την ιδιότητα, σημαίνει ότι είναι πιθανό το specification της να αλλάξει σημαντικά με τέτοιο τρόπο ώστε κομμάτια κώδικα που έχουν χρησιμοποιηθεί σε μία εφαρμογή να μην μπορούν να λειτουργήσουν με τις νέες αλλαγές. Αυτό, δείχνει πως μπορεί να βρίσκεται στο στάδιο προετοιμασίας για την μετατροπή της σε Standard, αλλά δεν αποκλείει και το γεγονός τελικά να μην χρησιμοποιηθεί ποτέ. Για τον λόγο αυτό, συστήνεται να μην γίνεται χρήση τέτοιων τεχνολογιών σε εφαρμογές οι οποίες θα βασίζονται πάνω στις λειτουργίες τους, αλλά αντίθετα είναι ασφαλέστερο να χρησιμοποιούνται σε demo ή πειράματα. Παρά το γεγονός ότι το Web Speech API παρέχει μεγάλες δυνατότητες, η ταμπέλα του experimental έχει ως αποτέλεσμα πολλοί προγραμματιστές να αποφεύγουν την χρήση του αξιοποιώντας άλλα εργαλεία τα οποία προς το παρόν αποτελούν μία περισσότερο σταθερή λύση. Το γεγονός σύμφωνα με το οποίο μία τεχνολογία μπορεί να βρίσκεται σε πειραματικό στάδιο, δεν πρέπει να απωθεί τους

προγραμματιστές από το να αλληλεπιδράσουν με αυτήν, αφού μέσω των δημιουργημάτων τους θα φανεί το πόσο πρακτικές είναι οι λειτουργίες της, παρέχοντας έτσι σημαντική γνώση τόσο στους προγραμματιστές που έχουν ενσωματώσει την εκάστοτε τεχνολογία σε ένα πρόγραμμα περιήγησης, όσο και στους δημιουργούς της ίδιας της τεχνολογίας, βλέποντας αν αυτή χρειάζεται αλλαγές ή προσθήκες που τελικά θα έχουν ως αποτέλεσμα την βελτίωση της.

Ένα ακόμη χαρακτηριστικό που αξίζει να σημειωθεί, αφορά την δυνατότητα χρήσης του Web Speech API μόνο σε https σελίδες. Στην αντίθετη περίπτωση, όταν δηλαδή μία σελίδα βρίσκεται κάτω από http, το API δεν επιτρέπει την χρήση του. Μία τέτοια απόφαση είναι απολύτως λογική, αφού σήμερα το μεγαλύτερο ποσοστό στο διαδίκτυο χρησιμοποιεί https, ενώ παράλληλα προγράμματα περιήγησης, όπως το Google Chrome, μαρκάρουν τις http σελίδες ως επικίνδυνες και επιτρέπουν στον χρήστη να αλληλεπιδράσει μαζί τους μόνο μετά από συναίνεση του. Η χρήση του Web Speech API με http, θα αύξανε σημαντικά την επικινδυνότητα σε ότι έχει να κάνει με την διακίνηση των δεδομένων που ο χρήστης μπορεί για παράδειγμα να παρέχει στην σελίδα ως είσοδο (input) μέσω της φωνής του. Τέλος, αναφορικά με την λειτουργικότητα της αναγνώρισης ομιλίας και σύμφωνα με το MDN, σε προγράμματα περιήγησης, όπως το Google Chrome, η διαδικασία της αναγνώρισης περιλαμβάνει την μεταβίβαση της εισόδου που δίνει ο χρήστης μέσω της φωνής του σε ένα web service, το οποίο θα πραγματοποιήσει την αναγνώριση. Συνεπώς, σε τέτοιες περιπτώσεις η χρήση της αναγνώρισης ομιλίας απαιτεί online σύνδεση, διαφορετικά η χρήση της δεν καθίσταται εφικτή.

Στην παρούσα ενότητα, πραγματοποιήθηκε μία επισκόπηση των χαρακτηριστικών του Web Speech API. Αναφέρθηκε ότι χωρίζεται σε δύο βασικά interfaces, όπου κάθε ένα από αυτά διαθέτει με την σειρά του ένα πλήθος από μεθόδους, γεγονότα και ιδιότητες, τα οποία είναι έτοιμα για χρήση σε όποιον επιθυμεί να ενσωματώσει δυνατότητες αναγνώρισης και σύνθεσης ομιλίας στην εφαρμογή του. Το Web Speech API είναι γραμμένο εξ ολοκλήρου σε Javascript και έχει την δυνατότητα να λειτουργεί ασύγχρονα χάρη στην event – based φιλοσοφία του. Ένας προγραμματιστής που επιθυμεί να χρησιμοποιήσει το API, δεν χρειάζεται να γνωρίζει τους μηχανισμούς που βρίσκονται κάτω από την τεχνολογία της αναγνώρισης και σύνθεσης ομιλίας, αρκεί να κάνει χρήση των έτοιμων μεθόδων, γεγονότων και ιδιοτήτων που προσφέρει το εκάστοτε interface, πράγμα που κάνει πολύ απλή την χρήση του API. Επιπλέον σημειώθηκε πως το Web Speech API βρίσκεται ακόμα σε πειραματική κατάσταση και διατυπώθηκε η σημασία της αλληλεπίδρασης των προγραμματιστών με τέτοια API. Τέλος, περιγράφηκαν δύο περιορισμοί του API, σύμφωνα με τους οποίους δεν είναι εφικτή η χρήση του κάτω από http σελίδες για λόγους ασφάλειας, ενώ παράλληλα σε προγράμματα περιήγησης, όπως το Google Chrome, για τις λειτουργίες αναγνώρισης ομιλίας απαιτείται online σύνδεση. Αφού λοιπόν, περιγράφηκαν τα βασικά χαρακτηριστικά του Web Speech API, στην επόμενη ενότητα αναλύεται το SpeechRecognition interface με τις μεθόδους, τα

γεγονότα και τις ιδιότητες που αυτό κατέχει για την δυνατότητα της ενσωμάτωσης αναγνώρισης ομιλίας σε web εφαρμογές.

2.5 SpeechRecognition Interface

Όπως αναδείχτηκε από τις προηγούμενες ενότητες, ένα πρόγραμμα περιήγησης στο οποίο έχει υλοποιηθεί το Web Speech API δίνει στους προγραμματιστές την ευκαιρία για ενσωμάτωση speech χαρακτηριστικών στις εφαρμογές τους, παρέχοντας έτσι στους τελικούς χρήστες ένα νέο μοντέλο αλληλεπίδρασης. Ποια είναι όμως τα βήματα που πρέπει να ακολουθήσει ένας προγραμματιστής και τι επιλογές δίνονται, ώστε αυτός να φτιάξει το μοντέλο αλληλεπίδρασης που επιθυμεί; Στην παρούσα και την επόμενη ενότητα, δίνονται απαντήσεις σε αυτά τα ερωτήματα μέσω της ανάλυσης των δομικών συστατικών του Web Speech API, ξεκινώντας με τις δυνατότητες για αναγνώριση ομιλίας.

Για την καλύτερη εμπειρία των χρηστών καθώς επίσης και την συγκρότηση ενός διαδικτύου με όσο το δυνατόν λιγότερους κινδύνους, τα περισσότερα προγράμματα περιήγησης σήμερα παρεμβαίνουν ως ενδιάμεσοι μεταξύ της εφαρμογής και του τελικού χρήστη όταν απαιτείται λειτουργικότητα η οποία επιτυγχάνεται με μέσα όπως για παράδειγμα το μικρόφωνο, η κάμερα κ.α. Το εκάστοτε πρόγραμμα περιήγησης, θωρακίζει τον χρήστη επιτρέποντας τέτοιες λειτουργίες μόνο εφόσον δώσει ο ίδιος την άδεια του. Διαφορετικά, δεν θα απείχαν πολύ από την πραγματικότητα σενάρια κατά τα οποία καθημερινά εφαρμογές θα άνοιγαν το μικρόφωνο ή την κάμερα του χρήστη χωρίς ο ίδιος να το γνωρίζει, κάτι το οποίο μπορεί να έχει καταστροφικές συνέπειες. Για παράδειγμα, στην περίπτωση του προγράμματος περιήγησης Google Chrome, ανάλογα με τις ρυθμίσεις που έχει πραγματοποιήσει ο εκάστοτε χρήστης σχετικά με την ασφάλεια και την ιδιωτικότητα, μία εφαρμογή που θα προσπαθήσει να αποκτήσει πρόσβαση στο μικρόφωνο μπορεί να οδηγήσει σε δύο αποτελέσματα. Το πρώτο, αφορά την περίπτωση κατά την οποία θα εμφανιστεί εκτός της εφαρμογής ένα παραθυράκι το οποίο θα ενημερώνει τον χρήστη πως η συγκεκριμένη σελίδα θέλει να χρησιμοποιήσει το μικρόφωνο του και παράλληλα θα του δίνει την επιλογή να δεχτεί ή να αρνηθεί. Πρόσβαση στο μικρόφωνο θα δοθεί αν και εφόσον ο χρήστης επιλέξει στο συγκεκριμένο παραθυράκι ότι επιτρέπει την χρήση του μικροφώνου από την εφαρμογή. Το δεύτερο αποτέλεσμα, αφορά την περίπτωση κατά την οποία το πρόγραμμα περιήγησης θα μπλοκάρει αυτόματα την χρήση του μικροφώνου στην συγκεκριμένη εφαρμογή και ο μόνος τρόπος για να λειτουργήσει είναι έπειτα από παρέμβαση του χρήστη στις ρυθμίσεις του προγράμματος περιήγησης, έτσι ώστε να δίνεται πρόσβαση στο μικρόφωνο όταν αυτό ζητείται από την συγκεκριμένη εφαρμογή.

Όταν πρόκειται να ενσωματωθούν σε μία εφαρμογή λειτουργίες οι οποίες απαιτούν την χρήση μικροφώνου, αποτελεί απαραίτητη προϋπόθεση η γνώση των παραπάνω πληροφοριών. Αυτό, γιατί είναι σημαντικό να αποφευχθούν περιπτώσεις όπως η μη λειτουργία ορισμένων χαρακτηριστικών της εφαρμογής τα οποία απαιτούν την χρήση μικροφώνου, επειδή το πρόγραμμα περιήγησης έχει μπλοκάρει

αυτή την δυνατότητα από προεπιλογή, χωρίς ο ίδιος ο χρήστης να το γνωρίζει. Η συγκεκριμένη περίπτωση είναι ακόμα πιο σοβαρή, όταν βασικό μέρος της εφαρμογής χρειάζεται πρόσβαση στο μικρόφωνο για να λειτουργήσει και υπάρχει ο κίνδυνος της έκθεσης ενός χρήστη σε μία σελίδα η οποία δεν φαίνεται να είναι λειτουργική. Το ερώτημα που προκύπτει λοιπόν, αφορά τον τρόπο με τον οποίο καθίσταται εφικτή προγραμματιστικά η διαχείριση της άδειας για πρόσβαση στο μικρόφωνο έτσι ώστε να είναι δυνατή μίας μορφής επικοινωνία με τον χρήστη που θα αποσκοπεί είτε στην προτροπή του να δώσει την άδεια για χρήση του μικροφώνου από την εφαρμογή, είτε αντίστοιχα στην ενημέρωση του σε περίπτωση που το μικρόφωνο έχει μπλοκαριστεί από το πρόγραμμα περιήγησης. Αφού λοιπόν στην περίπτωση της αναγνώρισης ομιλίας είναι απαραίτητη η χρήση του μικροφώνου, το παραπάνω ερώτημα πρέπει να είναι το πρώτο πράγμα που θα απασχολήσει έναν προγραμματιστή ο οποίος επιθυμεί να σχεδιάσει μια διεπαφή με μέσο αλληλεπίδρασης την ομιλία.

Την απάντηση στο συγκεκριμένο ερώτημα έρχεται να δώσει η μέθοδος `getUserMedia`. Με την χρήση της, εμφανίζεται εκτός της εφαρμογής ένα παραθυράκι μέσα στο οποίο ενημερώνεται ο χρήστης ότι ζητείται πρόσβαση σε ένα media device όπως για παράδειγμα το μικρόφωνο και καλείται ο ίδιος να απαντήσει εάν το επιτρέπει η όχι. Παρακάτω, δίνεται σαν εικόνα ένα παράδειγμα κατά το οποίο ζητείται η άδεια χρήσης του μικροφώνου μέσω της συγκεκριμένης συνάρτησης και έπειτα ακολουθεί η περιγραφή της. Να σημειωθεί, πως η αναλυτική περιγραφή του κώδικα που εμφανίζεται στην εικόνα είναι εκτός του εύρους της παρούσας πτυχιακής, αφού δεν σχετίζεται με το Web Speech API αλλά μπορεί να χρησιμοποιηθεί για να δώσει στην συνέχεια την δυνατότητα στο API να κάνει χρήση του μικροφώνου.

```
navigator.mediaDevices.getUserMedia({ audio: true })
  .then(function(stream) {
    // Page can use the mic
  })
  .catch(function(err) {
    // Page can't use the mic
  });
```

Εικόνα 1: Κώδικας για την άδεια χρήσης μικροφώνου

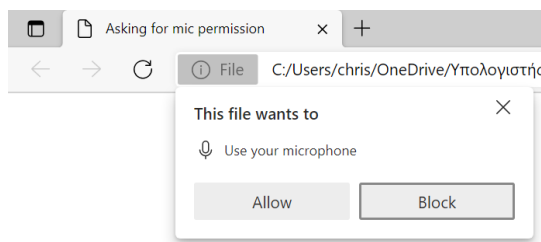
Αρχικά, παρατηρούμε το `navigator` αντικείμενο (object) από το οποίο χρησιμοποιείται την ιδιότητα (property) `mediaDevices`. Η συγκεκριμένη ιδιότητα, θα επιστρέψει ένα `MediaDevices` αντικείμενο το οποίο δίνει την δυνατότητα πρόσβασης σε media devices όπως είναι το μικρόφωνο και η κάμερα. Αυτό, πετυχαίνεται μέσω της πληθώρας συναρτήσεων που διαθέτει μεταξύ των οποίων περιλαμβάνεται και η `getUserMedia`. Η συγκεκριμένη συνάρτηση, θα εμφανίσει το παραθυράκι στο οποίο θα ζητείται η άδεια χρήσης του εκάστοτε media device από τον χρήστη. Όπως φαίνεται στην εικόνα 1, η συνάρτηση δέχεται μία παράμετρο. Μέσω αυτής, ορίζονται από τον προγραμματιστή τα media types που είναι επιθυμητό να χρησιμοποιηθούν σε περίπτωση που ο χρήστης δώσει την άδεια του. Η παράμετρος, αποτελεί ένα `MediaStreamConstraints` αντικείμενο το οποίο μπορεί να περιλαμβάνει ως media types το audio, το video

ή και τα δύο. Για παράδειγμα, στην περίπτωση που ήταν επιθυμητή η χρήση του μικροφώνου και της κάμερας, θα δινόταν ως παράμετρος ένα dictionary που θα είχε την μορφή μορφή:

```
{ audio: true, video: true }
```

Μετά την κλήση της `getUserMedia` συνάρτησης, θα εκτελεστεί είτε το block κώδικα του `then`, είτε του `catch`. Στην πρώτη περίπτωση, ο χρήστης έχει δώσει την άδεια του για χρήση του εκάστοτε media device και αυτό μπορεί να χρησιμοποιηθεί χωρίς κανένα πρόβλημα. Σε αυτό σενάριο, θα επιστραφεί από την κλήση της συνάρτησης ένα `MediaStream` αντικείμενο με βάση τα media types που ζητήθηκαν κατά την κλήση. Στην δεύτερη περίπτωση, είτε ο χρήστης αρνήθηκε την λειτουργία του ζητούμενου media device, είτε συνέβη κάποιο σφάλμα όπως για παράδειγμα η έλλειψη μικροφώνου στην συσκευή του τελικού χρήστη. Σε αυτό το σενάριο, η συνάρτηση θα επιστρέψει ένα `error` περιεχόμενο του οποίου θα αποτελούν πληροφορίες σχετικά με την αιτία που δεν είναι εφικτή η χρήση του εκάστοτε media device. Χαρακτηριστικά παραδείγματα αποτελούν τα `errors` `NotAllowedError` και `NotFoundError`. Το πρώτο, αφορά περιπτώσεις όπως την μη έγκριση του χρήστη για πρόσβαση στο μικρόφωνο του. Το δεύτερο, αφορά την περίπτωση κατά την οποία ο χρήστης έχει δώσει άδεια για πρόσβαση π.χ. στο μικρόφωνο, αλλά η συσκευή του δεν διαθέτει ένα. Με την κατάλληλη διαχείριση των `errors`, οι προγραμματιστές έχουν την δυνατότητα να ενημερώσουν τον τελικό χρήστη σχετικά με τον λόγο για τον οποίο δεν είναι εφικτή η χρήση του εκάστοτε media device, καθώς επίσης και ποιες λειτουργίες της εφαρμογής δεν μπορεί να χρησιμοποιήσει σε αυτή την περίπτωση.

Στην παρακάτω εικόνα, φαίνεται το παραθυράκι που εμφανίζεται στο πρόγραμμα περιήγησης Microsoft Edge εάν γίνει η κλήση της συνάρτησης `getUserMedia` με στόχο την απόκτηση πρόσβασης στο μικρόφωνο. Να σημειωθεί, πως για την εκτέλεση του κώδικα δημιουργήθηκε τοπικά ένα αρχείο με κατάληξη `.html` το οποίο ανοίχτηκε στον Microsoft Edge.



Εικόνα 2: Παραθυράκι για την παραχώρηση άδειας χρήσης μικροφώνου στο πρόγραμμα περιήγησης Microsoft Edge

Τέλος, σε περίπτωση που για παράδειγμα ο χρήστης δώσει πρόσβαση στο μικρόφωνο, τότε με κάθε επίσκεψη του στην εφαρμογή δεν θα ξανά ζητείται η άδεια χρήσης, αλλά θα θεωρείται αυτόματα ότι αυτή έχει δοθεί.

Μέχρι στιγμής, αναφέρθηκε η μέθοδος με την οποία είναι εφικτή προγραμματιστικά η επικοινωνία με τον χρήστη σε περιπτώσεις που η εφαρμογή χρειάζεται πρόσβαση σε μέσα όπως είναι το μικρόφωνο και η κάμερα. Η χρήση της συγκεκριμένης μεθόδου είναι απαραίτητη, καθώς δεν θα αποτελούσε σε καμία περίπτωση καλή πρακτική η υπόθεση των προγραμματιστών πως το μικρόφωνο ή η κάμερα του τελικού χρήστη θα είναι διαθέσιμα. Ωστόσο, μία τέτοια ενέργεια αποτελεί το πρώτο βήμα στο μονοπάτι που πρέπει να ακολουθήσει ένας προγραμματιστής για την δημιουργία εφαρμογών στις οποίες ένα από τα μέσα αλληλεπίδρασης μπορεί να είναι η ομιλία. Συγκεκριμένα, με την επιλογή του χρήστη να δοθεί πρόσβαση στο μικρόφωνο του, δεν σημαίνει ότι πλέον η εφαρμογή διαθέτει δυνατότητες αναγνώρισης ομιλίας. Για να επιτευχθεί αυτό, απαιτείται η χρήση ενός εργαλείου αναγνώρισης ομιλίας το οποίο στην περίπτωση μας αποτελεί το Web Speech API και συγκεκριμένα το SpeechRecognition interface που αυτό διαθέτει.

Το SpeechRecognition interface αποτελεί τον πυρήνα του Web Speech API σε ότι έχει να κάνει με την αναγνώριση ομιλίας. Το συγκεκριμένο interface, απαρτίζεται από έναν κατασκευαστή (constructor) και μία πληθώρα από μεθόδους (methods), γεγονότα (events) και ιδιότητες (properties). Ο κατασκευαστής, χρησιμοποιείται για την δημιουργία ενός SpeechRecognition αντικείμενου (object) το οποίο αποτελεί το μέσο για την πρόσβαση σε όλες τις μεθόδους, τα γεγονότα και τις ιδιότητες που παρέχει το Web Speech API για την αναγνώριση ομιλίας. Στην συνέχεια, γίνεται χρήση του κατασκευαστή για την δημιουργία ενός SpeechRecognition αντικείμενου με όνομα myFirstRecognitionObject.

```
var myFirstRecognitionObject = new window.webkitSpeechRecognition();
```

Εικόνα 3: Κώδικας για την δημιουργία ενός SpeechRecognition αντικείμενου

Η παραπάνω εκχώρηση, θα δημιουργήσει ένα SpeechRecognition αντικείμενο το οποίο μπορεί να χρησιμοποιηθεί για την ενσωμάτωση σε μία εφαρμογή δυνατοτήτων αναγνώρισης ομιλίας. Στην παραπάνω σύνταξη, παρατηρείται η χρήση του window αντικείμενου καθώς επίσης και η προσθήκη του προθέματος webkit στο SpeechRecognition. Το SpeechRecognition interface, ζει μέσα στο αντικείμενο window ως μία από τις ιδιότητες (properties) του και συνεπώς αποκτάται πρόσβαση σε αυτό με την παραπάνω σύνταξη. Σε ότι έχει να κάνει με την χρήση του προθέματος webkit, αναφέρθηκε στην προηγούμενη ενότητα πως το Web Speech API θεωρείται έως και σήμερα ένα πειραματικό (experimental) API. Σύμφωνα με το MDN, πολλές φορές σε πειραματικά APIs υπάρχει η πρακτική να τοποθετούνται προθέματα (prefixes). Αυτό, έτσι ώστε οι προγραμματιστές να μπορούν να πειραματιστούν χωρίς να υπάρχει ο κίνδυνος της δημιουργίας προβλημάτων στον κώδικα τους, αφού σε ένα experimental API μπορεί να πραγματοποιηθούν σημαντικές αλλαγές μέχρι να θεωρηθεί πρότυπο (Standard), ενώ υπάρχει και η πιθανότητα να καταργηθεί. Αν και εφόσον το εκάστοτε API σταματήσει να θεωρείται πειραματικό (experimental), τότε οι προγραμματιστές μπορούν να αφαιρέσουν το εκάστοτε πρόθεμα. Αξίζει να σημειωθεί, πως όταν ένα ολόκληρο interface

θεωρείται experimental τότε θα χρησιμοποιηθεί το εκάστοτε πρόθεμα μόνο στο όνομα του και όχι στις ιδιότητες και τις μεθόδους που αυτό διαθέτει. Έτσι λοιπόν και στην περίπτωση του SpeechRecognition, οι ιδιότητες και οι μέθοδοι του δεν χρησιμοποιούνται με πρόθεμα και η μόνη φορά που απαιτείται η προσθήκη του είναι κατά την δημιουργία νέων αντικειμένων μέσω του κατασκευαστή. Κάθε πρόγραμμα περιήγησης, μπορεί να χρησιμοποιεί διαφορετικό πρόθεμα σε ένα πειραματικό API. Ωστόσο, όπως θα φανεί σε επόμενη ενότητα κατά την οποία παρουσιάζεται η συμβατότητα (compatibility) του SpeechRecognition σε διαφορετικά προγράμματα περιήγησης, σε όσα από αυτά έχει υλοποιηθεί είναι υποχρεωτική η χρήση του προθέματος webkit.

Αναφέρθηκε παραπάνω ότι οι προγραμματιστές μπορούν να σταματήσουν να κάνουν χρήση ενός προθέματος όταν το εκάστοτε API πάψει να θεωρείται πειραματικό. Για τον λόγο αυτό, σε πολλά παραδείγματα χρήσης του SpeechRecognition interface στο διαδίκτυο χρησιμοποιείται η εξής σύνταξη:

```
var SpeechRecognition = SpeechRecognition || webkitSpeechRecognition;  
var myFirstRecognitionObject = new window.SpeechRecognition();
```

Εικόνα 4: Κώδικας για την δημιουργία ενός SpeechRecognition αντικειμένου σε περίπτωση που το Web Speech API σταματήσει να είναι πειραματικό

Παραπάνω, με την χρήση του τελεστή διάζευξης αρχικά εκχωρείται στο SpeechRecognition ένα από τα δύο interfaces, καλύπτοντας έτσι το σενάριο κατά το οποίο το SpeechRecognition interface στο μέλλον δεν θα θεωρείται experimental και συνεπώς δεν θα χρησιμοποιείται το πρόθεμα webkit. Έπειτα, στην δεύτερη εκχώρηση γίνεται η χρήση του κατασκευαστή όπως ακριβώς περιγράφηκε στο προηγούμενο παράδειγμα της δημιουργίας ενός SpeechRecognition αντικείμενου.

Σε πολλές περιπτώσεις, αποτελεί βέλτιστη πρακτική η γνώση του κατά πόσο το πρόγραμμα περιήγησης που επιλέγει ο τελικός χρήστης υποστηρίζει την τεχνολογία που χρησιμοποιείται από την εκάστοτε εφαρμογή. Με αυτό τον τρόπο, ο τελικός χρήστης θα μπορεί να ενημερωθεί κατάλληλα όταν το πρόγραμμα περιήγησης που χρησιμοποιεί δεν μπορεί να υποστηρίξει συγκεκριμένες λειτουργίες της εφαρμογής. Η ανάγκη για την συγκεκριμένη πρακτική είναι ακόμα μεγαλύτερη όταν η τεχνολογία που χρησιμοποιείται στην εφαρμογή είναι πειραματική. Το γεγονός αυτό, έχει ως πιθανό αποτέλεσμα να μην είναι υλοποιημένη σε πολλά προγράμματα περιήγησης σε αντίθεση με μία τεχνολογία η οποία θεωρείται πρότυπο (Standard). Έτσι λοιπόν και στην περίπτωση του SpeechRecognition interface, συστήνεται πριν την δημιουργία ενός αντικείμενου με την χρήση του κατασκευαστή να πραγματοποιηθεί πρώτα έλεγχος για το κατά πόσο υποστηρίζεται το interface από το τρέχον πρόγραμμα περιήγησης. Παρακάτω, παρατίθεται ένας από τους πιθανούς τρόπους με τον οποίο μπορεί να διαπιστωθεί εάν το τρέχον πρόγραμμα περιήγησης υποστηρίζει το SpeechRecognition interface.

```

if ('webkitSpeechRecognition' in window) {
  // This browser support SpeechRecognition interface
} else {
  // This browser does not support SpeechRecognition interface
  // and the user should be aware
}

```

Εικόνα 5: Κώδικας για τον έλεγχο εάν το τρέχον πρόγραμμα περιήγησης υποστηρίζει το *SpeechRecognition interface*

Όπως έχει προαναφερθεί, το *SpeechRecognition interface* ζει μέσα στο αντικείμενο *window* ως μία από τις ιδιότητες του. Έτσι λοιπόν, με την χρήστη του τελεστή (operator) *in*, ελέγχουμε αν το *window* αντικείμενο του συγκεκριμένου προγράμματος περιήγησης περιλαμβάνει την ιδιότητα *SpeechRecognition*. Ανάλογα με τον αν περιλαμβάνεται η όχι, η συνθήκη θα αποτιμηθεί σε αληθής (*true*) ή ψευδής (*false*), επιτρέποντας στον εκάστοτε προγραμματιστή να ορίσει τις αντίστοιχες ενέργειες σε κάθε περίπτωση.

Μέχρι στιγμής, περιγράφηκε ο τρόπος με τον οποίο είναι εφικτή η δημιουργία ενός *SpeechRecognition* αντικείμενου. Ωστόσο, αυτό που καθιστά χρήσιμο ένα αντικείμενο είναι τα περιεχόμενα του τα οποία απαρτίζονται από μεθόδους, ιδιότητες και γεγονότα. Έτσι λοιπόν και στην περίπτωση ενός *SpeechRecognition* αντικείμενου, παρέχεται με την δημιουργία του μία πληθώρα από ιδιότητες οι οποίες μπορούν να χρησιμοποιηθούν για την παραμετροποίηση των λειτουργιών της αναγνώρισης, ενώ παράλληλα προσφέρονται μέθοδοι για την εκκίνηση ή αντίστοιχα τον τερματισμό της. Τέλος, δεν θα μπορούσε να παραλειφθεί η πληθώρα των γεγονότων (*events*) τα οποία μπορούν να χρησιμοποιηθούν με την δημιουργία ενός *SpeechRecognition* αντικείμενου, προσφέροντας πολλαπλά οφέλη όπως την ασύγχρονη διαχείριση των αποτελεσμάτων της αναγνώρισης, καθώς επίσης και την εκτέλεση λειτουργιών ανάλογα με την κατάσταση στην οποία αυτή βρίσκεται. Παρακάτω, ακολουθεί η περιγραφή των χαρακτηριστικών που προαναφέρθηκαν ξεκινώντας με τις ιδιότητες ενός *SpeechRecognition* αντικείμενου.

Μέσω των ιδιοτήτων του, ένα *SpeechRecognition* αντικείμενο επιτρέπει την ρύθμιση πληθώρας παραμέτρων σχετικά με την αναγνώριση ομιλίας που πρόκειται να πραγματοποιηθεί στην εκάστοτε σελίδα. Συγκεκριμένα, στην παρούσα φάση, με την δημιουργία ενός *SpeechRecognition* αντικείμενου ο εκάστοτε προγραμματιστής έχει πρόσβαση στις εξής πέντε ιδιότητες:

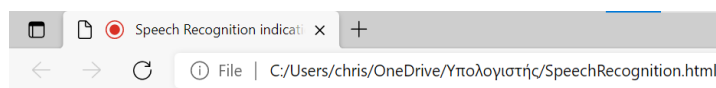
- *lang*: Η συγκεκριμένη ιδιότητα, χρησιμοποιείται για να καθοριστεί η γλώσσα την οποία το εκάστοτε *web service* πρόκειται να αναγνωρίζει σε κάθε είσοδο (*input*) του τελικού χρήστη μετατρέποντας τελικά την ομιλία σε κείμενο. Η τιμή που δέχεται η συγκεκριμένη ιδιότητα είναι ένα *DOMString* το οποίο αντιπροσωπεύει μία *BCP 47* ετικέτα γλώσσας (*language tag*). Σε περίπτωση που δεν εκχωρηθεί τιμή σε αυτή την ιδιότητα, τότε θα πάρει από προεπιλογή την τιμή

του HTML χαρακτηριστικού lang, ενώ σε περίπτωση που ούτε αυτό έχει οριστεί θα χρησιμοποιηθεί η γλώσσα που έχει ο εκάστοτε user agent.

- continuous: Με την χρήση της συγκεκριμένης ιδιότητας, καθορίζεται εάν κάθε φορά θα επιστρέφεται από την αναγνώριση ομιλίας μόνο ένα αποτέλεσμα ή σε αντίθετη περίπτωση πολλαπλά επιτρέποντας στον τελικό χρήστη να δώσει ως είσοδο περισσότερες από μία φράσεις. Η συγκεκριμένη ιδιότητα είναι τύπου boolean και σε περίπτωση που δεν της δοθεί θα πάρει από προεπιλογή την τιμή false.
- interimResults: Η συγκεκριμένη ιδιότητα είναι τύπου boolean και σε περίπτωση που της δοθεί η τιμή true τότε κατά την διάρκεια μιας αναγνώρισης θα επιστρέφονται ενδιάμεσα αποτελέσματα μέχρι την στιγμή που θα επιστραφεί ένα αποτέλεσμα το οποίο θα έχει οριστεί ως το τελικό. Σε περίπτωση που δεν δοθεί τιμή στην συγκεκριμένη ιδιότητα τότε θα πάρει από προεπιλογή την τιμή false.
- maxAlternatives: Σε κάθε αποτέλεσμα που επιστρέφεται από την υπηρεσία αναγνώρισης, περιλαμβάνεται ένας πλήθος με εναλλακτικές όπου κάθε μία από αυτές συνοδεύεται από έναν αριθμό που εκφράζει το πόσο σίγουρη είναι η υπηρεσία ότι το αποτέλεσμα που επιστράφηκε είναι ακριβές. Το πλήθος των εναλλακτικών αυτών ορίζεται με την χρήση της ιδιότητας maxAlternatives η οποία σύμφωνα με το specification είναι τύπου unsigned long και σε περίπτωση που δεν καθοριστεί παίρνει από προεπιλογή την τιμή ένα.
- grammars: Μία από τις δυνατότητες που παρέχει το Web Speech API σε ότι έχει να κάνει με την αναγνώριση ομιλίας αποτελεί η χρήση grammars. Συγκεκριμένα, ένα grammar αποτελεί ένα σύνολο λέξεων οι οποίες είναι επιθυμητό να αναγνωριστούν από το σύστημα που θα πραγματοποιήσει την αναγνώριση πετυχαίνοντας έτσι καλύτερα αποτελέσματα. Για παράδειγμα, σε μία περίπτωση κατά την οποία ο τελικός χρήστης μπορεί να δίνει ως είσοδο (input) μέσω της ομιλίας μάρκες αυτοκινήτων, τα αποτελέσματα της αναγνώρισης μπορεί να μην είναι ακριβή λόγω της δυσκολίας του συστήματος αναγνώρισης να καταλάβει τους περίεργους συνδυασμούς λέξεων που μπορεί να περιέχει μία μάρκα. Με την χρήση ωστόσο ενός grammar, το σύστημα που πραγματοποιεί την αναγνώριση θα μπορεί να δει πιθανά δεδομένα εισόδου με αποτέλεσμα να μπορεί να δώσει ακριβή αποτελέσματα. Για την αντιστοίχιση ενός οι περισσότερων grammar σε ένα SpeechRecognition αντικείμενο, χρησιμοποιείται η ιδιότητα grammars η οποία είναι τύπου SpeechGrammarList. Ένα SpeechGrammarList περιλαμβάνει ένα οι περισσότερα SpeechGrammar αντικείμενα καθένα από τα οποία αναπαριστά ένα grammar με το σύνολο των λέξεων που έχει προσθέσει ο εκάστοτε προγραμματιστής σε αυτό. Να σημειωθεί, ότι η χρήση της των grammars δεν είναι υποχρεωτική και συνεπώς η υπηρεσία που θα χρησιμοποιηθεί για αναγνώριση θα προσπαθήσει να αναγνωρίσει οτιδήποτε της δοθεί ως είσοδο (input) από τον τελικό χρήστη.

Αναλύθηκαν παραπάνω οι ιδιότητες μέσω των οποίων μπορεί να παραμετροποιηθεί κατάλληλα το εκάστοτε SpeechRecognition αντικείμενο έτσι ώστε η αναγνώριση ομιλίας να μπορεί να προσαρμοστεί κατάλληλα σύμφωνα με τις ανάγκες κάθε εφαρμογής. Αφού λοιπόν ολοκληρωθεί η διαδικασία απόδοσης τιμών σε όποιες από τις παραπάνω ιδιότητες κρίνεται απαραίτητο, ένας προγραμματιστής μπορεί πλέον να ξεκινήσει ή αντίστοιχα να σταματήσει την αναγνώριση ομιλίας μέσω της χρήσης τριών μεθόδων που παρέχει το SpeechRecognition interface για αυτόν ακριβώς τον σκοπό. Συγκεκριμένα:

- `start()`: Όπως φαίνεται και από το όνομα της, με την χρήση της συγκεκριμένης μεθόδου η εκάστοτε υπηρεσία που χρησιμοποιείται από το πρόγραμμα περιήγησης ξεκινάει την αναγνώριση ομιλίας. Σε περίπτωση που έχουν εκχωρηθεί στο χρησιμοποιούμενο SpeechRecognition αντικείμενο ένα ή περισσότερα grammars, τότε με την κλήση της `start` η υπηρεσία αναγνώρισης θα προσπαθήσει να κάνει αντιστοιχίσεις των δεδομένων εισόδου με βάση αυτά τα grammars. Διαφορετικά, θα προσπαθήσει να πραγματοποιήσει την αναγνώριση μόνο με τα δεδομένα εισόδου χωρίς την αξιοποίηση ενός ή περισσότερων grammars. Η μέθοδος `start`, δεν δέχεται παραμέτρους και με την ολοκλήρωση της δεν επιστρέφεται κάποια τιμή. Βέβαια, αποτελεί λογικό συνειρμό πως με την ολοκλήρωση της συνάρτησης θα επιστραφεί και το εκάστοτε αποτέλεσμα που έχει αναγνωριστεί. Ωστόσο, όπως θα φανεί στην συνέχεια το αποτέλεσμα ή αντίστοιχα τα αποτελέσματα επιστρέφονται με την χρήση γεγονότων (events). Αυτό, αφού όπως αναφέρθηκε και παραπάνω το Web Speech API δίνει την δυνατότητα κατά την διάρκεια της αναγνώρισης να επιστρέφονται ενδιάμεσα αποτελέσματα και όχι μόνο ένα τελικό, κάτι το οποίο δεν θα ήταν εφικτό αν το αποτέλεσμα επιστρεφόταν με την ολοκλήρωση της συνάρτησης `start`. Τέλος, όπως έχει προαναφερθεί στην ενότητα κατά την οποία μελετήθηκε το specification του Web Speech API, ένα πρόγραμμα περιήγησης που έχει υλοποιήσει το API πρέπει να πληροί ορισμένα κριτήρια όπως για παράδειγμα μία ένδειξη μέσω της οποίας ο τελικός χρήστης θα αντιλαμβάνεται ότι η αναγνώριση ομιλίας είναι ενεργή. Χαρακτηριστικά παραδείγματα αποτελούν τα προγράμματα περιήγησης Google Chrome και Microsoft Edge, στα οποία με την κλήση της συνάρτησης `start` εμφανίζεται ένα κόκκινο κυκλάκι στο παράθυρο της σελίδας, ενημερώνοντας έτσι τον εκάστοτε χρήστη ότι πραγματοποιείται αναγνώριση ομιλίας. Παρακάτω, φαίνεται η συγκεκριμένη ένδειξη με την χρήση της συνάρτησης `start` στο πρόγραμμα περιήγησης Microsoft Edge. Όπως και σε προηγούμενο παράδειγμα, για την εκτέλεση του κώδικα χρησιμοποιήθηκε ένα HTML αρχείο τοπικά.



Εικόνα 6: Ένδειξη ότι πραγματοποιείται αναγνώριση ομιλίας στο πρόγραμμα περιήγησης Microsoft Edge

- `stop()`: Όπως είναι λογικό, αφού υπάρχει συνάρτηση για την ενεργοποίηση της αναγνώρισης ομιλίας δεν θα μπορούσε να λείπει και η συμπληρωματική της με την χρήση της οποίας η αναγνώριση θα σταματά. Με την κλήση της μεθόδου `stop`, η εκάστοτε υπηρεσία αναγνώρισης σταματάει να δέχεται δεδομένα εισόδου από το μικρόφωνο του χρήστη και επιχειρεί να επιστρέψει ένα αποτέλεσμα με βάση τα δεδομένα που είχε λάβει μέχρι πριν την κλήση της συνάρτησης. Παρόμοια με την `start`, η συγκεκριμένη συνάρτηση δεν δέχεται παραμέτρους καθώς επίσης δεν επιστρέφει και καμία τιμή.
- `abort()`: Η χρήση της `abort`, έχει παρόμοιο αποτέλεσμα με την μέθοδο `stop` με την βασική διαφορά ότι δεν επιχειρεί να επιστρέψει κανένα αποτέλεσμα το οποίο μπορεί να έχει προκύψει κατά την αναγνώριση με τα δεδομένα που είχε λάβει η υπηρεσία μέχρι και πριν την κλήση της συνάρτησης. Αυτονόητο είναι πως και στην συγκεκριμένη συνάρτηση δεν δίνονται ούτε παράμετροι ούτε επιστρέφεται κάποια τιμή.

Μέχρι στιγμής, ορίστηκε ο τρόπος με τον οποίο είναι εφικτή η παραμετροποίηση ενός `SpeechRecognition` αντικειμένου μέσω των ιδιοτήτων του και στην συνέχεια η χρήση αυτού για την αναγνώριση ομιλίας μέσω τριών μεθόδων. Ωστόσο, μία εφαρμογή που υποστηρίζει την αλληλεπίδραση μέσω της ομιλίας είναι απαραίτητο να διαθέτει στον πυρήνα της μία λογική η οποία θα της επιτρέπει να εκτελεί διαφορετικές λειτουργίες με βάση το εκάστοτε αποτέλεσμα που έχει αναγνωριστεί. Ο τρόπος λοιπόν με τον οποίο είναι εφικτή η ανάκτηση των αποτελεσμάτων αναγνώρισης για την χρήση τους στην λογική της εκάστοτε εφαρμογής είναι μέσω μίας πληθώρας γεγονότων. Μέσω αυτών, ένας προγραμματιστής έχει την δυνατότητα να ανακτήσει τα αποτελέσματα της εκάστοτε αναγνώρισης καθώς επίσης και να εκτελέσει διαφορετικά μπλοκ κώδικα ανάλογα με την κατάσταση στην οποία βρίσκεται μία αναγνώριση π.χ. μόλις ξεκινήσει, μόλις ολοκληρωθεί κλπ. Συγκεκριμένα, με την δημιουργία ενός `SpeechRecognition` αντικειμένου αποκτάται πρόσβαση στα εξής έντεκα γεγονότα (events):

- `audiostart`: Το συγκεκριμένο γεγονός (event), εκτελείται όταν ξεκινήσει η καταγραφή ήχου από το μικρόφωνο του τελικού χρήστη έτσι ώστε να πραγματοποιηθεί η αναγνώριση ομιλίας. Με άλλα λόγια, το `onaudiostart` είναι αναμενόμενο να εκτελεστεί μόλις γίνει η χρήση της μεθόδου `start`.
- `audioend`: Αντίστοιχα με το `audiostart`, το συγκεκριμένο γεγονός θα εκτελεστεί όταν σταματήσει η καταγραφή ήχου από το μικρόφωνο του τελικού χρήστη.
- `start`: Το `start` γεγονός, πυροδοτείται όταν η υπηρεσία αναγνώρισης ξεκινήσει να ακούει τα δεδομένα που παρέχονται από το μικρόφωνο του τελικού χρήστη με στόχο την αναγνώριση ομιλίας.
- `end`: Αντίστοιχα με το `start` γεγονός, το `end` θα εκτελεστεί όταν η υπηρεσία που χρησιμοποιείται για την αναγνώριση ομιλίας αποσυνδεθεί.

- **error:** Όπως προδίδει και το όνομα του, το error γεγονός θα εκτελεστεί όταν συμβεί κάποιο σφάλμα κατά την αναγνώριση.
- **nomatch:** Το nomatch γεγονός, εκτελείται όταν η υπηρεσία αναγνώρισης επιστρέφει ένα τελικό αποτέλεσμα το οποίο είτε θα περιέχει μία ή περισσότερες εναλλακτικές με πολύ χαμηλό βαθμό σιγουριάς, είτε η τιμή τους θα είναι null.
- **result:** Το συγκεκριμένο γεγονός, όπως αναφέρει και το όνομα του θα εκτελεστεί όταν επιστραφεί από την υπηρεσία αναγνώρισης ένα αποτέλεσμα το οποίο περιέχει πιθανές λέξεις ή φράσεις (η αντίστοιχα μία λέξη ή φράση αν η ιδιότητα maxAlternatives έχει την τιμή ένα) τις οποίες μπορεί να έδωσε ο τελικός χρήστης ως είσοδο (input).
- **soundstart:** Το soundstart γεγονός, πυροδοτείται όταν ανιχνευθεί από το μικρόφωνο του τελικού χρήστη οποιοσδήποτε ήχος είτε αυτός αποτελεί κάτι το οποίο ειπώθηκε από τον ίδιο τον χρήστη είτε αποτελεί ένα θόρυβο.
- **soundend:** Αντίστοιχα, το soundend γεγονός πυροδοτείται όταν ένας ήχος που ανιχνευόταν έπαψε να είναι ανιχνεύσιμος.
- **speechstart:** Το speechstart γεγονός, πυροδοτείται όταν ανιχνευθεί ένας ήχος από την υπηρεσία αναγνώρισης ο οποίος χαρακτηρίζεται ως ομιλία και όχι απλά σαν ένας θόρυβος.
- **speechend:** Αντίστοιχα, το speechend γεγονός πυροδοτείται όταν η ομιλία που αναγνωρίζεται από την υπηρεσία αναγνώρισης πάψει να είναι ανιχνεύσιμη.

Να σημειωθεί, πως κάθε γεγονός από τα παραπάνω μπορεί να χρησιμοποιηθεί είτε με την χρήση της συνάρτησης `addEventListener` επάνω στο εκάστοτε `SpeechRecognition` αντικείμενο, είτε με την προσθήκη στο όνομα του του προθέματος οπ όπως για παράδειγμα `onresult`, `onstart` κλπ. Με αυτό τον τρόπο, δίνεται η δυνατότητα εκτέλεσης διαφορετικού μπλοκ κώδικα ανάλογα με το γεγονός που θα εκτελεστεί κάθε φορά. Με την εκτέλεση ορισμένων γεγονότων, παρέχεται μία πληθώρα δεδομένων τα οποία μπορούν να αξιοποιηθούν από τον κώδικα ο οποίος είναι αντιστοιχισμένος στο συγκεκριμένο γεγονός. Για παράδειγμα, στην περίπτωση του `SpeechRecognition` interface, τα δεδομένα που έχουν προκύψει από μία αναγνώριση επιστρέφονται μέσω του `result` event και ο εκάστοτε προγραμματιστής μπορεί να τα αξιοποιήσει στην εφαρμογή του κατάλληλα.

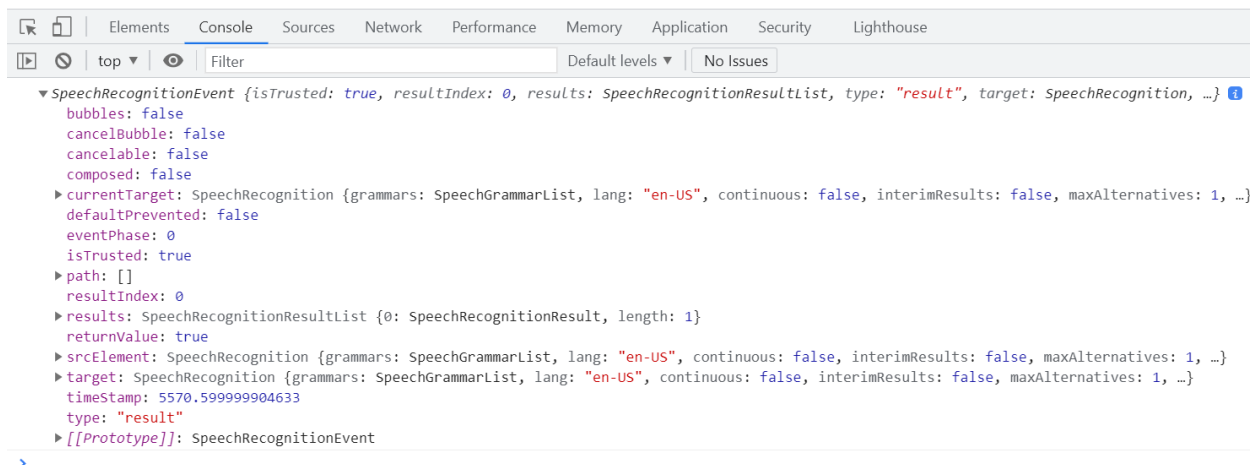
Μέχρι στιγμής, μελετήθηκαν τα περιεχόμενα του `SpeechRecognition` interface και χαρακτηρίστηκε ως ο βασικός πυρήνας σε ότι αφορά την αναγνώριση ομιλίας. Ωστόσο, το συγκεκριμένο interface δεν είναι το μοναδικό που χρησιμοποιείται για την αναγνώριση ομιλίας αφού σύμφωνα με το `Web Speech API` κατά την διαδικασία της αναγνώρισης ένας προγραμματιστής καλείται να διαχειριστεί και δεδομένα τα οποία είναι προσβάσιμα μέσω άλλων interfaces. Παρακάτω, δίνεται ένα παράδειγμα χρήσης του `SpeechRecognition` interface μέσω του οποίου θα περιγραφούν στην συνέχεια κάποια από τα interfaces

που αυτό αξιοποιεί, ενώ παράλληλα θα γίνει περισσότερο κατανοητή η μεθοδολογία που περιγράφηκε μέχρι τώρα. Να σημειωθεί, πως για την εκτέλεση του παρακάτω κώδικα δημιουργήθηκε ένα HTML αρχείο τοπικά, το οποίο ανοίχτηκε με το πρόγραμμα περιήγησης Google Chrome.

```
const myFirstRecognition = new window.webkitSpeechRecognition();
myFirstRecognition.lang = 'en-US';
myFirstRecognition.addEventListener('result', function(event) {
    console.log(event);
});
myFirstRecognition.start();
```

Εικόνα 7: Κώδικας για την πραγματοποίηση αναγνώριση ομιλίας στα αγγλικά και την τύπωση του αποτελέσματος στην κονσόλα του προγράμματος περιήγησης

Με την χρήση του παραπάνω κώδικα, δίνεται η δυνατότητα της αναγνώρισης μίας λέξης ή φράσης με την χρήση του μικροφώνου και στην συνέχεια εφόσον αυτό που ειπώθηκε αναγνωριστεί από την υπηρεσία που χρησιμοποιεί το πρόγραμμα περιήγησης Google Chrome, επιστρέφονται τα αντίστοιχα δεδομένα της αναγνώρισης τα οποία τελικά τυπώνονται στην κονσόλα (console). Συγκεκριμένα, αρχικά γίνεται η χρήση του κατασκευαστή (constructor) για την δημιουργία ενός SpeechRecognition αντικειμένου με όνομα myFirstSpeechRecognition. Έπειτα, χρησιμοποιείται η ιδιότητα lang ορίζοντας ως γλώσσα αναγνώρισης στο αντικείμενο τα αγγλικά. Παρατηρείται, πως δεν γίνεται χρήση των ιδιοτήτων maxAlternatives, interimResults και continuous κάτι το οποίο όπως έχει αναφερθεί παραπάνω θα έχει ως αποτέλεσμα να λάβουν τις τιμές που έχουν οριστεί για το καθένα ως προεπιλογή. Επιπλέον, δεν γίνεται χρήση της ιδιότητας grammars, κάτι το οποίο θα έχει ως αποτέλεσμα η υπηρεσία αναγνώρισης να προσπαθήσει να αναγνωρίσει τα δεδομένα εισόδου χωρίς την βοήθεια ενός ή περισσοτέρων grammars τα οποία μπορούν να περιέχουν λέξεις που είναι πιθανό να χρησιμοποιηθούν από τον τελικό χρήστη. Στην συνέχεια, γίνεται χρήση της συνάρτησης addEventListener για την αντιστοίχιση του αντικειμένου myFirstRecognition με το γεγονός result. Ως πρώτη παράμετρος, δίνεται το όνομα του event (στην περίπτωση του παραδείγματος result) που είναι επιθυμητό και ως δεύτερη δίνεται μία συνάρτηση η οποία θα εκτελεστεί όταν ενεργοποιηθεί το συγκεκριμένο event. Παρατηρείται στην συνάρτηση η παράμετρος με όνομα event. Η συγκεκριμένη παράμετρος, αποτελεί ένα αντικείμενο με περιεχόμενο μία πληθώρα από interfaces μέσα στα οποία περιλαμβάνεται και το αποτέλεσμα που επιστράφηκε από την υπηρεσία αναγνώρισης. Το μπλοκ της συνάρτησης περιλαμβάνει μία γραμμή κώδικα με την χρήση της οποίας θα εκτυπωθούν στην κονσόλα του Google Chrome τα περιεχόμενα του αντικειμένου event. Τέλος, μετά την δήλωση του event result γίνεται χρήση της μεθόδου start, η οποία θα ενεργοποιήσει το μικρόφωνο του τελικού χρήστη ξεκινώντας την αναγνώριση ομιλίας. Παρακάτω, φαίνεται το αποτέλεσμα το οποίο εκτυπώνεται στην κονσόλα του Google Chrome εάν εκτελεστεί ο παραπάνω κώδικας. Να σημειωθεί, πως για αναγνώριση ομιλίας δόθηκε η φράση “hello world” (Γειά σου κόσμε).



Εικόνα 8: Το αποτέλεσμα της τύπωσης του `SpeechRecognitionEvent` στην κονσόλα του προγράμματος περιήγησης Google Chrome

Αρχικά, όπως φαίνεται στην εικόνα με την εκτέλεση του παραπάνω κώδικα επιστρέφεται και τυπώνεται στην κονσόλα ένα `SpeechRecognitionEvent`. Το συγκεκριμένο interface, διαθέτει ιδιότητες οι οποίες μπορούν να αξιοποιηθούν για την ανάκτηση δεδομένων τα οποία έχουν επιστραφεί από την υπηρεσία αναγνώρισης. Συγκεκριμένα, με την χρήση της ιδιότητας `results`, θα επιστραφεί ένα αντικείμενο `SpeechRecognitionResultList` το οποίο διαθέτει ένα ή περισσότερα αποτελέσματα αναγνώρισης. Κάθε τέτοιο αποτέλεσμα αναγνώρισης, απεικονίζεται ως ένα `SpeechRecognitionResult` αντικείμενο το οποίο για να ανακτηθεί χρησιμοποιείται σύνταξη παρόμοια με αυτή ενός πίνακα. Έτσι λοιπόν, με την σύνταξη `[0]` θα επιστραφεί το `SpeechRecognitionResult` το οποίο βρίσκεται στην θέση μηδέν. Για τον συνολικό αριθμό των `SpeechRecognitionResult` αντικειμένων το `SpeechRecognitionResultList` interface παρέχει την ιδιότητα `length`. Με την ανάκτηση ενός από τα διαθέσιμα `SpeechRecognitionResult` αντικείμενα, δίνεται πρόσβαση σε ένα ή περισσότερα `SpeechRecognitionAlternative` αντικείμενα (ανάλογα με την τιμή που έχει δοθεί στην ιδιότητα `maxAlternatives`). Για την ανάκτηση ενός από αυτά, χρησιμοποιείται εξίσου παρόμοια σύνταξη με έναν πίνακα και συνεπώς το αντικείμενο στην θέση μηδέν θα ανακτηθεί με την χρήση του `[0]`. Ένα `SpeechRecognitionAlternative` αντικείμενο, διαθέτει δύο ιδιότητες. Η πρώτη, ονομάζεται `transcript` και επιστρέφει ως κείμενο αυτό που αναγνώρισε η υπηρεσία αναγνώρισης από το μικρόφωνο του τελικού χρήστη. Η δεύτερη, ονομάζεται `confidence` και επιστρέφει έναν αριθμό μεταξύ του 0 και του 1 ο οποίος αντιπροσωπεύει το πόσο σίγουρη είναι η υπηρεσία αναγνώρισης ότι αυτό που έχει αναγνωρισθεί είναι όντως αυτό που είπε ο τελικός χρήστης στο μικρόφωνο του. Παρακάτω, απεικονίζεται η δομή που περιγράφηκε μέχρι τώρα.

```

▼ results: SpeechRecognitionResultList
  ▼ 0: SpeechRecognitionResult
    ▼ 0: SpeechRecognitionAlternative
      confidence: 0.5469405651092529
      transcript: "hello world"
      ► [[Prototype]]: SpeechRecognitionAlternative
        isFinal: true
        length: 1
      ► [[Prototype]]: SpeechRecognitionResult
        length: 1
      ► [[Prototype]]: SpeechRecognitionResultList
        returnValue: true

```

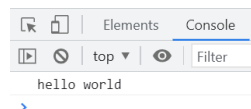
Εικόνα 9: Οι τιμές των ιδιοτήτων *confidence* και *transcript* του *SpeechRecognitionAlternative* αντικείμενου

Όπως αποδεικνύεται και από το παραπάνω αποτέλεσμα, επειδή δεν ορίστηκαν στο *SpeechRecognition* αντικείμενο οι ιδιότητες *continuous*, *maxAlternatives* και *interimResults*, στην απάντηση που έδωσε η υπηρεσία αναγνώρισης περιλαμβάνεται μόνο ένα *SpeechRecognitionResult* αντικείμενο, το οποίο διαθέτει μόνο μία εναλλακτική που στην προκειμένη περίπτωση περιλαμβάνει την λέξη “hello world”, κάτι το οποίο σημαίνει πως η υπηρεσία αναγνώρισε σωστά τα δεδομένα εισόδου που έδωσε ο τελικός χρήστης μέσω της φωνής του. Έτσι λοιπόν, με βάση την παραπάνω δομή, σε περίπτωση που ήταν επιθυμητή η ανάκτηση του κειμένου που αναγνωρίστηκε και η τύπωση του στην κονσόλα θα μπορούσε να τροποποιηθεί ο ίδιος κώδικας σε:

```
console.log(event.results[0][0].transcript);
```

Εικόνα 10: Τύπωση στην κονσόλα ενός προγράμματος περιήγησης του κειμένου που επιστρέφεται από την υπηρεσία αναγνώρισης

Με την χρήση της παραπάνω εντολής, αντί να τυπωθούν στην κονσόλα οι πληροφορίες που περιλαμβάνονται μέσα στο event τυπώνεται κατευθείαν η λέξη που αναγνωρίστηκε από την εκάστοτε υπηρεσία αναγνώρισης που χρησιμοποιείται από το τρέχον πρόγραμμα περιήγησης. Παρακάτω, φαίνεται το αποτέλεσμα στην κονσόλα:



Εικόνα 11: Το αποτέλεσμα του αλλαγμένου κώδικα εφόσον δοθεί ως είσοδος η λέξη *hello world*

Με παρόμοια λογική, κάθε προγραμματιστής μπορεί να διαχειριστεί τα δεδομένα που επιστρέφονται από την υπηρεσία αναγνώρισης μέσω γεγονότων όπως του *result*. Ακόμα ένα τέτοιο παράδειγμα αποτελεί και η χρήση του *error* event, το οποίο θα επιστρέψει ένα *SpeechRecognitionErrorEvent* αντικείμενο μέσω του οποίου δίνεται πρόσβαση σε δύο ιδιότητες, την *error* και την *message*. Στην πρώτη περίπτωση, επιστρέφεται ένα *DomString* το οποίο αντιπροσωπεύει τον τύπο του σφάλματος που συνέβη, όπως για παράδειγμα *no-speech* για την περίπτωση που δεν ανιχνεύθηκε καμία ομιλία, *not-allowed* κατά το οποίο δεν δόθηκε η δυνατότητα της χρήσης ομιλίας ως είσοδο (*input*) για λόγους ασφάλειας κλπ. Αντίστοιχα, με

την χρήση της ιδιότητας `message` επιστρέφεται ένα `DomString` το οποίο αποτελεί μία πιο λεπτομερή περιγραφή για το σφάλμα που συνέβη.

Στην παρούσα ενότητα, μελετήθηκε το `SpeechRecognition` interface βλέποντας τις δυνατότητες που αυτό παρέχει σε ότι έχει να κάνει με την αναγνώριση φωνής. Αρχικά, αναλύθηκε η συνάρτηση `getUserMedia` η οποία δεν περιλαμβάνεται στο `Web Speech API` αλλά μέσω αυτής είναι εφικτή η επικοινωνία με τον τελικό χρήστη ζητώντας του άδεια χρήσης του μικροφώνου. Στην συνέχεια, περιγράφηκαν τα περιεχόμενα του interface και συγκεκριμένα ο κατασκευαστής με τον οποίο είναι εφικτή η δημιουργία `SpeechRecognition` αντικειμένων, οι ιδιότητες με την χρήση των οποίων παραμετροποιείται το εκάστοτε αντικείμενο που έχει δημιουργηθεί, οι μέθοδοι με τις οποίες ελέγχεται η ενεργοποίηση ή αντίστοιχα απενεργοποίηση της αναγνώρισης και τέλος τα γεγονότα (events) τα οποία επιτρέπουν την εκτέλεση διαφορετικού κώδικα ανάλογα με την κατάσταση στην οποία βρίσκεται η αναγνώριση καθώς επίσης και την χρήση δεδομένων από τον κώδικα τα οποία έχουν επιστραφεί από την υπηρεσία αναγνώρισης. Όπως φάνηκε παραπάνω, για την χρήση της αναγνώρισης ομιλίας το `SpeechRecognition` interface αλληλεπιδρά και με άλλα interfaces του `Web Speech API` όπως για παράδειγμα το `SpeechRecognitionEvent`, `SpeechRecognitionErrorEvent` κ.α. Κάθε interface από αυτά περιέχει αντίστοιχα τις δικές του ιδιότητες, μεθόδους κλπ. Ωστόσο, λόγω του μεγάλου μεγέθους περιεχομένων που έχουν κάποια interfaces, στην παρούσα πτυχιακή αναλύονται τα βασικότερα από αυτά για την ενσωμάτωση σε μία εφαρμογή δυνατότητες αναγνώρισης και σύνθεσης ομιλίας, ενώ παράλληλα ο τρόπος αλληλεπίδρασης σε όλα είναι παρόμοιος. Το σύνολο όλων των interfaces μαζί με τα περιεχόμενα τους και παραδείγματα χρήσης μπορεί να τα βρει κάποιος στο διαδίκτυο σε πηγές όπως το MDN.

Για την δημιουργία μιας εμπειρίας κατά την οποία ο χρήστης θα αλληλεπιδρά εξ ολοκλήρου μέσω της φωνής του, χρειάζονται δύο δομικά στοιχεία. Το πρώτο, επιτρέπει την είσοδο δεδομένων μέσω της φωνής. Ωστόσο, για την απάντηση από την πλευρά της εφαρμογής είναι απαραίτητη η χρήση της σύνθεσης ομιλίας η οποία αποτελεί και το δεύτερο δομικό στοιχείο. Μέσω της χρήσης της, μία συσκευή μπορεί εκφωνήσει πληροφορίες στον χρήστη χωρίς αυτός να χρειάζεται να κοιτάει την οθόνη του. Στην ενότητα που ακολουθεί, αναλύονται τα εργαλεία που παρέχει το `Web Speech API` για την ενσωμάτωση σε μία εφαρμογή δυνατότητες συνθετικής ομιλίας.

2.6 SpeechSynthesis Interface

Όπως αποδείχτηκε στην προηγούμενη ενότητα, το `Web Speech API` δίνει μία πληθώρα δυνατοτήτων γύρω από την αναγνώριση ομιλίας επιτρέποντας έτσι την δημιουργία πρωτότυπων εφαρμογών στις οποίες ένας χρήστης θα μπορεί να δίνει δεδομένα εισόδου μέσω της φωνής του. Ωστόσο, σε περιπτώσεις εφαρμογών που δεν είναι εφικτή ή δεν συνίσταται η χρήση της οθόνης καθόλου, το κομμάτι της αναγνώρισης ομιλίας δεν αρκεί για την δημιουργία της διεπαφής. Εκεί, έρχεται να δώσει λύση η

τεχνολογία για την σύνθεση ομιλίας (speech synthesis). Για παράδειγμα, σε περίπτωση που ένας χρήστης χρησιμοποιεί μία εφαρμογή την ώρα της οδήγησης η οποία διαθέτει μόνο δυνατότητες αναγνώρισης ομιλίας, τότε θα πρέπει να αλληλεπιδράσει με την οθόνη υποχρεωτικά εάν επιθυμεί να δει τα αποτελέσματα των ενεργειών του, κάτι το οποίο είναι επικίνδυνο λόγω των συνθηκών. Προσθέτοντας όμως την δυνατότητα συνθετικής ομιλίας, αυτόματα δεν απαιτείται ο τελικός χρήστης να έρθει σε επαφή με το γραφικό περιβάλλον αφού τα αποτελέσματα των ενεργειών που πραγματοποίησε μέσω της φωνής του θα μπορούν να εκφωνούνται από την ίδια την εφαρμογή. Καταλαβαίνει κανείς λοιπόν την κρισιμότητα της συνθετικής ομιλίας σε ένα φωνητικό μοντέλο αλληλεπίδρασης. Για τον λόγο αυτό, το Web Speech API αντίστοιχα με τις δυνατότητες για αναγνώρισης ομιλίας παρέχει μία πληθώρα από interfaces μέσω των οποίων ένας προγραμματιστής έχει την δυνατότητα να ενσωματώσει συνθετική ομιλία στην εφαρμογή του. Παρακάτω, μελετάται το συγκεκριμένο κομμάτι του Web Speech API.

Όπως έχει προαναφερθεί, για την αναγνώριση ομιλίας κύριο interface αποτελεί το SpeechRecognition. Αντίστοιχα, σε ότι αφορά την συνθετική ομιλία τον πυρήνα αποτελεί το SpeechSynthesis interface. Με την χρήση του συγκεκριμένου, δίνεται η δυνατότητα ανάκτησης πληροφοριών για τις διαθέσιμες φωνές που μπορούν να αξιοποιηθούν, ενώ παράλληλα είναι εφικτή και η διαχείριση της συνθετικής ομιλίας μέσω μιας πληθώρας ιδιοτήτων και μεθόδων. Παρόμοια με το SpeechRecognition interface, το SpeechSynthesis μπορεί να χρησιμοποιηθεί μέσω του window αντικείμενου, αφού αποτελεί μία ιδιότητα αυτού μέσω της οποίας επιστρέφεται ένα SpeechSynthesis αντικείμενο. Να σημειωθεί, πως το όνομα της ιδιότητας αυτής είναι ακριβώς ίδιο με του interface, με μόνη διαφορά ότι το πρώτο γράμμα είναι πεζό.

Πριν από την ανάλυση των περιεχομένων του SpeechSynthesis interface, αξίζει να σημειωθούν κάποιες διαφορές μεταξύ της σύνθεσης και της αναγνώρισης ομιλίας του Web Speech API αναφορικά με τον τρόπο χρήσης τους. Συγκεκριμένα, σε αντίθεση με την αναγνώριση, στην περίπτωση της σύνθεσης ομιλίας όπως είναι προφανές, δεν απαιτείται η άδεια χρήσης του μικροφώνου από τον τελικό χρήστη. Η μόνη προϋπόθεση είναι να υποστηρίζεται το SpeechSynthesis από το εκάστοτε πρόγραμμα περιήγησης, κάτι το οποίο έχει αναφερθεί στην προηγούμενη ενότητα πως μπορεί να διαπιστωθεί. Ακόμη μία διαφορά αποτελεί και το γεγονός σύμφωνα με το οποίο για το SpeechSynthesis interface δεν παρέχεται κάποιος κατασκευαστής (constructor) και κατά συνέπεια για την χρήση του δεν πραγματοποιείται η δημιουργία ενός στιγμιότυπου όπως στην περίπτωση του SpeechRecognition interface, όπου είναι απαραίτητο να δημιουργηθούν ένα ή περισσότερα διαφορετικά SpeechRecognition αντικείμενα τα οποία μπορούν να έχουν διαφορετικές ρυθμίσεις σε παραμέτρους όπως η γλώσσα αναγνώρισης κλπ. Τέλος, όπως θα φανεί και στην επόμενη ενότητα σχετικά με την συμβατότητα του Web Speech API σε διαφορετικά προγράμματα περιήγησης, στην περίπτωση του SpeechSynthesis δεν απαιτείται για την χρήση του η προσθήκη του

προθέματος webkit, σε αντίθεση με το SpeechRecognition interface το οποίο προς το παρόν είναι προσβάσιμο μόνο με την χρήση του συγκεκριμένου προθέματος. Βέβαια, παρά το γεγονός αυτό, οι δυνατότητες για συνθετική ομιλία δεν παύουν να θεωρούνται μέρος ενός πειραματικού (experimental) API και κατά συνέπεια είναι πιθανή στο μέλλον η τροποποίηση ή ακόμα και η διαγραφή τους.

Με την χρήση της speechSynthesis ιδιότητας του αντικειμένου window, δίνεται πρόσβαση στα περιεχόμενα του SpeechSynthesis interface και συγκεκριμένα σε τρεις ιδιότητες (properties), πέντε μεθόδους (methods) και ένα γεγονός (event). Παρακάτω, αναλύονται τα συγκεκριμένα ξεκινώντας με τις μεθόδους.

Μέσω των μεθόδων του SpeechSynthesis interface, ένας προγραμματιστής έχει την δυνατότητα να διαχειριστεί την συνθετική ομιλία. Συγκεκριμένα, παρέχονται οι εξής πέντε μέθοδοι:

- **speak:** Με την χρήση της συνάρτησης speak, δίνεται ως παράμετρος ένα SpeechSynthesisUtterance αντικείμενο μέσα στα περιεχόμενα του οποίου βρίσκεται το κείμενο που είναι επιθυμητό να εκφωνηθεί. Το SpeechSynthesisUtterance interface αντιπροσωπεύει το κείμενο που θέλει ο προγραμματιστής να εκφωνηθεί αλλά και το πως αυτό πρέπει να εκφωνηθεί π.χ. σε τι γλώσσα, τι ταχύτητα κλπ. Το συγκεκριμένο interface μελετάται παρακάτω. Σε περίπτωση που χρησιμοποιηθεί η μέθοδος speak χωρίς να έχει ολοκληρωθεί η εκφώνηση ενός κειμένου, θα σχηματιστεί μία ουρά (queue) από SpeechSynthesisUtterance αντικείμενα από την οποία θα επιλέγεται ένα κάθε φορά μέχρι την στιγμή που θα εκφωνηθούν τα κείμενα όλων. Τέλος, με την χρήση της συνάρτησης speak δεν επιστρέφεται κάποια τιμή.
- **cancel:** Όταν γίνεται κλήση της μεθόδου cancel, διαγράφονται κατευθείαν όλα τα SpeechSynthesisUtterance αντικείμενα από την ουρά, ενώ σε περίπτωση που την τρέχουσα χρονική στιγμή εκφωνείται ένα κείμενο θα σταματήσει κατευθείαν. Η συγκεκριμένη μέθοδος δεν δέχεται παραμέτρους και δεν επιστρέφει κάποια τιμή με την ολοκλήρωση της.
- **pause:** Η μέθοδος pause, θα θέσει το SpeechSynthesis αντικείμενο σε κατάσταση paused. Σε περίπτωση που εκφωνείται ένα κείμενο και γίνει κλήση της μεθόδου pause, θα σταματήσει η εκφώνηση σε εκείνο το σημείο μέχρι την στιγμή που θα αλλάξει η κατάσταση pause όπου και θα ξανά ξεκινήσει από το σημείο που είχε σταματήσει. Σε περίπτωση που η ουρά δεν περιέχει κανένα SpeechSynthesisUtterance αντικείμενο, το SpeechSynthesis αντικείμενο μπορεί να μπει κανονικά σε κατάσταση paused, με αποτέλεσμα να μην εκφωνείται τίποτα μέσω της μεθόδου speak μέχρι να αλλάξει η κατάσταση του αντικειμένου από paused. Η συνάρτηση pause δεν δέχεται κάποια παράμετρο και δεν επιστρέφει κάποια τιμή.

- `resume`: Όπως είναι προφανές από το όνομα της, η μέθοδος `resume` θα βγάλει το `SpeechSynthesis` αντικείμενο από την κατάσταση `paused`. Η συγκεκριμένη μέθοδος δεν δέχεται παραμέτρους ούτε επιστρέφει κάποια τιμή.
- `getVoices`: Ένα από τα πλεονεκτήματα της σύνθεσης ομιλίας που προσφέρει το Web Speech API, αφορά την δυνατότητα χρήσης διαφορετικών φωνών για την εκφώνηση κειμένου. Αυτό, επιτυγχάνεται μέσω της συνάρτησης `getVoices` όπου με την κλήση της επιστρέφεται μία λίστα από `SpeechSynthesisVoice` αντικείμενα. Καθένα από αυτά, αντιπροσωπεύει μία φωνή η οποία μπορεί να χρησιμοποιηθεί για την εκφώνηση κειμένων. Όπως σε κάθε `interface`, έτσι και στην περίπτωση του `SpeechSynthesisVoice` παρέχονται ιδιότητες μέσω των οποίων επιστρέφονται πληροφορίες σχετικά με την εκάστοτε φωνή. Να σημειωθεί, ότι οι φωνές που επιστρέφονται μπορεί να διαφέρουν σε κάθε πρόγραμμα περιήγησης αφού εξαρτάται από το ίδιο ποιες θα είναι δυνατό να χρησιμοποιηθούν. Τέλος, η συνάρτηση `getVoices` δεν δέχεται κάποια παράμετρο.

Ενώ με την χρήση των παραπάνω μεθόδων είναι εφικτός ο καθορισμός του πότε θα εκφωνείται ένα κείμενο, σε πολλές περιπτώσεις είναι χρήσιμη η γνώση της κατάστασης στην οποία βρίσκεται το `SpeechSynthesis` αντικείμενο, αναφορικά με το αν εκφωνείται κάποιο κείμενο ή όχι την τρέχουσα χρονική στιγμή. Για την ανάκτηση τέτοιου είδους πληροφοριών, μπορούν να χρησιμοποιηθούν οι ιδιότητες του `SpeechSynthesis` και συγκεκριμένα:

- `paused`: Με την χρήση της συγκεκριμένης ιδιότητας, επιστρέφεται μία `boolean` τιμή η οποία αναφέρει αν το `SpeechSynthesis` αντικείμενο βρίσκεται σε κατάσταση `paused` ή όχι.
- `pending`: Η ιδιότητα `pending`, επιστρέφει μία `boolean` τιμή η οποία αναφέρει εάν βρίσκονται στην ουρά `SpeechSynthesisUtterance` αντικείμενα τα κείμενα των οποίων βρίσκονται σε αναμονή προς εκφώνηση
- `speaking`: Όπως είναι προφανές και από το όνομα της, με την χρήση της ιδιότητας `speaking` επιστρέφεται μία `boolean` τιμή η οποία αναφέρει εάν έχει ξεκινήσει να εκφωνείται το κείμενο ενός `SpeechSynthesisUtterance` αντικειμένου και δεν έχει ολοκληρωθεί. Να σημειωθεί, πως θα επιστραφεί `true` ακόμα και αν το `SpeechSynthesis` αντικείμενο τεθεί σε κατάσταση `paused` πριν την ολοκλήρωση της εκφώνησης.

Εκτός από τις ιδιότητες και τις μεθόδους, το `SpeechSynthesis` `interface` περιλαμβάνει και ένα γεγονός (`event`) το οποίο χρησιμοποιείται για την ανάκτηση των διαθέσιμων φωνών σε συνδυασμό με την μέθοδο `getVoices`. Συγκεκριμένα, λόγω του ότι σε πολλές περιπτώσεις οι διαθέσιμες φωνές φορτώνονται ασύγχρονα, με την κλήση της `getVoices` μεθόδου δεν επιστρέφεται κανένα `SpeechSynthesisVoice` αντικείμενο. Για την αντιμετώπιση αυτού του θέματος, μπορεί να χρησιμοποιηθεί το γεγονός `voiceschanged` το οποίο ενεργοποιείται όταν η λίστα με τα `SpeechSynthesisVoice` αντικείμενα αλλάζει,

κάτι το οποίο σηματοδοτεί την φόρτωση των διαθέσιμων φωνών προς χρήση. Έτσι λοιπόν, το μπλοκ κώδικα που θα εκτελεστεί με την ενεργοποίηση του voiceschanged μπορεί να περιέχει μέσα του την κλήση της μεθόδου getVoices, ανακτώντας έτσι τις διαθέσιμες φωνές που παρέχει το εκάστοτε πρόγραμμα περιήγησης. Να σημειωθεί, ότι το συγκεκριμένο γεγονός μπορεί να χρησιμοποιηθεί είτε μέσω της συνάρτησης addEventListener, είτε με την χρήση του προθέματος on στο όνομα του.

Παρόλο που το SpeechSynthesis interface αποτελεί τον πυρήνα σε ότι έχει να κάνει με την συνθετική ομιλία, χωρίς την χρήση του SpeechSynthesisUtterance δεν θα ήταν εφικτή η εκφώνηση ενός κειμένου. Συγκεκριμένα, ένα SpeechSynthesisUtterance αντικείμενο αντιπροσωπεύει δύο πράγματα. Το πρώτο, αφορά το κείμενο που είναι επιθυμητό να εκφωνηθεί, ενώ το δεύτερο αφορά το πως αυτό πρέπει να εκφωνηθεί από την εκάστοτε υπηρεσία που θα χρησιμοποιηθεί. Όπως είναι προφανές λοιπόν, το πρώτο βήμα για την ενσωμάτωση συνθετικής ομιλίας σε μία εφαρμογή αποτελεί η δημιουργία ενός τέτοιου αντικειμένου και στην συνέχεια η παραμετροποίηση του έτσι ώστε τελικά να χρησιμοποιηθεί ως παράμετρο στην μέθοδο speak. Για την δημιουργία παρέχεται ένας κατασκευαστής (constructor), ενώ στην συνέχεια για τον καθορισμό του κειμένου και των ρυθμίσεων για το πως αυτό πρέπει να εκφωνηθεί παρέχονται έξι ιδιότητες. Τέλος, δεν θα μπορούσαν να παραλειφθούν τα γεγονότα (events) με την χρήση των οποίων μπορεί να εκτελεστεί διαφορετικός κώδικας ανάλογα με την κατάσταση στην οποία βρίσκεται το κείμενο του εκάστοτε αντικειμένου που έχει δοθεί προς εκφώνηση. Παρακάτω, πραγματοποιείται η δημιουργία ενός αντικειμένου με όνομα myFirstSynthesisObject με την χρήση του κατασκευαστή.

```
var myFirstSynthesisObject = new SpeechSynthesisUtterance();
```

Εικόνα 12: Κώδικας για την δημιουργία ενός SpeechSynthesisUtterance αντικειμένου

Όπως φαίνεται και από την παραπάνω εκχώρηση, για την δημιουργία ενός SpeechSynthesisUtterance αντικειμένου δεν απαιτείται η χρήση του προθέματος webkit στο όνομα του κατασκευαστή. Εκτός από τον παραπάνω τρόπο δημιουργίας, υπάρχει και η επιλογή της αρχικοποίησης του αντικειμένου με ένα κείμενο το οποίο είναι επιθυμητό να εκφωνηθεί στην συνέχεια με την χρήση της μεθόδου speak. Παρακάτω, φαίνεται η δημιουργία ενός αντικειμένου με τον συγκεκριμένο τρόπο δίνοντας ως κείμενο το “Hello world” (Γεια σου κόσμο).

```
var myFirstSynthesisObject = new SpeechSynthesisUtterance("Hello world");
```

Εικόνα 13: Δημιουργία ενός SpeechSynthesisUtterance αντικειμένου και αρχικοποίηση του με το κείμενο hello world

Με την δημιουργία ενός SpeechSynthesisUtterance αντικειμένου, δίνεται πρόσβαση σε έξι ιδιότητες μέσω των οποίων μπορεί να δοθεί από τον προγραμματιστή το κείμενο που είναι επιθυμητό να

εκφωνηθεί, ενώ παράλληλα ο ίδιος καλείται να ρυθμίσει τον τρόπο με τον οποίο πρέπει να εκφωνήσει το κείμενο η εκάστοτε υπηρεσία που θα χρησιμοποιηθεί. Συγκεκριμένα, οι ιδιότητες είναι οι εξής:

- **lang:** Όπως και στην περίπτωση του SpeechRecognition interface, έτσι και εδώ η ιδιότητα lang δέχεται ένα DOMString το οποίο αντιπροσωπεύει μία BCP 47 ετικέτα γλώσσας (language tag) η οποία αναφέρει στην υπηρεσία σε τι γλώσσα πρέπει να εκφωνηθεί το εκάστοτε κείμενο. Αν δεν εκχωρηθεί τιμή στην συγκεκριμένη ιδιότητα, θα πάρει από προεπιλογή την τιμή του HTML χαρακτηριστικού lang, ενώ σε περίπτωση που δεν έχει οριστεί ούτε αυτό, θα δοθεί η γλώσσα που έχει ο εκάστοτε user agent.
- **text:** Μέσω της συγκεκριμένης ιδιότητας, ορίζεται ένα DOMString το οποίο αντιπροσωπεύει το κείμενο που είναι επιθυμητό να εκφωνηθεί. Να σημειωθεί, ότι εκτός από απλό κείμενο μπορεί να δοθεί και SSML (Speech Synthesis Markup Language). Ωστόσο, απαραίτητη προϋπόθεση είναι να υποστηρίζεται από την εκάστοτε υπηρεσία που θα χρησιμοποιηθεί για την εκφώνηση καθώς σε διαφορετική περίπτωση θα αφαιρεθεί η συγκεκριμένη σύνταξη κρατώντας μόνο το κείμενο.
- **voice:** Η ιδιότητα voice, χρησιμοποιείται για τον καθορισμό της φωνής που είναι επιθυμητό να εκφωνήσει το εκάστοτε κείμενο και συνεπώς η τιμή που παίρνει είναι ένα SpeechSynthesisVoice αντικείμενο. Σε περίπτωση που δεν δοθεί τιμή στην συγκεκριμένη ιδιότητα, όταν γίνει χρήση της μεθόδου speak θα χρησιμοποιηθεί μία φωνή από προεπιλογή ανάλογα με το τρέχον πρόγραμμα περιήγησης η οποία θα είναι συμβατή με την τιμή της ιδιότητας lang.
- **volume:** Όπως φανερώνει και το όνομα, μέσω της συγκεκριμένης ιδιότητας ορίζεται η ένταση στην οποία πρέπει να πραγματοποιηθεί η εκφώνηση. Η τιμή που λαμβάνει η συγκεκριμένη ιδιότητα είναι float και έχει όρια μεταξύ 0 και 1 με το 0 να αφορά την πιο χαμηλή ένταση ενώ το 1 αντίστοιχα την μέγιστη. Να σημειωθεί, πως στην περίπτωση που χρησιμοποιηθεί SSML σύνταξη η συγκεκριμένη τιμή θα αντικατασταθεί από τις ετικέτες prosody (prosody tags). Τέλος, σε περίπτωση που δεν δοθεί τιμή στην συγκεκριμένη ιδιότητα θα πάρει από προεπιλογή την τιμή 1.
- **rate:** Με την χρήση της συγκεκριμένης ιδιότητας, καθορίζεται η ταχύτητα με την οποία θα γίνεται η εκφώνηση του εκάστοτε κειμένου. Ως τιμή, δίνεται ένας αριθμός float με χαμηλότερο όριο το 0.1 και μεγαλύτερο το 10. Σε περίπτωση που δεν καθοριστεί η συγκεκριμένη ιδιότητα θα πάρει από προεπιλογή την τιμή 1 η οποία αντιπροσωπεύει την κανονική ταχύτητα ομιλίας. Να σημειωθεί, πως η φωνή ή η μηχανή για την σύνθεση ομιλίας (speech synthesis engine) που θα χρησιμοποιηθεί είναι πιθανό να έχει θέσει διαφορετικά όρια για την ταχύτητα τα οποία όμως δεν θα ξεπερνούν τα παραπάνω. Τέλος, όπως και στην περίπτωση της volume ιδιότητας αν

χρησιμοποιηθεί SSML σύνταξη η τιμή του rate θα αντικατασταθεί από τις ετικέτες prosody (prosody tags).

- pitch: Μέσω της συγκεκριμένης ιδιότητας, καθορίζεται το pitch στο οποίο θα εκφωνηθεί το εκάστοτε κείμενο. Η τιμή του pitch είναι ένας float αριθμός με χαμηλότερο όριο το 0 και μεγαλύτερο το 2. Σε περίπτωση που δεν αποδοθεί τιμή στην συγκεκριμένη ιδιότητα θα χρησιμοποιηθεί από προεπιλογή η τιμή 1 ενώ παράλληλα όπως και στην περίπτωση του rate τα όρια μπορεί να διαφέρουν ανάλογα με την φωνή ή την μηχανή για την σύνθεση ομιλίας που χρησιμοποιείται. Να σημειωθεί, ότι και στην περίπτωση της ιδιότητας pitch αν χρησιμοποιηθεί SSML σύνταξη η τιμή του pitch θα αντικατασταθεί από τις ετικέτες prosody (prosody tags).

Αφού οριστούν οι παραπάνω ιδιότητες, το επόμενο βήμα μπορεί να αποτελέσει ο καθορισμός του κώδικα που θα εκτελείται ανάλογα με την κατάσταση στην οποία βρίσκεται το κείμενο του αντικειμένου το οποίο έχει δοθεί προς εκφώνηση μέσω της μεθόδου speak. Για να επιτευχθεί κάτι τέτοιο, το SpeechSynthesisUtterance interface παρέχει μία πληθώρα από γεγονότα (events). Παρακάτω, περιγράφονται κάποια βασικά από αυτά και συγκεκριμένα:

- start: Το συγκεκριμένο γεγονός, πυροδοτείται όταν ξεκινήσει να εκφωνείται το κείμενο που βρίσκεται αποθηκευμένο στην ιδιότητα text του εκάστοτε SpeechSynthesisUtterance αντικειμένου.
- end: Παρόμοια με το γεγονός start, το end πυροδοτείται όταν έχει εκφωνηθεί το κείμενο που υπάρχει μέσα στην ιδιότητα text του αντικειμένου.
- pause: Όπως φανερώνει και το όνομα του, το γεγονός pause θα εκτελεστεί όταν η εκφώνηση του κειμένου σταματήσει πριν την ολοκλήρωση του.
- resume: Αντίθετα με το γεγονός pause, το resume πυροδοτείται όταν η εκφώνηση ενός κειμένου είχε διακοπεί και πλέον συνεχίζεται από εκείνο το σημείο. Να σημειωθεί, ότι το συγκεκριμένο γεγονός δεν θα εκτελεστεί εάν προστεθεί στην ουρά ένα SpeechSynthesisUtterance αντικείμενο ενώ το SpeechSynthesis αντικείμενο βρίσκεται σε κατάσταση paused. Σε αυτή την περίπτωση, μόλις αλλάξει η κατάσταση από paused θα εκτελεστεί το γεγονός start.
- error: Το συγκεκριμένο γεγονός, πυροδοτείται όταν συμβεί κάποιο σφάλμα το οποίο εμπόδισε την εκφώνηση του κειμένου που περιέχεται στην ιδιότητα text του εκάστοτε SpeechSynthesisUtterance αντικειμένου.

Παρόμοια με τα γεγονότα του SpeechRecognition interface, έτσι και εδώ με την εκτέλεση των παραπάνω γεγονότων επιστρέφονται πληροφορίες σχετικά με την κατάσταση του τρέχοντος SpeechSynthesisUtterance αντικειμένου το οποίο έχει υποστεί επεξεργασία από την εκάστοτε υπηρεσία

που χρησιμοποιείται για την σύνθεση ομιλίας. Συγκεκριμένα, σε όλα τα παραπάνω γεγονότα που περιγράφηκαν με εξαίρεση το error, με την εκτέλεση τους επιστρέφεται ένα SpeechSynthesisEvent αντικείμενο. Το συγκεκριμένο interface, περιέχει μία πληθώρα από ιδιότητες οι οποίες επιστρέφουν πληροφορίες για το τρέχον SpeechSynthesisUtterance αντικείμενο. Χαρακτηριστικά παραδείγματα αποτελούν οι ιδιότητες utterance και elapsedTime, όπου με την χρήση της πρώτης θα επιστραφεί το τρέχον SpeechSynthesisUtterance αντικείμενο, ενώ με την χρήση της δεύτερης θα επιστραφεί μία float τιμή η οποία αντιπροσωπεύει τα δευτερόλεπτα που πέρασαν από την στιγμή που ξεκίνησε η εκφώνηση του κειμένου του τρέχον αντικειμένου μέχρι την ενεργοποίηση του εκάστοτε event. Παραπάνω, αναφέρθηκε ότι στην περίπτωση του error event δεν θα επιστραφεί ένα SpeechSynthesisEvent αντικείμενο. Αυτό, διότι με την ενεργοποίηση του error θα επιστραφεί ένα SpeechSynthesisError αντικείμενο. Όπως είναι αναμενόμενο, το περιεχόμενο ενός τέτοιου αντικειμένου αφορά πληροφορίες για οποιοδήποτε σφάλμα προκύψει. Συγκεκριμένα, το SpeechSynthesisError interface περιλαμβάνει μία ιδιότητα με όνομα error και η οποία θα επιστρέφει κάθε φορά έναν κωδικό σφάλματος ανάλογα με το τι συνέβη. Για παράδειγμα, όταν επιστρέφεται ο κωδικός canceled σημαίνει πως έγινε η χρήση της μεθόδου cancel κάτι το οποίο οδήγησε στην διαγραφή του SpeechSynthesisUtterance αντικειμένου από την ουρά πριν να ξεκινήσει η εκφώνηση του. Το πλήθος των διαθέσιμων κωδικών σφαλμάτων μπορεί να το βρει κάποιος σε μία πηγή όπως το MDN ή το specification του Web Speech API. Να σημειωθεί, πως για την χρήση όλων των γεγονότων μπορεί να χρησιμοποιηθεί είτε η συνάρτηση addEventListener, είτε με την προσθήκη του προθέματος on στο όνομα του εκάστοτε event όπως για παράδειγμα onstart. Παρακάτω, δίνεται ένα παράδειγμα χρήσης της συνθετικής ομιλίας με στόχο την καλύτερη κατανόηση όσων περιγράφηκαν ως τώρα. Για την εκτέλεση του κώδικα, δημιουργήθηκε ένα html αρχείο το οποίο στην συνέχεια ανοίχτηκε με το πρόγραμμα περιήγησης Microsoft edge.

```
var myFirstSynthesis = window.speechSynthesis;
var myFirstUtterance = new SpeechSynthesisUtterance();

myFirstUtterance.text = "Καλώς ήρθες στο website μας, ελπίζουμε να έχεις μία ευχάριστη πλοήγηση";
myFirstUtterance.lang = "el-GR";

myFirstSynthesis.addEventListener('voiceschanged', function() {
    var voices = myFirstSynthesis.getVoices();
    for(i = 0; i < voices.length; i++) {
        if (voices[i].lang == "el-GR") {
            myFirstUtterance.voice = voices[i];
            break;
        }
    }
});

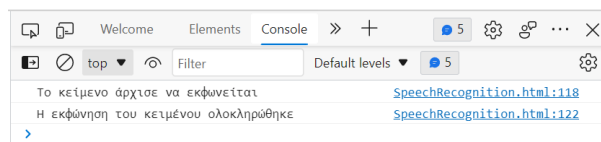
myFirstUtterance.addEventListener('start', function(event) {
    console.log('Το κείμενο άρχισε να εκφωνείται');
});

myFirstUtterance.addEventListener('end', function(event) {
    console.log('Η εκφώνηση του κειμένου ολοκληρώθηκε');
});

myFirstSynthesis.speak(myFirstUtterance);
```

Εικόνα 14: Κώδικας για την εκφώνηση ενός κειμένου καλωσορίσματος στα ελληνικά και αξιοποίηση των γεγονότων start και end για την τύπωση αντίστοιχου μηνύματος στην κονσόλα του προγράμματος περιήγησης

Με την χρήση του παραπάνω κώδικα, θα εκφωνηθεί σε οποιοδήποτε χρήστη επισκεφτεί την συγκεκριμένη σελίδα το κείμενο 'Καλώς ήρθες στο website μας, ελπίζουμε να έχεις μία ευχάριστη πλοήγηση'. Αρχικά, δημιουργείται μία αναφορά στο speechSynthesis interface με την χρήση της μεταβλητής myFirstSynthesis. Να σημειωθεί, πως δεν είναι απαραίτητη η χρήση μιας μεταβλητής καθώς θα μπορούσε να χρησιμοποιηθεί κατευθείαν η σύνταξη window.speechSynthesis για την χρήση των περιεχομένων του speechSynthesis interface. Στην συνέχεια, γίνεται χρήση του κατασκευαστή (constructor) που παρέχει το SpeechSynthesisUtterance interface για την δημιουργία ενός αντικειμένου με όνομα myFirstUtterance. Αφού δημιουργηθεί το αντικείμενο, αποκτάται πρόσβαση μέσω αυτού στις ιδιότητες text και lang. Στην πρώτη ιδιότητα αποθηκεύεται το κείμενο που είναι επιθυμητό να εκφωνηθεί ενώ με την χρήση της δεύτερης ορίζεται η γλώσσα. Οι υπόλοιπες ιδιότητες του αντικειμένου θα λάβουν τις τιμές που δίνονται από προεπιλογή. Στην συνέχεια, χρησιμοποιείται το γεγονός voiceschanged το οποίο θα εκτελεστεί όταν οι διαθέσιμες φωνές που παρέχονται έχουν φορτωθεί. Με την εκτέλεση του voiceschanged, το πρώτο πράγμα που γίνεται είναι η κλήση της συνάρτησης getVoices και η αποθήκευση του πίνακα που επιστρέφεται εντός της μεταβλητής voices. Συγκεκριμένα, ο πίνακας που θα επιστραφεί έχει ως περιεχόμενο μία πληθώρα από SpeechSynthesisVoice αντικείμενα καθένα από τα οποία αντιπροσωπεύει μία διαθέσιμη προς χρήση φωνή. Αφού γίνει ανάκτηση του συγκεκριμένου πίνακα, χρησιμοποιείται η επαναληπτική δομή for για την διάσχιση του με στόχο την εύρεση της ελληνικής φωνής. Να σημειωθεί, ότι η χρήση του length επιστρέφει το μέγεθος του πίνακα και δεν αποτελεί μία ιδιότητα του SpeechSynthesisVoice. Μέσω της ιδιότητα lang γίνεται αναζήτηση της BCP 47 ετικέτας που αντιστοιχεί στην Ελληνική φωνή. Αφού βρεθεί η συγκεκριμένη φωνή, γίνεται χρήση του break έτσι ώστε να μην συνεχιστεί άσκοπα η αναζήτηση μέσα στον πίνακα. Αφού έχει οριστεί η επιθυμητή φωνή, στην συνέχεια χρησιμοποιούνται δύο γεγονότα. Συγκεκριμένα, γίνεται χρήση του start και του end. Το μπλοκ κώδικα που περιέχει καθένα από αυτά θα τυπώσει ένα συγκεκριμένο μήνυμα στην κονσόλα του προγράμματος περιήγησης με την ενεργοποίηση του αντίστοιχου event. Στην πρώτη περίπτωση, όταν αρχίσει να εκφωνείται το κείμενο που περιλαμβάνει το SpeechSynthesisUtterance αντικείμενο θα ενεργοποιηθεί το γεγονός start με αποτέλεσμα να τυπωθεί στην οθόνη η φράση 'Το κείμενο άρχισε να εκφωνείται'. Αντίστοιχα, μόλις ολοκληρωθεί η εκφώνηση θα ενεργοποιηθεί το γεγονός end με αποτέλεσμα να τυπωθεί στην κονσόλα η φράση 'Η εκφώνηση του κειμένου ολοκληρώθηκε'. Παρακάτω, φαίνεται η κονσόλα του προγράμματος περιήγησης Microsoft edge μετά την εκτέλεση του παραπάνω κώδικα και αφού έχει εκφωνηθεί το κείμενο που δόθηκε.



Εικόνα 15: Η κονσόλα του προγράμματος περιήγησης Microsoft Edge μετά την εκτέλεση του κώδικα που εμφανίζεται στην εικόνα 14

Στην παρούσα ενότητα, μελετήθηκαν οι δυνατότητες που δίνει το Web Speech API σε ότι έχει να κάνει με την συνθετική ομιλία. Αρχικά, αναφέρθηκαν διαφορές που υπάρχουν σε σχέση με την αναγνώριση ομιλίας αναφορικά με τον τρόπο χρήσης. Στην συνέχεια, μελετήθηκε το SpeechSynthesis interface το οποίο αποτελεί τον πυρήνα σε ότι έχει να κάνει με την χρήση των δυνατοτήτων συνθετικής ομιλίας. Εκτός από αυτό, μελετήθηκαν και άλλα interfaces τα οποία χρησιμοποιούνται για την επίτευξη συνθετικής ομιλίας όπως για παράδειγμα το SpeechSynthesisUtterance το οποίο αντιπροσωπεύει το κείμενο που είναι επιθυμητό να εκφωνηθεί αλλά και το πως αυτό πρέπει να εκφωνηθεί. Τέλος, δόθηκε ένα παράδειγμα για την καλύτερη κατανόηση της θεωρίας που περιγράφηκε στην συγκεκριμένη ενότητα. Να σημειωθεί, πως παρόμοια με την ενότητα σχετικά με την αναγνώριση ομιλίας δεν μελετήθηκε όλο το περιεχόμενο των interfaces που χρησιμοποιούνται. Αυτό, αφού το μέγεθος του είναι αρκετά μεγάλο και η πλήρης ανάλυση του είναι εκτός του εύρους της παρούσας πτυχιακής, καθώς κύριο στόχο αποτελεί η περιγραφή των βασικών στοιχείων του Web Speech API πολλά από τα οποία χρησιμοποιήθηκαν για την δημιουργία του proof of concept. Ωστόσο, τα υπόλοιπα στοιχεία χρησιμοποιούνται με παρόμοιο τρόπο και συνεπώς οποιοσδήποτε έχει κατανοήσει τις παραπάνω ενότητες δεν θα έχει πρόβλημα να τα ενσωματώσει στην εφαρμογή του. Τα περιεχόμενα των interfaces στο σύνολο τους μαζί με παραδείγματα χρήσης μπορεί να τα βρει κάποιος σε μία πηγή όπως το MDN.

Επειδή το Web Speech API αποτελεί μέχρι και σήμερα ένα πειραματικό (experimental) API, πολλά προγράμματα περιήγησης είναι φυσικό να μην το έχουν υλοποιήσει, κάτι το οποίο καθιστά κρίσιμη την μελέτη από την πλευρά ενός προγραμματιστή σε ποια από αυτά μπορεί να αξιοποιηθεί. Στην επόμενη ενότητα, λοιπόν, σχολιάζεται η συμβατότητα του.

2.7 Συμβατότητα

Ένα από τα μεγαλύτερα οφέλη των εφαρμογών του διαδικτύου αποτελεί η δυνατότητα της προσέγγισης ενός αριθμού χρηστών ο οποίος δεν χαρακτηρίζεται από κάποιο όριο. Ωστόσο, αποτελεί ευθύνη κάθε προγραμματιστή η συγκρότηση ενός διαδικτύου το οποίο χαρακτηρίζεται από ποιότητα και σταθερότητα. Για να επιτευχθούν αυτά τα δύο χαρακτηριστικά, απαραίτητη είναι η συγγραφή κώδικα ο οποίος θα είναι ικανός να προσαρμόζει την εκάστοτε εφαρμογή ανάλογα με το πρόγραμμα περιήγησης που χρησιμοποιεί ο τελικός χρήστης με στόχο να μην αλλάζει η εμπειρία πλοήγησης. Αυτό, λόγω του ότι η υλοποίηση κάθε τεχνολογίας από ένα πρόγραμμα περιήγησης μπορεί να φέρει κάποιες διαφορές οι οποίες

τελικά να οδηγήσουν σε μία μη λειτουργική εφαρμογή. Βέβαια, το πρώτο βήμα πριν από την συγγραφή τέτοιου κώδικα, αποτελεί η μελέτη σκοπιμότητας για το αν η τεχνολογία που πρόκειται να χρησιμοποιηθεί είναι συμβατή σε διαφορετικά προγράμματα περιήγησης, αφού δεν έχει νόημα η δημιουργία μιας εφαρμογής η οποία θα είναι προσβάσιμη από μία μόνο μερίδα χρηστών. Έτσι λοιπόν και στην περίπτωση του Web Speech API, πριν την δημιουργία ή την ενσωμάτωση του σε μία ήδη υπάρχουσα εφαρμογή, συστήνεται η μελέτη της συμβατότητας του σε διαφορετικά προγράμματα περιήγησης. Για την ανάκτηση τέτοιου είδους δεδομένων, μπορούν να χρησιμοποιηθούν διάφορες υπηρεσίες όπως το caniuse.com, οι πίνακες συμβατότητας του MDN και η λίστα συνεχίζεται. Παρακάτω, παρατίθεται ένα κομμάτι από τους πίνακες συμβατότητας του MDN για τα δύο βασικά interfaces του Web Speech API, το SpeechRecognition και το SpeechSynthesis.

	🖥️						📱					
	Chrome	Edge	Firefox	Internet Explorer	Opera	Safari	WebView Android	Chrome Android	Firefox for Android	Opera Android	Safari on iOS	Samsung Internet
SpeechRecognition 🗨️	33 -X- ★ ▼	79 -X- ★ ▼	No	No	No	14.1 -X- ▼	4.4.3 -X- ★ ▼	33 -X- ★ ▼	No	No	14.5 -X- ▼	2.0 -X- ★ ▼

Εικόνα 16: Πίνακας συμβατότητας του MDN για το SpeechRecognition interface

	🖥️						📱					
	Chrome	Edge	Firefox	Internet Explorer	Opera	Safari	WebView Android	Chrome Android	Firefox for Android	Opera Android	Safari on iOS	Samsung Internet
SpeechSynthesis 🗨️	33	18	49	No	21	7	No ★ ▼	33	62	No	7	3.0

Εικόνα 17: Πίνακας συμβατότητας του MDN για το SpeechSynthesis interface

Με βάση το MDN, τα προγράμματα περιήγησης που έχουν πράσινο χρώμα στην στήλη τους υποστηρίζουν το εκάστοτε interface. Παρατηρείται, ότι στην περίπτωση του SpeechSynthesis το Firefox και το Opera το υποστηρίζουν σε αντίθεση με το SpeechRecognition. Παράλληλα, το εικονίδιο δίπλα στο όνομα κάθε interface αναφέρει ότι θεωρείται πειραματικό (experimental), κάτι το οποίο σημαίνει πως είναι πιθανό να τροποποιηθεί μελλοντικά η λειτουργία του. Επιπλέον, το X σε κάθε στήλη αναφέρει ότι για την χρήση του συγκεκριμένου interface είναι απαραίτητη η προσθήκη του προθέματος `webkit`. Η γνώση του αν υποστηρίζονται τα βασικά interfaces είναι το πρώτο πράγμα που πρέπει να κοιτάξει ένας προγραμματιστής. Ωστόσο, κάθε interface μπορεί να περιέχει μία πληθώρα από ιδιότητες, μεθόδους αλλά και γεγονότα, για τα οποία πρέπει να γίνει ξεχωριστή αναζήτηση για το αν αυτά υποστηρίζονται στο εκάστοτε πρόγραμμα περιήγησης καθώς επίσης και αν υπάρχει κάποια διαφορά σε σχέση με τον τρόπο λειτουργία τους.

Χαρακτηριστικό παράδειγμα, αποτελεί η αναφορά του MDN στον πίνακα συμβατότητας του SpeechSynthesis, σύμφωνα με την οποία στο πρόγραμμα περιήγησης Google Chrome για Android η χρήση της μεθόδου `pause` έχει το ίδιο αποτέλεσμα με την `cancel`. Ακόμη ένα παράδειγμα, αποτελεί η μη υποστήριξη της ιδιότητας `grammars` στα προγράμματα περιήγησης Safari και Safari σε iOS το οποίο αναφέρεται στον πίνακα συμβατότητας του SpeechRecognition interface. Τέλος, δεν αρκεί ένας προγραμματιστής να δει τους πίνακες με την συμβατότητα όλων των interfaces και των περιεχομένων αυτών, αφού χωρίς να πραγματοποιήσει ο ίδιος πειράματα δεν μπορεί να είναι σίγουρος για το πόσο επικαιροποιημένα και αληθή είναι τα στοιχεία που παρουσιάζονται. Για παράδειγμα, το caniuse.com αναφέρει πως το πρόγραμμα περιήγησης Microsoft Edge φαίνεται να υποστηρίζει το SpeechRecognition interface χωρίς να λειτουργούν τα γεγονότα που αυτό διαθέτει, κάτι το οποίο προς το παρόν φαίνεται να μην ισχύει, αφού σε πειράματα που διεξήχθησαν στα πλαίσια της παρούσας πτυχιακής φαίνεται να λειτουργούν κανονικά.

Εκτός από τον έλεγχο για την συμβατότητα του Web Speech API σε διάφορα προγράμματα περιήγησης, πρέπει να ληφθεί υπόψη ότι ακόμα και αν το API είναι υλοποιημένο σε όλα, καθένα από αυτά μπορεί να χρησιμοποιεί διαφορετικές πρακτικές όπως για παράδειγμα η υπηρεσία αναγνώρισης που χρησιμοποιείται να διαθέτει διαφορετικό αλγόριθμο μετατροπής ομιλίας σε κείμενο. Κάτι τέτοιο, μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα τα οποία επηρεάζουν την λειτουργικότητα της εφαρμογής από ένα πρόγραμμα περιήγησης σε ένα άλλο. Για παράδειγμα, ενώ στο πρόγραμμα περιήγησης Google Chrome επιστρέφεται η φράση που αναγνωρίστηκε όπως ακριβώς ειπώθηκε με όλα πεζά (εκτός από ορισμένες εξαιρέσεις όπως π.χ. ονόματα όπου το πρώτο είναι κεφαλαίο), στην περίπτωση του Microsoft Edge προστίθεται στο τέλος μία τελεία και το πρώτο γράμμα είναι κεφαλαίο. Κάτι τέτοιο, μπορεί να επηρεάσει την λογική της εκάστοτε εφαρμογής σε περιπτώσεις όπως ο έλεγχος για το αν αυτό που ειπώθηκε από τον χρήστη αποτελεί μία προκαθορισμένη εντολή, αφού όπως είναι γνωστό στο πεδίο της πληροφορικής ένα αλφαριθμητικό όπως το “`read products`” είναι διαφορετικό από το “`Read products.`” Μία λύση στο συγκεκριμένο πρόβλημα, θα μπορούσε να αποτελέσει η συγγραφή κώδικα ο οποίος να αφαιρεί την τελεία και να μετατρέπει όλα τα γράμματα σε πεζά από κάθε αποτέλεσμα που επιστρέφεται. Παρόμοια προβλήματα μπορούν να εμφανιστούν και στην μορφή που θα επιστρέφει μία υπηρεσία αναγνώρισης αποτελέσματα όπως για παράδειγμα μία ημερομηνία ή έναν αριθμό, για τα οποία ο προγραμματιστής καλείται εξίσου να γράψει κώδικα ο οποίος να είναι ικανός να διαχειριστεί κάθε περίπτωση. Παράλληλα, στην περίπτωση της συνθετικής ομιλίας, δεν αποτελεί καλή πρακτική η επιλογή μιας συγκεκριμένης φωνής, αφού δεν είναι απαραίτητο ότι αυτή θα είναι διαθέσιμη και σε άλλα προγράμματα περιήγησης.

Με βάση τα παραπάνω είναι αυτονόητο πως ένας προγραμματιστής που επιθυμεί να κάνει χρήση του Web Speech API, θα πρέπει αρχικά να μελετήσει την συμβατότητα αυτού σε διαφορετικά

προγράμματα περιήγησης έτσι ώστε να είναι σίγουρος ότι τα interfaces και τα περιεχόμενα αυτών που επιθυμεί να χρησιμοποιήσει θα είναι διαθέσιμα. Παράλληλα, θα πρέπει να γράψει κώδικα ο οποίος να προστατεύει την λειτουργικότητα της εφαρμογής όταν αυτή ανοιχτεί σε διαφορετικά προγράμματα περιήγησης, αφού για παράδειγμα είναι πιθανό να επιστρέφονται με ανόμοιο τρόπο τα αποτελέσματα που αναγνωρίζονται κάθε φορά (όπως με το πρόγραμμα περιήγησης Microsoft Edge στο οποίο προστίθεται μία τελεία). Συμπερασματικά, το Web Speech API υποστηρίζεται από τα μεγαλύτερα προγράμματα περιήγησης, κάτι το οποίο φαίνεται να δίνει αρκετές προοπτικές για δημιουργία πληθώρας εφαρμογών με την χρήση του. Ωστόσο, το γεγονός σύμφωνα με το οποίο μέχρι και σήμερα θεωρείται πειραματικό, πρέπει να ωθεί τους προγραμματιστές να πειραματιστούν με τις δυνατότητες του, αλλά η χρήση του σε υπάρχουσες εφαρμογές θα πρέπει να γίνεται με στόχο την βελτίωση της υπάρχουσας υπηρεσίας και όχι την αντικατάστασή της.

2.8 Συμπεράσματα

Το Web Speech API, αποτελεί μία από τις πρώτες προσπάθειες ενσωμάτωσης στις εφαρμογές του διαδικτύου δυνατότητες αναγνώρισης και σύνθεσης ομιλίας. Μέσω της χρήσης του, καθίσταται εφικτή η δημιουργία πληθώρας εφαρμογών ικανές να αλλάζουν τον τρόπο αλληλεπίδρασης έτσι ώστε αυτός να μην αποκλείει ομάδες όπως οι χρήστες με αναπηρία. Στο παρόν κεφάλαιο, μελετήθηκαν οι δυνατότητες του API τόσο για την αναγνώριση όσο και την σύνθεση ομιλίας. Αναλύθηκαν βασικά interfaces και τα περιεχόμενα αυτών, ενώ πραγματοποιήθηκε μία συνοπτική περιγραφή για τις ενέργειες που έγιναν πριν την δημιουργία του αλλά και για το περιεχόμενο του specification του. Τέλος, μελετήθηκε η συμβατότητα του σε διαφορετικά προγράμματα περιήγησης καθώς επίσης και οι προκλήσεις με τις οποίες πρέπει να έρθει αντιμέτωπος ένας προγραμματιστής που επιθυμεί να το χρησιμοποιήσει. Όπως φάνηκε, οι δυνατότητες του Web Speech API δεν αφορούν μόνο μία απλή αναγνώριση και σύνθεση ομιλίας. Δίνεται η ικανότητα της παραμετροποίησης αλλά και της εκτέλεσης κώδικα ανάλογα με την κατάσταση στην οποία βρίσκεται η αναγνώριση η αντίστοιχα η σύνθεση ομιλίας. Όλα αυτά τα χαρακτηριστικά, συνθέτουν ένα εργαλείο το οποίο μπορεί να χρησιμοποιηθεί για την δημιουργία εμπειριών αλληλεπίδρασης υψηλών προδιαγραφών. Το γεγονός ότι βρίσκεται σε πειραματική κατάσταση, πρέπει να αποτελεί κίνητρο για τους προγραμματιστές έτσι ώστε να δείξουν μέσω της δημιουργικότητάς τους τι μπορεί να φτιαχτεί με την χρήση του, ενώ παράλληλα το ίδιο μπορεί να ενσωματωθεί σε ήδη υπάρχουσες εφαρμογές βελτιώνοντας την παρεχόμενη υπηρεσία. Στο επόμενο κεφάλαιο, περιγράφεται αναλυτικά ένα proof of concept ηλεκτρονικό κατάστημα που δημιουργήθηκε με στόχο την ανάδειξη του τρόπου με τον οποίο μέσω της χρήσης τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας μπορούν να βοηθηθούν άτομα με προβλήματα όρασης για την πλοήγηση τους σε ένα τέτοιο περιβάλλον. Το συγκεκριμένο ηλεκτρονικό κατάστημα, αξιοποιεί τις δυνατότητες του Web Speech API και συνεπώς ενσωματώνει το μεγαλύτερο μέρος των στοιχείων που αναλύθηκαν στο παρόν κεφάλαιο.

3. Ηλεκτρονικό κατάστημα με δυνατότητες υποβοήθησης ατόμων με προβλήματα όρασης για πλοήγηση μέσω φωνητικών εντολών

3.1 Εισαγωγή

Στα πλαίσια της παρούσας πτυχιακής εργασίας, δημιουργήθηκε το menShoes. Ένα demo με στόχο την προσομοίωση του τρόπου με τον οποίο χρήστες με προβλήματα όρασης μπορούν να πλοηγηθούν σε ένα ηλεκτρονικό κατάστημα μέσω της ομιλίας τους με την χρήση ενός φωνητικού βοηθού (voice assistant). Αξιοποιώντας τις δυνατότητες του Web Speech API, το menShoes αποτελεί ένα proof of concept μέσω του οποίου αποδεικνύεται ότι με την χρήση τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας καθίσταται εφικτή προγραμματιστικά η δυνατότητα αλληλεπίδρασης ενός χρήστη με ένα ηλεκτρονικό κατάστημα ακόμα και αν ο ίδιος έχει μηδενική όραση. Στο menShoes, ένας χρήστης μπορεί μέσω της φωνής του να πλοηγηθεί στις διαθέσιμες κατηγορίες βάζοντας στο καλάθι του προϊόντα που επιθυμεί, με βάση κριτήρια όπως είναι η περιγραφή και η τιμή κάθε προϊόντος. Εκτός από την αλληλεπίδραση με τις υπάρχουσες κατηγορίες, σημαντική είναι και η δυνατότητα που δίνεται για αναζήτηση προϊόντων μέσω της φωνής. Ανά πάσα στιγμή, ο χρήστης έχει την ικανότητα να μεταβεί στο καλάθι αγορών του, περιεχόμενο του οποίου αποτελεί οποιοδήποτε προϊόν έχει βάλει ο ίδιος προηγουμένως. Από εκεί, καθίστανται εφικτές λειτουργίες όπως η αλλαγή στην ποσότητα ή η διαγραφή ενός προϊόντος, ενώ παράλληλα είναι δυνατή και η γνώση της συνολικής τιμής του καλαθιού. Επιπλέον, το menShoes δίνει την δυνατότητα στον χρήστη να συμπληρώσει μία φόρμα επικοινωνίας εκφράζοντας οποιαδήποτε ανάγκη ή παράπονο μπορεί να έχει. Όλα αυτά, πραγματοποιούνται μέσω της χρήσης ενός φωνητικού βοηθού ο οποίος αποτελείται από τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας.

Κατά την υλοποίηση του συγκεκριμένου proof of concept, αποτέλεσε κρίσιμο παράγοντα η αλληλεπίδραση ενός χρήστη να γίνεται με τον πιο εύκολο και ταυτόχρονα αποδοτικό τρόπο. Για την επίτευξη του συγκεκριμένου στόχου, κάθε σελίδα του ηλεκτρονικού καταστήματος σχεδιάστηκε έτσι ώστε να περιλαμβάνει τρεις εντολές, οι οποίες καλύπτουν τις βασικότερες ανάγκες ενός χρήστη. Πρώτον, την ανάγκη για γνώση των περιεχομένων του μενού πλοήγησης έτσι ώστε ο εκάστοτε χρήστης να γνωρίζει οποιαδήποτε στιγμή που μπορεί να μεταβεί. Δεύτερον, την ανάγκη για γνώση του είδους του περιεχομένου που υπάρχει στην τρέχουσα σελίδα αλλά και τον τρόπο με τον οποίο ο χρήστης μπορεί να αλληλεπιδράσει με αυτό και τρίτον, την ανάγκη για γνώση των διαθέσιμων εντολών που μπορούν να χρησιμοποιηθούν στην σελίδα αυτή. Με την συγκεκριμένη φιλοσοφία, ένας χρήστης μπορεί να πλοηγηθεί στο ηλεκτρονικό κατάστημα χωρίς να χρειάζεται να θυμάται μία μεγάλη πληθώρα εντολών ανάλογα με την φύση της σελίδας στην οποία βρίσκεται, κάτι το οποίο μπορεί να μετατραπεί σε μία δυσάρεστη εμπειρία η οποία τελικά θα οδηγήσει στην έξοδο του από το κατάστημα. Αντίθετα, με την συγκεκριμένη μέθοδο ο χρήστης μπορεί να πλοηγηθεί με την χρήση τριών βασικών εντολών μέσω των οποίων αποκτά την απαραίτητη

γνώση για να αλληλεπιδράσει με κάθε σελίδα του καταστήματος, είτε αυτή περιλαμβάνει προϊόντα μίας κατηγορίας, είτε αποτελεί το καλάθι αγορών κλπ. Η ανάγκη για την δημιουργία μίας διεπαφής η οποία θα παρέχει ανά πάσα στιγμή τις απαραίτητες πληροφορίες στον τελικό χρήστη για το που βρίσκεται αλλά και τι επιλογές του δίνονται σε εκείνο το σημείο έχει μεγάλη βαρύτητα, αφού σε αντίθεση με έναν χρήστη που βλέπει κανονικά δεν έχει την επιλογή να πλοηγηθεί με καθιερωμένους τρόπους όπως το ποντίκι. Συνεπώς, η πρόκληση της δημιουργίας μίας διεπαφής η οποία θα επιτρέπει την εξ ολοκλήρου αλληλεπίδραση με την φωνή είναι μεγάλη και θα πρέπει να θωρακίζει την εμπειρία του τελικού χρήστη έτσι ώστε αυτός να μην φτάσει ποτέ στο σημείο που δεν γνωρίζει πως πρέπει να ενεργήσει.

Το συγκεκριμένο κεφάλαιο χωρίζεται σε τρεις ενότητες. Στην πρώτη, παρουσιάζονται φωτογραφίες του ηλεκτρονικού καταστήματος σε συνδυασμό με την ανάλυση των λειτουργιών του. Στην δεύτερη, αναλύεται η τεχνική υποδομή του ηλεκτρονικού καταστήματος αναφέροντας τις μεθόδους υλοποίησης που έκαναν εφικτή την δημιουργία του συγκεκριμένου proof of concept. Στην τρίτη και τελευταία ενότητα, αναφέρονται τα προβλήματα που αντιμετωπίστηκαν κατά την διάρκεια της δημιουργίας.

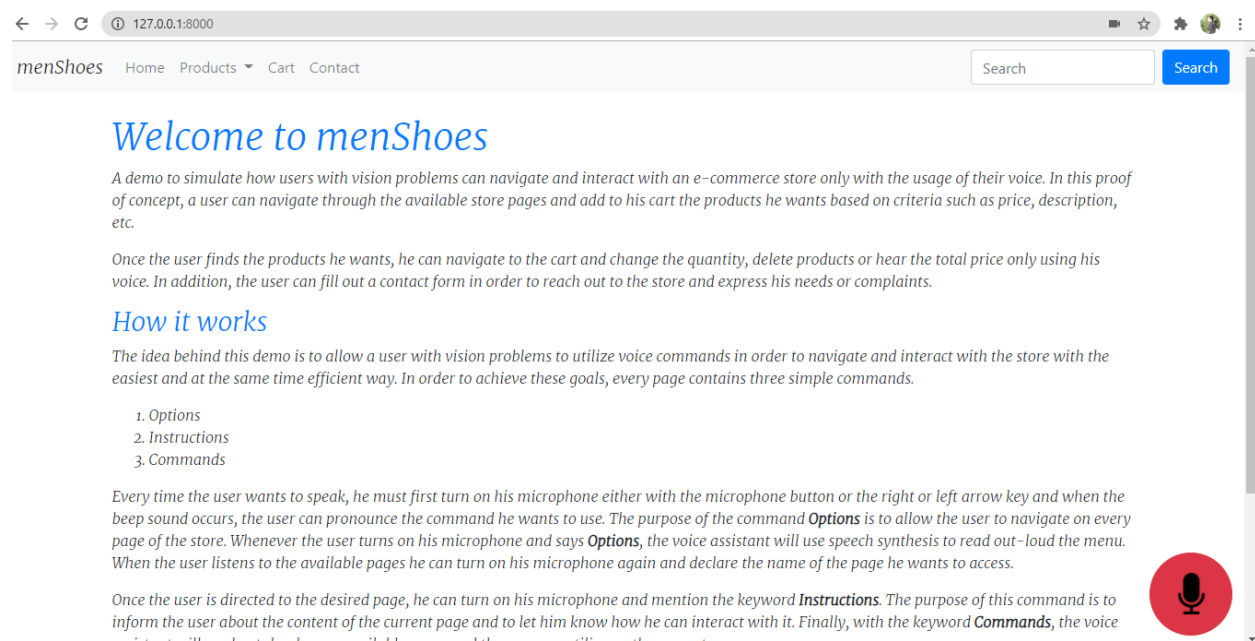
3.2 Αναλυτική περιγραφή των λειτουργιών του ηλεκτρονικού καταστήματος

Όπως αναφέρθηκε στην εισαγωγή, το menShoes αποτελεί ένα proof of concept βάση του οποίου ένας χρήστης με προβλήματα όρασης μπορεί να πλοηγηθεί εξ ολοκλήρου μέσω της φωνής του σε ένα ηλεκτρονικό κατάστημα. Για την καλύτερη περιγραφή των λειτουργιών που παρέχονται στον τελικό χρήστη, η συγκεκριμένη ενότητα χωρίζεται σε έξι υποενότητες. Σε κάθε μία από αυτές, παρατίθεται μία πτυχή του ηλεκτρονικού καταστήματος με την χρήση εικόνων, αλλά και περιγράφοντας αναλυτικά τις λειτουργίες που περιλαμβάνονται σε αυτή. Με την ολοκλήρωση ενός έργου λογισμικού, ένα από τα μεγαλύτερα εμπόδια για την διάδοση του στον κόσμο μπορεί να αποτελέσει σε πολλές περιπτώσεις η γλώσσα. Για τον λόγο αυτό, αποτελεί συχνή πρακτική η επιλογή των Αγγλικών ως βασική γλώσσα, αφού αυτή έχει επιλεγεί ως μέσο για την λεκτική επικοινωνία μεταξύ διαφορετικών χωρών. Έτσι λοιπόν και εδώ, η γλώσσα που επιλέχτηκε για την δημιουργία του proof of concept είναι η αγγλική, λόγω του διεθνούς εύρους χρηστών στο οποίο μπορεί να απευθυνθεί. Ωστόσο, όπως αναφέρεται και στο κεφάλαιο 4, ένα κομμάτι της μελλοντικής έρευνας αποτελεί η μελέτη του τρόπου με τον οποίο είναι εφικτή η μετατροπή της εφαρμογής σε πολύγλωσση, έτσι ώστε ο γλωσσολογικός παράγοντας να μην αποτελεί εμπόδιο για κανέναν.

3.2.1 Αρχική σελίδα

Η αρχική σελίδα στο menShoes ή αλλιώς Home, έχει ως στόχο της να ενημερώσει τον εκάστοτε επισκέπτη για τον σκοπό της δημιουργίας του συγκεκριμένου proof of concept, αλλά και να τον κατευθύνει

για το πως μπορεί να πλοηγηθεί στο ηλεκτρονικό κατάστημα εξ ολοκλήρου μέσω της φωνής του. Παρακάτω, παρατίθεται φωτογραφία από την συγκεκριμένη σελίδα.



Εικόνα 18: Αρχική σελίδα του menShoes

Όπως προαναφέρθηκε, έγινε προσπάθεια για την εύρεση ενός τρόπου πλοήγησης ο οποίος να είναι όσο το δυνατόν πιο απλός και ταυτόχρονα αποδοτικός γίνεται. Η ευχέρεια στην χρήση του διαδικτύου από άτομα τα οποία αντιμετωπίζουν προβλήματα όρασης δεν πρέπει να θεωρείται δεδομένη και συνεπώς είναι απαραίτητη η παροχή όσο το δυνατό περισσότερων πληροφοριών γίνεται για τους διαθέσιμους τρόπους αλληλεπίδρασης. Για έναν ηλικιωμένο με προβλήματα όρασης ή ένα άτομο το οποίο αντιμετωπίζει τέτοια προβλήματα από μικρή ηλικία είναι αρκετά πιθανό μεγάλο μέρος της τεχνολογίας του διαδικτύου να αποτελεί μαύρο κουτί, το οποίο είναι στο χέρι των προγραμματιστών να συμβάλουν έτσι ώστε αυτά τα άτομα να ανακαλύψουν το περιεχόμενο του με τρόπο ο οποίος να μην είναι χαώδης. Έτσι λοιπόν, στην περίπτωση του menShoes, το μόνο που απαιτείται από έναν χρήστη για την πλοήγηση του είναι να θυμάται τρεις βασικές εντολές και συγκεκριμένα:

1. Instructions: Με την χρήση της συγκεκριμένης εντολής, θα εκφωνηθούν από τον φωνητικό βοηθό οδηγίες σχετικά με το περιεχόμενο της τρέχουσας σελίδας στην οποία βρίσκεται ο χρήστης, αλλά και του τρόπου με τον οποίο μπορεί να αλληλεπιδράσει με αυτό. Για παράδειγμα, σε μία σελίδα η οποία περιλαμβάνει τα προϊόντα μιας κατηγορίας, ο χρήστης καλείται να τοποθετήσει ορισμένα από αυτά στο καλάθι του, ενώ στην σελίδα με την φόρμα επικοινωνίας καλείται να συμπληρώσει

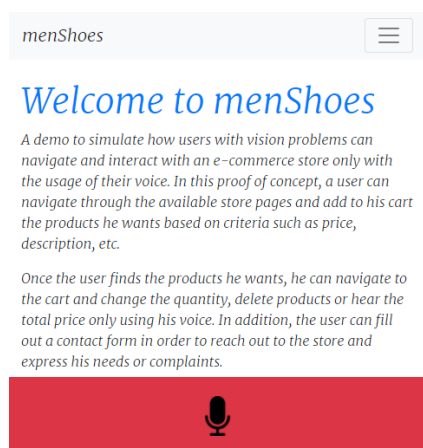
ορισμένα πεδία. Συνεπώς, σε κάθε σελίδα ο χρήστης αλληλεπιδρά διαφορετικά με βάση το περιεχόμενο της, κάτι το οποίο γεννά την ανάγκη για γνώση του τρόπου με τον οποίο πρέπει να ενεργήσει κάθε φορά.

2. **Commands:** Στο menShoes, ένας χρήστης μπορεί να χρησιμοποιήσει μία πληθώρα εντολών ανάλογα με το περιεχόμενο της σελίδας στην οποία βρίσκεται. Για παράδειγμα, σε μία σελίδα η οποία περιλαμβάνει τα προϊόντα μίας κατηγορίας, ο χρήστης μπορεί να χρησιμοποιήσει εντολές οι οποίες θα διαβάσουν τα στοιχεία των προϊόντων, θα τοποθετήσουν αυτά που επιθυμεί στο καλάθι του κλπ. Αντίθετα, στην σελίδα του καταστήματος η οποία αφορά το καλάθι αγορών, ο χρήστης έχει διαθέσιμες διαφορετικές εντολές οι οποίες μπορούν για παράδειγμα να αλλάξουν την ποσότητα ενός προϊόντος, να διαγράψουν ένα άλλο κλπ. Συνεπώς, σε κάθε σελίδα ο χρήστης μπορεί να χρησιμοποιήσει μία πληθώρα διαφορετικών εντολών, κάτι το οποίο γεννά την ανάγκη για συνεχή γνώση αυτών που μπορεί να αξιοποιήσει, χωρίς να χρειάζεται να θυμάται το σύνολο τους, αλλά ούτε και να περιμένει την εκφώνηση όλων κάθε φορά που επιθυμεί να βρει αυτή που χρειάζεται. Αντίθετα, μέσω της χρήσης της εντολής commands και ανάλογα με την σελίδα στην οποία βρίσκεται, ο φωνητικός βοηθός θα εκφωνήσει τις αντίστοιχες εντολές που μπορεί να αξιοποιήσει σε αυτήν.
3. **Options:** Κάθε κατάσταση που σέβεται τους πελάτες του, οφείλει να οργανώνει τον χώρο του με τρόπο ο οποίος να είναι φιλικός προς τους καταναλωτές. Μέσω ενός υπαλλήλου ή πινακίδων, οι πελάτες μπορούν να κατατοπιστούν για το σημείο στο οποίο υπάρχει το εκάστοτε προϊόν που αναζητούν. Διαφορετικά, σε ένα κατάστημα του οποίου τα προϊόντα δεν είναι κατηγοριοποιημένα ή οι πελάτες δεν γνωρίζουν που βρίσκονται τα δοκιμαστήρια ή το ταμείο είναι επόμενο να συγχιστούν με αποτέλεσμα να αποχωρήσουν. Με παρόμοιο τρόπο, τα ηλεκτρονικά καταστήματα είναι απαραίτητο να έχουν οργανωμένο το περιεχόμενο τους διευκολύνοντας τους καταναλωτές κατά την πλοήγηση τους. Σε αυτή την περίπτωση, τον ρόλο ενός υπαλλήλου ή των πινακίδων για το που βρίσκεται τι, τον έχει το μενού πλοήγησης. Μέσω αυτού, οι καταναλωτές έχουν την ικανότητα να περιηγηθούν στο ηλεκτρονικό κατάστημα με οργανωμένο τρόπο, βρίσκοντας κάθε φορά αυτό που αναζητούν χωρίς να χρειάζεται να ψάχνουν εξαντλητικά σε όλες τις σελίδες. Στην περίπτωση ενός χρήστη με προβλήματα όρασης, η ανάγκη ύπαρξης ενός οργανωμένου μενού είναι ακόμα μεγαλύτερη, αφού αυτό θα αποτελέσει το μέσο για τον κατατοπισμό αλλά και την μεταφορά του σε οποιαδήποτε σελίδα του καταστήματος. Παράλληλα, θα αποτελέσει την πηγή γνώσης σχετικά με το αν αυτό που αναζητά παρέχεται στο συγκεκριμένο κατάστημα, κάτι το οποίο δεν είναι εύκολο να διαπιστώσει οπτικά, αφού είναι πιθανό να έχει από πολύ λίγη έως και μηδενική όραση. Βέβαια, όσο οργανωμένο και αν είναι ένα μενού πλοήγησης, καθοριστικό ρόλο έχει ο τρόπος με τον οποίο θα σχεδιαστεί η αλληλεπίδραση με αυτό μέσω της χρήσης τεχνολογιών

αναγνώρισης και σύνθεσης ομιλίας, έτσι ώστε ένας χρήστης με προβλήματα όρασης να μπορεί να πληροφορηθεί σχετικά με τις διαθέσιμες σελίδες, αλλά και να επιλέξει μέσω της φωνής του αυτή στην οποία επιθυμεί να μεταβεί. Στην περίπτωση του menShoes, ο τρόπος διάσχισης του μενού μπορεί να παρομοιαστεί με ένα δέντρο, όπου με την χρήση της εντολής Options εκφωνείται από τον φωνητικό βοηθό σειριακά το πρώτο επίπεδο ή αλλιώς η ρίζα. Στο συγκεκριμένο επίπεδο, περιλαμβάνονται δύο ειδών επιλογές. Αυτές οι οποίες αφορούν σελίδες στις οποίες ο χρήστης μπορεί να μεταβεί κατευθείαν όπως για παράδειγμα το καλάθι αγορών και αυτές που αποτελούν ένα υπό-μενού τα περιεχόμενα του οποίου θα εκφωνηθούν μόνο εάν ο χρήστης αιτηθεί μέσω της φωνής του την συγκεκριμένη επιλογή. Με παρόμοιο τρόπο, μία από τις επιλογές ενός υπό-μενού μπορεί να περιλαμβάνει αντίστοιχα ένα επιπλέον υπο-μενού, τα περιεχόμενα του οποίου θα εκφωνηθούν μόνο εφόσον ο χρήστης αναφέρει το όνομα της. Η συγκεκριμένη λογική, μπορεί να εφαρμοστεί για όσο βάθος είναι επιθυμητό από το εκάστοτε κατάστημα. Με αυτό τον σχεδιασμό, επιτυγχάνονται δύο σημαντικά πλεονεκτήματα. Σύμφωνα με το πρώτο, ένας χρήστης δεν αναγκάζεται να ακούσει όλες τις επιλογές του μενού, παρά μόνο του πρώτου επιπέδου και στην συνέχεια με βάση αυτές μπορεί να αλληλεπιδράσει με το κομμάτι του μενού το οποίο σχετίζεται με αυτό που αναζητά κάθε φορά. Το δεύτερο πλεονέκτημα, αφορά την δυνατότητα ενός χρήστη να μεταβεί κατευθείαν σε μία σελίδα όσο βαθιά στο μενού και αν αυτή περιέχεται μόνο με την χρήση του ονόματος της. Έτσι, αποφεύγεται μία κατάσταση κατά την οποία ο χρήστης θα ήταν υποχρεωμένος να διασχίζει κάθε φορά ολόκληρο το μονοπάτι του μενού που αντιστοιχεί στην επιλογή που επιθυμεί, ενώ ήδη γνωρίζει το όνομα αυτής από προηγούμενη αλληλεπίδραση. Με βάση τον συγκεκριμένο σχεδιασμό, το menShoes επιχειρεί να μετατρέψει την διαδικασία της πλοήγησης ενός χρήστη στις σελίδες του καταστήματος σε όσο το δυνατόν πιο απλή και ταυτόχρονα αποδοτική γίνεται.

Μέχρι στιγμής, αναφέρθηκαν οι τρεις βασικές εντολές με την χρήση των οποίων καθίσταται εφικτή η πλοήγηση ενός χρήστη στο κατάστημα εξ ολοκλήρου μέσω της φωνής του. Ωστόσο, δεν αναφέρθηκε ο τρόπος με τον οποίο είναι εφικτή η αλληλεπίδραση με τον φωνητικό βοηθό έτσι ώστε να αυτός να ακούει τις εντολές που επιθυμεί να δώσει ο τελικός χρήστης. Στα πλαίσια της δημιουργίας του συγκεκριμένου proof of concept, υλοποιήθηκαν δύο δυνατοί τρόποι για την ενεργοποίηση του μικροφώνου. Ο πρώτος, αφορά την χρήση του κουμπιού που εμφανίζεται κάτω δεξιά στην εικόνα 18. Το συγκεκριμένο κουμπί, υπάρχει σε κάθε σελίδα του καταστήματος και με την χρήση του ενεργοποιείται το μικρόφωνο και συνεπώς η αναγνώριση ομιλίας. Το κουμπί αυτό, ενσωματώθηκε λόγω της πιθανότητας που υπάρχει ένας χρήστης με προβλήματα όρασης να χρησιμοποιεί για την πλοήγηση του στο διαδίκτυο εργαλεία όπως ο είναι ο αναγνώστης οθόνης (screen reader). Με την χρήση τέτοιου εργαλείου, ο χρήστης μπορεί να επιλέξει το κουμπί ενεργοποίησης του μικροφώνου οποιαδήποτε στιγμή επιθυμεί να αλληλεπιδράσει με τον φωνητικό

βοηθό. Εναλλακτικά, για τους χρήστες που δεν έχουν την δυνατότητα ή δεν επιθυμούν να αξιοποιήσουν εργαλεία όπως είναι οι αναγνώστες οθόνης, δίνεται η δυνατότητα ενεργοποίησης του μικροφώνου με την χρήση του αριστερού ή του δεξιού βέλους του πληκτρολογίου. Να σημειωθεί, ότι θα μπορούσαν να επιλεγούν διάφοροι συνδυασμοί πλήκτρων για την ενεργοποίηση του μικροφώνου. Ωστόσο, η εύρεση του καταλληλότερου συνδυασμού δεν αποτέλεσε κομμάτι της παρούσας πτυχιακής, αφού αυτή επικεντρώθηκε στον τρόπο με τον οποίο ένας χρήστης με προβλήματα όρασης μπορεί μέσω της χρήσης τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας να αλληλεπιδράσει με ένα ηλεκτρονικό κατάστημα. Έτσι λοιπόν, ο αποδοτικότερος τρόπος αλληλεπίδρασης ενός χρήστη με προβλήματα όρασης με μία συσκευή όπως είναι το πληκτρολόγιο ή μία οθόνη κινητού δεν ερευνηθήκε στα πλαίσια υλοποίησης του συγκεκριμένου proof of concept. Βέβαια, για την περίπτωση που ο χρήστης πλοηγείται από μία συσκευή όπως είναι το κινητό ή το τάμπλετ, γράφηκε κώδικας έτσι ώστε σε τέτοια μεγέθη το κουμπί της ενεργοποίησης του μικροφώνου να τοποθετείται στο κάτω μέρος της οθόνης πιάνοντας όλο το πλάτος της, καθιστώντας με αυτό τον τρόπο όσο το δυνατόν ευκολότερη γίνεται την αλληλεπίδραση μαζί του όταν δεν είναι εφικτή η χρήση συσκευών όπως το πληκτρολόγιο. Παρακάτω, φαίνεται το συγκεκριμένο κουμπί όταν το κατάστημα είναι ανοιχτό σε μικρότερη οθόνη.



Εικόνα 19: Αρχική σελίδα του menShoes όταν είναι ανοιχτό σε μικρότερη οθόνη

Με την ενεργοποίηση του μικροφώνου, γράφηκε κώδικας έτσι ώστε να ακούγεται πάντα ένας χαρακτηριστικός ήχος. Στόχος της συγκεκριμένης προσθήκης, αποτελεί η ειδοποίηση του τελικού χρήστη ότι το μικρόφωνο του έχει ενεργοποιηθεί και συνεπώς μπορεί να εκφωνήσει οποιαδήποτε εντολή επιθυμεί. Μία τέτοια προσθήκη είναι κρίσιμη, αφού ο τελικός χρήστης λόγω του προβλήματος όρασης που αντιμετωπίζει πρέπει να είναι βέβαιος ότι έχει ενεργοποιήσει το μικρόφωνο του, προβαίνοντας έτσι στην εκφώνηση των δεδομένων εισόδου. Με την ολοκλήρωση της εκάστοτε εντολής ή φράσης που θα εκφωνήσει ο τελικός χρήστης, η διαδικασία αναγνώρισης ομιλίας ολοκληρώνεται και το μικρόφωνο του απενεργοποιείται, ενώ στην συνέχεια χρησιμοποιείται συνθετική ομιλία ανάλογα με τα δεδομένα εισόδου

που έδωσε ο ίδιος μέσω της φωνής του, ακούγοντας έτσι τα αποτελέσματα των ενεργειών του. Με την απενεργοποίηση του μικροφώνου του χρήστη, υπάρχουν τρία πιθανά σενάρια. Το πρώτο, αφορά την περίπτωση κατά την οποία η υπηρεσία που χρησιμοποιείται για την αναγνώριση ομιλίας δεν κατάφερε να αναγνωρίσει κάτι από τα δεδομένα εισόδου και συνεπώς με την χρήση συνθετικής ομιλίας θα εκφωνηθεί από τον φωνητικό βοηθό το κείμενο “I didn’t understand anything. Please try again”. Με το συγκεκριμένο κείμενο, ο τελικός χρήστης ενημερώνεται για την αδυναμία αναγνώρισης των λεγόμενων του και εφόσον επιθυμεί μπορεί να ξανά προσπαθήσει. Το δεύτερο σενάριο, αφορά την περίπτωση κατά την οποία ο χρήστης θα εκφωνήσει μία εντολή η οποία δεν υπάρχει, με αποτέλεσμα να εκφωνηθεί με την χρήση συνθετικής ομιλίας το κείμενο “No command with name [name of the command the user said]. Please try again”. Το τρίτο και τελευταίο σενάριο, αφορά την περίπτωση κατά την οποία ο χρήστης θα εκφωνήσει μία εντολή η οποία υπάρχει. Αυτό, θα οδηγήσει στην χρήση συνθετικής ομιλίας για την εκφώνηση μίας φράσης η οποία αντιστοιχεί στο αποτέλεσμα της ενέργειας που πραγματοποιήθηκε μέσω της εντολής. Βέβαια, σε πολλές περιπτώσεις ο χρήστης καλείται να δώσει δεδομένα εισόδου μέσω της φωνής του τα οποία δεν αφορούν κάποια εντολή, αλλά αντίθετα χρειάζονται για την διεκπεραίωση μιας λειτουργίας η οποία αιτήθηκε αρχικά μέσω μιας εντολής. Για παράδειγμα, όταν ο χρήστης κάνει χρήση μίας εντολής όπως η “Add to cart”, ο φωνητικός βοηθός θα ρωτήσει με την χρήση συνθετικής ομιλίας την ποσότητα που επιθυμεί ο χρήστης να αγοράσει. Έτσι, ο χρήστης καλείται να ενεργοποιήσει ξανά το μικρόφωνο του δίνοντας ως είσοδο αυτή την φορά έναν αριθμό ποσότητας. Σε αυτή την περίπτωση, η ποσότητα δεν αποτελεί κάποια εντολή αλλά έναν ακέραιο αριθμό. Ωστόσο, ο φωνητικός βοηθός γνωρίζει ότι ο αριθμός αυτός είναι απαραίτητος για την ολοκλήρωση της εντολής “Add to cart” και συνεπώς δεν θα προβεί στην εκφώνηση του κειμένου “No command with name [name of the command the user said]. Please try again”.

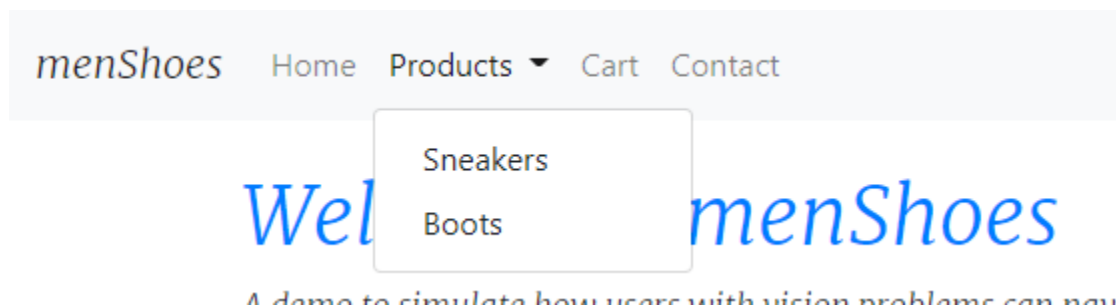
Όπως αναφέρθηκε παραπάνω, σε κάθε σελίδα του menShoes ένας χρήστης μπορεί να χρησιμοποιήσει τις εντολές Instructions, Commands και Options. Ωστόσο, λόγω του ότι η σελίδα Home έχει ενημερωτικό χαρακτήρα σχετικά με το τι αφορά το συγκεκριμένο proof of concept, αλλά και το πως μπορεί κάποιος να το χρησιμοποιήσει, με την ενεργοποίηση του μικροφώνου την πρώτη φορά αυτό δεν θα ανοίξει, αλλά θα εκφωνηθεί με την χρήση συνθετικής ομιλίας το εξής κείμενο:

“Welcome to menshoes demo store. Every time you want to speak, you must first open your microphone either with the microphone button or the right or left arrow key. To start shopping with your voice, open your microphone and when the beep sound occurs, use the command options.”

Συνεπώς, στην αρχική σελίδα η μόνη διαθέσιμη εντολή είναι η Options, με την χρήση της οποίας θα εκφωνηθούν από τον φωνητικό βοηθό οι επιλογές του μενού. Συγκεκριμένα, το κείμενο που θα εκφωνηθεί είναι το εξής:

“Available options: home, products, shopping cart, contact, search for. Say the option you want”

Με την εκφώνηση του παραπάνω κειμένου, παρατηρούνται δύο πράγματα. Αρχικά, ότι οι επιλογές Shopping cart και Search for διατυπώνονται με διαφορετικό τρόπο στο γραφικό περιβάλλον. Κάτι τέτοιο, δεν αποτελεί πρόβλημα, αφού η φωνητική διεπαφή είναι ανεξάρτητη από αυτή που έχει σχεδιαστεί για το γραφικό περιβάλλον. Συνεπώς, είτε επιλεγεί από το μενού με μία συσκευή όπως είναι το ποντίκι η λέξη Cart, είτε χρησιμοποιηθεί με την χρήση της φωνής η λέξη Shopping cart, το αποτέλεσμα θα είναι ακριβώς το ίδιο, το οποίο στην προκυμμένη περίπτωση αφορά την μετάβαση του χρήστη στην σελίδα η οποία αποτελεί το καλάθι αγορών του. Κάτι τέτοιο, επιτρέπει την χρήση περισσότερο περιγραφικών ονομάτων όπου είναι επιθυμητό, με στόχο την καλύτερη δυνατή πληροφόρηση ενός χρήστη που αντιμετωπίζει προβλήματα όρασης, χωρίς να είναι παράλληλα απαραίτητη η χρήση του ίδιου ονόματος και στο γραφικό περιβάλλον. Η δεύτερη παρατήρηση, αφορά το γεγονός σύμφωνα με το οποίο η επιλογή Products αποτελεί ένα υπό-μενού, τα περιεχόμενα του οποίου δεν εκφωνήθηκαν. Παρακάτω, φαίνεται στην εικόνα το περιεχόμενο του υπό-μενού.



Εικόνα 20: Το υπό-μενού που βρίσκεται κάτω από την επιλογή Products

Όπως φαίνεται στην εικόνα, το menShoes διαθέτει δύο κατηγορίες υποδημάτων. Η πρώτη, περιέχει παπούτσια τύπου sneakers, ενώ αντίστοιχα η δεύτερη τύπου Boots. Ωστόσο, όπως φάνηκε από το παραπάνω κείμενο, ο φωνητικός βοηθός εκφώνησε μόνο την επιλογή Products χωρίς να αναφέρει τις κατηγορίες που βρίσκονται κάτω από αυτή. Για την εκφώνηση του συγκεκριμένου υπό-μενού, ο χρήστης πρέπει να ενεργοποιήσει το μικρόφωνο του και να εκφωνήσει την λέξη products. Όπως αναλύθηκε και παραπάνω, με την συγκεκριμένη μέθοδο κάθε φορά που ο χρήστης αξιοποιεί την εντολή options, δεν θα αναγκάζεται να ακούει ολόκληρο το μενού. Βέβαια, στην συγκεκριμένη υλοποίηση η εκφώνηση όλων των επιλογών συμπεριλαμβανομένων και αυτών που βρίσκονται στο υπό-μενού δεν θα αποτελούσε μεγάλο πρόβλημα λόγω του μικρού μεγέθους τους. Ωστόσο, σε ένα πλήρως λειτουργικό κατάστημα το οποίο

παρέχει δεκάδες κατηγορίες και υποκατηγορίες, η αποδοτικότητα του τρόπου πλοήγησης για τον τελικό χρήστη θα μειωνόταν δραματικά.

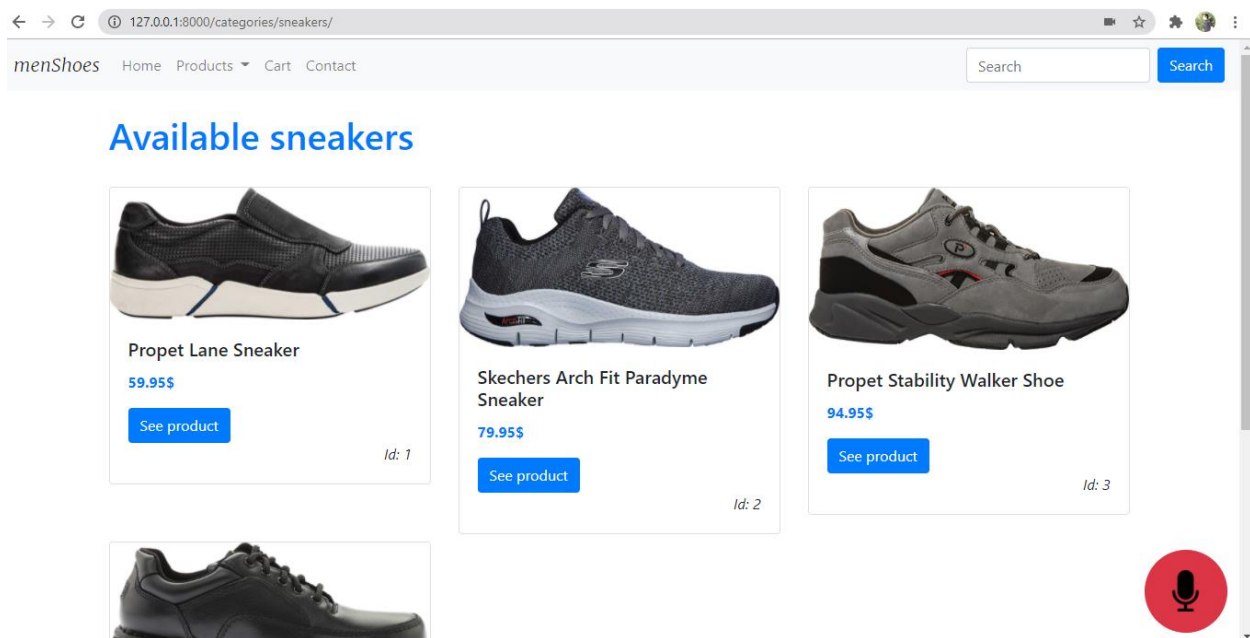
Στις υποενότητες που ακολουθούν, περιγράφονται αναλυτικά οι λειτουργίες κάθε σελίδας που περιλαμβάνεται στο μενού του καταστήματος, ξεκινώντας με τις επιλογές που δίνονται στο υπο-μενού που βρίσκεται κάτω από την επιλογή Products. Συγκεκριμένα, με την ενεργοποίηση του μικροφώνου και την χρήση της λέξης Products, ο φωνητικός βοηθός θα εκφωνήσει το εξής κείμενο:

“Available categories: sneakers, boots. Say the category you want”

Με βάση το παραπάνω κείμενο, εάν ενεργοποιηθεί το μικρόφωνο και στην συνέχεια εκφωνηθεί μία από τις διαθέσιμες κατηγορίες, το αποτέλεσμα θα είναι η ανακατεύθυνση του χρήστη στην αντίστοιχη σελίδα, περιεχόμενο της οποίας θα αποτελούν προϊόντα της κατηγορίας που επέλεξε.

3.2.2 Προϊόντα

Όπως αναφέρθηκε στην προηγούμενη υποενότητα, το menShoes διαθέτει δύο κατηγορίες προϊόντων. Η πρώτη, αφορά υποδήματα τύπου sneakers, ενώ η δεύτερη τύπου boots. Επιλέγοντας μία από τις δύο κατηγορίες, ο χρήστης έχει την δυνατότητα με την χρήση του φωνητικού βοηθού να αξιοποιήσει εντολές με στόχο την ανάκτηση πληροφοριών σχετικά με τα προϊόντα της τρέχουσας σελίδας, έτσι ώστε τελικά να προσθέσει στο καλάθι του αυτά που επιθυμεί. Παρακάτω, φαίνεται η σελίδα με τα προϊόντα της κατηγορίας sneakers. Να σημειωθεί, ότι η ίδια σελίδα χρησιμοποιείται και για τα προϊόντα της κατηγορίας boots δυναμικά (θα αλλάξει στο URL η λέξη sneakers σε boots), οπότε παρατίθεται φωτογραφία με τα προϊόντα της μίας κατηγορίας, αφού σε κάθε περίπτωση οι λειτουργίες που θα περιγραφούν είναι ίδιες με μόνη διαφορά τα προϊόντα.



Εικόνα 21: Προϊόντα της κατηγορίας sneakers

Με την φόρτωση της παραπάνω σελίδας, θα εκφωνηθεί από τον φωνητικό βοηθό κατευθείαν το εξής κείμενο:

“Welcome to sneakers category. If you don’t know what to do, use the command instructions”

Μέσω του συγκεκριμένου μηνύματος, ο χρήστης θα καταλάβει ότι έχει μεταβεί στην επιθυμητή κατηγορία προϊόντων με επιτυχία, ενώ παράλληλα του υπενθυμίζεται ότι σε περίπτωση που δεν γνωρίζει τι να κάνει γιατί π.χ. επισκέπτεται πρώτη φορά το συγκεκριμένο κατάστημα, μπορεί να χρησιμοποιήσει την εντολή “Instructions” έτσι ώστε να κατανοήσει τι υπάρχει στην συγκεκριμένη σελίδα, αλλά και το πως μπορεί να αλληλεπιδράσει με αυτό. Διαφορετικά, μπορεί να αξιοποιήσει κατευθείαν τις διαθέσιμες εντολές που έχει ξαναχρησιμοποιήσει κατά την πλοήγηση του στο συγκεκριμένο κατάστημα. Όταν γίνει χρήση της εντολής “Instructions” στην σελίδα με τα προϊόντα, θα εκφωνηθεί από τον φωνητικό βοηθό το εξής κείμενο:

“Instructions: This page contains all the products you have asked for. You can find the products you want and add them to your shopping cart. The first thing you have to do is to use the command read all. With this command, you will listen to all the available product titles and IDs one by one. When you listen to a product you like, use the command select ID. With this command, you will select the product you want by providing its ID. Once the product is selected you can use one of the available commands such as read product, description, add to cart, etc. If you want to listen to the available commands for this page use the

keyword commands. When you finish with this product, you can use the command continue to listen to the rest of the products on this page. Lastly, if you want to change page you can use the command options. Use the command read all to start shopping now”

Μέσω του παραπάνω κειμένου, ο χρήστης μπορεί να κατανοήσει τον ρόλο της συγκεκριμένης σελίδας αλλά και το τι μπορεί να επιτύχει μέσω της χρήσης της.

Κατά τον σχεδιασμό του συγκεκριμένου proof of concept, μία από τις σημαντικότερες αποφάσεις που έπρεπε να ληφθεί αφορούσε τον τρόπο με τον οποίο ένας χρήστης με προβλήματα όρασης θα έχει την ικανότητα αλληλεπίδρασης με τα προϊόντα με όσο το δυνατό πιο ευέλικτο και αποδοτικό τρόπο γίνεται. Ένας ευέλικτος τρόπος αλληλεπίδρασης, αναφέρεται στην ικανότητα του εκάστοτε χρήστη για γρήγορη μετακίνηση σε μία σελίδα με προϊόντα έτσι ώστε να μην είναι δεσμευμένος μόνο στην σειριακή διάσχιση της πληροφορίας.

Ενώ η σειριακή λογική αποτελεί έναν μη αποδοτικό τρόπο αναζήτησης όταν πρόκειται για μεγάλο πλήθος δεδομένων, η χρήση της σε συνδυασμό με μία επιλεκτική λογική μπορεί να βελτιώσει την διαδικασία της εύρεσης του επιθυμητού προϊόντος δραματικά. Συγκεκριμένα, όπως φαίνεται στην παραπάνω εικόνα, κάθε προϊόν διαθέτει ένα μοναδικό αναγνωριστικό (id) το οποίο το ξεχωρίζει από όλα τα υπόλοιπα προϊόντα του ηλεκτρονικού καταστήματος. Με αυτή την μέθοδο, επιτυγχάνονται ορισμένα οφέλη αναφορικά με την εμπειρία πλοήγησης του χρήστη. Ένα από αυτά, αφορά το γεγονός σύμφωνα με το οποίο ο χρήστης έχει την ικανότητα μέσω της φωνής του να μεταβεί κατευθείαν στο προϊόν με το οποίο επιθυμεί να αλληλεπιδράσει εκφωνώντας το αναγνωριστικό του. Βέβαια, όπως είναι λογικό ένας χρήστης δεν μπορεί να γνωρίζει τα διαθέσιμα αναγνωριστικά εάν αυτά δεν εκφωνηθούν πρώτα από τον φωνητικό βοηθό. Αυτός είναι και ο λόγος που αναφέρθηκε προηγουμένως πως η σειριακή διάσχιση της πληροφορίας σε συνδυασμό με μία επιλεκτική ικανότητα μπορούν να βελτιώσουν την εμπειρία της αναζήτησης προϊόντος. Συγκεκριμένα, με την χρήση της εντολής “Read all”, ο φωνητικός βοηθός θα ξεκινήσει την εκφώνηση του τίτλου και του αναγνωριστικού κάθε προϊόντος σειριακά από τα αριστερά προς τα δεξιά. Η εκφώνηση της τιμής δεν περιλαμβάνεται, με στόχο την γρηγορότερη διάσχιση των προϊόντων, αφού σε περίπτωση που ένας χρήστης ακούσει ένα τίτλο που τον ενδιαφέρει, μπορεί να μεταβεί στο συγκεκριμένο προϊόν μαθαίνοντας την τιμή του κ.α. Την στιγμή που ο φωνητικός βοηθός εκφωνεί τα προϊόντα, ο χρήστης έχει δύο επιλογές. Είτε να ενεργοποιήσει το μικρόφωνο του, είτε να περιμένει μέχρι ο φωνητικός βοηθός να τελειώσει με την εκφώνηση όλων των διαθέσιμων προϊόντων της σελίδας. Σε περίπτωση που χρήστης διαλέξει την πρώτη επιλογή επειδή άκουσε ένα προϊόν που τον ενδιαφέρει, τότε με την ενεργοποίηση του μικροφώνου ο φωνητικός βοηθός θα σταματήσει αυτόματα και θα ακουστεί ο χαρακτηριστικός ήχος που έχει ως στόχο την ενημέρωση για το ότι η αναγνώριση ομιλίας έχει ξεκινήσει. Με το άκουσμα του συγκεκριμένου ήχου, ο χρήστης μπορεί να χρησιμοποιήσει την εντολή “Select id”. Αυτό, θα έχει ως

αποτέλεσμα ο φωνητικός βοηθός να ρωτήσει για το αναγνωριστικό του προϊόντος στο οποίο επιθυμεί ο χρήστης να μεταβεί για την ανάκτηση περισσότερων πληροφοριών σχετικά με αυτό. Εφόσον δοθεί ένα αναγνωριστικό το οποίο είναι έγκυρο, ο φωνητικός βοηθός θα ενημερώσει ότι το προϊόν έχει επιλεγθεί με επιτυχία και ο χρήστης πλέον μπορεί να χρησιμοποιήσει μία από τις διαθέσιμες εντολές με στόχο την ανάκτηση πληροφοριών σχετικά με το προϊόν έτσι ώστε τελικά αν επιθυμεί να το τοποθετήσει στο καλάθι του.

Με την παραπάνω μέθοδο, ο χρήστης αποδεσμεύεται από την σειριακή λογική καθώς έχει την δυνατότητα να μετακινηθεί στα προϊόντα που τον ενδιαφέρουν μέσω της εντολής “Select id”, με μόνη προϋπόθεση την γνώση του αναγνωριστικού από το προϊόν που τον ενδιαφέρει. Επιπλέον, μέσω της χρήσης των αναγνωριστικών, ο χρήστης απαιτείται να θυμάται έναν ή περισσότερους αριθμούς για να μεταβεί στο προϊόν ή στα προϊόντα που επιθυμεί. Σε διαφορετική περίπτωση, θα έπρεπε να χρησιμοποιήσει ένα χαρακτηριστικό όπως ο τίτλος του προϊόντος, κάτι το οποίο σε πολλές περιπτώσεις είναι αδύνατο να θυμάται επακριβώς.

Η δυνατότητα μετακίνησης ανάμεσα σε προϊόντα που δίνεται μέσω της εντολής “Select id” δεν αποτελεί τον μοναδικό τρόπο μετάβασης σε ένα προϊόν. Συγκεκριμένα, ο φωνητικός βοηθός μπορεί να εκφωνεί ένα προϊόν την φορά, επιλέγοντας ο ίδιος ο χρήστης μέσω αντίστοιχης εντολής αν θέλει να του εκφωνηθεί το επόμενο ή αντίστοιχα το προηγούμενο, διασχίζοντας με αυτό τον τρόπο τα προϊόντα με σειριακό τρόπο. Σε κάθε προϊόν που του εκφωνείται, έχει την ικανότητα όπως προηγουμένως να χρησιμοποιήσει μία πληθώρα εντολών.

Με βάση τα παραπάνω, ένας χρήστης κατά την πλοήγηση του στο menShoes έχει δύο εναλλακτικές αναφορικά με τον τρόπο διάσχισης των προϊόντων μίας κατηγορίας. Έτσι, επιχειρείται η παροχή μίας εμπειρίας πλοήγησης η οποία να χαρακτηρίζεται από ευελιξία, με τελικό στόχο την αποφυγή καταστάσεων κατά τις οποίες ο τελικός χρήστης θα αισθανθεί περιορισμένος αναφορικά με την διαδικασίας εύρεσης του επιθυμητού προϊόντος. Παρακάτω, περιγράφεται ένα σενάριο κατά το οποίο ένας χρήστης έχει ως στόχο την τοποθέτηση στο καλάθι αγορών ενός ζευγαριού sneakers. Μέσω του συγκεκριμένου σεναρίου, θα περιγραφούν αναλυτικά οι διαθέσιμες εντολές καθώς επίσης και τα κείμενα που εκφωνεί ο φωνητικός βοηθός σε κάθε περίπτωση.

Αφού ο χρήστης έχει μεταβεί στην σελίδα με τα διαθέσιμα sneakers και έχει κάνει χρήση της εντολής “Instructions”, ενεργοποιεί το μικρόφωνο του και εκφωνεί την εντολή “Read all”. Αμέσως, ο φωνητικός βοηθός ξεκινάει να εκφωνεί από τα αριστερά προς τα δεξιά τα διαθέσιμα sneakers. Με την εκφώνηση του τίτλου και του αναγνωριστικού του δεύτερου προϊόντος, ο χρήστης επιθυμεί να μάθει περισσότερες πληροφορίες για αυτό και συνεπώς ενεργοποιεί το μικρόφωνο εκφωνώντας την εντολή

“Select id”. Με την χρήση της συγκεκριμένης εντολής, θα εκφωνηθεί από τον φωνητικό βοηθό το εξής κείμενο:

“Tell me the product id”

Ο χρήστης, ενεργοποιεί το μικρόφωνο του και εκφωνεί τον αριθμό δύο. Αυτό, θα έχει ως αποτέλεσμα την εκφώνηση από τον φωνητικό βοηθό του μηνύματος:

“Product with id 2 selected use one of the available commands. If you want to listen to the available commands use the keyword commands”

Σε αυτό το σημείο, ο φωνητικός βοηθός έχει επιλέξει το προϊόν με αναγνωριστικό τον αριθμό δύο και ο χρήστης μπορεί να χρησιμοποιήσει μία πληθώρα εντολών για την αλληλεπίδραση μαζί του. Λόγω του ότι ο χρήστης πλοηγείται πρώτη φορά στο menShoes, δεν γνωρίζει τις διαθέσιμες εντολές για την αλληλεπίδραση με ένα προϊόν και συνεπώς αφού ενεργοποιήσει το μικρόφωνο του και ακούσει τον χαρακτηριστικό ήχο που δηλώνει ότι έχει ξεκινήσει η αναγνώριση ομιλίας θα αξιοποιήσει την εντολή “Commands”. Με την χρήση της συγκεκριμένης εντολής, ο φωνητικός βοηθός θα εκφωνήσει όλες τις διαθέσιμες εντολές της τρέχουσας σελίδας. Συγκεκριμένα, οι εντολές που θα εκφωνηθούν στον χρήστη είναι οι εξής έντεκα:

1. Read product
2. Next
3. Back
4. Add to cart
5. Select id
6. Description
7. Open product
8. Read all
9. Continue
10. Instructions
11. Commands

Να σημειωθεί, πως στην παραπάνω λίστα δεν περιλαμβάνεται η εντολή Options, λόγω του ότι αυτή δεν αφορά μία εντολή την οποία ο χρήστης θα αξιοποιήσει για την αλληλεπίδραση του με την τρέχουσα σελίδα. Αντίθετα, θα την χρησιμοποιήσει όταν επιθυμεί να αποχωρήσει από αυτή. Αφού λοιπόν ο χρήστης ακούσει τις διαθέσιμες εντολές για την αλληλεπίδραση του με την συγκεκριμένη σελίδα, ενεργοποιεί το μικρόφωνο του και εκφωνεί την εντολή “Read product”. Με την χρήση της συγκεκριμένης εντολής, ο φωνητικός

βοηθός θα εκφωνήσει τον τίτλο, την τιμή και το αναγνωριστικό του προϊόντος που έχει επιλεγεί νωρίτερα με χρήση της εντολής “Select id”. Ο χρήστης, επιθυμεί να μάθει περισσότερες πληροφορίες για το συγκεκριμένο προϊόν και συνεπώς ενεργοποιεί το μικρόφωνο του και εκφωνεί την εντολή “Description”. Όπως φαίνεται από την εικόνα με τα διαθέσιμα sneakers που παρατέθηκε παραπάνω, καθένα από αυτά περιλαμβάνει ένα όνομα, μία τιμή και τέλος το αναγνωριστικό που το χαρακτηρίζει μοναδικά, χωρίς να εμφανίζεται κάπου μία περιγραφή (description). Αυτό, αφού στα περισσότερα ηλεκτρονικά καταστήματα σήμερα δεδομένα όπως η περιγραφή του προϊόντος, αξιολογήσεις, παραλλαγές (variations) και πολλά ακόμη, βρίσκονται μέσα στην σελίδα του συγκεκριμένου προϊόντος. Ωστόσο, με την χρήση της εντολής Description, θα πραγματοποιηθεί ένα αίτημα στον εξυπηρετητή του menShoes με στόχο την ανάκτηση της περιγραφής του προϊόντος με αναγνωριστικό δύο από την βάση δεδομένων και στην συνέχεια αυτή θα εκφωνηθεί από τον φωνητικό βοηθό στον τελικό χρήστη. Να σημειωθεί, ότι στο συγκεκριμένο proof of concept όπως θα φανεί και στην συνέχεια κάθε προϊόν έχει την δική του σελίδα, η οποία στην ουσία είναι η ίδια και λειτουργεί δυναμικά. Ωστόσο, για την αποδοτικότερη αλληλεπίδραση του χρήστη με τα προϊόντα, υλοποιήθηκε η δυνατότητα της ανάκτησης πληροφοριών οι οποίες δεν απεικονίζονται στην τρέχουσα σελίδα κατευθείαν από την βάση δεδομένων, με αποτέλεσμα ο χρήστης να μην χρειάζεται να αλλάξει σελίδα για την εύρεση τους. Για παράδειγμα, σε περίπτωση που δεν υπήρχε η εντολή “Description” στην σελίδα με τα sneakers, ο χρήστης θα ήταν υποχρεωμένος να ανοίξει το προϊόν με αναγνωριστικό δύο (αναφέρεται παρακάτω ο τρόπος για την μετάβαση στην σελίδα ενός προϊόντος) και έπειτα να αλληλεπιδράσει με την νέα σελίδα για να έρθει τελικά σε επαφή με την περιγραφή (description) του. Επιπλέον, σε περίπτωση που επιθυμούσε να ακούσει και τα υπόλοιπα προϊόντα, θα έπρεπε να επιστρέψει ξανά πίσω και να επαναλάβει την διαδικασία. Κάτι τέτοιο, μπορεί να αποφευχθεί με μία φωνητική διεπαφή, αφού δεν απαιτείται η απεικόνιση του συνόλου των πληροφοριών ενός προϊόντος στο γραφικό περιβάλλον για να εκφωνηθούν στον τελικό χρήστη. Αυτό, αποδείχθηκε με την υλοποίηση της εντολής “Description”, μέσω της οποίας ο χρήστης μπορεί να ακούσει την περιγραφή του εκάστοτε προϊόντος χωρίς να φύγει από την τρέχουσα σελίδα. Αφού εκφωνηθεί από τον φωνητικό βοηθό η περιγραφή του προϊόντος με αναγνωριστικό δύο, ο χρήστης καταλαβαίνει ότι δεν τον ενδιαφέρει και αποφασίζει να ενεργοποιήσει το μικρόφωνο του και να κάνει χρήση της εντολής “Continue”. Με την συγκεκριμένη εντολή, ο φωνητικός βοηθός θα συνεχίσει την εκφώνηση προϊόντων τα οποία βρίσκονται μετά το προϊόν με αναγνωριστικό δύο. Έτσι, θα εκφωνηθεί ο τίτλος και το αναγνωριστικό των υπόλοιπων διαθέσιμων προϊόντων στην σελίδα. Επειδή ο χρήστης δεν ανοίγει το μικρόφωνο του μέχρι και το τελευταίο διαθέσιμο προϊόν, ο φωνητικός βοηθός με την ολοκλήρωση τους θα εκφωνήσει το εξής κείμενο:

“You have listen to all the available products. If you want to listen to them from the beginning use the command read all or you can select one of them with the command select ID. Additionally, you can use the command next or back in order to select and listen to the next product or the previous one respectively.”

Με την χρήση του παραπάνω κειμένου, ο χρήστης θα ενημερωθεί ότι άκουσε όλα τα διαθέσιμα προϊόντα και θα κληθεί να επιλέξει ανάμεσα σε τρεις επιλογές. Είτε να ξανά ακούσει από την αρχή τα προϊόντα, είτε να επιλέξει ένα από αυτά, είτε να επιλέξει κατευθείαν το επόμενο ή το προηγούμενο του προϊόντος με αναγνωριστικό δύο. Επειδή ο χρήστης ενδιαφέρεται για το προϊόν με αναγνωριστικό τρία, κάνει χρήση της εντολής “Next”, αφού το συγκεκριμένο προϊόν είναι το αμέσως επόμενο. Να σημειωθεί, ότι με την χρήση των εντολών “Next” και “Back” εκφωνείται κατευθείαν το όνομα, η τιμή και το αναγνωριστικό του εκάστοτε προϊόντος χωρίς να απαιτείται η χρήση της εντολής “Read product”. Σε περίπτωση που το προϊόν που ενδιέφερε τον χρήστη βρισκόταν στην τελευταία θέση, τότε θα μπορούσε να μεταβεί κατευθείαν σε αυτό με την χρήση της εντολής “Select id”. Στην συγκεκριμένη περίπτωση, η μετάβαση στο τελευταίο προϊόν της σελίδας μέσω της εντολής “Next” δεν θα έπαιρνε πολύ χρόνο, αφού θα έπρεπε να χρησιμοποιηθεί μόνο δύο φορές. Ωστόσο, σε ένα πλήρως λειτουργικό κατάστημα στο οποίο είναι πιθανό να περιέχονται σε μία σελίδα δεκάδες προϊόντα είναι προφανές ότι ο χρόνος θα αυξανόταν σημαντικά και συνεπώς μία εντολή όπως η Select id μπορεί να δώσει την λύση. Να σημειωθεί, πως σε περίπτωση που ο χρήστης κάνει χρήση της εντολής “Next” ενώ βρίσκεται στο τελευταίο προϊόν, τότε θα εκφωνηθεί από τον φωνητικό βοηθό το εξής κείμενο:

“That was the last product of the page. If you want to listen to them from the beginning use the command Read product or Read all. Additionally, you can use the command select id in order to select the product you want.”

Παρόμοια με την περίπτωση κατά την οποία ολοκληρώνονται τα προϊόντα που εκφωνεί ο φωνητικός βοηθός μετά την χρήση της εντολής “Read all” ή “Continue”, έτσι και με την εντολή “Next” αν αυτή χρησιμοποιηθεί ενώ ο χρήστης βρίσκεται στο τελευταίο προϊόν, τότε ο ίδιος θα ενημερωθεί σχετικά με τις επιλογές που έχει σε εκείνο το σημείο. Ωστόσο, στην περίπτωση της εντολής “Next” υπάρχει μια διαφορά εάν αυτή χρησιμοποιηθεί όταν ο χρήστης βρίσκεται στο τελευταίο προϊόν. Συγκεκριμένα, η τιμή του δείκτη που αφορά το προϊόν το οποίο είναι επιλεγμένο κάθε φορά, θα τροποποιηθεί έτσι ώστε να δείχνει το πρώτο προϊόν της σελίδας και συνεπώς εάν ο χρήστης κάνει χρήση της εντολής “Read product” θα ακούσει τα περιεχόμενα εκείνου, έχοντας στην ουσία την δυνατότητα να χρησιμοποιήσει ξανά την εντολή “Next” από την αρχή της σελίδας. Παρόμοια λογική ακολουθείται και στην περίπτωση της εντολής “Back” αν αυτή χρησιμοποιηθεί στο πρώτο προϊόν, με την διαφορά ότι δεν θα αλλάξει ο δείκτης και θα εκφωνηθεί το μήνυμα:

“That was the first product of the page. If you want to listen to the available products you can use the command next or read all. Additionally, you can use the command select id in order to select the product you want.”

Να σημειωθεί, ότι με την μετάβαση ενός χρήστη σε μία κατηγορία με προϊόντα, η τιμή του δείκτη που αναφέρθηκε παραπάνω θα δείχνει από προεπιλογή στο πρώτο προϊόν και συνεπώς αν ο χρήστης κάνει κατευθείαν χρήση της εντολής Read product τότε θα του διαβαστούν τα περιεχόμενα του πρώτου προϊόντος ακόμα και αν δεν το έχει επιλέξει ο ίδιος με την εντολή “Select id”. Σε συνέχεια του προηγούμενου σεναρίου, με την μετάβαση του μέσω της εντολής “Next” στο επόμενο προϊόν με αναγνωριστικό τρία και την χρήση της εντολής “Description”, ο χρήστης αποφασίζει ότι θέλει να το τοποθετήσει στο καλάθι αγορών του. Έτσι, ενεργοποιεί το μικρόφωνο του και εκφωνεί την εντολή “Add to cart”. Με την χρήση της συγκεκριμένης εντολής, ο φωνητικός βοηθός θα ζητήσει από τον χρήστη να ορίσει την ποσότητα που επιθυμεί να αγοράσει εκφωνώντας το εξής κείμενο:

“Tell me the number of quantity you want me to add in your cart”

Ο χρήστης ενεργοποιεί το μικρόφωνο του και εκφωνεί τον αριθμό ένα. Αυτό, θα έχει ως αποτέλεσμα να δημιουργηθεί μία νέα εγγραφή στον αντίστοιχο πίνακα της βάσης δεδομένων ο οποίος αφορά το καλάθι αγορών και στην συνέχεια εάν δεν υπάρξει κάποιο σφάλμα θα εκφωνηθεί από τον φωνητικό βοηθό ένα κείμενο επιβεβαίωσης και συγκεκριμένα:

“Product with name Propet Stability Walker Shoe and quantity one added successfully to your cart. If you want to visit the cart page use the keyword shopping cart. Additionally, you can use the command continue to listen to the rest of the products or the commands next or back to select or listen to the next product or the previous one respectively”

Όπως φαίνεται από το παραπάνω κείμενο, ο χρήστης θα ενημερωθεί για το προϊόν το οποίο τοποθετήθηκε στο καλάθι αγορών του, την ποσότητα που ο ίδιος όρισε ότι επιθυμεί να αγοράσει, καθώς επίσης και τον τρόπο με τον οποίο είναι εφικτή η μετάβαση του στην σελίδα που αφορά το καλάθι αγορών. Να σημειωθεί, ότι σε περίπτωση που δεν αναφερόταν στο παραπάνω κείμενο ο τρόπος μετάβασης στο καλάθι αγορών και ο χρήστης δεν τον θυμόταν, τότε θα μπορούσε να κάνει χρήση της εντολής “Options” η οποία όπως έχει προαναφερθεί αποτελεί το μέσο για την μετακίνηση του στις σελίδες του καταστήματος.

Για την επικοινωνία ενός χρήστη με οποιοδήποτε λογισμικό, οι γλώσσες προγραμματισμού παρέχουν συναρτήσεις οι οποίες κάνουν εφικτή την είσοδο δεδομένων τα οποία μπορεί να λάβει το πρόγραμμα εκτελώντας μία σειρά λειτουργιών βασισμένες σε αυτά. Ωστόσο, το γεγονός ότι ο χρήστης έχει την συγκεκριμένη δυνατότητα, δεν σημαίνει αυτόματα ότι τα δεδομένα που θα δώσει θα είναι συμβατά με αυτά που αναμένονται από το πρόγραμμα. Έτσι λοιπόν, αποτελεί ευθύνη του εκάστοτε προγραμματιστή να διαχειριστεί το σενάριο της εισαγωγής ενός μη συμβατού τύπου δεδομένων από τον χρήστη, έτσι ώστε να διατηρείται η ομαλή λειτουργία του λογισμικού ενημερώνοντας τον κατάλληλα για το λάθος του. Στην περίπτωση του menShoes, θα μπορούσε να δημιουργηθεί ένα τέτοιο πρόβλημα κατά την εκφώνηση

αριθμών όταν για παράδειγμα πρόκειται ο χρήστης να δώσει το επιθυμητό αναγνωριστικό προϊόντος ή να ορίσει τον επιθυμητό αριθμό ποσότητας που επιθυμεί να αγοράσει. Στην πρώτη περίπτωση, εάν δηλαδή ο χρήστης αιτηθεί ένα αναγνωριστικό το οποίο δεν αποτελεί αριθμό, τότε θα εκφωνηθεί από τον φωνητικό βοηθό το κείμενο “Please provide a valid id”. Για παράδειγμα, με την εκφώνηση της φράσης “I want the id 9”, ο χρήστης μπορεί να έχει δώσει σωστά έναν αριθμό, ωστόσο έχει συμπεριλάβει στην φράση του και το κείμενο “I want the id”, κάτι το οποίο θα έχει ως αποτέλεσμα το πρόγραμμα να μην μπορεί να αναγνωρίσει τον αριθμό εννιά μέσα στο κείμενο, ενημερώνοντας έτσι μέσω της συνθετικής ομιλίας τον χρήστη ότι πρέπει να ξανά εκφωνήσει ένα αναγνωριστικό. Αντίστοιχα, σε περίπτωση που δώσει σωστά έναν αριθμό αλλά δεν υπάρχει κάποιο αναγνωριστικό στην τρέχουσα σελίδα στο οποίο να αντιστοιχεί, τότε θα εκφωνηθεί το κείμενο “No product on this page with id [id the user said]”. Με παρόμοια λογική, σε περίπτωση που ο χρήστης ορίσει μία ποσότητα η οποία δεν αποτελεί αριθμό, τότε θα του εκφωνηθεί το κείμενο “Please provide an appropriate number”. Επιπλέον, στην περίπτωση της ποσότητας, κρίνεται απαραίτητο να προφυλαχτεί το πρόγραμμα και από το σενάριο στο οποίο ένας χρήστης εκφωνεί ως ποσότητα τον αριθμό μηδέν, κάτι το οποίο θα προκαλέσει προβλήματα αργότερα κατά την διαδικασία της αγοράς. Για τον λόγο αυτό, στην περίπτωση της εκφώνησης του αριθμού μηδέν θα χρησιμοποιηθεί το ίδιο κείμενο με πριν το οποίο ζητάει από τον χρήστη να ενεργοποιήσει ξανά το μικρόφωνο του εκφωνώντας έναν κατάλληλο αριθμό.

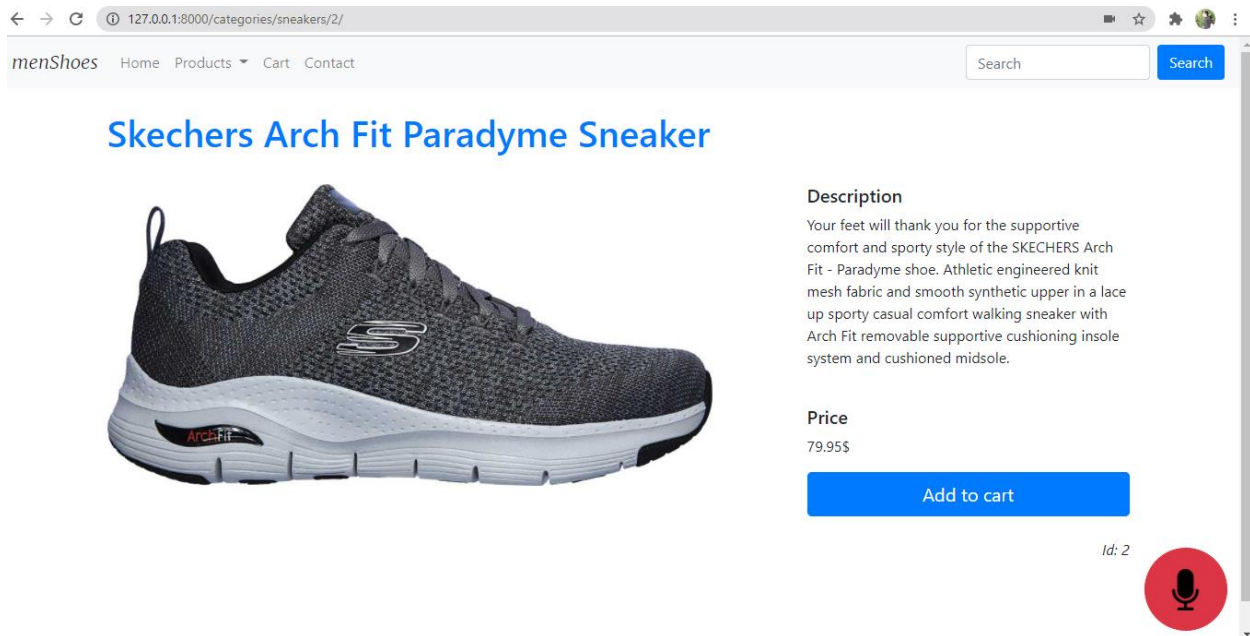
Μέχρι τώρα, αναλύθηκαν οι δέκα από τις έντεκα εντολές που μπορεί να χρησιμοποιήσει ένας χρήστης για την αλληλεπίδραση του με τα προϊόντα μιας κατηγορίας. Όπως προαναφέρθηκε, αποτελεί συχνή πρακτική ένα ηλεκτρονικό κατάστημα να διαθέτει διαφορετικές κατηγορίες όπου σε κάθε μία από αυτές περιλαμβάνεται ένας αριθμός προϊόντων. Κάθε ένα από αυτά, υπάρχει ως εγγραφή στην βάση δεδομένων με έναν αριθμό κελιών τα οποία αφορούν χαρακτηριστικά όπως είναι η τιμή, περιγραφή κλπ. Ωστόσο, όταν ένας χρήστης βρίσκεται σε μία κατηγορία προϊόντων, βλέπει για κάθε ένα από αυτά ορισμένες πληροφορίες του και όχι το σύνολο των διαθέσιμων πληροφοριών που αντιστοιχίζονται στο συγκεκριμένο προϊόν. Για να έρθει σε επαφή με το σύνολο των πληροφοριών, ο χρήστης επιλέγει από την κατηγορία το προϊόν που επιθυμεί και αυτόματα ανακατευθύνεται σε μία σελίδα η οποία περιλαμβάνει όλες τις πληροφορίες για αυτό. Με βάση την συγκεκριμένη φιλοσοφία, το menShoes παρέχει την εντολή “Open product”. Συγκεκριμένα, όταν ένας χρήστης έχει επιλέξει ένα προϊόν π.χ. μέσω της εντολής “Select id”, τότε με την ενεργοποίηση του μικροφώνου του και την εκφώνηση της εντολής “Open product”, θα ανακατευθυνθεί σε μία σελίδα περιεχόμενο της οποίας θα αποτελούν όλες οι πληροφορίες που σχετίζονται με το συγκεκριμένο προϊόν. Η δομή και οι διαθέσιμες εντολές της συγκεκριμένης σελίδας περιγράφονται στην υποενότητα που ακολουθεί.

Ο πυρήνας ενός ηλεκτρονικού καταστήματος είναι τα προϊόντα του. Αυτά είναι ο λόγος για τον οποίο καθημερινά καταναλωτές πλοηγούνται σε ηλεκτρονικά καταστήματα με στόχο την κάλυψη των εκάστοτε αναγκών τους. Συνεπώς, στην περίπτωση χρηστών που αντιμετωπίζουν προβλήματα όρασης, αποτελεί επιτακτική ανάγκη η δημιουργία λειτουργικότητας που να επιτρέπει την εύκολη αλλά ταυτόχρονα αποδοτική αλληλεπίδραση με αυτά. Ένας τέτοιος χρήστης, πρέπει να μπορεί με οργανωμένο τρόπο να ανακτήσει πληροφορίες για τα διαθέσιμα προϊόντα, αλλά στην συνέχεια να έχει και την ικανότητα για την πραγματοποίηση της αγοράς τους. Με την υλοποίηση των εντολών που αναλύθηκαν στην παρούσα υποενότητα, αποδείχτηκε ότι ένας χρήστης με προβλήματα όρασης μπορεί εξ ολοκλήρου μέσω της φωνής του να αλληλεπιδράσει με τα διαθέσιμα προϊόντα με ευέλικτο και αποδοτικό τρόπο, ο οποίος είναι ικανός να εξισώσει την καθιερωμένη μέθοδο πλοήγησης που παρέχει το γραφικό περιβάλλον, το οποίο πολλές φορές άτομα με μειωμένη ή μηδενική οπτική απόδοση είναι αδύνατο να αξιοποιήσουν.

3.2.3 Σελίδα προϊόντος

Όπως σε κάθε ηλεκτρονικό κατάστημα, έτσι και στο menShoes η σελίδα ενός προϊόντος περιλαμβάνει το σύνολο των δεδομένων που σχετίζονται άμεσα με αυτό. Χαρακτηριστικά παραδείγματα αποτελούν στοιχεία όπως το όνομα του προϊόντος, η τιμή, η περιγραφή, διαθέσιμες παραλλαγές και η λίστα συνεχίζεται. Βέβαια, συχνό φαινόμενο αποτελεί σε μία τέτοια σελίδα και η εύρεση δεδομένων τα οποία μπορούν να αφορούν αξιολογήσεις χρηστών, οι οποίες μπορούν να έχουν καθοριστικό ρόλο στο ταξίδι ενός καταναλωτή από την εύρεση ενός προϊόντος μέχρι και την αγορά του. Συνεπώς, λόγω της σημαντικότητας που έχουν για τον καταναλωτή τα δεδομένα που περιλαμβάνονται σε μία σελίδα προϊόντος αναφορικά με την λήψη αποφάσεων, γεννιέται η ανάγκη για μία φωνητική διεπαφή η οποία να δίνει την δυνατότητα σε έναν χρήστη με προβλήματα όρασης να αλληλεπιδράσει με το περιεχόμενο μιας τέτοιας σελίδας.

Στο συγκεκριμένο proof of concept, μία σελίδα προϊόντος περιλαμβάνει την εικόνα του εκάστοτε προϊόντος, την τιμή, την περιγραφή, το αναγνωριστικό, καθώς επίσης και ένα κουμπί το οποίο δίνει την δυνατότητα της προσθήκης αυτού στο καλάθι αγορών. Παρακάτω, φαίνεται η σελίδα ενός προϊόντος στο menShoes.



Εικόνα 22: Σελίδα προϊόντος

Η φιλοσοφία που ακολουθήθηκε αναφορικά με τον τρόπο αλληλεπίδρασης μέσω της ομιλίας σε μία σελίδα προϊόντος, αφορούσε τον διαχωρισμό του είδους των δεδομένων που περιλαμβάνονται σε αυτή και στην συνέχεια την δημιουργία εντολών με στόχο την ανάκτηση του συνόλου των δεδομένων με οργανωμένο τρόπο. Συγκεκριμένα, με την είσοδο ενός χρήστη σε μία σελίδα προϊόντος το πρώτο πράγμα που θα συμβεί είναι η εκφώνηση από τον φωνητικό βοηθό του εξής κειμένου:

“Welcome to the page for the product [name of the product]. If you don’t know what to do, use the command instructions”

Αφού ένας χρήστης επιβεβαιωθεί με το παραπάνω κείμενο ότι έχει μεταβεί στην σελίδα του προϊόντος που τον ενδιαφέρει, μπορεί στην συνέχεια να ανοίξει το μικρόφωνο του και να εκφωνήσει την εντολή “Instructions”, με στόχο να μάθει για το περιεχόμενο που περιλαμβάνει η τρέχουσα σελίδα αλλά και το πως μπορεί να αλληλεπιδράσει με αυτό. Με την χρήση λοιπόν της εντολής “Instructions”, θα εκφωνηθεί από τον φωνητικό βοηθό το εξής κείμενο:

“Instructions: This page contains all the information for the product [name of the product]. Simply use one of the available commands such as product name, product description, add to cart, etc. If you want to listen to the available commands for this page use the keyword commands. Lastly, if you want to change page use the command options.”

Όπως γίνεται κατανοητό και από το παραπάνω κείμενο, σε μία σελίδα προϊόντος ο χρήστης δεν καλείται να επιλέξει κάποιο προϊόν και στην συνέχεια να χρησιμοποιήσει εντολές για την αλληλεπίδραση με αυτό.

Αντίθετα, μπορεί να χρησιμοποιήσει κατευθείαν εντολές για την ανάκτηση πληροφοριών σχετικά με το συγκεκριμένο προϊόν, έτσι ώστε να σχηματίσει όσο το δυνατόν καλύτερη εικόνα γίνεται για το αν αυτό πληροί τις προδιαγραφές που επιθυμεί ή όχι.

Με την χρήση της εντολής “Commands”, θα εκφωνηθούν από τον φωνητικό βοηθό οι εξής επτά εντολές:

1. Product name
2. Product price
3. Product description
4. Product id
5. Add to cart
6. Commands
7. Instructions

Αναφορικά με τις πρώτες τέσσερις εντολές, αυτονόητο είναι ότι θα επιστρέψουν το αντίστοιχο κείμενο που φανερώνεται με το όνομα τους. Για παράδειγμα, ένας χρήστης με την ενεργοποίηση του μικροφώνου του και την εκφώνηση της εντολής “Product price”, θα λάβει ως απάντηση από τον φωνητικό βοηθό την τιμή του προϊόντος. Μία παρατήρηση η οποία αξίζει να σημειωθεί, αφορά την δυνατότητα που δίνεται στον χρήστη να προσθέσει ένα προϊόν στο καλάθι αγορών με δύο τρόπους. Ο πρώτος, αφορά την χρήση της εντολής “Add to cart” όταν βρίσκεται σε μία σελίδα η οποία περιλαμβάνει τα προϊόντα μιας κατηγορίας (αναφέρθηκε στην προηγούμενη υποενότητα), ενώ ο δεύτερος αφορά την χρήση της συγκεκριμένης εντολής από την σελίδα ενός προϊόντος. Σε κάθε ηλεκτρονικό κατάστημα σήμερα, ενώ ο δεύτερος τρόπος θεωρείται δεδομένος, κάτι τέτοιο δεν ισχύει και για την περίπτωση του πρώτου. Ωστόσο, στην περίπτωση μίας φωνητικής διεπαφής η οποία σχεδιάζεται με στόχο την υποβοήθηση ενός χρήστη με προβλήματα όρασης, η παροχή και του πρώτου τρόπου αποτελεί σημαντική προϋπόθεση. Αυτό, αφού σε πολλές περιπτώσεις ο χρήστης μπορεί να αποφασίσει να τοποθετήσει στο καλάθι του ένα προϊόν με μοναδικό κριτήριο το όνομα και την τιμή του. Συνεπώς, με την έλλειψη της εντολής “Add to cart” στην σελίδα με τα προϊόντα της εκάστοτε κατηγορίας, ο χρήστης θα αναγκαστεί να ανακατευθυνθεί στην σελίδα του συγκεκριμένου προϊόντος για την τοποθέτηση του στο καλάθι, κάτι το οποίο θα μπορούσε να είχε αποφευχθεί προσφέροντας μία αποδοτικότερη ικανότητα πλοήγησης.

Με βάση τα παραπάνω, γίνεται κατανοητό πως σε μία σελίδα προϊόντος περιλαμβάνεται οποιαδήποτε πληροφορία μπορεί να σχετίζεται με αυτό, είτε αυτή αφορά μία από τις ιδιότητες του, είτε αφορά για παράδειγμα αξιολογήσεις που έχουν κάνει οι καταναλωτές για αυτό. Συνεπώς, προκύπτει η ανάγκη για την οργανωμένη ανάκτηση του περιεχομένου που περιλαμβάνεται σε μία τέτοια σελίδα, κάτι

το οποίο είναι εφικτό με την φιλοσοφία που ακολουθούν οι εντολές που αναλύθηκαν στην παρούσα υποενότητα. Με την χρήση τους, ένας χρήστης με προβλήματα όρασης έχει την δυνατότητα να ανακτήσει οποιαδήποτε πληροφορία είναι διαθέσιμη για το προϊόν εξ ολοκλήρου μέσω της φωνής του, χωρίς να απαιτείται η χρήση του γραφικού περιβάλλοντος.

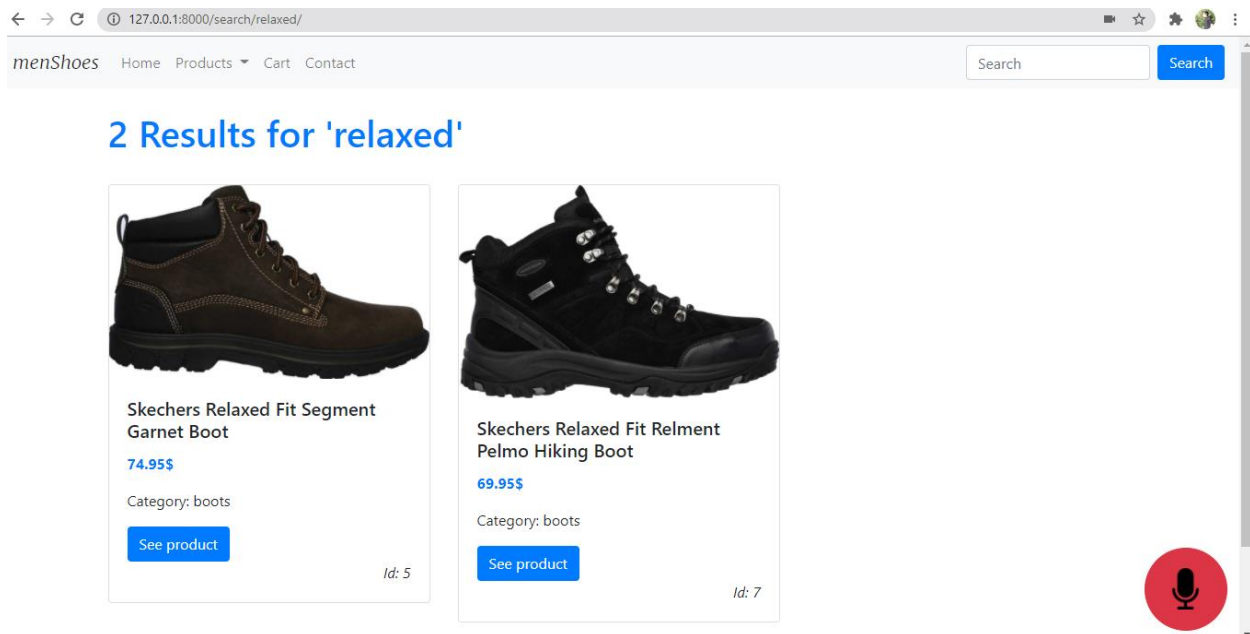
3.2.4 Αναζήτηση προϊόντων

Ένα από τα σημαντικότερα χαρακτηριστικά σε ένα ηλεκτρονικό κατάστημα, αποτελεί η δυνατότητα της αναζήτησης προϊόντων με την χρήση λέξεων κλειδιών. Για παράδειγμα, ένας χρήστης ο οποίος γνωρίζει εξ αρχής το προϊόν που επιθυμεί να αγοράσει, δεν αναγκάζεται να ψάξει μέσα σε μία πληθώρα κατηγοριών και προϊόντων για αυτό. Αντίθετα, το μόνο που χρειάζεται να κάνει είναι η συμπλήρωση λέξεων κλειδιών σε μία μπάρα αναζήτησης, οι οποίες περιγράφουν κάποιο ή κάποια χαρακτηριστικά του προϊόντος. Από εκείνο το σημείο, αναλαμβάνει το ηλεκτρονικό κατάστημα με βάση τα δεδομένα εισόδου να κάνει μια αναζήτηση στην βάση δεδομένων, επιστρέφοντας οποιοδήποτε προϊόν ταιριάζει με κάποια από τις λέξεις κλειδιά που έδωσε ο χρήστης. Καταλαβαίνει κανείς, πως η ανάγκη για την συγκεκριμένη λειτουργικότητα είναι ακόμα μεγαλύτερη όταν πρόκειται για έναν χρήστη που αντιμετωπίζει προβλήματα όρασης. Χωρίς την δυνατότητα της αναζήτησης, ένα τέτοιο άτομο δεν μπορεί να βγάλει εύκολα συμπέρασμα για το αν το κατάστημα περιλαμβάνει το προϊόν που επιθυμεί. Για να γίνει αυτό, θα πρέπει να πλοηγηθεί στις πιθανές κατηγορίες που μπορεί να έχουν το προϊόν και στην συνέχεια σε κάθε μία από αυτές να διασχίσει το πλήθος των προϊόντων της. Ενώ για ένα χρήστη που δεν αντιμετωπίζει προβλήματα όρασης η συγκεκριμένη διαδικασία μπορεί να ολοκληρωθεί αρκετά γρήγορα, για ένα χρήστη με μειωμένη ή μηδενική οπτική απόδοση η διαδικασία αυτή μετατρέπεται σε χρονοβόρα και μη αποδοτική.

Όπως περιγράφηκε στην υποενότητα που αφορούσε την σελίδα Home, μέσω της εντολής “Options”, ο φωνητικός βοηθός θα εκφωνήσει το πρώτο επίπεδο του μενού πλοήγησης. Η τελευταία επιλογή του συγκεκριμένου επιπέδου, εκφωνείται ως “Search for” και ένας χρήστης μπορεί να την χρησιμοποιήσει για την αναζήτηση προϊόντων. Σε αντίθεση με μία εντολή όπως η “Add to cart” την οποία ο χρήστης μπορεί να χρησιμοποιήσει μόνο σε συγκεκριμένες σελίδες, η χρήση της “Search for” είναι εφικτή από παντού. Αυτό, αφού η ίδια περιλαμβάνεται ως μία από τις πιθανές επιλογές του μενού πλοήγησης. Με την χρήση της συγκεκριμένης εντολής, ο φωνητικός βοηθός θα εκφωνήσει το εξής κείμενο:

“Tell the phrase you want to search”

Έστω ότι ένας χρήστης επιθυμεί να κάνει μία αναζήτηση για την λέξη κλειδί “relaxed”. Με το άνοιγμα του μικρόφωνου του και εκφωνώντας την συγκεκριμένη λέξη, ο χρήστης θα μεταβεί στην σελίδα που απεικονίζεται παρακάτω.



Εικόνα 23: Σελίδα για την αλληλεπίδραση με τα αποτελέσματα μιας αναζήτησης

Το πρώτο πράγμα που θα συμβεί στην συγκεκριμένη σελίδα είναι η εκφώνηση από τον φωνητικό βοηθό του κειμένου που φαίνεται στην αρχή της εικόνας και συγκεκριμένα:

“2 Results for relaxed”

Με αυτό τον τρόπο, ο χρήστης ενημερώνεται κατευθείαν για τον αριθμό των αποτελεσμάτων που επιστράφηκαν, έτσι ώστε σε περίπτωση που ο αριθμός αυτός δεν είναι ικανοποιητικός να δοκιμάσει την αναζήτηση με διαφορετική λέξη – φράση. Αντίστοιχα, στην περίπτωση που ο χρήστης θα δώσει μια λέξη η οποία δεν αντιστοιχίστηκε με κανένα προϊόν, τότε ο φωνητικός βοηθός θα εκφωνήσει ότι βρέθηκαν μηδέν αποτελέσματα για αυτή την λέξη, ενώ στην συνέχεια θα του αναφέρει πως μπορεί να χρησιμοποιήσει την εντολή “Options” για να αλλάξει σελίδα ή να ξανά κάνει μια αναζήτηση μέσω της εντολής “Search for”. Παρακάτω, φαίνεται η σελίδα των αποτελεσμάτων αναζήτησης όταν ο χρήστης δώσει ως είσοδο την λέξη “Nike”.

0 Results for 'Nike'

Search for another keyword



Εικόνα 24: Σελίδα αποτελεσμάτων αναζήτησης όταν η λέξη που θα δοθεί ως είσοδος δεν αντιστοιχεί σε κάποιο προϊόν

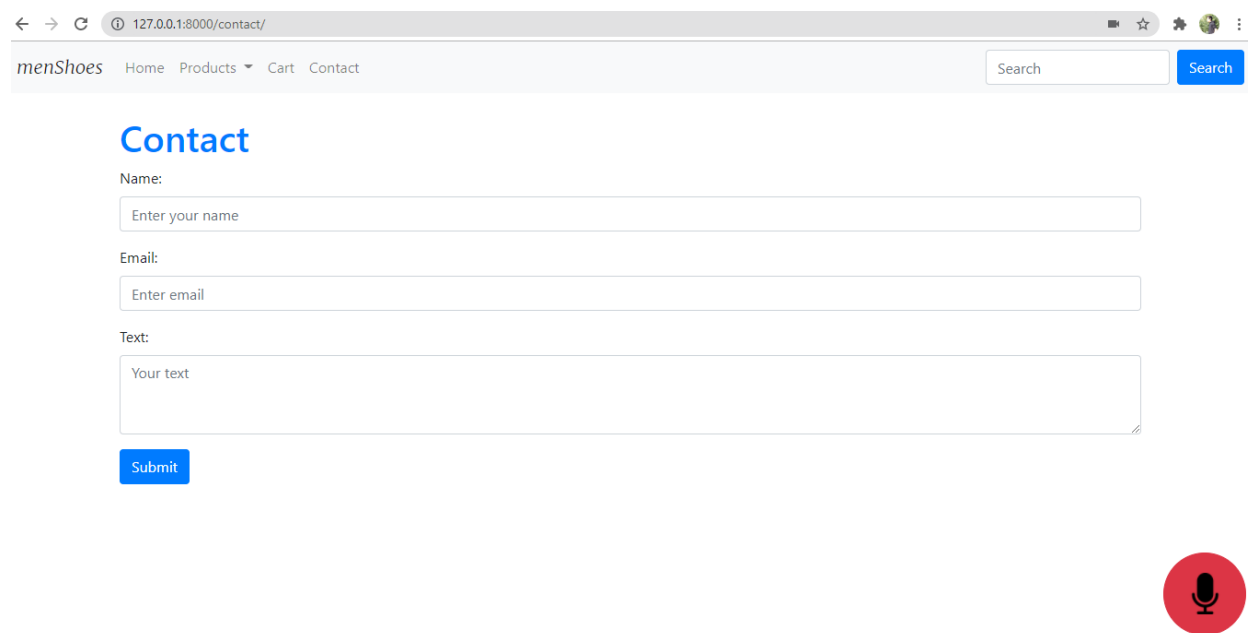
Στην εικόνα 23, η σελίδα που περιλαμβάνει τα προϊόντα που επιστράφηκαν για την λέξη “relaxed”, παρατηρείται ότι είναι σχεδόν ίδια με αυτή που περιλαμβάνει τα προϊόντα μιας κατηγορίας π.χ. sneakers. Η μόνη διαφορά στις λειτουργίες των δύο αυτών σελίδων αφορά την ύπαρξη της κατηγορίας (Category) στην πρώτη. Με αυτό τον τρόπο, ένας χρήστης ο οποίος έχει πραγματοποιήσει μία αναζήτηση και έχει λάβει ορισμένα αποτελέσματα, μπορεί να ακούσει σε ποια κατηγορία ανήκει καθένα από αυτά έτσι ώστε στην συνέχεια αν θέλει να μεταβεί σε αυτή για να βρει περισσότερα. Κατά τα άλλα, οι εντολές για την αλληλεπίδραση με τα προϊόντα είναι ακριβώς οι ίδιες με αυτές που περιγράφηκαν στην σελίδα με τα προϊόντα της κατηγορίας sneakers.

Η δυνατότητα της αναζήτησης, επιτρέπει σε οποιοδήποτε χρήστη να ανακτήσει σε δευτερόλεπτα το κομμάτι της πληροφορίας που τον ενδιαφέρει μέσα από ένα πλήθος δεδομένων. Συνεπώς, η συγκεκριμένη ικανότητα είναι απαραίτητο να μπορεί να αξιοποιηθεί από ένα άτομο με προβλήματα όρασης. Αυτό, λόγω του ότι με αποδοτικό τρόπο θα έχει την ικανότητα να έρθει σε επαφή με αυτό που επιθυμεί ή θα γνωρίζει αν αυτό είναι διαθέσιμο, χωρίς να χρειάζεται να έρθει αντιμέτωπος με ένα πλήθος δεδομένων τα οποία θα μετατρέψουν την διαδικασία της αναζήτησης σε χρονοβόρα και μη αποτελεσματική. Αποτελεί μεγάλη ανάγκη τα άτομα με προβλήματα όρασης να διευκολύνονται όσο το δυνατό περισσότερο γίνεται με απλές, αλλά ταυτόχρονα αποδοτικές πρακτικές, κάτι το οποίο είναι εφικτό με μία εντολή όπως η Search for.

3.2.5 Φόρμα επικοινωνίας

Σε πολλές περιπτώσεις, ένας καταναλωτής πριν προβεί σε οποιαδήποτε αγορά επιθυμεί να λάβει περαιτέρω πληροφορίες τις οποίες δεν μπορεί να βρει σε μία από τις διαθέσιμες σελίδες του ηλεκτρονικού καταστήματος. Επιπλέον, πιθανή είναι και η δημιουργία επιλοκών με αποτέλεσμα ο χρήστης να παραλάβει ένα προϊόν το οποίο δεν ήταν το αναμενόμενο ή ακόμα και να αργήσει να παραλάβει το προϊόν του. Λόγω των παραπάνω σεναρίων, τα ηλεκτρονικά καταστήματα παρέχουν στον χρήστη την δυνατότητα επικοινωνίας έτσι ώστε αυτός να μην νιώθει ότι πραγματοποιεί αγορές με μια απρόσωπη επιχείρηση. Επομένως, κρίνεται απαραίτητη η υλοποίηση ενός τρόπου με τον οποίο ένας χρήστης με προβλήματα όρασης θα μπορεί να αλληλεπιδράσει με μία φόρμα επικοινωνίας εξ ολοκλήρου μέσω της φωνής του. Βέβαια, με την δημιουργία της συγκεκριμένης λειτουργικότητας είναι προφανές ότι αυτή δεν θα εξυπηρετεί μόνο την ανάγκη για την συμπλήρωση μίας φόρμας επικοινωνίας, αφού θα μπορεί να έχει εφαρμογές σε οποιοδήποτε σημείο περιέχονται πεδία εισαγωγής (input fields), όπως για παράδειγμα την σελίδα συμπλήρωσης στοιχείων για την αγορά ενός προϊόντος.

Για την μετάβαση του στην σελίδα με την φόρμα επικοινωνίας, ένας χρήστης μπορεί να εκφωνήσει την λέξη Contact. Βέβαια, για να μάθει για την συγκεκριμένη επιλογή, αυτονόητο είναι πως πρώτα θα πρέπει να χρησιμοποιήσει την εντολή “Options”. Παρακάτω, απεικονίζεται η σελίδα με την φόρμα επικοινωνίας του καταστήματος menShoes.



The screenshot shows a web browser window with the URL 127.0.0.1:8000/contact/. The website is 'menShoes' and has a navigation bar with links: Home, Products, Cart, and Contact. A search bar is also present. The main content area is titled 'Contact' and contains a form with three input fields: 'Name' (with placeholder 'Enter your name'), 'Email' (with placeholder 'Enter email'), and 'Text' (with placeholder 'Your text'). A blue 'Submit' button is located below the text field. In the bottom right corner of the browser window, there is a red circular icon with a white microphone symbol inside.

Εικόνα 25: Σελίδα για την επικοινωνία μέσω φόρμας

Με την φόρτωση της συγκεκριμένης σελίδας, ο φωνητικός βοηθός θα εκφωνήσει το εξής κείμενο:

“Welcome to the contact page. If you don’t know what to do, use the command instructions”

Παρόμοια με κάθε σελίδα, σε περίπτωση που ο χρήστης δεν ξέρει πως να αλληλεπιδράσει με το περιεχόμενο μπορεί να κάνει χρήση της εντολής “Instructions”. Με την χρήση της συγκεκριμένης εντολής θα εκφωνηθεί το εξής μήνυμα:

“Instructions: This page contains a contact form which you can fill out in order to ask the store owner anything you want. The first thing you have to do is use the command fill out form. With this command, you will start providing the required information such as name, email etc. Every time you fill out a field of the form you have to open your microphone and use the command next. Once the form is filled out you can use one of the available commands such as read form, modify form etc. If you want to listen to the available commands for this page simply use the keyword commands. Lastly, if you want to change page use the command options.”

Όπως φαίνεται και από το παραπάνω κείμενο, η φιλοσοφία που ακολουθήθηκε αφορά αρχικά την συμπλήρωση των πεδίων της φόρμας και έπειτα την αξιοποίηση εντολών για την αλληλεπίδραση με αυτά. Για την κατανόηση των λειτουργιών της συγκεκριμένης σελίδας, περιγράφεται ένα σενάριο κατά το οποίο ο χρήστης αρχικά συμπληρώνει την φόρμα, στην συνέχεια εκφωνούνται από τον φωνητικό βοηθό τα πεδία που συμπληρώθηκαν, έπειτα τροποποιεί ένα από αυτά και στην συνέχεια υποβάλει την φόρμα.

Με βάση τις οδηγίες που εκφωνήθηκαν από τον φωνητικό βοηθό, ο χρήστης ενεργοποιεί το μικρόφωνο του και χρησιμοποιεί την εντολή “Fill out form”. Με την χρήση της συγκεκριμένης εντολής ενεργοποιείται ο φωνητικός βοηθός ζητώντας από τον χρήστη να δώσει το όνομα του. Συγκεκριμένα, εκφωνείται η φράση “Tell me your name”. Ο χρήστης, ενεργοποιεί το μικρόφωνο του και αναφέρει το όνομα John. Έπειτα, όπως αναφέρθηκε από τις οδηγίες μέσω της εντολής “Instructions”, αφού ο χρήστης έδωσε το όνομα του, ανοίγει ξανά το μικρόφωνο του και εκφωνεί την εντολή “Next”, έτσι ώστε ο φωνητικός βοηθός να ζητήσει τα στοιχεία που αφορούν το επόμενο πεδίο. Συγκεκριμένα, ο φωνητικός βοηθός αυτή την φορά εκφωνεί την φράση “Tell me your email” και συνεπώς ο χρήστης παρέχει την διεύθυνση ηλεκτρονικού ταχυδρομείου john@gmail.com. Η ίδια διαδικασία επαναλαμβάνεται μέχρι και την συμπλήρωση του πεδίου Text. Όταν ο χρήστης έχει εκφωνήσει το κείμενο που επιθυμεί και κάνει χρήση της εντολής “Next”, τότε επειδή το πεδίο Text αποτελεί το τελευταίο κομμάτι της φόρμας, ο φωνητικός βοηθός θα εκφωνήσει το εξής μήνυμα:

“You have filled out all the form fields. Use one of the available commands such as submit form, modify form etc. If you want to listen to the available commands use the keyword commands”

Αφού έχουν συμπληρωθεί όλα τα πεδία της φόρμας, ο χρήστης επιθυμεί να ακούσει τις διαθέσιμες εντολές που μπορεί να αξιοποιήσει και συνεπώς χρησιμοποιεί την εντολή “Commands”. Αμέσως, ο φωνητικός βοηθός εκφωνεί τις εξής έξι εντολές:

1. Fill out form
2. Submit form
3. Read
4. Modify form
5. Commands
6. Instructions

Όπως είναι φυσικό, το πρώτο πράγμα που επιθυμεί να κάνει ο χρήστης είναι να ακούσει τα στοιχεία που έχει δώσει σε κάθε πεδίο. Συνεπώς, η εντολή που χρησιμοποιεί είναι η “Read”. Με την χρήση της συγκεκριμένης εντολής, ο φωνητικός βοηθός θα διαβάσει για κάθε πεδίο το όνομα του συνοδευόμενο από την τιμή που έχει δώσει ο χρήστης. Για παράδειγμα, για τα πρώτα δύο πεδία θα εκφωνηθεί το κείμενο “Name: John, email: john@gmail.com” και με την ίδια λογική θα εκφωνηθούν όσα πεδία και αν είχε συμπληρώσει ο χρήστης. Μετά την εκφώνηση των στοιχείων της φόρμας, ο χρήστης θυμάται ότι η διεύθυνση του ηλεκτρονικού ταχυδρομείου που έχει δώσει είναι λάθος. Για τον λόγο αυτό, αξιοποιεί την εντολή “Modify form”. Με την χρήση της συγκεκριμένης εντολής, ο φωνητικός βοηθός θα εκφωνήσει το εξής κείμενο:

“Tell me the name of the field you want to change.”

Ο χρήστης, καλείται να εκφωνήσει το όνομα του πεδίου στο οποίο επιθυμεί να αλλάξει την τιμή και συνεπώς ενεργοποιεί το μικρόφωνο του και δίνει ως είσοδο την λέξη Email. Αμέσως, θα γίνει έλεγχος αν το πεδίο αυτό υπάρχει στην φόρμα επικοινωνίας και εφόσον βρεθεί, θα ζητηθεί από τον χρήστη η νέα διεύθυνση ηλεκτρονικού ταχυδρομείου με την χρήση του κειμένου “Tell me the new email”. Σε περίπτωση που το πεδίο που έδωσε ο χρήστης δεν βρεθεί στην φόρμα, τότε θα εκφωνηθεί το εξής κείμενο:

“No field with name [keyword the user said]. Please try again”

Αφού ζητήθηκε από τον χρήστη να δώσει την νέα διεύθυνση ηλεκτρονικού ταχυδρομείου, σημαίνει ότι έδωσε σωστά το όνομα του πεδίου και συνεπώς ενεργοποιεί το μικρόφωνο του και εκφωνεί την νέα διεύθυνση. Με την απενεργοποίηση του μικροφώνου του, ο φωνητικός βοηθός εκφωνεί ένα μήνυμα επιτυχίας και συγκεκριμένα:

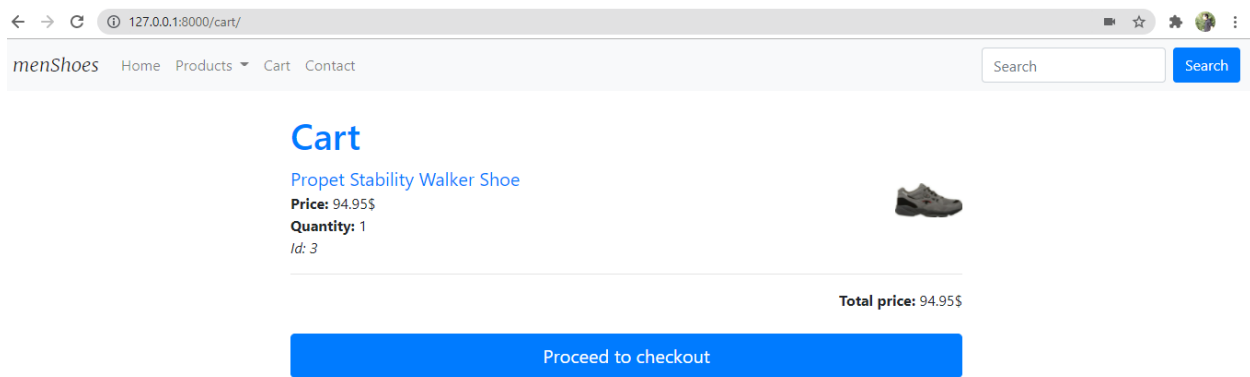
“The value of the field changed to [new email address]”

Πλέον, ο χρήστης έχει συμπληρώσει την φόρμα επικοινωνίας με τα επιθυμητά στοιχεία και συνεπώς το μόνο που απομένει είναι η υποβολή της. Για την συγκεκριμένη λειτουργία είναι προφανές ότι θα χρησιμοποιηθεί η εντολή “Submit”. Να σημειωθεί, ότι σε περίπτωση που ένας χρήστης έχει συμπληρώσει την φόρμα κάνει χρήση της εντολής “Fill out form”, τότε αυτό θα έχει ως αποτέλεσμα την διαγραφή των στοιχείων και την επανάληψη της διαδικασίας από την αρχή.

Με την χρήση των παραπάνω εντολών, ένας χρήστης με προβλήματα όρασης αποκτά την δυνατότητα επικοινωνίας με το εκάστοτε κατάστημα εξ ολοκλήρου μέσω της φωνής του, κάτι το οποίο συμβάλει στην παροχή μίας περισσότερο προσωποποιημένης εμπειρίας πλοήγησης. Η δυνατότητα της επικοινωνίας αποτελεί σημαντικό χαρακτηριστικό στα καταστήματα και συνεπώς ένας χρήστης που αντιμετωπίζει προβλήματα με την όραση του δεν πρέπει να αποκλείεται από το δικαίωμα αυτό. Παράλληλα, το συγκεκριμένο μοντέλο μπορεί να εφαρμοστεί σε οποιοδήποτε σημείο χρησιμοποιούνται πεδία εισαγωγής, επιτρέποντας στον χρήστη να εισάγει δεδομένα με οργανωμένο και απλό τρόπο.

3.2.6 Καλάθι αγορών

Ένα από τα βασικότερα συστατικά ενός ηλεκτρονικού καταστήματος αποτελεί το καλάθι αγορών, το οποίο αποτελεί την προτελευταία στάση ενός καταναλωτή στο ψηφιακό ταξίδι του για την αγορά ενός ή περισσότερων προϊόντων. Μέσω της συγκεκριμένης σελίδας, ένας χρήστης μπορεί να έρθει σε επαφή με τα προϊόντα που έχει προσθέσει στο καλάθι κατά την περιήγηση του στο ηλεκτρονικό κατάστημα, έχοντας δυνατότητες όπως η προβολή του συνολικού ποσού, η τροποποίηση των ποσοτήτων, καθώς επίσης και η αφαίρεση προϊόντων τα οποία τελικά δεν είναι επιθυμητό να αποκτηθούν. Συνεπώς, καταλαβαίνει κανείς την αναγκαιότητα για την δημιουργία μίας φωνητικής διεπαφής μέσω της οποίας ένας χρήστης με προβλήματα όρασης θα μπορεί να διαχειριστεί το περιεχόμενο του καλαθιού με απλό, αλλά ταυτόχρονα αποδοτικό τρόπο. Παρακάτω, απεικονίζεται η σελίδα η οποία αφορά το καλάθι αγορών του menShoes. Να σημειωθεί, ότι μέσα σε αυτό περιλαμβάνεται το προϊόν που τοποθέτησε ο χρήστης με την εντολή “Add to cart” ως συνέχεια του σεναρίου που περιγράφηκε στην υποενότητα με τα προϊόντα της κατηγορίας sneakers.



Εικόνα 26: Καλάθι αγορών

Για την μετάβαση του στο καλάθι αγορών, ένας χρήστης πρέπει να χρησιμοποιήσει την λέξη Shopping cart η οποία δίνεται σαν επιλογή όταν εκφωνούνται τα περιεχόμενα του μενού ως αποτέλεσμα της εντολής “Options”. Το πρώτο πράγμα που θα ακούσει ένας χρήστης με την μετάβαση του στο καλάθι αγορών εφόσον αυτό περιέχει τουλάχιστον ένα προϊόν είναι το εξής κείμενο:

“Welcome to your shopping cart. Your cart has [number of products in cart] products. If you don’t know what to do, use the command instructions.”

Να σημειωθεί, ότι σε περίπτωση που το καλάθι δεν περιείχε κανένα προϊόν, ο φωνητικός βοηθός θα εκφωνούσε διαφορετικό κείμενο και συγκεκριμένα:

“Welcome to your shopping cart. Your cart is currently empty. Use the command options to start shopping now.”

Εξετάζοντας την περίπτωση που απεικονίζεται στην εικόνα 26, ο χρήστης κάνει χρήση της εντολής “Instructions”, με στόχο την ενημέρωση του για το είδος του περιεχομένου και του τρόπου με τον οποίο μπορεί να αλληλεπιδράσει με αυτό. Αμέσως, ο φωνητικός βοηθός εκφωνεί το αντίστοιχο κείμενο και συγκεκριμένα:

“Instructions: This page contains all the products you have added to your cart. From your cart, you can listen to the total price, change the quantity of products or delete them. The first thing you have to do if you don’t remember what products you have added to your cart is to use the command read all. With this command, you will listen to all the products in cart one by one. When you listen to the product you want,

use the command select ID. With this command, you will select this product by providing its ID. Once the product is selected you can use one of the available commands such as read product, new quantity, delete product, etc. If you want to listen to the available commands for this page simply use the keyword commands. When you finish with this product you can use the command continue to listen to the rest of the products in cart. If you want to listen to the total price you don't have to select a specific product, the only thing you have to do is use the command total price. Lastly, if you want to change page use the command options.”

Από το παραπάνω κείμενο, γίνεται κατανοητό πως ο τρόπος διάσχισης των προϊόντων που περιέχονται μέσα στο καλάθι αγορών είναι ίδιος με αυτόν που χρησιμοποιείται στα προϊόντα μιας κατηγορίας. Κάτι τέτοιο, προσφέρει στον χρήστη έναν ομοιόμορφο τρόπο αλληλεπίδρασης, ο οποίος όταν γίνει κατανοητός μπορεί να εφαρμοστεί οπουδήποτε υπάρχουν ένα ή περισσότερα προϊόντα, χωρίς να αναγκάζεται ο ίδιος να προσαρμόζεται σε νέες μεθόδους κάθε φορά ανάλογα με την σελίδα την οποία επισκέπτεται. Αφού ο χρήστης κατανόησε μέσω της εντολής “Instructions” την σκοπιμότητα της συγκεκριμένης σελίδας, επιλέγει την εντολή “Commands” έτσι ώστε να του εκφωνηθούν οι διαθέσιμες εντολές με τις οποίες μπορεί να αλληλεπιδράσει με τα προϊόντα του καλαθιού. Συγκεκριμένα, οι εντολές που εκφωνούνται είναι οι εξής δώδεκα:

1. Read product
2. Next
3. Back
4. Total price
5. New quantity
6. Delete product
7. Select id
8. Read all
9. Continue
10. Open product
11. Commands
12. Instructions

Όπως φαίνεται από την παραπάνω λίστα, οι μόνες εντολές που δεν είναι διαθέσιμες στην σελίδα με τα προϊόντα μιας κατηγορίας είναι η 4, 5, 6. Αυτό, αφού όπως προαναφέρθηκε οι υπόλοιπες χρησιμοποιούνται με στόχο την επιλογή ενός από τα προϊόντα και λειτουργούν ακριβώς με τον ίδιο τρόπο που περιγράφηκε στην ενότητα με τα προϊόντα της κατηγορίας sneakers. Λόγω του ότι ο χρήστης δεν θυμάται την τιμή του συγκεκριμένου προϊόντος, θα το επιλέξει και στην συνέχεια θα χρησιμοποιήσει την εντολή “Read product”

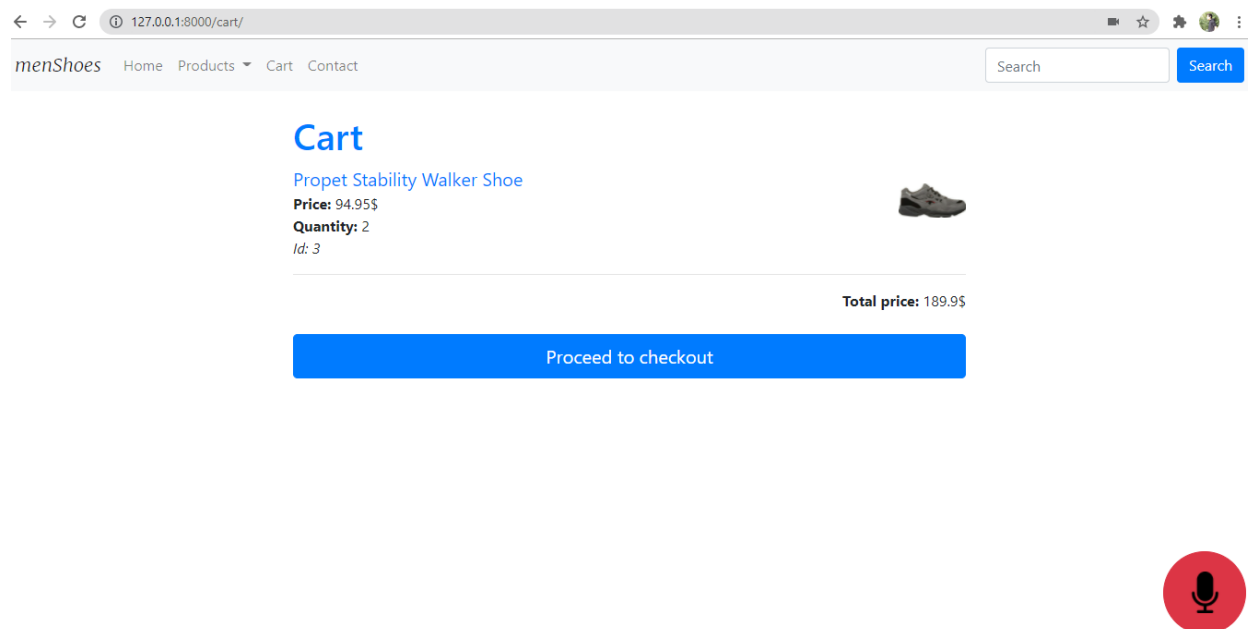
μέσω της οποίας θα εκφωνηθεί το όνομα, η τιμή, η ποσότητα, καθώς επίσης και το αναγνωριστικό του. Ο χρήστης, αποφασίζει ότι θέλει να αυξήσει την ποσότητα του προϊόντος και συνεπώς κάνει χρήση της εντολής “New quantity”. Αμέσως, ο φωνητικό βοηθός θα εκφωνήσει το εξής μήνυμα:

“Tell me the number of new quantity”

Επομένως, ο χρήστης ενεργοποιεί το μικρόφωνο του και εκφωνεί τον αριθμό δύο. Με την εισαγωγή της επιθυμητής ποσότητας, ο φωνητικός βοηθός θα επιβεβαιώσει ότι η αλλαγή πραγματοποιήθηκε με επιτυχία μέσω του παρακάτω κειμένου:

“Quantity for product [name of the selected product] has changed to: [number the user said]. Total price is now [new total price]”

Να σημειωθεί, πως σε περίπτωση που ο χρήστης δεν δώσει έναν σκέτο αριθμό ή ο αριθμός είναι το μηδέν, τότε θα εκφωνηθεί αντίστοιχο μήνυμα το οποίο θα τον ενημερώνει ότι πρέπει να δώσει ένα έγκυρο νούμερο. Αυτονόητο είναι πως με την αλλαγή της ποσότητας θα αλλάξει και η συνολική τιμή του καλαθιού. Συνεπώς, ο χρήστης όποτε επιθυμεί μπορεί να κάνει χρήση της εντολής “Total price” η οποία θα δώσει ως αποτέλεσμα την εκφώνηση από τον φωνητικό βοηθό της συνολικής τιμής του καλαθιού και συγκεκριμένα του κειμένου “Total price: [number of total price]”. Παρακάτω, δίνεται μία εικόνα η οποία απεικονίζει το καλάθι μετά την αλλαγή της ποσότητας. Παρατηρείται, πως έχει τροποποιηθεί η ποσότητα από 1 σε 2, ενώ παράλληλα η συνολική τιμή του καλαθιού όπως είναι αναμενόμενο διπλασιάστηκε.

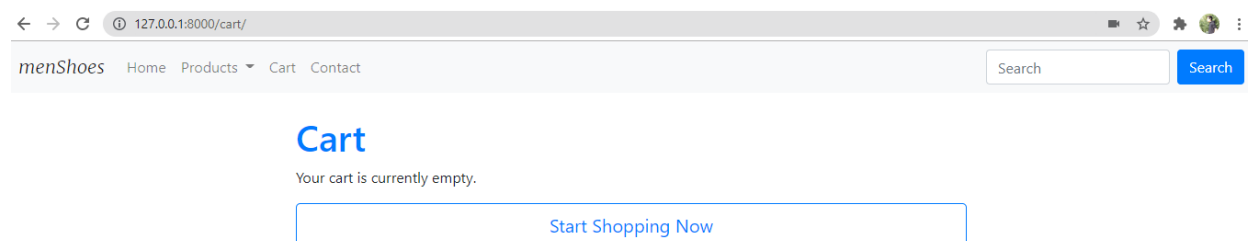


Εικόνα 27: Το καλάθι αγορών όταν η ποσότητα του προϊόντος που περιέχει αυξήθηκε από 1 σε 2

Σε πολλές περιπτώσεις, ένας χρήστης αλλάζει γνώμη σχετικά με την αγορά ενός προϊόντος το οποίο έχει τοποθετήσει στο καλάθι του. Συνεπώς, δεν θα μπορούσε να λείπει από τις διαθέσιμες εντολές αυτή που θα αφαιρεί ένα επιλεγμένο προϊόν από το καλάθι αγορών. Συγκεκριμένα, με την χρήση της εντολής “Remove product”, το προϊόν το οποίο έχει επιλεγεί από τον χρήστη αφαιρείται από το καλάθι ενώ στην συνέχεια προκύπτουν δύο πιθανά σενάρια και συγκεκριμένα:

1. Σε περίπτωση που το καλάθι έχει δύο ή περισσότερα προϊόντα, τότε με την διαγραφή ενός από αυτά θα τροποποιηθεί η συνολική τιμή του καλαθιού ενώ παράλληλα θα εκφωνηθεί το κείμενο “Product with name [name of the product that was deleted] has deleted successfully from your cart. Total price has changed to [new total price]”.
2. Σε περίπτωση που το καλάθι περιέχει ένα προϊόν και αυτό αφαιρεθεί, τότε θα εκφωνηθεί το μήνυμα “Your cart is now empty. Use the command options to start shopping now”.

Να σημειωθεί, πως στην πρώτο σενάριο με την διαγραφή ενός προϊόντος η τιμή του δείκτη που δείχνει το τρέχον προϊόν θα τροποποιηθεί έτσι ώστε να δείχνει το πρώτο προϊόν του καλαθιού. Παρακάτω, απεικονίζεται το καλάθι αγορών όταν διαγραφούν όλα τα προϊόντα από αυτό.



Εικόνα 28: Το καλάθι αγορών όταν διαγραφούν όλα τα προϊόντα που περιλαμβάνει

Παρατηρείται, ότι η συνολική τιμή του καλαθιού εξαφανίστηκε, ενσωματώθηκε το κείμενο Your cart is currently empty, ενώ παράλληλα το κουμπί που έγραφε Proceed to checkout αντικαταστάθηκε από ένα που γράφει Start Shopping Now.

Με την χρήση των παραπάνω εντολών, γίνεται εφικτή η αλληλεπίδραση εξ ολοκλήρου μέσω της φωνής με το καλάθι αγορών, κάτι το οποίο αποτελεί ένα βασικό βήμα πριν την πραγματοποίηση οποιασδήποτε αγοράς. Έτσι, ένας χρήστης με προβλήματα όρασης διαθέτει ακριβώς τις ίδιες δυνατότητες με έναν που χρησιμοποιεί το γραφικό περιβάλλον, μέσω της αξιοποίησης ενός μικρού αριθμού εντολών οι οποίες του προσφέρουν τα απαραίτητα εφόδια για την προετοιμασία του βήματος κατά το οποίο πραγματοποιείται η συναλλαγή.

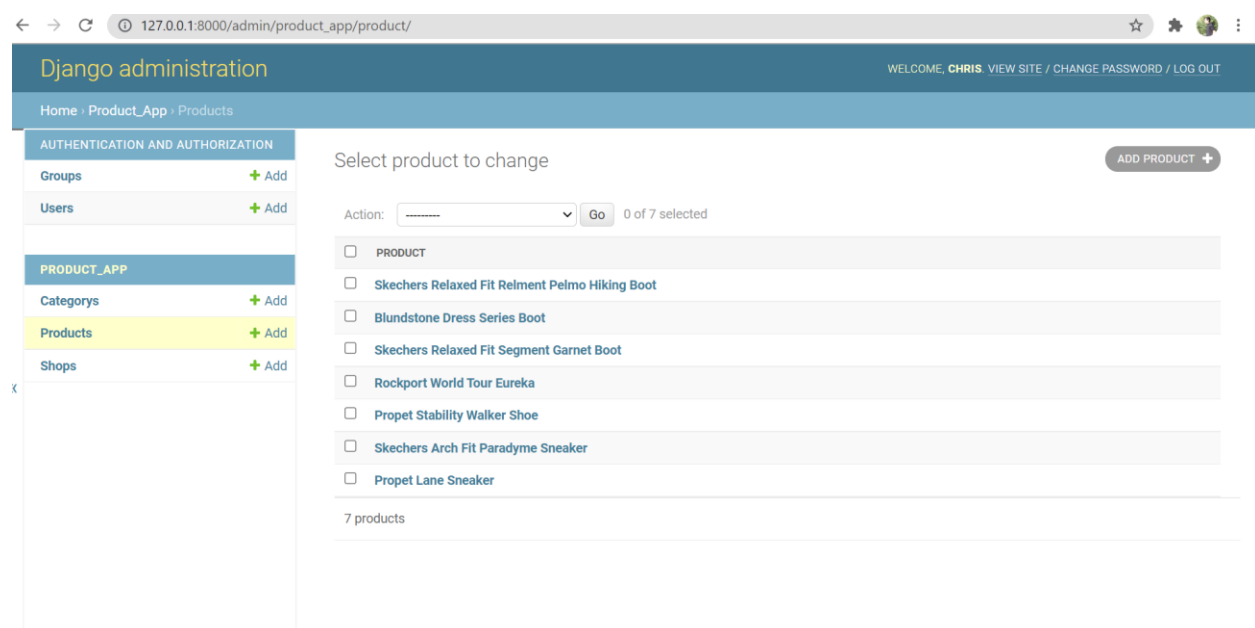
3.3 Τεχνική υποδομή

Σε μία ερευνητική εργασία που έχει ως στόχο την συνεισφορά μέσω της δημιουργίας λογισμικού, αποτελεί απαραίτητη προϋπόθεση τόσο η αναφορά στις τεχνολογίες με τις οποίες αλληλεπίδρασε, όσο και η περιγραφή του τρόπου με τον οποίο αυτές χρησιμοποιήθηκαν έτσι ώστε τελικά να παραχθεί το επιθυμητό αποτέλεσμα. Μέσω της παροχής τέτοιου είδους πληροφοριών, ερευνητές του ίδιου πεδίου μπορούν να λάβουν ιδέες για την δημιουργία νέων έργων, ή αντίστοιχα την βελτίωση ήδη υπαρχόντων. Παράλληλα, οποιοσδήποτε επιθυμεί μπορεί ανά πάσα στιγμή να σχηματίσει μία εικόνα για τις προσπάθειες που έχουν γίνει σχετικά με την εύρεση λύσεων σε ένα συγκεκριμένο πρόβλημα μέσω της χρήσης τεχνολογιών πληροφορικής. Έτσι λοιπόν, στόχος της παρούσας ενότητας αποτελεί η περιγραφή των τεχνολογιών που αξιοποιήθηκαν για την δημιουργία του συγκεκριμένου proof of concept, ενώ παράλληλα αναλύονται οι μέθοδοι με τις οποίες υλοποιήθηκε το φωνητικό μοντέλο αλληλεπίδρασης. Για την καλύτερη δυνατή περιγραφή των παραπάνω πληροφοριών, κρίθηκε απαραίτητος ο διαχωρισμός τους σε δύο υποενότητες. Στην πρώτη, αναφέρονται οι τεχνολογίες που χρησιμοποιήθηκαν τόσο για το σχεδιαστικό, όσο και για το λειτουργικό κομμάτι του menShoes, συνθέτοντας όλες μαζί το τελικό αποτέλεσμα. Στην δεύτερη, περιγράφονται τα συστατικά του menShoes που κάνουν εφικτή την πλοήγηση ενός χρήστη εξ ολοκλήρου μέσω της φωνής του.

3.3.1 Τεχνολογίες

Για την δημιουργία διαδικτυακών εφαρμογών σήμερα, ένας προγραμματιστής δεν χρειάζεται να ανακαλύψει ξανά τον τροχό. Αυτό, λόγω της πληθώρας διαθέσιμων frameworks τα οποία αναλαμβάνουν την δημιουργία των βασικών συστατικών πάνω στα οποία στηρίζεται μία εφαρμογή, αφήνοντας έτσι τον προγραμματιστή να προβεί άμεσα στο κομμάτι της δημιουργίας. Στην περίπτωση του menShoes, χρησιμοποιήθηκε το Django. Ένα framework γραμμένο σε Python. Με την χρήση του, έγινε εφικτή η παροχή της εκάστοτε σελίδας του καταστήματος ανάλογα με το αίτημα του χρήστη. Κάθε σελίδα που επιστρέφεται μπορεί να περιλαμβάνει εκτός από τον κώδικα HTML περιεχόμενο από την βάση δεδομένων, ενώ παράλληλα είναι προφανές πως επιτρέπονται λειτουργίες όπως η εγγραφή, η αφαίρεση και η ενημέρωση της π.χ. προσθήκη στοιχείων ενός προϊόντος στα πεδία που αφορούν το καλάθι αγορών. Αναφορικά με την βάση δεδομένων, το Django δίνει ως προεπιλογή την SQLite. Για την υλοποίηση του συγκεκριμένου proof of concept, η συγκεκριμένη επιλογή είναι υπέρ αρκετή, αφού δημιουργήθηκαν μόνο

τρεις πίνακες. Συγκεκριμένα, για την αλληλεπίδραση με την βάση δεδομένων στο Django, δημιουργούνται models. Κάθε model αφορά μία κλάση, η οποία στην ουσία αντιστοιχεί σε έναν πίνακα. Για την περίπτωση του menShoes, δημιουργήθηκαν τρία models και συγκεκριμένα ένα για τις διαθέσιμες κατηγορίες με όνομα Category, ένα για τα προϊόντα με όνομα Product και τέλος ένα το οποίο αφορά το καλάθι αγορών με όνομα Shop. Ένα από τα πλεονεκτήματα του Django, αφορά το γεγονός κατά το οποίο με την δημιουργία ενός οι περισσότερων μοντέλων δίνεται πρόσβαση σε ένα admin interface, μέσω του οποίου είναι εφικτές λειτουργίες όπως η δημιουργία ή η επεξεργασία του περιεχομένου της βάσης δεδομένων, η διαχείριση των χρηστών μιας εφαρμογής κ.α. Παρακάτω, απεικονίζεται μία εικόνα από το συγκεκριμένο interface που δημιουργήθηκε για το menShoes. Παρατηρούνται στο αριστερό μέρος οι τρεις πίνακες που αναφέρθηκαν προηγουμένως με ονόματα Products, Shops και Categories (το admin interface μετατρέπει τα ονόματα των πινάκων στον πληθυντικό προσθέτοντας ένα s).



Εικόνα 29: Το admin interface του Django

Με την χρήση ενός URL του καταστήματος, θα σταλεί ένα αίτημα (request) στον εξυπηρετητή (server) και το πρώτο πράγμα που θα γίνει είναι η ανακατεύθυνση του στην αντίστοιχη συνάρτηση ή αλλιώς view η οποία αφού ολοκληρώσει τις επιθυμητές λειτουργίες όπως π.χ. ανάκτηση δεδομένων από την βάση θα επιστρέψει τελικά μία HTML σελίδα. Μέσα στην συγκεκριμένη σελίδα, το Django δίνει την δυνατότητα της χρήσεως συγκεκριμένης σύνταξης έτσι ώστε να τυπώνονται τα επιθυμητά δεδομένα που ανακτήθηκαν από την βάση. Για παράδειγμα, με την χρήση του URL /categories/sneakers/ το Django θα ανακατευθύνει το αίτημα στην συνάρτηση που είναι υπεύθυνη για την σελίδα με τα προϊόντα οποιασδήποτε κατηγορίας και αυτή θα ανακτήσει από την βάση όλα όσα ανήκουν στην κατηγορία sneakers, επιστρέφοντας τα μαζί

με τον κώδικα HTML της συγκεκριμένης σελίδας. Με αυτό τον τρόπο, γίνεται εφικτή η υλοποίηση μιας εφαρμογής η οποία να λειτουργεί με δυναμικό τρόπο. Αναφέρθηκε προηγουμένως μία ειδική σύνταξη που χρησιμοποιεί το Django για την τύπωση δεδομένων σε HTML σελίδες. Για παράδειγμα, στην περίπτωση με τα προϊόντα της κατηγορίας sneakers χρησιμοποιήθηκε η εξής σύνταξη για την εμφάνιση τους.

```
<div class="row">
  {% for pr in products %}
    <div class="col-md-4">
      <div class="card mb-2">
        
        <div class="card-body">
          <h5 class="card-title name">{{ pr.name }}</h5>
          <p class="card-text price blue-price"><b>{{ pr.price }}$</b></p>
          <a href="{% url 'product_detail' pr.category pr.pk %}"
            class="btn btn-primary button-link">
            See product
          </a>
          <div class="id">Id: {{ pr.id }}</div>
        </div>
      </div>
    </div>
  {% endfor %}
</div>
```

Εικόνα 30: Κώδικας για την τύπωση των προϊόντων μιας κατηγορίας δυναμικά

Παρατηρείται, ότι ανάμεσα στον κώδικα HTML εμφανίζεται η γραμμή `{% for pr in products %}`. Στην συγκεκριμένη περίπτωση, το `products` διαθέτει όλα τα προϊόντα που ανήκουν στην κατηγορία `sneakers` και με την χρήση του `for` θα τυπωθούν τα επιθυμητά δεδομένα καθενός από αυτά με επαναληπτικό τρόπο. Έτσι, ως αποτέλεσμα κάθε προϊόν που εμφανίζεται στην σελίδα έχει τον ίδιο HTML κώδικα με μόνη διαφορά τα δεδομένα του.

Μία εφαρμογή που χρησιμοποιεί μόνο κώδικα HTML δεν θα μπορούσε να είναι λειτουργική, αφού αυτός είναι υπεύθυνος κυρίως για τον σκελετό μιας σελίδας. Για την συμπεριφορά και την εμφάνιση της, οι προγραμματιστές καλούνται να χρησιμοποιήσουν επιπλέον τεχνολογίες ο συνδυασμός των οποίων θα δώσει μία ολοκληρωμένη εφαρμογή. Συγκεκριμένα, για την υλοποίηση του `menShoes`, σε ότι έχει να κάνει με την εμφάνιση χρησιμοποιήθηκε η γλώσσα CSS, ενώ παράλληλα έγινε χρήση του framework Bootstrap το οποίο παρέχει μία πληθώρα έτοιμων συστατικών τα οποία μπορεί κάποιος να παραμετροποιήσει έτσι ώστε να δώσει το επιθυμητό style στην εφαρμογή του. Χαρακτηριστικά παραδείγματα της χρήσης του Bootstrap αποτελεί το μενού πλοήγησης του `menShoes` καθώς επίσης και η εμφάνιση των προϊόντων στις κατηγορίες. Ωστόσο, η δημιουργία μίας αισθητικά άρτιας εφαρμογής, δεν θα είχε κανένα νόημα εάν αυτή δεν παρείχε ορισμένη λειτουργικότητα στον χρήστη. Έτσι λοιπόν και στην περίπτωση του `menShoes`, δεν θα μπορούσε να λείπει η χρήση της γλώσσας JavaScript, μέσω της οποίας υλοποιείται το μεγαλύτερο κομμάτι της λειτουργικότητας του ηλεκτρονικού καταστήματος στην πλευρά του πελάτη (client side). Μέσω της χρήσης της, επιτεύχθηκε η δημιουργία δύο βασικών πυλώνων στην εφαρμογή. Ο πρώτος, αφορά την δυνατότητα ενός χρήστη για αλληλεπίδραση εξ ολοκλήρου μέσω της φωνής του. Αυτό, επιτεύχθηκε μέσω της χρήσης του Web Speech API του οποίου οι δυνατότητες χρησιμοποιήθηκαν τόσο για την αναγνώριση, όσο και για την σύνθεση ομιλίας. Συγκεκριμένα, αναφορικά με την αναγνώριση ομιλίας αξιοποιήθηκαν τα περιεχόμενα των εξής interfaces:

- `SpeechRecognition`
- `SpeechRecognitionEvent`
- `SpeechRecognitionResultList`
- `SpeechRecognitionResult`
- `SpeechRecognitionAlternative`

Αντίστοιχα, σε ότι έχει να κάνει με την περίπτωση της σύνθεσης ομιλίας, χρησιμοποιήθηκαν τα περιεχόμενα των παρακάτω interfaces και συγκεκριμένα:

- `SpeechSynthesis`
- `SpeechSynthesisEvent`
- `SpeechSynthesisUtterance`
- `SpeechSynthesisVoice`

Κάποιος που επιθυμεί να κατανοήσει τον τρόπο λειτουργίας αλλά και τις δυνατότητες των παραπάνω interfaces, μπορεί να ανατρέξει στο κεφάλαιο 2 όπου και παρουσιάζεται αναλυτικά το Web Speech API.

Ο δεύτερος πυλώνας, αφορά την δυνατότητα της ανταλλαγής δεδομένων με τον εξυπηρετητή χωρίς να χρειάζεται η ανανέωση της σελίδας. Με την συγκεκριμένη ικανότητα, ένας χρήστης μπορεί να ανακτήσει ή να προσθέσει περιεχόμενο στην βάση δεδομένων, χωρίς να επηρεάζεται η τρέχουσα κατάσταση που μπορεί να έχει σχηματιστεί μεταξύ αυτού και του φωνητικού βοηθού. Για παράδειγμα, με την επιλογή ενός προϊόντος και την χρήση της εντολής “Add to cart”, αυτό θα προστεθεί στο καλάθι αγορών, χωρίς να χρειάζεται να γίνει ανανέωση της σελίδας, κάτι το οποίο θα είχε ως αποτέλεσμα την λειτουργία του φωνητικού βοηθού από την αρχή, βάζοντας έτσι τον τελικό χρήστη να επαναλάβει έναν αριθμό βημάτων για να φτάσει στην προηγούμενη κατάσταση. Ακόμα ένα παράδειγμα, αποτελεί η δυνατότητα του χρήστη να ανακτήσει την περιγραφή ενός προϊόντος από την βάση δεδομένων ενώ ο ίδιος βρίσκεται στην σελίδα με τα προϊόντα μιας κατηγορίας. Κάτι τέτοιο, δεν θα ήταν εύκολο αν η πραγματοποίηση του αιτήματος (request) προκαλούσε ανανέωση της σελίδας, ενώ σε διαφορετική περίπτωση ο χρήστης θα έπρεπε να μπει στην σελίδα ενός προϊόντος για την εύρεση της περιγραφής του, κάτι το οποίο αυξάνει τα βήματα που πρέπει να πραγματοποιήσει για την αλληλεπίδραση του με το κατάστημα, ενώ μειώνει την αποτελεσματικότητα που μπορεί να έχει ένα φωνητικό μοντέλο αλληλεπίδρασης. Για την επίτευξη της συγκεκριμένης λειτουργικότητας, χρησιμοποιήθηκε AJAX (Asynchronous JavaScript and XML) και συγκεκριμένα αξιοποιήθηκε η συνάρτηση `ajax` που παρέχει η βιβλιοθήκη `jQuery`. Μέσω της συγκεκριμένης συνάρτησης, έγινε εφικτή η πραγματοποίηση ασύγχρονων `http` αιτημάτων στον εξυπηρετητή, χωρίς να πραγματοποιείται ανανέωση της εκάστοτε σελίδας. Ανάλογα με την φύση του αιτήματος (request), δημιουργήθηκαν αντίστοιχες συναρτήσεις ή σε ορολογία του Django

views, οι οποίες λαμβάνουν τα εκάστοτε δεδομένα και αφού αλληλεπιδράσουν με την βάση επιστρέφουν μία κατάλληλη απάντηση η οποία θα εκφωνηθεί τελικά από τον φωνητικό βοηθό στον χρήστη.

3.3.2 Μεθοδολογία ενσωμάτωσης φωνητικού βοηθού

Το σημαντικότερο κομμάτι από την συνεισφορά της παρούσας εργασίας πέρα από την προσπάθεια για την απόδειξη ότι η πλοήγηση σε ένα ηλεκτρονικό κατάστημα εξ ολοκλήρου με την φωνή είναι εφικτή, έγκειται στο φωνητικό μοντέλο αλληλεπίδρασης που προτείνεται. Συγκεκριμένα, όπως παρατηρείται στον τίτλο της συγκεκριμένης υποενότητας, επιλέχθηκε η λέξη ενσωμάτωση αντί για δημιουργία. Αυτό, αφού επίκεντρο κατά την φάση του σχεδιασμού αποτέλεσε η εύρεση ενός τρόπου με τον οποίο ηλεκτρονικά καταστήματα δεν θα χρειάζεται να κάνουν ριζικές και χρονοβόρες αλλαγές, οι οποίες παράλληλα θα απαιτούν την καταβολή μεγάλων χρηματικών ποσών. Αντίθετα, το φωνητικό μοντέλο αλληλεπίδρασης σχεδιάστηκε έτσι ώστε να έχει την ικανότητα ενσωμάτωσης σε ένα ηλεκτρονικό κατάστημα, με τρόπο ο οποίος να μην απαιτεί αλλαγές στην ήδη υπάρχουσα δομή του, παρά μόνο ορισμένες προσθήκες. Με αυτό τον τρόπο, ηλεκτρονικά καταστήματα θα μπορούν να αποκτήσουν σε μικρό χρονικό διάστημα δυνατότητες αναγνώρισης και σύνθεσης ομιλίας, προσεγγίζοντας έτσι ένα επιπλέον κοινό το οποίο έχει ανάγκη από μεθόδους υποβοήθησης αναφορικά με τον τρόπο της πλοήγησης. Παρακάτω, αναλύονται τα δομικά συστατικά που συνθέτουν τον φωνητικό βοηθό του menShoes.

Όπως έχει προαναφερθεί, κάθε σελίδα του menShoes διαθέτει τις δικές της εντολές, αλλά και οδηγίες για το πως μπορεί να αλληλεπιδράσει ένας χρήστης με το περιεχόμενο της. Ωστόσο, όλες παρουσιάζουν ένα κοινό το οποίο αφορά το μενού πλοήγησης καθώς επίσης και τον τρόπο με τον οποίο ενεργοποιείται το μικρόφωνο. Για τον λόγο αυτό, σε κάθε HTML σελίδα του menShoes περιέχονται δύο βασικά Javascript αρχεία. Το πρώτο, έχει όνομα base και είναι κοινό σε όλες τις σελίδες, ενώ αποτελείται από τέσσερα βασικά μέρη.

Αρχικά, το base αρχείο περιλαμβάνει την συνάρτηση getUserMedia η οποία θα επιχειρήσει να ζητήσει άδεια χρήσης του μικροφώνου έτσι ώστε να είναι εφικτή στην συνέχεια η χρήση των δυνατοτήτων αναγνώρισης ομιλίας. Σε περίπτωση που δοθεί άδεια χρήσης του μικροφώνου, θα γίνει κλήση σε μία συνάρτηση η οποία είναι υπεύθυνη για την εκφώνηση ενός μηνύματος καλωσορίσματος στον χρήστη όπως για παράδειγμα “Welcome to the contact page. If you don’t know what to do, use the command instructions”. Να σημειωθεί, ότι στα πλαίσια της υλοποίησης του συγκεκριμένου proof of concept θεωρείται δεδομένο ότι ο χρήστης θα δώσει άδεια χρήσης μικροφώνου. Αυτό, λόγω του ότι η ίδια η εφαρμογή δεν μπορεί να συμβάλει στην υποβοήθηση του έτσι ώστε να αλληλεπιδράσει με το παραθυράκι της άδειας χρήσης μικροφώνου, αφού αυτό προέρχεται από το πρόγραμμα περιήγησης και συνεπώς βρίσκεται εκτός της εφαρμογής. Η μόνη επέμβαση που θα μπορούσε να έχει μία τέτοια εφαρμογή θα ήταν

με την χρήση συνθετικής ομιλίας έτσι ώστε να ενημερώσει τον χρήστη ότι πρέπει να δώσει άδεια χρήσης μικροφώνου εάν επιθυμεί να αλληλεπιδράσει με τον φωνητικό βοηθό.

Το δεύτερο μέρος του αρχείου base, αφορά τις δηλώσεις με στόχο την αξιοποίηση του Web Speech API. Συγκεκριμένα, πραγματοποιείται η δημιουργία ενός SpeechRecognition και ενός SpeechSynthesisUtterance αντικειμένου. Στην περίπτωση του πρώτου, η ιδιότητα lang θα λάβει την τιμή 'en-US', ενώ οι υπόλοιπες θα λάβουν τις τιμές από προεπιλογή που έχουν αναφερθεί στο κεφάλαιο περιγραφής του Web Speech API, αφού αυτές αρκούν για τον σκοπό του proof of concept. Στην περίπτωση του δεύτερου, οι ιδιότητες του θα λάβουν παρόμοια τις τιμές από προεπιλογή εκτός από τις ιδιότητες που αφορούν την επιλογή της φωνής και της γλώσσας. Συγκεκριμένα, αξιοποιήθηκε το γεγονός onvoiceschanged σε συνδυασμό με την μέθοδο getVoices, έτσι ώστε να γίνει αναζήτηση στις διαθέσιμες φωνές με στόχο την εύρεση μίας η οποία θα αφορά την αγγλική γλώσσα, ενώ παράλληλα η ιδιότητα της με όνομα localService θα επιστρέφει την τιμή true. Στην ενότητα με τα προβλήματα περιγράφεται γιατί είναι σημαντικό η συγκεκριμένη ιδιότητα να είναι true.

Το τρίτο μέρος του αρχείου base, αφορά την χρήση τριών γεγονότων τα δύο εκ των οποίων αφορούν τον τρόπο ενεργοποίησης του μικροφώνου. Συγκεκριμένα, όπως έχει προαναφερθεί, ένας χρήστης μπορεί να ενεργοποιήσει το μικρόφωνο του είτε με την χρήση ενός κουμπιού, είτε με το πάτημα του αριστερού ή δεξιού βέλους του πληκτρολογίου. Για κάθε μία από αυτές τις περιπτώσεις, δημιουργήθηκε ένα γεγονός το οποίο περιλαμβάνει τον ίδιο κώδικα με μόνη διαφορά τον τρόπο πυροδότησης του. Με την πυροδότηση ενός από αυτών, δηλαδή είτε με την χρήση του κουμπιού είτε με την χρήση του δεξιού ή αριστερού βέλους, θα ξεκινήσει η αναγνώριση ομιλίας με την χρήση της μεθόδου start του SpeechRecognition αντικειμένου που δημιουργήθηκε προηγουμένως. Να σημειωθεί, πως μέσα στο μπλοκ κώδικα γίνεται χρήση της ιδιότητας speaking σε συνδυασμό με την συνάρτηση cancel του SpeechSynthesis interface, έτσι ώστε σε περίπτωση που εκφωνείται κάτι την στιγμή πυροδότησης του γεγονότος να σταματήσει αμέσως η σύνθεση ομιλίας πριν ξεκινήσει η αναγνώριση. Επιπλέον, πριν από την ενεργοποίηση της αναγνώρισης ομιλίας γίνεται χρήση της συνάρτησης play πάνω σε ένα αντικείμενο το οποίο δημιουργήθηκε με χρήση του audio κατασκευαστή (constructor) και έχει ως αποτέλεσμα την αναπαραγωγή ενός χαρακτηριστικού ήχου που αναφέρει ότι ξεκινάει η αναγνώριση ομιλίας. Επιπρόσθετα, αναφορικά με τον τρόπο ενεργοποίησης του μικροφώνου μέσω του κουμπιού, λόγω της ύπαρξης πάνω σε αυτό ενός εικονιδίου αντί για κείμενο, χρησιμοποιήθηκε πάνω στο button στοιχείο το χαρακτηριστικό aria-label. Το συγκεκριμένο χαρακτηριστικό, χρησιμοποιείται όταν δεν είναι ορατό κάποιο κείμενο π.χ. σε ένα στοιχείο όπως είναι το button με αποτέλεσμα κάποιος χρήστης που χρησιμοποιεί έναν αναγνώστη οθόνης να μην γνωρίζει ποια είναι η λειτουργία του κουμπιού αυτού. Με την χρήση όμως του χαρακτηριστικού aria-label, ο αναγνώστης οθόνης θα εκφωνήσει την τιμή του δίνοντας στον χρήστη να καταλάβει τι

επιτυγχάνεται με την χρήση του. Το τρίτο γεγονός στο αρχείο base, αποτελεί ένα από τα περιεχόμενα του SpeechRecognition interface. Συγκεκριμένα, γίνεται χρήση του end event μέσα στο μπλοκ κώδικα του οποίου χρησιμοποιείται μία Boolean μεταβλητή, στόχος της οποίας είναι ο έλεγχος για το αν η υπηρεσία αναγνώρισης επέστρεψε κάποιο αποτέλεσμα πριν την αποσύνδεση της. Να σημειωθεί, ότι η τιμή της συγκεκριμένης μεταβλητής γίνεται true μέσα στο μπλοκ κώδικα του result event, κάτι το οποίο σημαίνει πως αν δεν αλλάξει η τιμή της από false είναι προφανές ότι η υπηρεσία αναγνώρισης δεν κατάφερε να αναγνωρίσει κάτι και συνεπώς θα εκφωνηθεί αντίστοιχο μήνυμα. Όπως θα φανεί και παρακάτω, το result event περιλαμβάνεται κάθε φορά σε διαφορετικό αρχείο ανάλογα με την σελίδα του καταστήματος.

Το τέταρτο και τελευταίο μέρος του base αρχείου, αφορά το περιεχόμενο του μενού πλοήγησης. Όπως προαναφέρθηκε, ο σχεδιασμός του φωνητικού μοντέλου αναφορικά με την χρήση του μενού, έγινε έτσι ώστε αρχικά να διαβάζεται το πρώτο επίπεδο των επιλογών και στην συνέχεια ο χρήστης με βάση την επιλογή που θα εκφωνήσει είτε να μεταβαίνει σε μία σελίδα, είτε να ακούει τις επιπλέον επιλογές σε περίπτωση που η λέξη που εκφώνησε αφορά ένα υπό-μενού. Για να είναι εφικτή η συγκεκριμένη λειτουργικότητα, χρησιμοποιήθηκαν δύο συστατικά. Το πρώτο, αφορά την δημιουργία ενός αντικειμένου με όνομα menu, περιεχόμενο του οποίου αφορά μία πληθώρα από ιδιότητες κάθε μία από τις οποίες μπορεί κάποιος μπορεί να δει ως ένα ζεύγος key – value. Κάθε ένα από αυτά τα κλειδιά, αφορά μία επιλογή του μενού ενώ η τιμή που αντιστοιχεί σε αυτό μπορεί να είναι είτε ένας σύνδεσμος σε μία σελίδα του καταστήματος, είτε κείμενο το οποίο αφορά την περίπτωση που το κλειδί αντιστοιχεί σε ένα υπό-μενού. Σε αυτή την περίπτωση, το κείμενο περιλαμβάνει τις διαθέσιμες επιλογές του υπό-μενού, ενώ αντίστοιχα μέσα στο αντικείμενο menu υπάρχουν κλειδιά τα οποία αφορούν αυτές τις επιλογές και ως τιμή τους έχουν τον αντίστοιχο σύνδεσμο. Το δεύτερο συστατικό, αφορά την δημιουργία ενός μονοδιάστατου πίνακα με όνομα pages. Περιεχόμενο του συγκεκριμένου πίνακα, αποτελούν τα κλειδιά του αντικειμένου menu τα οποία έχουν ως τιμή τους ένα σύνδεσμο και στην ουσία αφορούν τις διαθέσιμες σελίδες του καταστήματος και όχι ένα υπό-μενού. Σε επόμενο μέρος της εργασίας, περιγράφεται ο τρόπος με τον οποίο χρησιμοποιούνται τα δύο παραπάνω συστατικά, έτσι ώστε ένας χρήστης να μπορεί να χρησιμοποιήσει το μενού για την πλοήγηση στις σελίδες του menShoes εξ ολοκλήρου μέσω της φωνής του. Παρακάτω, ακολουθεί εικόνα με περιεχόμενο το αντικείμενο menu και τον πίνακα pages.

```
var menu = { 'options' : 'Available options: home, products, shopping cart, contact, search for. Say the option you want.',
            'contact' : '/contact/',
            'home' : '/',
            'products' : 'Available categories: sneakers, boots. Say the category you want.',
            'sneakers' : '/categories/sneakers/',
            'boots' : '/categories/boots/',
            'search for' : '/search/',
            'shopping cart' : '/cart/'};

var pages = ["sneakers", "boots", "contact", "home", "shopping cart", "search for"];
```

Εικόνα 31: Το αντικείμενο `menu` και ο πίνακας `pages` τα οποία χρησιμοποιούνται για την αλληλεπίδραση του χρήστη με το μενού πλοήγησης μέσω της φωνής του

Όπως προαναφέρθηκε, η χρήση του αρχείου `base` περιλαμβάνεται σε κάθε HTML σελίδα του `menShoes`, αφού αυτό περιλαμβάνει κώδικα ο οποίος είναι απαραίτητος για την αλληλεπίδραση μέσω της φωνής και συνεπώς με την τοποθέτηση του σε ένα μόνο αρχείο γίνεται εφικτή η επαναχρησιμοποίηση του. Ωστόσο, κάτι τέτοιο δεν μπορεί να εφαρμοστεί και στην περίπτωση του κώδικα που αφορά το υπόλοιπο κομμάτι της αλληλεπίδρασης μέσω φωνητικών εντολών, αφού κάθε σελίδα δίνει στον εκάστοτε χρήστη διαφορετικές δυνατότητες. Έτσι λοιπόν, με κάθε σελίδα που επιστρέφεται από το Django, εκτός από το `base` αρχείο περιλαμβάνεται και ένα ακόμα με περιεχόμενο τις αντίστοιχες δυνατότητες που μπορεί να αξιοποιήσει ένας χρήστης σε αυτή. Ωστόσο, τα κύρια συστατικά αυτών των αρχείων ακολουθούν παρόμοια λογική.

Κάθε τέτοιο αρχείο, αποτελείται από τρία βασικά μέρη. Το πρώτο, αφορά υλοποίηση των συναρτήσεων κάθε μία από τις οποίες αντιστοιχεί σε μία εντολή που μπορεί να δώσει ο χρήστης. Έτσι λοιπόν, έχοντας ως παράδειγμα την σελίδα με τα προϊόντα της κατηγορίας `sneakers`, για κάθε μία από τις εντολές που είναι διαθέσιμες σε αυτή υπάρχει μέσα στο Javascript αρχείο μία αντίστοιχη συνάρτηση. Βέβαια, σε κάποιες περιπτώσεις μία συνάρτηση είναι δυνατό να χρησιμοποιείται και σε άλλο αρχείο είτε αυτούσια, είτε με την χρήση διαφορετικού κειμένου προς εκφώνηση στον χρήστη όταν αυτή ολοκληρωθεί. Για παράδειγμα, σε κάθε σελίδα του `menShoes` είναι εφικτή η χρήση της εντολής `Commands`. Η συγκεκριμένη εντολή, αντιστοιχίζεται σε μία συνάρτηση η οποία χρησιμοποιείται σε κάθε σελίδα όμοια, με μόνη διαφορά το περιεχόμενο που θα εκφωνηθεί όταν αυτή καλείται. Παρακάτω, δίνεται εικόνα με περιεχόμενο της συγκεκριμένης συνάρτησης.

```
function readCommands(){
    var i;
    var j = 0;
    speech.text = 'Available commands for this page. ';
    for(i = 0; i < productCommands.length; i++){
        ++j;
        speech.text += "Command " + j + ": " + productCommands[i] + ". ";
    }
    window.speechSynthesis.speak(speech);
}
```

Εικόνα 32: Κώδικας για την εκφώνηση από τον φωνητικό βοηθό των διαθέσιμων εντολών μιας σελίδας

Με την εκτέλεση του παραπάνω κώδικα, θα εκφωνηθούν οι διαθέσιμες εντολές της σελίδας με τα προϊόντα sneakers ως “Available commands for this page. Command 1: [name of the command]. Command 2: [name of the command]...”. Παρατηρείται, πως το όνομα κάθε εντολής ανακτάται από έναν μονοδιάστατο πίνακα. Συγκεκριμένα, κάθε Javascript αρχείο του menShoes εκτός του base διαθέτει διαθέτει έναν πίνακα με όνομα productCommands, ο οποίος όπως προδίδει και το όνομα του περιλαμβάνει σε κάθε θέση του μία εντολή. Μέσω της συγκεκριμένης μεθόδου πετυχαίνονται δύο πλεονεκτήματα. Το πρώτο, αφορά την ικανότητα διαχείρισης των εντολών με οργανωμένο τρόπο χωρίς να απαιτείται η συγγραφής τους ως ένα μεγάλο αλφαριθμητικό το οποίο κάθε φορά θα τροποποιείται ανάλογα με την προσθήκη ή αντίστοιχα την αφαίρεση εντολών. Το δεύτερο πλεονέκτημα, αναφέρεται κατά την περιγραφή του επόμενου βασικού μέρους που περιέχει ένα Javascript αρχείο.

Η υλοποίηση των συναρτήσεων, αφορά ένα μεγάλο μέρος της διαδικασίας ενσωμάτωσης ενός φωνητικού μοντέλου αλληλεπίδρασης, αφού αυτές είναι που κάνουν εφικτή την λειτουργικότητα που αιτείται ένας χρήστης μέσω των διαθέσιμων εντολών. Ωστόσο, αυτή η διαδικασία αποτελεί την μία όψη του νομίσματος. Συγκεκριμένα, ένα εξίσου σημαντικό κομμάτι είναι ο καθορισμός του τρόπου με τον οποίο θα καλείται κάθε φορά η κατάλληλη συνάρτηση ανάλογα με την εντολή που θα χρησιμοποιείται από τον χρήστη. Όπως ακριβώς με την περίπτωση του framework Django, κατά την οποία ένα αίτημα (request) φτάνει στον εξυπηρετητή (server) και από εκεί ανακατευθύνεται στην κατάλληλη συνάρτηση με στόχο την επιστροφή ως απάντηση μία HTML σελίδα, έτσι και στην περίπτωση του φωνητικού μοντέλου ανάλογα με την είσοδο του χρήστη θα πρέπει να εκτελείται η αντίστοιχη συνάρτηση. Μία τέτοια λειτουργία, αυτονόητο είναι πως θα πρέπει να υλοποιηθεί με τρόπο ο οποίος να είναι εύκολα διαχειρίσιμος αλλά και επεκτάσιμος. Στο συγκεκριμένο proof of concept, η συγκεκριμένη διαδικασία πραγματοποιείται μέσα στο μπλοκ κώδικα που εκτελείται με την πυροδότηση του event result, το οποίο όπως περιγράφηκε στο κεφάλαιο 2 ανήκει στο SpeechRecognition interface. Ειδικότερα, στο Javascript αρχείο κάθε σελίδας του menShoes, χρησιμοποιείται το γεγονός result με την πυροδότηση του οποίου επιστρέφεται οποιοδήποτε αποτέλεσμα έχει αναγνωριστεί από την ομιλία του χρήστη μέσω ενός SpeechRecognitionEvent αντικείμενου. Αφού ανακτηθεί από το συγκεκριμένο αντικείμενο το κείμενο που έχει αναγνωριστεί, στην συνέχεια ακολουθεί ένας αριθμός συνθηκών από τις οποίες θα εκτελεστεί αυτή που ταιριάζει στο αποτέλεσμα που έχει αναγνωριστεί. Συγκεκριμένα, οι συνθήκες αυτές αντιπροσωπεύουν τρία διαφορετικά σενάρια.

Το πρώτο σενάριο, αφορά την περίπτωση κατά την οποία ο χρήστης θα δώσει ως είσοδο με την φωνή του μία από τις διαθέσιμες εντολές η οποία δεν σχετίζεται με το μενού πλοήγησης. Για παράδειγμα, έστω ότι ο χρήστης βρίσκεται στην σελίδα ενός προϊόντος και κάνει χρήση της εντολής “Add to cart”.

Παρακάτω, φαίνεται ένα κομμάτι του μπλοκ κώδικα που εκτελείται όταν πυροδοτείται το γεγονός result στην σελίδα ενός προϊόντος.

```
if (productCommands.includes(transcript) || cart == true) {  
    search = false;  
    if (transcript == "add to cart" || cart == true){  
        addToCart(transcript);  
    } else if (transcript == "product id"){  
        readId();  
    } else if (transcript == "product name"){  
        readName();  
    } else if (transcript == "commands"){  
        readCommands();  
    } else if (transcript == "instructions"){  
        readInstructions();  
    } else if (transcript == "product price"){  
        readPrice();  
    } else if (transcript == "product description"){  
        readDescription();  
    }  
}
```

Εικόνα 33: Κώδικας για την περίπτωση που ο χρήστης θα εκφωνήσει μία από τις διαθέσιμες εντολές η οποία δεν σχετίζεται με το μενού

Αναφέρθηκε προηγουμένως ότι με την χρήση ενός μονοδιάστατου πίνακα για την αποθήκευση των εντολών μιας σελίδας έχει δύο πλεονεκτήματα. Ωστόσο, έγινε σχολιασμός μόνο του πρώτου. Αυτό, αφού η χρησιμότητα του δεύτερου φαίνεται κατά την εκτέλεση του μπλοκ κώδικα που εκτελείται με την πυροδότηση του result event. Συγκεκριμένα, όπως φαίνεται από την παραπάνω εικόνα, στην πρώτη συνθήκη που εκτελείται γίνεται έλεγχος για το αν ο πίνακας με τις εντολές περιλαμβάνει την λέξη που έδωσε ο χρήστης ως είσοδο μέσω του μικροφώνου του. Αυτονόητο είναι πως η είσοδος του χρήστη έχει αποθηκευτεί μέσα στην μεταβλητή με όνομα transcript. Αφού λοιπόν το περιεχόμενο αυτής της μεταβλητής είναι η λέξη “add to cart”, θα πραγματοποιηθεί το πρώτο σενάριο κατά το οποίο ο χρήστης έχει χρησιμοποιήσει μία από τις διαθέσιμες εντολές για την αλληλεπίδραση του με την σελίδα και συνεπώς θα εκτελεστεί το μπλοκ κώδικα της πρώτης συνθήκης. Έπειτα, όπως φαίνεται από την εικόνα μέσα στο μπλοκ της συγκεκριμένης συνθήκης υπάρχει άλλη μία, η οποία έχει ως στόχο της να εντοπίσει πια από τις διαθέσιμες εντολές της σελίδας χρησιμοποίησε ο χρήστης, καλώντας έτσι την αντίστοιχη συνάρτηση. Σε πολλές περιπτώσεις, η χρήση μιας εντολής θα καλέσει μία συνάρτηση η οποία θα απαιτεί την παροχή επιπλέον δεδομένων από τον χρήστη. Για παράδειγμα, μελετώντας πάλι την περίπτωση της εντολής “Add to cart”, όταν με την χρήση της καλεστεί η αντίστοιχη συνάρτηση που είναι υπεύθυνη για την τοποθέτηση ενός προϊόντος στο καλάθι αγορών, θα εκφωνηθεί από τον φωνητικό βοηθό ένα κείμενο το οποίο θα ζητάει από τον χρήστη τον αριθμό της ποσότητας που επιθυμεί να τοποθετηθεί στο καλάθι του. Η συγκεκριμένη διαδικασία λοιπόν, απαιτεί την παροχή δεδομένων στην ίδια συνάρτηση έτσι ώστε να ολοκληρωθεί η διαδικασία για την οποία είναι υπεύθυνη. Η συγκεκριμένη απαίτηση ωστόσο δημιουργεί ένα εμπόδιο το οποίο πρέπει να ξεπεραστεί. Η ποσότητα που θα δοθεί από τον χρήστη ως είσοδο, δεν αποτελεί κάποια εντολή και συνεπώς πρέπει να βρεθεί ένας τρόπος με τον οποίο να μπορεί αυτή να σταλεί ως παράμετρος στην ίδια συνάρτηση. Για την λύση του συγκεκριμένου προβλήματος, στις συναρτήσεις που απαιτείται η

κλήση τους περισσότερες από μία φορά για την ολοκλήρωση μιας λειτουργίας, γίνεται η χρήση Boolean μεταβλητών. Συγκεκριμένα, με την πρώτη κλήση μιας τέτοιας συνάρτησης, θα μετατραπεί η τιμή της αντίστοιχης Boolean μεταβλητής σε true. Έτσι λοιπόν, χρησιμοποιώντας την συγκεκριμένη μεταβλητή στην συνθήκη που υπάρχει μέσα στο μπλοκ κώδικα που εκτελείται όταν πυροδοτείται το result γεγονός, γίνεται εφικτή η εκτέλεση της ίδιας συνάρτησης στέλνοντας τα επιπλέον δεδομένα που έχει δώσει ως είσοδο ο χρήστης. Οπότε, χρησιμοποιώντας ο χρήστης την εντολή “Add to cart”, αρχικά θα κληθεί η συνάρτηση addToCart μέσα στην οποία η Boolean μεταβλητή cart θα λάβει την τιμή true και στην συνέχεια θα ζητηθεί από τον χρήστη να δώσει την επιθυμητή ποσότητα. Αφού ο χρήστης δώσει την συγκεκριμένη πληροφορία, λόγω της τιμής true στην cart μεταβλητή θα εκτελεστεί ο κώδικας της πρώτης συνθήκης παρόλο που η είσοδος που έδωσε ο χρήστης δεν αφορά μία εντολή, ενώ στην συνέχεια θα εκτελεστεί ο κώδικας της αμέσως επόμενης συνθήκης όπου και θα γίνει ξανά κλήση της συνάρτησης addToCart, στέλνοντας αυτή την φορά την ποσότητα που έδωσε ο χρήστης. Τέλος, αφού η συνάρτηση κάνει τους κατάλληλους ελέγχους, θα τοποθετήσει το προϊόν στο καλάθι αλλάζοντας στην συνέχεια την τιμή της μεταβλητής cart σε false.

Το δεύτερο σενάριο, αφορά την περίπτωση κατά την οποία ένας χρήστης θα χρησιμοποιήσει μια εντολή η οποία αφορά το μενού πλοήγησης. Για παράδειγμα, έστω ότι ο χρήστης χρησιμοποιεί την εντολή “Options”. Παρακάτω, φαίνονται οι συνθήκες που θα ελεγχθούν στην συγκεκριμένη περίπτωση.

```
else if (transcript in menu || search == true) {  
    if (pages.includes(transcript) || search == true){  
        if (transcript == 'search for'){  
            search = true;  
            speech.text = "Tell the phrase you want to search";  
            window.speechSynthesis.speak(speech);  
        } else if (search == true) {  
            search = false;  
            location.replace(menu['search for'] + transcript + '/');  
        } else {  
            location.replace(menu[transcript]);  
        }  
    } else {  
        speech.text = menu[transcript];  
        window.speechSynthesis.speak(speech);  
    }  
}
```

Εικόνα 34: Κώδικας για την περίπτωση που ο χρήστης θα εκφωνήσει μία εντολή που σχετίζεται με το μενού

Όπως είναι προφανές, το πρώτο else if αποτελεί την συνέχεια της δομής ελέγχου που παρουσιάστηκε στην εικόνα 33. Όταν σχολιάστηκε το περιεχόμενο του base αρχείου, αναφέρθηκε η δομή του μενού πλοήγησης, η οποία υλοποιήθηκε με ένα αντικείμενο ονόματος menu και ένα πίνακα με όνομα pages. Μέσω της χρήσης αυτών σε συνδυασμό με τον κώδικα που φαίνεται στην παραπάνω εικόνα, γίνεται εφικτή η δυνατότητα αξιοποίησης του μενού πλοήγησης από έναν χρήστη εξ ολοκλήρου μέσω της φωνής του. Συγκεκριμένα, οι παραπάνω συνθήκες αντικατοπτρίζουν δύο διαφορετικές περιπτώσεις. Αρχικά, με την χρήση της πρώτης

συνθήκης ελέγχεται εάν η είσοδος που έδωσε ο χρήστης αφορά μία από τις επιλογές που βρίσκονται μέσα στο αντικείμενο menu ως ιδιότητες του. Εφόσον ισχύει κάτι τέτοιο, τότε σημαίνει πως ο χρήστης χρησιμοποίησε μία από τις επιλογές που αντιστοιχούν στο μενού πλοήγησης και συνεπώς θα ελεγχθεί η πρώτη περίπτωση από τις παραπάνω, σύμφωνα με την οποία μπορεί να έχει δώσει ως είσοδο μία από τις διαθέσιμες σελίδες του καταστήματος στην οποία επιθυμεί να μεταβεί ή να έχει κάνει χρήση της εντολής “Search for” (η εντολή search for περιλαμβάνεται μέσα στον πίνακα pages). Σε κάθε ένα από αυτά τα σενάρια, θα χρησιμοποιηθεί η τιμή της αντίστοιχης ιδιότητας του αντικειμένου menu, αφού η τιμή αυτής αφορά το URL της σελίδας που έχει αιτηθεί ο χρήστης, με τελικό αποτέλεσμα την μετάβαση του σε αυτή. Να σημειωθεί, πως η χρήση της Boolean μεταβλητής search έχει παρόμοια χρήση με την αντίστοιχη μεταβλητή cart που αναφέρθηκε παραπάνω, αφού και στην περίπτωση της εντολής “Search for” ο φωνητικός βοηθός θα ζητήσει την λέξη ή φράση για την οποία θα γίνει αναζήτηση και συνεπώς ο χρήστης καλείται να δώσει επιπλέον πληροφορία. Η δεύτερη περίπτωση, αφορά την αξιοποίηση από τον χρήστη μιας επιλογής του μενού η οποία θα έχει ως αποτέλεσμα την εκφώνηση περαιτέρω πληροφοριών οι οποίες αφορούν π.χ. τα περιεχόμενα ενός υπό-μενού. Για παράδειγμα, με την χρήση της λέξης Products, αφού διαπιστωθεί ότι αυτή δεν ανήκει στον πίνακα με τις σελίδες του καταστήματος αλλά και η μεταβλητή search έχει την τιμή false, τότε κατευθείαν θα χρησιμοποιηθεί ως κλειδί στο αντικείμενο menu, εκφωνώντας ο φωνητικός βοηθός την τιμή του και συγκεκριμένα το κείμενο “Available categories: sneakers, boots. Say the category you want”. Όπως είναι προφανές, με την χρήση λίγων σειρών κώδικα γίνεται εφικτή η δυνατότητα αλληλεπίδρασης με ένα μενού πλοήγησης μέσω φωνητικών εντολών, του οποίου το περιεχόμενο μπορεί να τροποποιηθεί έτσι ώστε να έχει το επιθυμητό βάθος, χωρίς να χρειάζεται να αλλάξουν οι συγκεκριμένες εντολές, παρά μόνο το περιεχόμενο του αντικειμένου menu και του πίνακα pages αντίστοιχα.

Το τρίτο και τελευταίο σενάριο, αφορά την περίπτωση κατά την οποία ο χρήστης αξιοποιεί μία εντολή η οποία δεν υπάρχει. Παρακάτω, δίνεται το κομμάτι κώδικα που αντιστοιχεί σε αυτή την περίπτωση.

```
} else {  
  speech.text = "No command with name: " + transcript + ". Please try again";  
  window.speechSynthesis.speak(speech);  
}
```

Εικόνα 35: Κώδικας για την περίπτωση που ο χρήστης θα εκφωνήσει μία εντολή η οποία δεν υπάρχει

Όπως είναι προφανές, το else που απεικονίζεται παραπάνω αφορά το τελευταίο κομμάτι της δομής ελέγχου που παρουσιάστηκε στις δύο προηγούμενες εικόνες, μέσα στον κώδικα του οποίου σχηματίζεται το κείμενο που θα ενημερώσει τον χρήστη για την μη ύπαρξη της εντολής που αιτήθηκε.

Παραπάνω, μελετήθηκαν τα δύο βασικά από τα τρία βασικά μέρη που κάνουν το φωνητικό μοντέλο αλληλεπίδρασης του menShoes λειτουργικό. Ωστόσο, μεγάλο κομμάτι της λειτουργικότητας

αυτής οφείλεται στην μέθοδο με την οποία ο φωνητικός βοηθός εκφωνεί στον τελικό χρήστη περιεχόμενο το οποίο βρίσκεται στην εκάστοτε σελίδα του ηλεκτρονικού καταστήματος. Ένας σχεδιασμός κατά τον οποίο μέσα στο Javascript αρχείο κάθε σελίδας θα υπάρχει στατικά το περιεχόμενο της είναι προφανές πως δεν αποτελεί μία λύση η οποία να χαρακτηρίζεται από εύκολη διαχείριση, αλλά και ικανότητα επεκτασιμότητας. Επιπλέον, μία τέτοια πρακτική θα είχε δυσκολίες και σε περιπτώσεις όπως τα προϊόντα στο καλάθι αγορών ενός ηλεκτρονικού καταστήματος, τα οποία δεν είναι ίδια σε κάθε χρήστη και συνεπώς το στατικό κείμενο δεν θα αποτελούσε μία επιλογή. Έτσι λοιπόν, σχηματίζεται η έντονη ανάγκη για ένα σχεδιασμό ο οποίος θα επιτρέπει ένα δυναμικό τρόπο λειτουργίας, κάνοντας εφικτή κάθε φορά την ανάκτηση του τρέχοντος περιεχομένου μίας σελίδας. Κάτι τέτοιο, επιτρέπει την λειτουργία του φωνητικού βοηθού χωρίς την ανάγκη για αλλαγές στο εκάστοτε Javascript αρχείο κάθε φορά που τροποποιείται, προστίθεται ή αντίστοιχα αφαιρείται περιεχόμενο από μία σελίδα. Για την ικανοποίηση της συγκεκριμένης ανάγκης, χρησιμοποιήθηκαν σε κάθε HTML σελίδα συγκεκριμένες κλάσεις στα στοιχεία που περιλαμβάνουν περιεχόμενο το οποίο είναι επιθυμητό να εκφωνηθεί στον χρήστη. Συγκεκριμένα, στην υποενότητα με τις τεχνολογίες που χρησιμοποιήθηκαν για την υλοποίηση του proof of concept, παρατέθηκε η εικόνα 30, περιεχόμενο της οποίας αποτελεί η σύνταξη του framework Django που χρησιμοποιείται για την τύπωση ενός ή περισσότερων προϊόντων μίας κατηγορίας με δυναμικό τρόπο. Εάν παρατηρήσει κανείς προσεκτικά το περιεχόμενο των HTML ετικετών της συγκεκριμένης εικόνας, θα δει την χρήση ενός αριθμού κλάσεων σε κάθε στοιχείο. Παρακάτω, φαίνονται οι κλάσεις αυτές που περιέχονται στις HTML σειρές κώδικα της εικόνας 30.

```
<h5 class="card-title name">{{ pr.name }}</h5>
<p class="card-text price blue-price"><b>{{ pr.price }}$</b></p>
<a href="{% url 'product_detail' pr.category pr.pk %}"
  class="btn btn-primary button-link">
  See product
</a>
<div class="id">Id: {{ pr.id }}</div>
```

Εικόνα 36: Οι κλάσεις *name*, *price* και *id* που χρησιμοποιούνται στα στοιχεία κάθε προϊόντος για την ανάκτηση των περιεχομένων του μέσω Javascript

Παρατηρείται, ότι στις HTML σειρές κώδικα που αφορούν την τύπωση του ονόματος, της τιμής και του αναγνωριστικού κάθε προϊόντος, υπάρχει ένας αριθμός κλάσεων. Μέσα σε αυτές, διακρίνονται η *name*, *price* και *id*. Με την χρήση των συγκεκριμένων κλάσεων, γίνεται εφικτή η ανάκτηση του περιεχομένου αυτών των στοιχείων κάθε προϊόντος μέσω της χρήσης του HTML DOM (Document Object Model). Συγκεκριμένα, όπως είναι γνωστό μέσω του DOM γίνεται εφικτή η αλληλεπίδραση της Javascript με το περιεχόμενο της HTML σελίδας. Κάτι τέτοιο, δίνει την δυνατότητα της ανάκτησης του επιθυμητού κάθε φορά περιεχομένου έτσι ώστε αυτό να εκφωνηθεί κατάλληλα στο χρήστη. Για παράδειγμα, θεωρώντας ότι ένας χρήστης βρίσκεται στην σελίδα με τα προϊόντα της κατηγορίας *sneakers* και κάνει χρήση της εντολής

“Read all”, τότε ο κώδικας που θα εκτελεστεί από το Javascript αρχείο σε αυτή την περίπτωση φαίνεται στην παρακάτω εικόνα.

```
else if (transcript == "read all"){
    var i;
    speech.text = '';
    for (i = 0; i < limit; i++){
        if(i == limit - 1){
            speech.text += "Product name: " + document.getElementsByClassName("name")[i].innerHTML + ". " + document.getElementsByClassName("id")[i].innerHTML + ": You hav
        } else {
            speech.text += "Product name: " + document.getElementsByClassName("name")[i].innerHTML + ". " + document.getElementsByClassName("id")[i].innerHTML + ": ";
        }
    }
    window.speechSynthesis.speak(speech);
}
```

Εικόνα 37: Κώδικας που αντιστοιχεί στην εντολή Read all ο οποίος θα έχει ως αποτέλεσμα την εκφώνηση του ονόματος και της τιμής κάθε προϊόντος από τα αριστερά προς τα δεξιά

Όπως προαναφέρθηκε στην υποενότητα με την παρουσίαση του menShoes, με την χρήση της εντολής “Read all” θα διαβαστεί ο τίτλος και το αναγνωριστικό κάθε προϊόντος σειριακά από αριστερά προς τα δεξιά. Όπως φαίνεται από την παραπάνω εικόνα, κάτι τέτοιο είναι εφικτό μόνο με την χρήση δύο κλάσεων στα αντίστοιχα HTML στοιχεία σε συνδυασμό με λίγες σειρές Javascript κώδικα ο οποίος εκμεταλλεύεται τις δυνατότητες που του δίνονται μέσω του DOM. Συγκεκριμένα, αφού κάθε προϊόν της κατηγορίας sneakers διαθέτει στο HTML στοιχείο που αφορά το όνομα του την κλάση name και σε αυτό που αφορά το αναγνωριστικό του την κλάση id, τότε είναι εφικτή η πρόσβαση στο περιεχόμενο αυτών των στοιχείων με σύνταξη πίνακα όπου ο τίτλος και το αναγνωριστικό του πρώτου προϊόντος βρίσκονται στην θέση μηδέν. Σημαντικό ρόλο για την σωστή λειτουργία του παραπάνω κώδικα έχει η χρήση της μεταβλητής limit. Η συγκεκριμένη καθολική μεταβλητή έχει ως περιεχόμενο της τον αριθμό των προϊόντων της σελίδας, ο οποίος μπορεί να ανακτηθεί με μία εντολή όπως η παρακάτω:

```
var limit = document.getElementsByClassName("name").length;
```

Εικόνα 38: Κώδικας για την ανάκτηση του αριθμού των στοιχείων που περιλαμβάνουν την κλάση name ο οποίος αντιπροσωπεύει το σύνολο των προϊόντων της σελίδας

Με την χρήση της παραπάνω εντολής, επιστρέφεται ο αριθμός των HTML στοιχείων που περιέχουν την κλάση name, ο οποίος είναι προφανές πως ισούται με τα διαθέσιμα προϊόντα που έχει κάθε φορά η σελίδα. Με παρόμοιο τρόπο, υλοποιούνται και οι λειτουργίες των υπόλοιπων εντολών που αιτείται ένας χρήστης κατά την πλοήγηση του στο menShoes. Για παράδειγμα, για την επιλογή ενός προϊόντος με την εντολή “Select ID”, αφού ο χρήστης δώσει το επιθυμητό αναγνωριστικό, θα γίνει αναζήτηση του με την χρήση της επαναληπτικής δομής for. Εφόσον αυτό βρεθεί, θα αποθηκευτεί σε μία καθολική μεταβλητή ο αριθμός επανάληψης κατά τον οποίο βρέθηκε, αφού ρόλος της συγκεκριμένης μεταβλητής είναι να έχει αποθηκευμένο τον αριθμό που αντιπροσωπεύει την θέση στην οποία βρίσκεται το προϊόν που είναι την τρέχουσα χρονική στιγμή επιλεγμένο. Έτσι, με την χρήση στην συνέχεια εντολών όπως είναι η “Read

product”, θα ανακτηθούν τα περιεχόμενα του συγκεκριμένου προϊόντος μέσω των δυνατοτήτων που παρέχει το HTML DOM, όπως ακριβώς έγινε και στην εικόνα 37, με την διαφορά ότι μέσα στις αγκύλες του πίνακα θα χρησιμοποιηθεί η τιμή της καθολικής μεταβλητής. Να σημειωθεί, πως παρόμοια λογική χρησιμοποιείται και στην περίπτωση της φόρμας επικοινωνίας, με την διαφορά ότι η τιμή της αντίστοιχης καθολικής μεταβλητής δεν δείχνει την θέση του επιλεγμένου προϊόντος, αλλά του πεδίου εισαγωγής. Για παράδειγμα, κάθε φορά που ο χρήστης χρησιμοποιεί την εντολή Next, η αντίστοιχη καθολική μεταβλητή θα αυξηθεί κατά ένα έχοντας ως αποτέλεσμα μέσω της χρήσης της λειτουργίες όπως η ανάγνωση αλλά και η συμπλήρωση του επόμενου πεδίου εισαγωγής. Καταλαβαίνει κανείς λοιπόν το μεγάλο πλεονέκτημα που δίνουν οι τεχνικές χειρισμού του HTML DOM για την δημιουργία ενός φωνητικού μοντέλου το οποίο θα λειτουργεί με δυναμικό τρόπο, χωρίς να υπάρχει η ανάγκη διαχείρισης στατικού περιεχομένου της εκάστοτε σελίδας εντός του αντίστοιχου Javascript αρχείου.

Έγινε λόγος αρκετές φορές για τα οφέλη που μπορεί φέρει σε ένα φωνητικό μοντέλο αλληλεπίδρασης η ασύγχρονη ανταλλαγή δεδομένων με τον εξυπηρετητή, χωρίς την ανάγκη για ανανέωση της εκάστοτε σελίδας. Ένα από αυτά, αφορά την δυνατότητα ανάκτησης πληροφοριών ενός προϊόντος χωρίς να χρειάζεται ο χρήστης να αλλάξει σελίδα για την εύρεση τους. Συγκεκριμένα, στην υποενότητα με τις τεχνολογίες που χρησιμοποιήθηκαν, αναφέρθηκε η συνάρτηση ajax της βιβλιοθήκης jQuery. Παρακάτω, φαίνεται η χρήση της συγκεκριμένης μεθόδου για την ανάκτηση της περιγραφής ενός προϊόντος όταν ο χρήστης βρίσκεται στην σελίδα με τα προϊόντα μιας κατηγορίας.

```
else if (transcript == "description"){
$.ajax(
{
type:"GET",
url: "/prinfo/",
data:{
prname: name
},
success: function(data)
{
speech.text = data;
window.speechSynthesis.speak(speech);
}
})
}
```

Εικόνα 39: Χρήση της συνάρτησης ajax για την ανάκτηση της περιγραφής ενός προϊόντος από την βάση δεδομένων και στην συνέχεια την εκφώνηση της από τον φωνητικό βοηθό στον χρήστη

Με την χρήση του παραπάνω κώδικα, θα πραγματοποιηθεί ένα Get αίτημα στον εξυπηρετητή ο οποίος θα ανακατευθύνει τις παραμέτρους που περιλαμβάνονται σε αυτό στην κατάλληλη συνάρτηση. Έπειτα, αφού η ίδια ανακτήσει την περιγραφή του ζητούμενου προϊόντος από την βάση δεδομένων θα την επιστρέψει, με τελικό αποτέλεσμα την εκφώνηση της από τον φωνητικό βοηθό. Να σημειωθεί, ότι η συνάρτηση ajax χρησιμοποιείται με παρόμοιο τρόπο και για την περίπτωση της τοποθέτησης, διαγραφής, αλλά και τροποποίησης της ποσότητας ενός προϊόντος.

Στην συγκεκριμένη υποενότητα, αναλύθηκαν τα βασικά συστατικά του φωνητικού μοντέλου αλληλεπίδρασης που δημιουργήθηκε στα πλαίσια της παρούσας πτυχιακής εργασίας. Όπως αποδείχτηκε, η ενσωμάτωση ενός φωνητικού βοηθού σε ένα σύγχρονο ηλεκτρονικό κατάστημα μπορεί να πραγματοποιηθεί με τρεις απλές προσθήκες. Αρχικά, την δημιουργία ενός βασικού αρχείου το οποίο θα περιέχει λειτουργίες οι οποίες απαιτούνται σε κάθε σελίδα του ηλεκτρονικού καταστήματος, όπως είναι οι δηλώσεις για την χρήση του Web Speech API και το περιεχόμενο του μενού πλοήγησης. Έπειτα, ένα Javascript αρχείο για κάθε σελίδα του καταστήματος στην οποία απαιτούνται δυνατότητες αναγνώρισης και σύνθεσης ομιλίας. Περιεχόμενο του συγκεκριμένου αρχείου, θα αποτελούν οι απαιτούμενες συναρτήσεις για κάθε διαθέσιμη εντολή, καθώς επίσης και ένα μπλοκ κώδικα για την διαχείριση των αποτελεσμάτων αναγνώρισης, στόχος του οποίου θα είναι η κλήση της κατάλληλης συνάρτησης ανάλογα με την εκάστοτε είσοδο που δίνει ο χρήστης μέσω της φωνής του αλλά και η δυνατότητα αλληλεπίδρασης του χρήστη με το μενού πλοήγησης. Τέλος, απαραίτητη είναι η προσθήκη κλάσεων στα HTML στοιχεία μιας σελίδας των οποίων το περιεχόμενο είναι επιθυμητό να ανακτάται μέσω της Javascript, με τελικό στόχο την οργανωμένη εκφώνηση του στον χρήστη.

3.3 Προβλήματα

Θα αποτελούσε ουτοπία η έλλειψη προβλημάτων κατά την διαδικασία παραγωγής ενός έργου λογισμικού. Ωστόσο, σε πραγματικά δεδομένα μεγάλο κομμάτι της συγγραφής κώδικα αφορά την αντιμετώπιση προβλημάτων που είτε είναι προφανή, είτε έρχονται στην επιφάνεια μετά την πραγματοποίηση επαναλαμβανόμενων δοκιμών. Έτσι λοιπόν και στην περίπτωση του menShoes, κατά την διάρκεια της υλοποίησης του διαπιστώθηκαν εμπόδια αναφορικά με την χρήση του Web Speech API για τα οποία έπρεπε να γραφεί κατάλληλος κώδικας έτσι ώστε να ξεπεραστούν.

Αρχικά, το proof of concept υλοποιήθηκε με κύριο στόχο την πλήρης συμβατότητα του στο πρόγραμμα περιήγησης Google Chrome σε desktop. Αυτό, λόγω του ότι το συγκεκριμένο πρόγραμμα περιήγησης είναι το πρώτο που υλοποίησε το Web Speech API, δίνει την δυνατότητα χρήσης του και από κινητό, ενώ το γεγονός σύμφωνα με το οποίο είναι το πρώτο πρόγραμμα περιήγησης στο οποίο υλοποιήθηκε έχει ως αποτέλεσμα με το πέρασμα των χρόνων να έχει μεγαλύτερη υποστήριξη και υλικό σχετικά με την χρήση του και τις ιδιαιτερότητες του. Αντίθετα, σε ένα πρόγραμμα περιήγησης το οποίο έχει υλοποιήσει πρόσφατα το Web Speech API είναι δυσκολότερη η εύρεση υλικού σχετικά με τυχόν προβλήματα που μπορεί να προκύψουν μέσω της χρήσης του.

Κατά την πραγματοποίηση δοκιμών του φωνητικού μοντέλου στο menShoes, διαπιστώθηκε μία ιδιαιτερότητα αναφορικά με τις δυνατότητες της σύνθεσης ομιλίας. Συγκεκριμένα, κάθε πρόγραμμα περιήγησης που έχει υλοποιήσει στο κομμάτι του Web Speech API που αφορά την σύνθεση ομιλίας, παρέχει μία πληθώρα διαθέσιμων φωνών οι οποίες μπορούν να αξιοποιηθούν για την εκφώνηση

οποιοδήποτε κειμένου στον τελικό χρήστη. Οι συγκεκριμένες φωνές, εκτός από την γλώσσα που μπορούν να υποστηρίξουν διαθέτουν και ένα ακόμα χαρακτηριστικό το οποίο τις διαχωρίζει σε δύο μέρη. Αυτές που είναι local και αυτές που παρέχονται από το εκάστοτε πρόγραμμα περιήγησης. Σε περίπτωση που ο κώδικας μιας εφαρμογής δεν περιλαμβάνει την δήλωση μίας συγκεκριμένης φωνής, τότε το πρόγραμμα περιήγησης θα επιλέξει μία από προεπιλογή η οποία μπορεί να είναι είτε local είτε όχι. Στην περίπτωση του menShoes, μέσω της πραγματοποίησης δοκιμών στο πρόγραμμα περιήγησης Google Chrome, παρατηρήθηκε κάποια στιγμή η αλλαγή στην φωνή που χρησιμοποιούταν από προεπιλογή, κάτι το οποίο οδήγησε στην εκφώνηση ενός περιορισμένου μεγέθους κειμένου, με αποτέλεσμα η εκφώνηση πολλές φορές να διακόπτεται όταν το κείμενο ξεπερνούσε αυτό το όριο. Μετά από αναζήτηση για την συγκεκριμένη ιδιαιτερότητα, ανακαλύφθηκε ότι ο συγκεκριμένος περιορισμός αφορά φωνές που δεν είναι local και παρέχονται ως επιλογή στο Google Chrome. Παρακάτω, φαίνεται ένα απόσπασμα της απάντησης που δίνει ένας χρήστης στο stackoverflow.com αναφορικά με την συγκεκριμένη ιδιαιτερότητα.

I believe this is an issue with the speech synthesizer chosen. If you choose a *local* voice - one for which the audio can be generated simply by running in-browser / in-OS code on your local computer - then there should be no character limit, unless the voice you chose has such a limit, which doesn't seem likely

For example, on Windows, the two local voices I have are Microsoft David Desktop and Microsoft Zira Desktop. If I select these voices, the whole text gets played, regardless of length.

In contrast, other voices such as those from Google (eg Google US English) are not local; they may need to make a network request to parse the text into audio. Google's speech API seems to have a 200 character or so limit on how much text can be translated at once.

If the character limit is a problem, select a voice which is local, rather than one from Google.

Εικόνα 40: Απάντηση χρήστη στο stackoverflow.com αναφορικά με τον περιορισμό στο μέγεθος του κειμένου εκφώνησης στο Google Chrome όταν μία φωνή δεν είναι local

Με βάση την συγκεκριμένη απάντηση, γράφτηκε κώδικας έτσι ώστε αντί να επιλέγεται μία φωνή από προεπιλογή, να γίνεται αναζήτηση στις διαθέσιμες φωνές και να χρησιμοποιείται μία η οποία να είναι αγγλική και local. Παρακάτω, φαίνεται το κομμάτι κώδικα που γράφτηκε για την επίτευξη της συγκεκριμένης λειτουργικότητας.

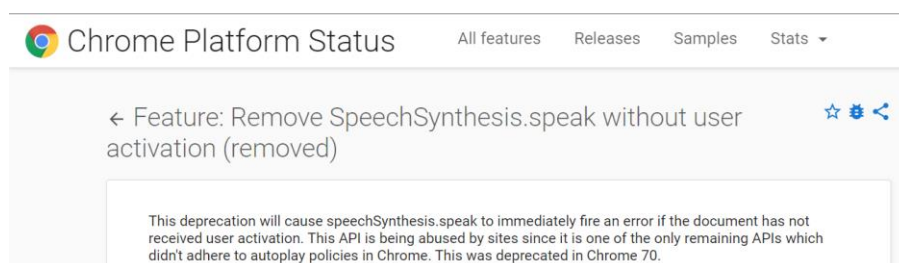
```
window.speechSynthesis.onvoiceschanged = function() {  
  voices = window.speechSynthesis.getVoices();  
  var i;  
  for(i=0; i<voices.length; i++){  
    if(voices[i].lang == 'en-US' && voices[i].localService){  
      speech.voice = voices[i];  
      speech.lang = voices[i].lang;  
      break;  
    }  
  }  
};
```

Εικόνα 41: Κώδικας για την επιλογή μίας φωνής η οποία θα εκφωνεί αγγλικά κείμενα και θα είναι local

Παρατηρείται η χρήση της ιδιότητας localService, η οποία θα επιστρέψει μία Boolean τιμή ανάλογα με το αν η εκάστοτε φωνή προέρχεται από local υπηρεσία σύνθεσης ομιλίας ή όχι. Μέσω του συγκεκριμένου

κώδικα λοιπόν, έγινε ξανά εφικτή η δυνατότητα εκφώνησης ολόκληρου του κειμένου χωρίς να διακόπτεται. Βέβαια, η συγκεκριμένη ιδιαιτερότητα διαπιστώθηκε μόνο στο Google Chrome, αφού οι φωνές που δεν είναι local και παρέχονται από πρόγραμμα περιήγησης Microsoft Edge φαίνεται να εκφωνούν μεγέθη κειμένων αντίστοιχα του menShoes χωρίς πρόβλημα.

Η παραπάνω ιδιαιτερότητα, αποτελεί ένα εμπόδιο το οποίο όπως φάνηκε μπορεί να ξεπεραστεί με την χρήση λίγων γραμμών κώδικα. Το πρόγραμμα περιήγησης Google Chrome ωστόσο, παρουσιάζει μία ακόμη ιδιαιτερότητα για την οποία προς το παρόν δεν υπάρχει λύση. Συγκεκριμένα, το Google Chrome δήλωσε ότι η πλειοψηφία των ιστοσελίδων κάνει κατάχρηση των δυνατοτήτων σύνθεσης ομιλίας με αποτέλεσμα την κατάργηση της δυνατότητας να εκφωνείται οποιοδήποτε κείμενο εάν ο χρήστης πρώτα δεν έχει αλληλεπιδράσει τουλάχιστον μία φορά με την σελίδα όπως για παράδειγμα με το πάτημα ενός κουμπιού. Παρακάτω, φαίνεται η συγκεκριμένη ανακοίνωση.

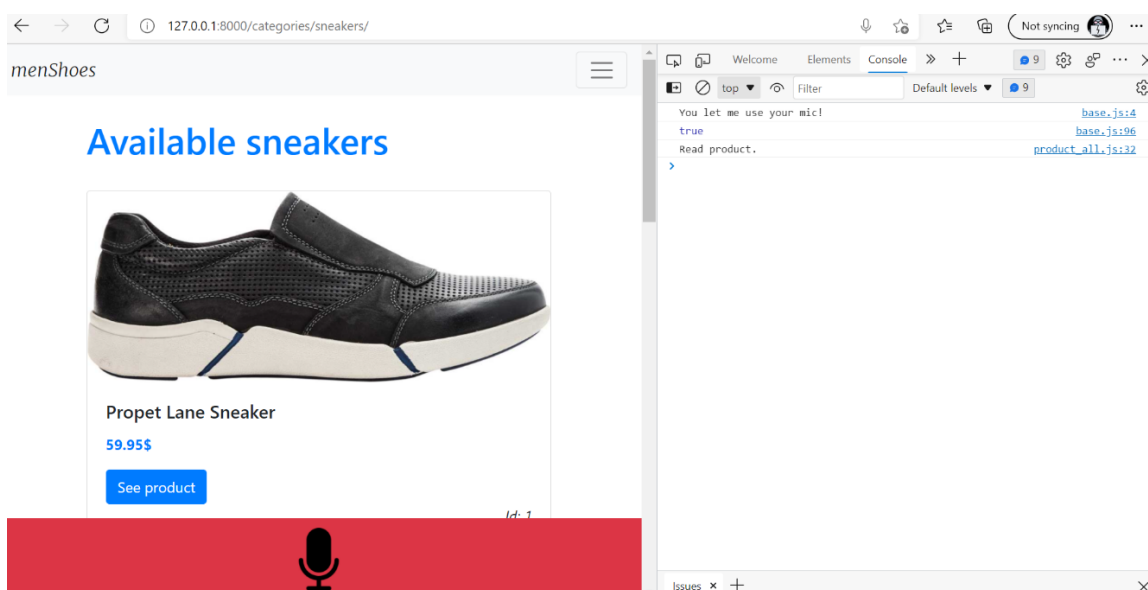


Εικόνα 42: Μήνυμα του Google Chrome αναφορικά με την μη δυνατή χρήση της σύνθεσης ομιλίας εάν δεν έχει αλληλεπιδράσει πρώτα ο χρήστης με την σελίδα

Όπως είναι προφανές, μία τέτοια κίνηση θέτει εμπόδια στους προγραμματιστές που αξιοποιούν τις δυνατότητες της σύνθεσης ομιλίας για την υποβοήθηση ατόμων με αναπηρίες, αφού για παράδειγμα ένας χρήστης ο οποίος αντιμετωπίζει προβλήματα όρασης δεν θα μπορεί να αξιοποιήσει τις δυνατότητες ενός φωνητικό βοηθού εάν πρώτα δεν έχει αλληλεπιδράσει με την σελίδα πατώντας π.χ. ένα κουμπί. Ωστόσο, σε Group της Google που δημιουργήθηκε για την συγκεκριμένη απόφαση αναφέρεται ότι εφόσον έχει πραγματοποιηθεί έστω και μία αλληλεπίδραση με μία σελίδα, τότε εάν ο χρήστης περιηγηθεί σε άλλες σελίδες κάτω από το ίδιο domain δεν θα υπάρχει ο παραπάνω περιορισμός. Μία τέτοια απόφαση δεν φαίνεται να έχει παρθεί στο πρόγραμμα περιήγησης Microsoft Edge, όπου η εκφώνηση ενός κειμένου γίνεται κανονικά ακόμα και αν δεν έχει υπάρξει πρώτα αλληλεπίδραση με την σελίδα.

Στο κεφάλαιο κατά το οποίο αναλύθηκαν οι δυνατότητες του Web Speech API, αναφέρθηκε στην υποενότητα σχετικά με την συμβατότητα του σε διαφορετικά προγράμματα περιήγησης ότι ένας προγραμματιστής πρέπει να λάβει υπόψη του το γεγονός σύμφωνα με το οποίο δεν χρησιμοποιείται παντού η ίδια υπηρεσία αναγνώρισης ομιλίας. Για τον λόγο αυτό, πρέπει να γράψει κώδικα ο οποίος να είναι ικανός να προσαρμόζει το αποτέλεσμα αναγνώρισης που επιστρέφεται κάθε φορά σε μία αυστηρά

καθορισμένη μορφή. Αναφέρθηκε παραπάνω, ότι το menShoes δημιουργήθηκε με στόχο την πλήρης συμβατότητα του στο πρόγραμμα περιήγησης Google Chrome. Ωστόσο, με την ολοκλήρωση του δοκιμάστηκε και στο πρόγραμμα περιήγησης Microsoft Edge. Ενώ με μία πρώτη ματιά φαινόταν αδύνατη η αναγνώριση των εντολών στο συγκεκριμένο πρόγραμμα περιήγησης, μέσω εντολής για την τύπωση των αποτελεσμάτων αναγνώρισης που επιστρέφονται, αποδείχτηκε ότι ενώ το Google Chrome επιστρέφει ένα αποτέλεσμα με όλα πεζά (εκτός από περιπτώσεις όπως π.χ. όνομα) και χωρίς καμία προσθήκη, στην περίπτωση του Microsoft Edge επιστρέφεται με το πρώτο γράμμα κεφαλαίο και την προσθήκη τελείας στο τέλος. Παρακάτω, απεικονίζεται η επιστροφή του αποτελέσματος αναγνώρισης στο πρόγραμμα περιήγησης Microsoft Edge όταν δόθηκε ως είσοδος η εντολή “Read product”.



Εικόνα 43: Η μορφή με την οποία επιστράφηκε από την υπηρεσία αναγνώρισης η λέξη *read all* στο πρόγραμμα περιήγησης Microsoft Edge

Η σύμβαση κατά την οποία όλες οι εντολές στις συνθήκες να είναι με πεζά, σε συνδυασμό με την προσθήκη της τελείας από την πλευρά του Microsoft Edge, είχε ως αποτέλεσμα την μη δυνατή αναγνώριση των εντολών στο συγκεκριμένο πρόγραμμα περιήγησης. Με την συγγραφή ωστόσο λίγων γραμμών κώδικα οι οποίες μετατρέπουν κάθε αποτέλεσμα αναγνώρισης που επιστρέφεται σε πεζά και αφαιρούν την τελεία σε περίπτωση που αυτή εμφανίζεται, τότε το menShoes λειτουργήσε κανονικά και στο πρόγραμμα περιήγησης Microsoft Edge. Καταλαβαίνει κανείς λοιπόν πόσο σημαντική είναι η παρατήρηση της μορφής στην οποία επιστρέφεται ένα αποτέλεσμα, καθώς αυτή μπορεί να έχει καθοριστικό ρόλο για την λειτουργία του φωνητικού μοντέλου αλληλεπίδρασης. Αξίζει να σημειωθεί, ότι η συμπεριφορά του SpeechRecognition στο πρόγραμμα περιήγησης Microsoft Edge τροποποιήθηκε κατά την διάρκεια διεξαγωγής της εργασίας.

Συγκεκριμένα, παρατηρήθηκε ότι αν η ιδιότητα `continuous` είναι `false`, τότε το μοναδικό αποτέλεσμα που πρόκειται να επιστραφεί κάθε φορά καθυστερεί κάποια δευτερόλεπτα, κάτι το οποίο δεν συνηθιζόταν. Επίσης, παρατηρήθηκε πως σε περίπτωση που δεν δοθεί τιμή στην ιδιότητα `lang`, τότε δεν επιστρέφεται κανένα αποτέλεσμα κάτι το οποίο επίσης δεν συνέβαινε. Ωστόσο, κάτι τέτοιο είναι αναμενόμενο, αφού όπως έχει προαναφερθεί η συμπεριφορά ενός πειραματικού API είναι πιθανό να τροποποιείται στην διάρκεια του χρόνου.

4. Συμπεράσματα και μελλοντική έρευνα

Η δημιουργία του συγκεκριμένου proof of concept, αποσκοπεί στη μελέτη του κατά πόσο είναι εφικτή η αξιοποίηση τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας σε ηλεκτρονικά καταστήματα και κατά συνέπεια σε εφαρμογές του διαδικτύου με στόχο την υποβοήθηση χρηστών με προβλήματα όρασης. Αποδείχτηκε ότι είναι εφικτή η πραγματοποίηση βασικών λειτουργιών εξ ολοκλήρου μέσω φωνητικών εντολών, αποδεσμεύοντας ένα χρήστη από την ανάγκη του γραφικού περιβάλλοντος. Με την χρήση ενός εργαλείου όπως το Web Speech API, γίνεται εφικτή η ενσωμάτωση σε διαδικτυακές εφαρμογές τεχνολογίες που χρησιμοποιούν σήμερα φωνητικοί βοηθοί όπως η Alexa της Amazon, ικανές να εξαλείψουν την ανάγκη για την χρήση επιπλέον εξοπλισμού και εκπαίδευσης σε χρήστες που αντιμετωπίζουν προβλήματα όρασης. Ένα εργαλείο όπως ο αναγνώστης οθόνης επιχειρεί να εξισώσει την εμπειρία πλοήγησης στο διαδίκτυο, ενώ η χρήση τεχνολογιών όπως η αναγνώριση και η σύνθεση ομιλίας έχουν την ικανότητα να την βελτιώσουν για τους χρήστες που έχουν χαμηλή ή μηδενική οπτική απόδοση, προσθέτοντας έναν επιλεκτικό, πέραν του σειριακού τρόπου ανάκτησης της πληροφορίας αλλά και μία προσωποποιημένη επικοινωνία. Αποτελεί σημαντικό πλεονέκτημα η πληροφορία να μπορεί να έρθει σε έναν χρήστη και όχι να πρέπει να κατευθυνθεί ο ίδιος προς αυτή. Με την χρήση ενός φωνητικού μοντέλου αλληλεπίδρασης όπως του menShoes, ένα ηλεκτρονικό κατάστημα αποκτά την δυνατότητα προσέγγισης ενός ευρύτερου κοινού, χωρίς την πραγματοποίηση ριζικών αλλαγών στην υπάρχουσα τεχνική υποδομή του. Παράλληλα, συμβάλει στην οικοδόμηση ενός διαδικτύου για όλους, ενισχύοντας το δικαίωμα στην πληροφορία η οποία σήμερα έχει καθοριστικό ρόλο για την ένταξη αλλά και συμβίωση οποιουδήποτε ατόμου στην κοινωνία. Τέλος, σημαντικό είναι πως ο σχεδιασμός του συγκεκριμένου φωνητικού μοντέλου αλληλεπίδρασης επιτρέπει την παραμετροποίηση του έτσι ώστε να χρησιμοποιείται και σε άλλου είδους εφαρμογές όπως είναι τα blogs, ειδησεογραφικές σελίδες, εφαρμογές παραγγελίας φαγητού και η λίστα συνεχίζεται.

Το W3C οραματίστηκε ένα διαδίκτυο με δυνατότητες ομιλίας και ένα από τα πρώτα λιθαράκια για την οικοδόμηση του αποτελεί το Web Speech API. Μέσω της χρήσης καθαρής Javascript σε συνδυασμό με κατάλληλες τεχνικές διαχείρισης DOM και AJAX, δίνεται η δυνατότητα δημιουργίας φωνητικών διεπαφών οι οποίες λειτουργούν δυναμικά και είναι ικανές να εξυπηρετήσουν μεγάλο εύρος σεναρίων με ένα από αυτά να αποτελεί η υποβοήθηση ατόμων με προβλήματα όρασης. Η ασύγχρονη event driven λογική του Web Speech API, σε συνδυασμό με την παροχή πληθώρας interfaces, κάνει εφικτή την δημιουργία ενός φωνητικού μοντέλου αλληλεπίδρασης το οποίο εδραιώνει μία νέα μορφή επικοινωνίας μεταξύ μηχανής και ανθρώπου. Το γεγονός σύμφωνα με το οποίο θεωρείται πειραματικό, έδωσε ένα παραπάνω κίνητρο για την αξιοποίηση των δυνατοτήτων του αναφορικά με την δημιουργία του proof of

concept, έτσι ώστε το τελικό αποτέλεσμα να αποτελέσει πηγή πληροφοριών στους δημιουργούς του, με τελικό στόχο την συνεχή βελτίωση του. Παράλληλα, σημαντικό κίνητρο αποτέλεσε η πολυδιάστατη δομή του, αφού δεν δίνεται απλώς η δυνατότητα αναγνώρισης και σύνθεσης ομιλίας με μονόδρομο τρόπο. Αντίθετα, προσφέρεται η ικανότητα της παραμετροποίησης αυτών των τεχνολογιών έτσι ώστε να καλύπτονται διαφορετικά σενάρια χρήσης, καθορίζοντας την εκτέλεση διαφορετικών μπλοκ κώδικα ανάλογα με την κατάσταση της εφαρμογής με οργανωμένο αλλά και αποδοτικό τρόπο, ο οποίος είναι πλήρως συμβατός με την προγραμματιστική λογική που χρησιμοποιείται στην πλευρά του πελάτη (client side).

Μέσω της δημιουργίας του συγκεκριμένου proof of concept, αποδείχτηκε ότι με την ενσωμάτωση σε ένα ηλεκτρονικό κατάστημα τεχνολογιών αναγνώρισης και σύνθεσης ομιλίας μπορούν να επωφεληθούν άτομα με προβλήματα όρασης. Ωστόσο, όπως είναι γνωστό μέσω ενός proof of concept δεν υλοποιείται ολοκληρωτικά μία ιδέα αλλά ένα κομμάτι της, με στόχο την διαπίστωση αν η εκάστοτε υπόθεση ισχύει ή όχι. Παρόμοια και στην περίπτωση του menShoes, για την απόδειξη της υπόθεσης δημιουργήθηκε ένα φωνητικό μοντέλο αλληλεπίδρασης το οποίο δίνει σε έναν χρήστη την δυνατότητα αλληλεπίδρασης με βασικά συστατικά ενός ηλεκτρονικού καταστήματος και όχι με όλα. Έτσι λοιπόν, ενώ η παρούσα εργασία θέτει τα θεμέλια για ένα ηλεκτρονικό κατάστημα με δυνατότητες αναγνώρισης και σύνθεσης ομιλίας, απαραίτητη είναι η μελλοντική έρευνα για την εφαρμογή των μεθόδων που έχουν αναλυθεί έτσι ώστε αυτές να περιλαμβάνουν όλα τα συστατικά από την εύρεση ενός προϊόντος μέχρι και την δυνατότητα αγοράς του. Συνεπώς, στα μελλοντικά βήματα αρχικά περιλαμβάνεται η δοκιμή του συγκεκριμένου proof of concept από άτομα με προβλήματα όρασης με στόχο την ανατροφοδότηση, έτσι ώστε στην συνέχεια να βελτιωθεί με βάση τις ανάγκες – απαιτήσεις που θα διαπιστωθούν. Έπειτα, απαραίτητη είναι η αλληλεπίδραση με άλλα εργαλεία αναγνώρισης και σύνθεσης ομιλίας για την δημιουργία του φωνητικού μοντέλου αλληλεπίδρασης, αφού το Web Speech API βρίσκεται ακόμα σε πειραματικό στάδιο και συνεπώς η συμβατότητα του στα προγράμματα περιήγησης είναι περιορισμένη, ενώ παράλληλα η συμπεριφορά του μπορεί να αλλάζει ανάλογα με την υλοποίηση του εκάστοτε προγράμματος περιήγησης. Συνεπώς, απαιτείται μία πιο σταθερή λύση, η οποία μπορεί να προσφερθεί από εργαλεία τα οποία παρέχονται από κολοσσούς στον χώρο της τεχνολογίας όπως η Google, Microsoft αλλά και IBM. Επιπλέον, θα δοκιμαστεί ένα μοντέλο αλληλεπίδρασης κατά το οποίο απαιτείται η ενεργοποίηση του μικροφώνου μόνο μία φορά στην αρχή, δίνοντας έτσι στην εφαρμογή δυνατότητες για συνεχή αναγνώριση την οποία ο χρήστης θα μπορεί να απενεργοποιεί όταν επιθυμεί μέσω αντίστοιχης εντολής. Παράλληλα, θα μελετηθεί η δυνατότητα υπογράμμισης του εκάστοτε κειμένου που εκφωνείται όταν αυτό είναι ορατό στην τρέχουσα σελίδα, αφού αυτή είναι μία δυνατότητα που δίνεται σε λογισμικά όπως ο αναγνώστης οθόνης και συνεπώς μπορεί να ωφελήσει άτομα που έχουν μειωμένη, αλλά όχι μηδενική όραση. Επιπρόσθετα, θα υλοποιηθούν εντολές για την δυνατότητα ρύθμισης από τον χρήστη της φωνής του φωνητικού βοηθού σχετικά με παραμέτρους

όπως είναι η ταχύτητα εκφώνησης αλλά και η ένταση. Εκτός από αυτές τις εντολές, αυτονόητη είναι η προσθήκη και άλλων για την δυνατότητα αλληλεπίδρασης με όλα τα απαραίτητα συστατικά ενός ηλεκτρονικού καταστήματος όπως είναι οι αξιολογήσεις, οι παραλλαγές, η δημιουργία και διαχείριση του προσωπικού λογαριασμού, η δυνατότητα φιλτραρίσματος και ταξινόμησης προϊόντων αλλά και η διαδικασία του checkout. Σημαντικές προσθήκες θα ενσωματωθούν και σε ήδη υπάρχουσες εντολές όπως η δυνατότητα εκφώνησης του πότε ένα προϊόν έχει έκπτωση, καθώς επίσης και η εκφώνηση μηνύματος λάθους όταν ο χρήστης έχει συμπληρώσει με μη κατάλληλο τρόπο ένα ή περισσότερα πεδία εισαγωγής. Σημαντικό κομμάτι θα αποτελέσει επιπλέον και η μελέτη για το πως μπορεί ένα τέτοιο φωνητικό μοντέλο αλληλεπίδρασης να υποστηρίξει περισσότερες από μία γλώσσες. Τέλος, θα μελετηθεί η δυνατότητα μέσω μίας σελίδας να γίνεται εφικτή η ανάκτηση και στην συνέχεια εκφώνηση κειμένων που περιλαμβάνονται συνήθως σε σελίδες που αναφέρονται στο footer, όπως για παράδειγμα σχετικά με εμάς, πολιτικές απορρήτου, χρόνοι παράδοσης, ενώ παράλληλα θα μπορούν να εκφωνούνται και κείμενα όπως τρέχουσες προσφορές. Κάτι τέτοιο, γίνεται εφικτό μέσω AJAX και κατάλληλου χειρισμού DOM, αφού όπως και στην περίπτωση με την περιγραφή ενός προϊόντος έτσι και εδώ μπορεί να ανακτάται και στην συνέχεια να εκφωνείται ένα κείμενο την φορά, με αποτέλεσμα ένας χρήστης να μπορεί να έρθει σε επαφή με μεγάλο κομμάτι στατικού περιεχόμενου του καταστήματος, χωρίς να χρειάζεται πλοηγηθεί σε πληθώρα σελίδων για την εύρεση αυτού.

Το πεδίο της τεχνητής νοημοσύνης, αποτελεί αδιαμφισβήτητο το μέσο για την ανέλιξη του τρόπου ζωής από κοινωνική, οικονομική και ανθρωπιστική σκοπιά. Συνεπώς, η χρήση των δυνατοτήτων της ως μαύρο κουτί μέσω της αξιοποίησης εργαλείων όπως το Web Speech API, δίνει την ικανότητα δημιουργίας λογισμικού ικανού να συμβάλει στην συγκρότηση ενός διαδικτύου προσβάσιμο από όλους. Αποτελεί λοιπόν ευθύνη των προγραμματιστών του διαδικτύου με την πραγματοποίηση του ψηφιακού μετασχηματισμού να περάσουν στην απέναντι όχθη όλες τις κοινωνικές ομάδες δίχως αποκλεισμούς. Η πρόσβαση στην πληροφορία είναι δικαίωμα και πρέπει να συνεχίσει να είναι ανεξάρτητα της μορφής με την οποία απεικονίζεται.

Βιβλιογραφία

- Y. Borodin et al, More than Meets the Eye: A Survey of Screen-Reader Browsing Strategies, Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A), 2010
- F. Algarni, Y. Cheung, V. Lee, An Intelligent Voice-Based eMarketplace for Visually Impaired People, Journal of Software Engineering and Applications, Vol. 6, p 91-96, 2013
- M. S. Kandhari, F. Zulkemine, H. Isah, A Voice Controlled E-Commerce Web Application, 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018
- K. Mohadikar, R. Nawkhare, ECOMMERCE BASED ONLINE SHOPPING FOR VISUALLY IMPAIRED PEOPLE USING SPEECH RECOGNITION, International Journal of Development Research, Vol. 07, p 14581-14584, 2017
- V. Iyer, S. Sheth, K. Shah, K. Devadkar, Virtual assistant for the visually impaired, 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020
- S. Rallabhandy, S. Rodda, KeyBoard-Less Online Shopping for the Visually Impaired Using Natural Language Processing and Face Recognition Mechanism, Smart Intelligent Computing and Applications, p 253-260, 2020
- M. N. Talib et al, E-Business Access for Blinds: A Semantic Approach, 2009 International Conference on E-Business and Information System Security, 2009
- C. Power et al, Guidelines are only half of the story: accessibility problems encountered by blind users on the web, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, p 433–442, 2012
- A. Libby, Introducing the HTML5 Web Speech API: Your Practical Introduction to Adding Browser-Based Speech Capabilities to your Websites and Online Applications, 2020
- J. Adorf, Web Speech API, KTH Royal Institute of Technology, Stockholm, 2013
- A. Pradhan, K. Mehta, L. Findlater, "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p 1-13, 2018

Δικτυογραφία

- Blindness and vision impairment. (2021). World Health Organization. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Digital voice assistants in use to triple to 8 billion by 2023, driven by smart home devices. (2018). JUNIPER RESEARCH. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.juniperresearch.com/press/digital-voice-assistants-in-use-to-8-million-2023>
- Django The web framework for perfectionists with deadlines. (χ.χ.). Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.djangoproject.com/>
- Bootstrap. (χ.χ.). Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://getbootstrap.com/>

Feature: Remove SpeechSynthesis.speak without user activation (removed). (2018). Chrome Platform Status. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.chromestatus.com/feature/5687444770914304>

HTML Speech Incubator Group Final Report. (2011). W3C Incubator Group Report. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.w3.org/2005/Incubator/htmlspeech/XGR-htmlspeech-20111206/>

W3C Web Accessibility Initiative WAI. (χ.χ.). Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.w3.org/WAI/>

MediaDevices.getUserMedia(). (χ.χ.). MDN Web Docs. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://developer.mozilla.org/en-US/docs/Web/API/MediaDevices/getUserMedia>

Speech API Community Group. (2012). W3C COMMUNITY & BUSINESS GROUPS. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.w3.org/community/speech-api/>

Speech synthesizer on chrome only play the first 200 characters. (2020). stackoverflow. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://stackoverflow.com/questions/63996511/speech-synthesizer-on-chrome-only-play-the-first-200-characters>

W3C HTML Speech Incubator Group. (χ.χ.). W3C Incubator Activity. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.w3.org/2005/Incubator/htmlspeech/>

The WebAIM Million: An annual accessibility analysis of the top 1.000.000 home pages. (2021). WebAIM web accessibility in mind. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://webaim.org/projects/million/>

W3schools. (χ.χ.). Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://www.w3schools.com/>

Web Speech API. (χ.χ.). MDN Web Docs. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API

Web Speech API: Draft Community Group Report. (χ.χ.). W3C Community Group Draft Report. Ημερομηνία ανάκτησης 26 Σεπτεμβρίου, από: <https://wicg.github.io/speech-api/>