

Εφαρμογές σοβαρών παιχνιδιών στη συλλογή και συναισθηματική επισημείωση δειγμάτων ομιλίας

Ματσουλιάδης Λάζαρος^{1*}, Χαδέλλης Λουκάς², Δημούλας Χαράλαμπος³

^{1*} matsouliadis.l@gmail.com (υπεύθυνου)

² lukehdlls@gmail.com

³ c.a.dimoulas@gmail.com

ΠΕΡΙΛΗΨΗ

Στον τομέα των σοβαρών παιχνιδιών, η ενσωμάτωση της τεχνολογίας Αναγνώρισης Συναισθημάτων Ομιλίας (Speech Emotion Recognition – SER) παρουσιάζει μια πολλά υποσχόμενη οδό για τη βελτίωση της συλλογής δεδομένων και την ενίσχυση της διαδραστικότητας των χρηστών. Αυτή η εργασία εξερευνά τις δυνατότητες των σοβαρών παιχνιδιών με δυνατότητα SER να συλλέγουν πλούσια δεδομένα ομιλίας από τους παίκτες, συμβάλλοντας στην πρόοδο τόσο της τεχνολογίας SER όσο και της ανάπτυξης σοβαρών παιχνιδιών.

Η έρευνα που παρουσιάζεται σε αυτή την εργασία επικεντρώνεται στον σχεδιασμό και την ανάπτυξη δύο διαφορετικών εφαρμογών σοβαρών παιχνιδιών ειδικά σχεδιασμένων για τη συλλογή δεδομένων χροιάς φωνής. Αυτά τα παιχνίδια, που δημιουργήθηκαν με χρήση της Unreal Engine 5, παρέχουν εμπειρίες παιχνιδιού, ενώ ταυτόχρονα καταγράφουν τις συναισθηματικές εκφράσεις των παικτών μέσω της ομιλίας τους.

Η ενσωμάτωση SER σε σοβαρά παιχνίδια προσφέρει πληθώρα πλεονεκτημάτων. Πρώτον, διευκολύνει τη συλλογή ποικίλων δεδομένων ομιλίας, τα οποία είναι απαραίτητα για την ανάπτυξη της τεχνολογίας SER. Δεύτερον, βελτιώνει την εμπειρία του χρήστη επιτρέποντας στα παιχνίδια να προσαρμόζονται και να ανταποκρίνονται στις συναισθηματικές καταστάσεις των παικτών, δημιουργώντας πιο εξατομικευμένη εμπειρία για κάθε χρήστη.

Πέρα από την ενσωμάτωση SER, η εργασία εστιάζει και στην υλοποίηση προηγμένης τεχνολογίας ήχου, αξιοποιώντας χωρικό ήχο (spatial sound), φαινόμενο Ντόπλερ (Doppler Effect), την διφωνικότητα του ήχου (Binaural) και μηχανισμούς τρισδιάστατης ηχητικής τοποθέτησης (3D sound). Η ενσωμάτωση αυτών των στοιχείων συμβάλλει στην υλοποίηση ρεαλιστικών ηχητικών περιβαλλόντων, ενισχύοντας την αίσθηση παρουσίας και την συναισθηματική σύνδεση των παικτών με το παιχνίδι.

Αυτή η εργασία αποδεικνύει την αποτελεσματικότητα των σοβαρών παιχνιδιών με την δυνατότητα SER ως εργαλείο για τη συλλογή δεδομένων και την ενίσχυση της διαδραστικότητας με τους χρήστες. Οι εφαρμογές παιχνιδιών που αναπτύχθηκαν και η υλοποίηση των τεχνολογιών ήχου ανοίγουν τον δρόμο για το μέλλον του σχεδιασμού παιχνιδιών, όπου η συναισθηματική νοημοσύνη και η ρεαλιστική ηχητική εμπειρία διαδραματίζουν καθοριστικό ρόλο στη δημιουργία παιχνιδιών με βαθύτερη απήχηση στους χρήστες και αισθητή επίδραση στην ψυχολογία και τη συμπεριφορά τους.

Applications of serious games in the collection and emotional labeling of speech samples

ABSTRACT

In the field of serious games, the integration of Speech Emotion Recognition (SER) technology presents a promising avenue for improving data collection and enhancing user interactivity. This research explores the potential of serious games integrated with SER technology to gather rich speech data from players, contributing to advancements in both SER technology and the development of serious games.

The research presented in this paper focuses on the design and development of two distinct serious game applications specifically tailored for collecting voice tone data. These games, created using Unreal Engine 5, provide engaging gameplay experiences while simultaneously capturing players' emotional expressions through their speech.

The integration of SER into serious games offers a multitude of advantages. Firstly, it facilitates the collection of diverse speech data, which is essential for the advancement of SER technology. Secondly, it enhances the user experience by allowing games to adapt and respond to players' emotional states, creating a more personalized experience for each user.

Beyond SER integration, the research also focuses on implementing advanced audio technology, utilizing spatial audio, Doppler, binaural, and 3D sound positioning mechanisms. The incorporation of these elements contributes to the creation of realistic sound environments, enhancing the sense of presence and emotional connection of players to the game.

This research demonstrates the effectiveness of serious games integrated with SER technology as a tool for data collection and enhancing user interactivity. The developed game applications and the implementation of audio technologies pave the way for the future of game design, where emotional intelligence and realistic audio experiences play a pivotal role in creating video games with a deeper impact on users and a noticeable effect on their psychology and behavior.

Εισαγωγή

Η διεπαφή ανθρώπου-υπολογιστή έχει εξελιχθεί ραγδαία χάρη στην τεχνητή νοημοσύνη, με την αναγνώριση συναισθημάτων να αποτελεί κεντρικό άξονα. Η εργασία αυτή εξερευνά τη χρήση σοβαρών παιχνιδιών ως εργαλείο για τη συλλογή δεδομένων φωνής και την ανάπτυξη μοντέλων αναγνώρισης συναισθημάτων.

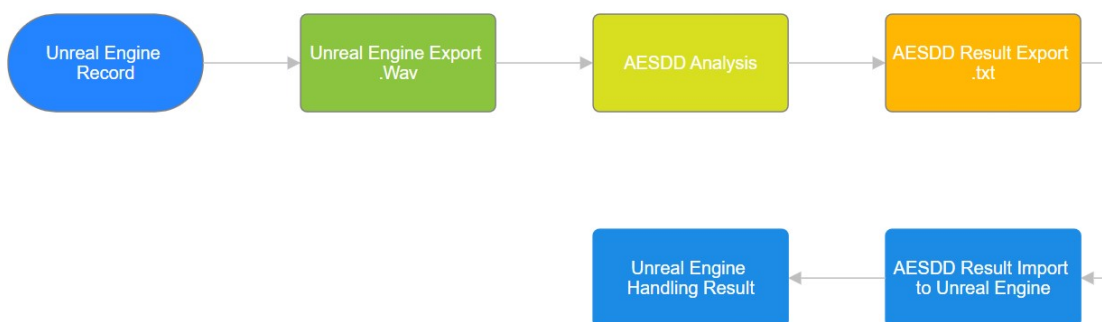
Συγκεκριμένα, σχεδιάστηκαν και υλοποιήθηκαν δύο παιχνίδια, το “Silent Kingdom” και το “Job Interview Simulator”, με στόχο να καταγράψουν μια ποικιλία συναισθηματικών εκφράσεων, όπως χαρά, θυμός, φόβος, λύπη και αηδία. Το “Silent Kingdom” είναι δομημένο έτσι ώστε να ζητάει την εκφορά όλων των συναισθημάτων σε μια ήδη προκαθορισμένη ιστορία που πάντα εξελίσσεται με τον ίδιο τρόπο. Σε αντίθεση το “Job Interview Simulator” δεν υπάρχει κάποια απαίτηση

εκφοράς συγκεκριμένου συναισθήματος, έτσι εξελίσσεται βασισμένο στις εκφορές του παίκτη.

Η εργασία αυτή αποτελεί ένα βήμα προς την κατεύθυνση της ανάπτυξης τεχνολογιών που μπορούν να κατανοούν και να ανταποκρίνονται στα ανθρώπινα συναισθήματα, ανοίγοντας νέους ορίζοντες για την αλληλεπίδραση ανθρώπου – υπολογιστή [1][2].

1. Ανάλυση συναισθημάτων ομιλίας μέσω Unreal Engine

Η ανάλυση της ομιλίας γίνεται μέσω της βάσης δεδομένων Acted Emotional Speech Dynamic Database (AESDD) μέσω Cloud, μετά από εξαγωγή των αρχείων .wav από την Unreal Engine, μετά την ανάλυση εισάγονται τα αποτελέσματα στην Unreal Engine που φέρουν την συνέχεια στην ιστορία στα δύο παιχνίδια σε πραγματικό χρόνο, όπως φαίνεται και από το παρακάτω σχεδιάγραμμα. Στην πραγματικότητα ο έλεγχος αυτός χρειάζεται 16 δευτερόλεπτα ώστε να γίνει η μεταφόρτωση του αρχείου .wav όπως βλέπουμε στο σχήμα (1.1)[3].



Σχήμα 1.1 Σχεδιάγραμμα ροής επικοινωνίας AESDD με Unreal Engine

Αναλυτικότερα, η καθυστέρηση έχει εισαχθεί στον κώδικα ως χρόνος ασφάλειας, ώστε να έχει μεταφορτωθεί το αρχείο και να έχει εισαχθεί το αποτέλεσμα πριν τον έλεγχο για αποφυγή προβλημάτων.

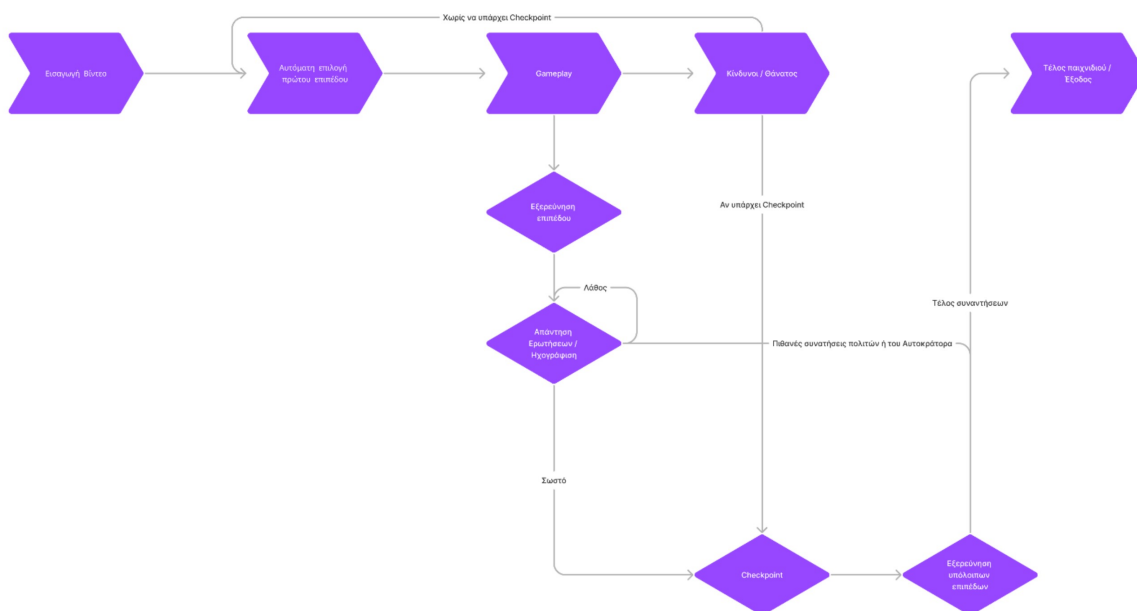
2. Σχεδιασμός – Υλοποίηση των παιχνιδιών

Με την ανάπτυξη δύο διαφορετικών βιντεοπαιχνιδιών, καλύπτονται ποικιλόμορφες προτιμήσεις, γεγονός που διευρύνει το κοινό των παικτών. Στο παιχνίδι πλατφόρμας (Silent Kingdom), η προσέγγιση είναι πιο κλασική (Arcade) ο παίκτης έχει ζωές και υπάρχουν πίστες. Επικεντρώνεται στην κίνηση του παίκτη στον χώρο, στην αλληλεπίδραση με το περιβάλλον, του ζητείται του παίκτη συγκεκριμένη εκφορά συναισθημάτων για να προχωρήσει την ιστορία στο παιχνίδι, και την αντιμετώπιση των εχθρών. Από την άλλη, το παιχνίδι προσομοίωσης (Job Interview Simulator) απευθύνεται σε εκείνος που απολαμβάνουν τα παιχνίδια ρόλων, τη λήψη αποφάσεων και την εξερεύνηση σεναρίων της καθημερινής ζωής, όπως και

δεν του ζητείται συγκεκριμένη εκφορά συναισθημάτων, προσφέροντας μια πιο ρεαλιστική εμπειρία[4].

2.1 Σχεδιασμός Silent Kingdom

Το παιχνίδι αυτό ανήκει στο είδος των 2D περιπετειών πλατφόρμας. Ο παίκτης πρέπει να χρησιμοποιήσει την ικανότητά του να κινείται για να ξεπεράσει διάφορα εμπόδια που συναντά. Επιπλέον, έχει τη δυνατότητα να αμυνθεί και να επιτεθεί σε ενδεχόμενους εχθρούς που μπορεί να βρεθούν μπροστά του. Ο στόχος του παιχνιδιού είναι να καταφέρει ο παίκτης, μέσω της έκφρασης συναισθημάτων, να διδάξει στους μη παίζοντες χαρακτήρες πώς να εκφράζουν τα αντίστοιχα συναισθήματα. Παρακάτω φαίνεται η εμπειρία του χρήστη σχήμα (2.2).



Σχήμα 2.2 Εμπειρία του παίκτη (User Experience – UX) στο Silent Kingdom

2.2 Υλοποίηση Silent Kingdom

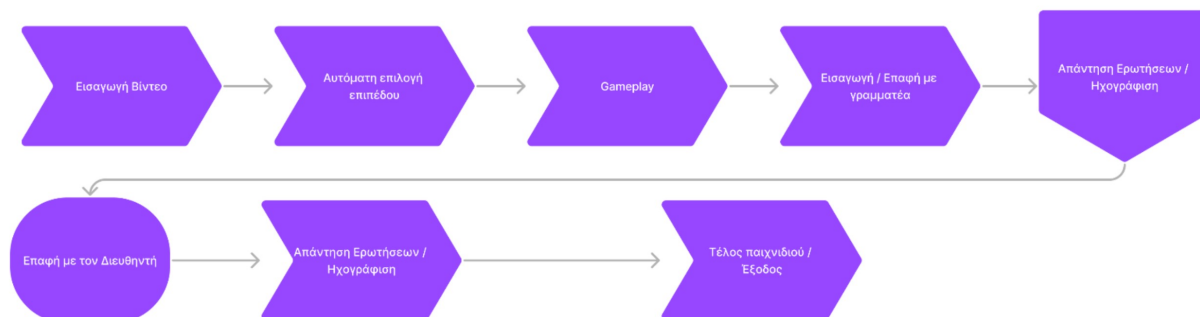
Οι χαρακτήρες στο Silent Kingdom είναι συνολικά 7, εκ των οποίων οι 6 είναι μη παίζοντες χαρακτήρες (NPC) που ενεργοποιούνται μόνο όταν ο παίκτης τους πλησιάσει. Αυτά τα NPC παρέχουν αποστολές για τον παίκτη, λειτουργώντας με προκαθορισμένα πρότυπα συμπεριφοράς. Τα NPC αλληλοεπιδρούν με τον παίκτη μέσω διαλόγων, όπου ο παίκτης ηχογραφεί τις απαντήσεις του και αυτές ελέγχονται με την βάση δεδομένων AESDD. Στο παιχνίδι υπάρχουν στοιχεία καθοδήγησης, όπως τα αστέρια που οδηγούν τον παίκτη και κινούμενες πλατφόρμες που προσθέτουν στοιχεία παζλ. Οι ακίδες προσθέτουν δυσκολία, προκαλώντας ζημιά στον παίκτη.

Η τεχνητή νοημοσύνη των NPC βασίζεται σε δέντρα συμπεριφοράς (behavior trees) που περιλαμβάνουν κόμβους εργασίας, όπως η τυχαία κίνηση και η επίθεση του «ακοντιστή». Ο «ακοντιστής» κινείται τυχαία εντός μια καθορισμένης περιοχής και όταν ο παίκτης πλησιάσει, ενεργοποιείται η επίθεση του, με σφαίρα ανίχνευσης που προκαλεί ζημιά αν έρθει σε επαφή με τον παίκτη. Το σύστημα δέντρου συμπεριφοράς επιτρέπει τη δημιουργία σύνθετων και ευέλικτων συμπεριφορών για τους χαρακτήρες τεχνητής νοημοσύνης, διευκολύνοντας τη διαδικασία ανάπτυξης του παιχνιδιού.

2.3 Σχεδιασμός Job Interview Simulator

Το “Job Interview Simulator” είναι ένα τρισδιάστατο παιχνίδι προσομοίωσης πρώτου προσώπου, σχεδιασμένο να καταγράφει και να αναλύει τη φωνητική χροιά του παίκτη. Ο παίκτης εισέρχεται σε μια εταιρεία όπου συναντά την γραμματέα, η οποία τον ενημερώνει για τον σκοπό της επίσκεψης: να συμμετάσχει σε μια συνέντευξη για εύρεση εργασίας.

Καθώς διασχίζει τον χώρο της εταιρείας, ο παίκτης ακούει τους υπαλλήλους να εργάζονται και να μιλούν με αμφιωτική ακοή, προσθέτοντας ρεαλισμό στην εμπειρία. Στη συνέχεια, επικοινωνεί με τον διευθυντή της εταιρείας, και η συνέντευξη ξεκινά με μια σειρά ερωτήσεων που έχουν ως στόχο την εκφορά συναισθημάτων του παίκτη. Οι απαντήσεις του παίκτη αξιολογούνται με κρυφό σκορ με βάση την σωστή ή λανθασμένη εκφορά των συναισθημάτων. Στο σχήμα (2.3) βλέπουμε την εμπειρία του παίκτη.



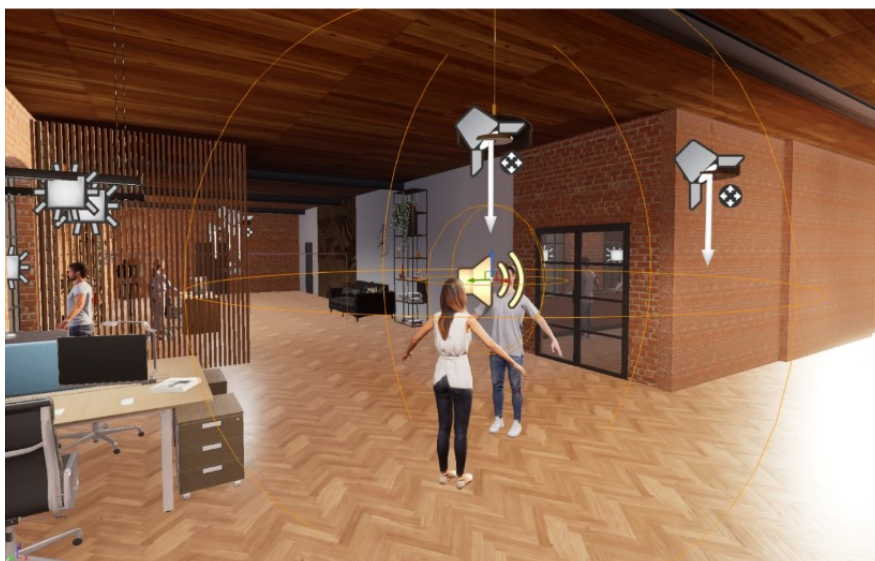
Σχήμα 2.3 Εμπειρία του παίκτη (User Experience – UX) στο Job Interview Simulator

2.4 Υλοποίηση Job Interview Simulator

Η δημιουργία των χαρακτήρων έγινε με τρισδιάστατα μοντέλα που προέρχονται από την κοινότητα της Unreal Engine, εξασφαλίζοντας ρεαλισμό και αποδοτικότητα στην χρήση των διαθέσιμων πόρων του υπολογιστή. Στον προγραμματισμό του παιχνιδιού, δεν υπάρχει τεχνητή νοημοσύνη, με αποτέλεσμα οι αυτόματες εκκινήσεις γεγονότων να ενεργοποιούνται μέσω αυτοματισμών και “trigger boxes”, επιτρέποντας στους διαλόγους και τις αλληλεπιδράσεις να προχωρούν χωρίς άμεσες

ενέργειες από τον παίκτη. Ο παίκτης έχει την ελευθερία να εκφράσει ό,τι επιθυμεί κατά τη διάρκεια της συνέντευξης, χωρίς το παιχνίδι να ζητά συγκεκριμένα συναισθήματα ή απαντήσεις, γεγονός που προσφέρει μια πιο αυθεντική και φυσική εμπειρία αλληλεπίδρασης.

Όσο αφορά την εφαρμογή της αμφιωτικής ακοής, η Unreal Engine 5 ενσωματώνει προηγμένες λειτουργίες αμφιωτικής ακοής, που επιτρέπουν στους προγραμματιστές να δημιουργούν ρεαλιστικά ηχητικά τοπία που αναπαράγουν τον ήχο όπως θα ακουγόταν από ανθρώπινα αυτιά. Μέσω της συλλογής συναρτήσεων HRTF (Head Related Transfer Function), η Unreal Engine 5 προσομοιώνει τις ακουστικές παραμέτρους σε πραγματικό χρόνο, προσφέροντας στους χρήστες μια ακριβή αναπαράσταση του ηχητικού περιβάλλοντος. Αυτό σημαίνει ότι οι ήχοι δεν χρειάζεται να έχουν ηχογραφηθεί με την τεχνική αμφιωτικής ηχογράφησης, καθώς η Unreal Engine μπορεί να προσαρμόσει τον ήχο δυναμικά, ανάλογα με τη θέση και τις κινήσεις του παίκτη. Επιπλέον, η Unreal Engine παρέχει ένα ισχυρό σύστημα διαχείρισης ήχου που περιλαμβάνει καμπύλες εξασθένησης, ρυθμίσεις αντήχησης και εφέ ηχοαπορρόφησης, διευκολύνοντας τη δημιουργία ρεαλιστικών ηχητικών αποτελεσμάτων, όπως βλέπουμε στο σχήμα (2.4)[5].



Σχήμα 2.4 Ρύθμιση καμπύλης εξασθένησης και τοποθέτηση του ήχου στον χώρο

3. Αποτελέσματα αξιολόγησης

Η συμμετοχή στα ερωτηματολόγια ήταν προαιρετική και ανώνυμη, με συνολικά 11 συμμετέχοντες και στα δύο παιχνίδια. Τα γραφήματα παρουσίασαν την ηλικιακή κατανομή των συμμετεχόντων, όπου οι περισσότεροι ήταν 26–30 ετών, με συμμετοχή από 18–25 ετών μόνο στο Silent Kingdom. Το μεγαλύτερο ποσοστό των χρηστών είχε τριτοβάθμια εκπαίδευση, ενώ δεν υπήρχε συμμετοχή από τα χαμηλότερα επίπεδα εκπαίδευσης.

Περισσότεροι από τους συμμετέχοντες ήταν ήδη έμπειροι με τα ηλεκτρονικά παιχνίδια, γεγονός που επηρεάζει τα ποσοστά δυσκολίας. Ο χρόνος παιχνιδιού ήταν μεγαλύτερος στο Silent Kingdom σε σύγκριση με το Job Interview Simulator, ενώ οι χρήστες δεν εντυπωσιάστηκαν με τα γραφικά του Silent Kingdom, τα οποία ήταν λιγότερο ρεαλιστικά. Σε γενικές γραμμές, κανένα από τα δύο παιχνίδια δεν ήταν πολύ δύσκολο, με 90.9% των συμμετεχόντων να τερματίζουν και τα δύο παιχνίδια.

Το Silent Kingdom είχε αυξημένα ποσοστά δυσκολίας όσον αφορά το χειρισμό, ενώ στο Job Interview Simulator παρατηρήθηκε απλοποίηση. Όσον αφορά την προτεινόμενη εμπειρία, το 100% των χρηστών του Silent Kingdom θα το πρότεινε, σε αντίθεση με το 90.9% του Job Interview Simulator. Και τα δύο παιχνίδια κατέγραψαν 100% ευχάριστης εμπειρίας, ενώ το Silent Kingdom θεωρήθηκε πολύ καλή εμπειρία από 18.2% των χρηστών, ενώ στο Job Interview Simulator το 27.3% το αξιολόγησε ως πάρα πολύ καλή.

Τέλος, υπήρξαν σχόλια που ανέφεραν ότι το Silent Kingdom ήταν διασκεδαστικό αλλά εύκολο, ενώ στο Job Interview Simulator προτάθηκε η προσθήκη επιλογών θεμάτων συνέντευξης.

4. Συμπεράσματα και μελλοντικές εξελίξεις

Η εργασία έχει ως σκοπό την ενσωμάτωση της αναγνώρισης συναισθημάτων ομιλίας σε πραγματικό χρόνο σε βιντεοπαιχνίδια. Αυτή η ενσωμάτωση ανοίγει νέες δυνατότητες για τη δημιουργία καθηλωτικών παιχνιδιών, συλλέγοντας δεδομένα φωνητικής χροιάς, που συμβάλλουν στη βελτίωση των μοντέλων αναγνώρισης συναισθημάτων. Τα βιντεοπαιχνίδια αποτελούν μια αποτελεσματική μέθοδο συλλογής δεδομένων, καθώς προσφέρουν ψυχαγωγία ενώ διευκολύνουν τη συμμετοχή στην έρευνα.

Η εργασία είναι διαθεματική, με πολλές περιοχές που μπορούν να βελτιωθούν. Μία από τις προκλήσεις είναι η καθυστέρηση 16 δευτερολέπτων κατά την ηχογράφηση, η οποία οφείλεται στις τρέχουσες ταχύτητες διαδικτύου. Για το Silent Kingdom, προτείνεται η προσθήκη περισσότερων επιπέδων, αποστολών και κινδύνων, καθώς και η εμπλουτισμένη ιστορία για μεγαλύτερη συλλογή ηχητικών δεδομένων. Στο Job Interview Simulator, θα μπορούσαν να προστεθούν εξειδικευμένα θέματα με δυνατότητα επιλογής επαγγέλματος και περισσότερες ερωτήσεις και χαρακτήρες για τη συλλογή ηχητικών δεδομένων.

Τέλος, η δυνατότητα επισημείωσης δειγμάτων ομιλίας και η ανάπτυξη της βάσης δεδομένων (Acted Emotional Speech Dynamic Database – AESDD) θα μπορούσαν να βελτιώσουν την εγκυρότητα των αποτελεσμάτων όπως και δεδομένου ότι πολλά στοιχεία του λόγου λειτουργούν ανεξάρτητα από τη σημασιολογία ή το πολιτισμικό πλαίσιο, αυτή η τεχνική θεωρείται μια πολλά υποσχόμενη κατεύθυνση για μελλοντική έρευνα[6].

5. Αναφορές

- [1] Ayadi, M. M. H. E., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3), 572–587. <https://doi.org/10.1016/j.patcog.2010.09.020>
- [2] Li, W., Zhang, Y., & Fu, Y. (2007). Speech Emotion Recognition in E-learning System Based on Affective Computing. <https://doi.org/10.1109/icnc.2007.677>
- [3] Vryzas, N., Kotsakis, R., Liatsou, A., Dimoulas, C., & Kalliris, G. (2018). Speech Emotion Recognition for Performance Interaction. *Journal of the Audio Engineering Society*, 66(6), 457–467. <https://doi.org/10.17743/jaes.2018.0036>
- [4] Scharping, N. (2022, August 15). *Gaming for Science: How Video Games Are Making Research Fun*. Discover Magazine. <https://www.discovermagazine.com/technology/gaming-for-science-how-video-games-are-making-research-fun>
- [5] Xu, X., Zhou, H., Liu, Z., Dai, B., Wang, X., & Lin, D. (2021). *Visually Informed Binaural Audio Generation without Binaural Audios*. <https://doi.org/10.1109/cvpr46437.2021.01523>
- [6] Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction. In *International journal of human-computer studies* (Vol. 59, Issues 1–2, pp. 1–32). Elsevier BV. [https://doi.org/10.1016/s1071-5819\(03\)00047-8](https://doi.org/10.1016/s1071-5819(03)00047-8)