

# ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΜΟΥΣΙΚΗΣ ΚΑΙ ΟΠΤΟΑΚΟΥΣΤΙΚΩΝ ΤΕΧΝΟΛΟΓΙΩΝ

Τμήμα Μουσικής Τεχνολογίας και Ακουστικής



## *Πτυχιακή εργασία*

**Συνεργατική επεξεργασία βίντεο παραγόμενου από  
χρήστες που παρακολουθούν το ίδιο συναυλιακό  
γεγονός**

**ΚΕΔΑΡΙΤΗΣ ΔΗΜΗΤΡΗΣ**

*Επιβλέπων:* **Νικόλαος Στεφανάκης**

*Επίκουρος Καθηγητής*

*Ρέθυμνο, Νοέμβριος 2022*



**ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ**

**ΣΧΟΛΗ ΜΟΥΣΙΚΗΣ ΚΑΙ ΟΠΤΟΑΚΟΥΣΤΙΚΩΝ ΤΕΧΝΟΛΟΓΙΩΝ**

**Τμήμα Μουσικής Τεχνολογίας και Ακουστικής**

***Πτυχιακή Εργασία***

**Συνεργατική επεξεργασία βίντεο παραγόμενου από  
χρήστες που παρακολουθούν το ίδιο συναυλιακό  
γεγονός**

***Του***

**Κεδαρίτη Δημήτρη**

**Επιβλέπων: Νικόλαος Στεφανάκης**

**Επίκουρος Καθηγητής**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την .....

.....

.....

.....

*Ρέθυμνο, Νοέμβριος 2022*

## **Πνευματικά δικαιώματα**

Copyright © Κεδαρίτης Δημήτρης, 2022

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Μουσικής Τεχνολογίας & Ακουστικής του Ελληνικού Μεσογειακού Πανεπιστημίου δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

Με το πέρας της παρούσας πτυχιακής εργασίας, θα ήθελα να πω ένα τεράστιο ευχαριστώ στην οικογένεια μου και στους φίλους μου, που μου στάθηκαν στα εύκολα και στα δύσκολα κατά την διάρκεια των σπουδών μου. Είναι ιδιαίτερα σημαντικό να υπάρχουν άνθρωποι στη ζωή μας που να μας υποστηρίζουν και είμαι πραγματικά ευγνώμων που τους έχω.

Θα ήταν παράλειψη μου να μην ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή της πτυχιακής κ. Στεφανάκη Νικόλαο, για την τεράστια ανταπόκριση του να δουλέψουμε μαζί το παρών θέμα.

## ΠΕΡΙΛΗΨΗ

Ο όρος User Generated Content (UGC) αναφέρεται σε πολυμεσικό υλικό όπως κείμενο, φωτογραφίες, ηχογραφήσεις και βίντεο που δημιουργείται από απλούς χρήστες φορητών συσκευών και το οποίο γίνεται διαθέσιμο στο ευρύ κοινό μέσα από τις διάφορες πολυμεσικές πλατφόρμες και τα κοινωνικά δίκτυα. Στην παρούσα πτυχιακή εργασία, μελετήσαμε την συνδυαστική αξιοποίηση βίντεο καταγραφών από χρήστες φορητών συσκευών που παρακολουθούν το ίδιο μουσικό γεγονός. Σκοπός μας είναι ο συγχρονισμός των καταγραφών αυτών και η συνεργατική επεξεργασία τους με βάση εργαλεία που έχουν προταθεί, για την δημιουργία ενός ενιαίου βελτιωμένου οπτικοακουστικού υλικού. Στην εργασία παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για τη συνεργατική επεξεργασία και παρουσιάζονται αποτελέσματα υποκειμενικής αξιολόγησης με βάση τέσσερα διαφορετικά συναυλιακά γεγονότα που επιλέχθηκαν από το YouTube. Τα αποτελέσματα αποδεικνύουν ότι η συνεργατική επεξεργασία των βίντεο μπορεί να οδηγήσει σε μια νέα πολυμεσική οντότητα που παρέχει βελτιωμένη εμπειρία θέασης και ακρόασης συγκριτικά με κάθε μία μεμονωμένη καταγραφή.

## **ABSTRACT**

User Generated Content (UGC) refers to digital content such as text, photos, sound and video recordings, produced by everyday users of mobile devices. Huge amounts of such content is produced every day and becomes accessible to the wide public through the various audio or video sharing platforms and through the social media. In this thesis, the focus is on collaborative processing of video content produced by users who attend the same musical event. We exploit already known tools in order to synchronize the recordings from the same event and to produce a new video stream that incorporates the information from more than one videos. Along with the methodology that was followed, we presented subjective results by conducting a survey based on four different musical events found on YouTube. The subjective results indicate that collaborative processing of audio content may result to a new video entity that improves the listening and viewing experience of the end user compared to the individual video entities.

# ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

|  |     |
|--|-----|
| ΠΕΡΙΛΗΨΗ.....  | iv  |
| ABSTRACT .....                                       | v   |
| ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ .....                           | vi  |
| ΕΙΣΑΓΩΓΗ.....  | xi  |
| 1.1 Αναφορά στη βιβλιογραφία.....                    | xii |
| 2 ΜΕΘΟΔΟΛΟΓΙΑ .....                                  | 1   |
| 2.1 Ηχητικά Αποτυπώματα .....                        | 1   |
| 2.2 Ταίριασμα και συγχρονισμός των αρχείων .....     | 3   |
| 2.3 Κανονικοποίηση .....                             | 4   |
| 2.4 Μίξη.....  | 5   |
| 3 ΥΛΟΠΟΙΗΣΗ.....                                     | 7   |
| 3.1 Εντοπισμός και κατέβασμα των βίντεο αρχείων..... | 7   |
| 3.2 Ο αλγόριθμος Python .....                        | 8   |
| 3.3 Μοντάζ .....                                     | 9   |
| 4 ΥΠΟΚΕΙΜΕΝΙΚΗ ΑΞΙΟΛΟΓΗΣΗ.....                       | 12  |
| 4.1 Υποθέσεις.....                                   | 12  |
| 4.2 Διατύπωση των ερωτήσεων.....                     | 12  |
| 4.2.1 Πρώτο μέρος ερωτήσεων .....                    | 13  |
| 4.2.2 Δεύτερο μέρος .....                            | 14  |
| 4.3 Επεξεργασία των απαντήσεων και συμπεράσματα..... | 16  |
| 4.3.1 Αποτελέσματα από το πρώτο μέρος.....           | 16  |
| 4.3.2 Αποτελέσματα από το δεύτερο μέρος .....        | 18  |
| 4.3.3 Ιδιαίτερες περιπτώσεις.....                    | 20  |
| 5 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ.....           | 22  |

ΒΙΒΛΙΟΓΡΑΦΙΑ..... 24



## ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

|   |    |
|---|----|
| Σχήμα 1 :Διεργασίες που εμφανίζονται κατά την συνεργατική επεξεργασία του ηχητικού υλικού.....  | 1  |
| Σχήμα 2: Αναπαράσταση της διαδικασίας του συγχρονισμού με βάση τις χρονικές διαφορές από την ηχογράφιση αναφοράς.....                         | 3  |
| Σχήμα 3: Τρεις μερικά επικαλυπτόμενες ηχογραφήσεις από τις οποίες πρέπει να παραχθεί μία ομοιόμορφη μίξη (Stefanakis 2019).....               | 6  |
| Σχήμα 4: Κατανομή των απαντήσεων όσων αφορά την οπτική πληροφορία για τις μεμονωμένες (αριστερά) και τις συνεργατικές (δεξιά) καταγραφές..... | 16 |
| Σχήμα 5: Κατανομή των απαντήσεων όσων αφορά την ποιότητα του ήχου σε μεμονωμένες (αριστερά) και συνεργατικές (δεξιά) καταγραφές.....          | 17 |
| Σχήμα 6: Κατανομή των απαντήσεων όσων αφορά την συνολική εμπειρία σε μεμονωμένες (αριστερά) και συνεργατικές (δεξιά) καταγραφές.....          | 17 |
| Σχήμα 7: Κατανομή απαντήσεων όσων αφορά την γενική προτίμηση ανάμεσα στις περιπτώσεις A (Μεμονωμένη) και B (Συνεργατική).....                 | 19 |
| Σχήμα 8: Κατανομή απαντήσεων όσων αφορά την προτίμηση αναφορικά με τον ήχο ανάμεσα στις περιπτώσεις A (Μεμονωμένη) και B (Συνεργατική).....   | 19 |
| Σχήμα 9: Οι απαντήσεις με την περισσότερη συμφωνία αναμεσά στις περιπτώσεις B (Συνεργατική) και A (Μεμονωμένη).....                           | 21 |
| <br>  |    |
| Πίνακας 1: Τα μεμονωμένα βίντεο που πήραμε από τους χρήστες του YouTube.....  | 7  |
| Πίνακας 2: Η τοποθεσία και η ημερομηνία που διεξάχθηκαν οι συναυλίες.....   | 8  |
| Πίνακας 3: Links από τα τελικά συνεργατικά βίντεο.....  | 10 |
| Πίνακας 4: Links από τα μεμονωμένα βίντεο μικρής διάρκειας.....   | 11 |
| Πίνακας 5: Όλα τα ερωτηματολόγια που διατέθηκαν στο κοινό.....  | 16 |
| <br>  |    |
| Εικόνα 1: Screenshot από την τη διαδικασία του τελικού μοντάζ.....  | 10 |

|  |    |
|--|----|
| Εικόνα 2: Δείγμα δομής ερωτηματολογίου στο πρώτο μέρος. ....   | 15 |
| Εικόνα 3: Δείγμα δομής ερωτηματολογίου στο δεύτερο μέρος. .... | 15 |

## ΑΠΟΔΟΣΗ ΟΡΩΝ

|                    |                                      |
|--------------------|--------------------------------------|
| Audio fingerprints | Ηχητικά αποτυπώματα                  |
| Audio matching     | Ηχητικό ταίριασμα                    |
| Match strength     | Δύναμη ταιριάσματος δύο ηχογραφήσεων |
| UGC                | User Generated Content               |
| UGV                | User Generated Video                 |
| UGRs               | User Generated Audio Recordings      |

## ΕΙΣΑΓΩΓΗ

Ζούμε στην εποχή της γρήγορης πληροφορίας, των φορητών συσκευών, της καθημερινής ενημέρωσης στα κοινωνικά δίκτυα, τύπου story και βρισκόμαστε μια ανάσα θα έλεγε κανείς πριν την εικονική πραγματικότητα. Κάθε ένα λεπτό, εκτιμάται ότι περίπου 600 ώρες βίντεο «ανεβαίνουν» στο κανάλι του Youtube και περίπου 695.000 «ιστορίες» κοινοποιούνται στο instagram. Λόγω αυτής της τάσης, μπορούμε σήμερα να βρούμε στο διαδίκτυο αμέτρητα πολυμεσικά αρχεία από πολιτισμικά γεγονότα όπως για παράδειγμα, μουσικές συναυλίες, θεατρικές παρατάσεις, χορό κλπ. Επίσης μεγάλος όγκος υλικού είναι διαθέσιμος από αθλητικά γεγονότα όπως το ποδόσφαιρο, το μπάσκετ ή το ράγκμπι. Τέλος, ένα μεγάλο μέρος των πολυμεσικών αρχείων που ανεβαίνει από χρήστες είναι από κοινωνικές εκδηλώσεις όπως γάμοι, βαφτίσεις, γενέθλια κλπ. Αυτή η κατακόρυφη αύξηση στο ρυθμό δημιουργίας οπτικοακουστικού υλικού ακολουθεί ουσιαστικά την παγκόσμια αύξηση στην αγορά και χρήση έξυπνων φορητών συσκευών, αφού πλέον κάθε smartphone, tablet ή laptop διαθέτει κάμερα και μικρόφωνο. Σημαντικό ρόλο παίζει προφανώς η παγκοσμιοποίηση και το γεγονός ότι οι συσκευές αυτές είναι πλέον προσιτές παντού και σε όλο και πιο χαμηλές τιμές. Στη βιβλιογραφία, το υλικό αυτό που προέρχεται από χρήστες είναι γνωστό ως User Generated Content (UGC) και αφορά κυρίως κείμενο, φωτογραφίες, ηχογραφήσεις και βίντεο. Οι τρόποι ανάλυσης, επεξεργασίας και αξιοποίησης του UGC είναι σήμερα ένα από τα πιο ταχεία αναπτυσσόμενα ερευνητικά πεδία στην επιστήμη της πληροφορίας.

Αν και ο όγκος των UGC βίντεο καταγραφών (User Generated Video – UGV) είναι τεράστιος, η ποιότητα συνήθως του οπτικού περιεχομένου είναι μέτρια έως κακή και το ακουστικό υλικό (ήχος) βρίσκεται πάλι στην ίδια μοίρα. Ναι μεν αν ψάξουμε στο διαδίκτυο θα βρούμε πολλαπλά μεμονωμένα βίντεο από το γεγονός ενδιαφέροντος, τις πιο πολλές φορές όμως αυτά έχουν μικρή διάρκεια και καταγράφουν ένα μικρό μέρος του όλου γεγονότος. Ωστόσο, αν με κάποιο τρόπο συγκεντρωθεί όλο το διαθέσιμο υλικό, θα μπορούσε ενδεχομένως να συντεθεί μια πιο ολοκληρωμένη οπτικοακουστική εμπειρία, καλύπτοντας μεγαλύτερο χρονικό εύρος του γεγονότος αλλά και παρέχοντας εικόνα από διαφορετικές γωνίες θέασης.

Στην παρούσα πτυχιακή εργασία, προσπαθούμε να αξιοποιήσουμε συνεργατικά οπτικοακουστικά αρχείων από το ίδιο δημόσιο γεγονός ώστε να δημιουργήσουμε μια νέα οντότητα βίντεο που να παρέχει βελτιωμένη εμπειρία ακρόασης και θέασης στον τελικό

χρήστη. Η συνεργατική επεξεργασία πολλών μεμονωμένων καταγραφών γίνεται με τη χρήση έτοιμων εργαλείων που έχουν προταθεί με βάση τον ήχο. Αν και αυτό θα μπορούσε να εφαρμοστεί σε δημόσια γεγονότα οποιουδήποτε τύπου, στην πτυχιακή αυτή εστιάζουμε σε υλικό που προέρχεται από μουσικές συναυλίες.

Στην εργασία παρουσιάζονται και αναλύονται τα στάδια επεξεργασίας από όπου περνάει το ηχητικό υλικό με σκοπό να ταιριάζουμε και να συγχρονίσουμε τις μεμονωμένες βίντεο καταγραφές με αυτόματο τρόπο, καθώς και η διεργασίες που εφαρμόζονται για τη μίξη του ηχητικού υλικού ώστε να δημιουργηθεί ένα συνεχές ηχητικό στρώμα που συνοδεύει το οπτικό υλικό. Για το οπτικό υλικό από την άλλη, βασιζόμαστε σε καθιερωμένες τεχνικές μοντάζ, οι οποίες όμως βοηθούνται σημαντικά από το αποτέλεσμα του αυτόματου συγχρονισμού.

Τέλος, για την τελική αξιολόγηση, δημιουργήσαμε ένα ερωτηματολόγιο με βίντεο υλικό και το κατανείμαμε στο ευρύ κοινό για να διαπιστώσουμε κατά πόσο η συνεργατική επεξεργασία του οπτικοακουστικού υλικού οδηγεί σε βελτίωση της εμπειρίας θέασης και ακρόασης σε σχέση με τις μεμονωμένες καταγραφές για το ίδιο συναυλιακό γεγονός.

## **1.1 Αναφορά στη βιβλιογραφία**

Αναφέροντας τον όρο UGC (User Generated Content) μιλάμε για οποιασδήποτε μορφής ψηφιακό περιεχόμενο που δημιουργείται από απλούς χρήστες και μπορεί να είναι κείμενο, φωτογραφίες, βίντεο, συνομιλίες, charts κλπ. Μέσω του ίντερνετ, το περιεχόμενο αυτό μπορεί να δημοσιευτεί σε διάφορα blogs, πλατφόρμες ψηφιακού υλικού ή στα κοινωνικά δίκτυα και να γίνει ορατό από εκατομμύρια ή δισεκατομμύρια άλλους χρήστες. Αν και το μεγαλύτερο μέρος από αυτήν την πληροφορία καταναλώνεται από ένα μικρό κοινό (πχ από αυτόν/αυτήν που το δημιούργησε και άτομα του στενού του/της κύκλου) μέρος του περιεχομένου που αναρτάται μπορεί να προβληθεί σε τεράστιο βαθμό και να ασκήσει τεράστια επιρροή, πχ επηρεάζοντας την κοινή γνώμη πάνω σε ένα κοινωνικό ή πολιτικό θέμα, διαμορφώνοντας μία νέα καλλιτεχνική τάση ή μόδα αλλά και προωθώντας νέα προϊόντα ή υπηρεσίες. Πράγματι, ένα από τα πιο σημαντικά σημεία όπου αναγνωρίζεται η δύναμη του UGC είναι οι δυνατότητες που παρέχει από πλευράς διαφήμισης (αλλά και δυσφήμισης) ενός νέου προϊόντος ή υπηρεσίας (Cheong 2008).

Μία άλλη σημαντική δυνατότητα που μελετάται αναφορικά με το UGC - και ειδικότερα με αρχεία ήχου και βίντεο - είναι η προβολή και on-demand κατανάλωση δημοσίων γεγονότων όπως συναυλίες, αθλητικά γεγονότα, κοινωνικές εκδηλώσεις κλπ. Οι εγγραφές που δημιουργήθηκαν από χρήστες φορητών συσκευών σε μια δημόσια εκδήλωση περιλαμβάνουν πολλές φορές πολλαπλές συσκευές καταγραφής σε συχνά διαφορετικές ανακρίβειες γνωστές χωρικές τοποθεσίες. Οι ηχογραφήσεις αυτές μπορεί να ξεκινήσουν και να σταματήσουν σε διαφορετική χρονική στιγμή, έχουν διαφορετική τεχνική μορφή προδιαγραφών και σπάνια θα έχουν επαρκώς αξιόπιστες πληροφορίες χρονικής σφραγίδας, για την ακρίβεια συγχρονισμού (Stefanakis 2018). Παρ'όλα αυτά, οι πολλαπλές ηχογραφήσεις θα μπορούσαν ενδεχομένως να συνδυαστούν για να αντληθεί πληροφορία που δε θα ήταν δυνατό να αντληθεί αν κάθε εγγραφή αναλυόταν ξεχωριστά.

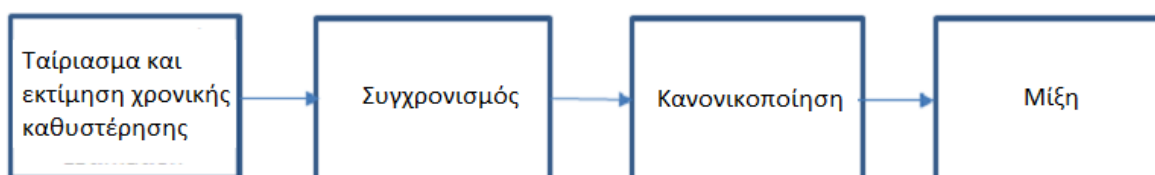
Δεδομένης μιας συλλογής από πολλά UGV, αρκετές προσεγγίσεις έχουν προταθεί αναφορικά με το πώς το οπτικό και ηχητικό περιεχόμενο από τις καταγραφές αυτές μπορεί να αξιοποιηθεί για να εκτιμηθεί ποια βίντεο είναι από το ίδιο δημόσιο γεγονός, να εκτιμηθεί αν δύο καταγραφές επικαλύπτονται χρονικά και εν τέλει, τα βίντεο που όντως είναι από το ίδιο χώρο και χρόνο να συγχρονιστούν πάνω σε ένα κοινό χρονικό άξονα. Οι περισσότερες προσεγγίσεις βασίζονται στην αξιοποίηση ηχητικών αποτυπωμάτων που απορρέουν από το ηχητικό υλικό (Shrestha 2007, Kennedy 2009, Cotton 2010, Bano 2015, Bryan 2012) αναδεικνύοντας έτσι την καταλληλότητα των ηχητικών αποτυπωμάτων σε μια διεργασία που είναι διαφορετική από αυτή που είχαν σχεδιαστεί αρχικά. Ωστόσο, σε εργασίες όπως αυτή του Stefanakis et al (2017) και των Kammerl et al (2014) αναγνωρίζεται η ανάγκη για περαιτέρω έρευνα πάνω στον τρόπο που θα πρέπει να δημιουργούνται τα ηχητικά αποτυπώματα αλλά και στον τρόπο που θα πρέπει να εξετάζονται οι ομοιότητες ανάμεσα στα ηχητικά αρχεία ώστε το ταίριασμα και ο συγχρονισμός να είναι πιο αξιόπιστος σε κάποια δημόσια γεγονότα.

Ένα επιπλέον κομμάτι της έρευνας σχετικά με τα UGV αφορά τη συνεργατική αξιοποίηση των καταγραφών που είναι από το ίδιο δημόσιο γεγονός με σκοπό την παραγωγή μιας νέας ηχητικής οντότητας με βελτιωμένα ακουστικά χαρακτηριστικά συγκριτικά με κάθε μία μεμονωμένη καταγραφή. Οι δουλειές των Kim et al (Kim 2013, Kim 2016) ήταν από τις πρώτες που επικεντρώθηκαν στο συγκεκριμένο ζήτημα, ωστόσο η έρευνά τους βασίστηκε σε συνθετικά δεδομένα που παρήχθησαν μέσα από προσομοιώσεις και η αξιολόγηση περιορίστηκε σε αντικειμενικά κριτήρια. Από την άλλη, οι δουλειές των Stefanakis et al

(Stefanakis 2017, 2018 and 2019) μελέτησαν τη συνεργατική επεξεργασία του ηχητικού υλικού βασιζόμενες σε πραγματικές ηχογραφήσεις ενώ στις (Stefanakis 2017, Valsamis 2018) η αξία της συνεργατικής επεξεργασίας επαληθεύτηκε με αποτελέσματα από υποκειμενικά τεστ (τεστ ακρόασης) . Στην εργασία (Stefanakis 2018) αναδεικνύεται η σημασία της κανονικοποίησης των ηχογραφήσεων ενώ στην (Stefanakis 2019) η κανονικοποίηση αξιολογείται για την πρόταση μίας ολοκληρωμένης διεργασίας για την επεξεργασία και μίξη των ηχογραφήσεων, περιορίζοντας ωστόσο τη μελέτη στην παραγωγή μονοφωνικού υλικού.

## 2 ΜΕΘΟΔΟΛΟΓΙΑ

Το ηχητικό περιεχόμενο είναι καθοριστικής σημασίας για την συνεργατική αξιοποίηση των UGV. Πρώτη προτεραιότητα στο εγχείρημα αυτό είναι το να επιβεβαιώσουμε ότι τα βίντεο είναι όντως από το ίδιο δημόσιο γεγονός και επίσης ότι καταγράφουν το χρονικό εύρος ενδιαφέροντος. Αφού επιβεβαιωθεί αυτό, πρέπει να βρεθούν οι χρονικές διαφορές με τις οποίες οι βίντεο καταγραφές από το ίδιο γεγονός σχετίζονται μεταξύ τους. Όπως θα αναλυθεί παρακάτω, ένα σημαντικό βήμα για τη δημιουργία του τελικού ήχου είναι η κανονικοποίηση των ηχογραφήσεων και τέλος η μίξη τους. Στο Σχήμα 1 απεικονίζεται με τη μορφή μπλοκ διαγράμματος η σειρά με την οποία εκτελείται η κάθε διεργασία.



Σχήμα 1 :Διεργασίες που εμφανίζονται κατά την συνεργατική επεξεργασία του ηχητικού υλικού

### 2.1 Ηχητικά Αποτυπώματα

Ηχητικό αποτύπωμα είναι η μειωμένης διάστασης πληροφορία που παίρνουμε από κάποιο ηχητικό υλικό με σκοπό να μπορούμε γρήγορα και αποτελεσματικά να βρούμε υλικό με παρόμοιο περιεχόμενο μέσα από μία μεγάλη βάση ηχητικών δεδομένων. Μια από τις πιο γνωστές εφαρμογές που βασίζονται στα ηχητικά αποτυπώματα είναι το Shazam. Ο χρήστης της εφαρμογής αυτής, παρέχει στο πρόγραμμα ένα μικρό σε διάρκεια ηχητικό απόσπασμα τραγουδιού και μέσα σε δευτερόλεπτα μαθαίνει τον καλλιτέχνη, τον παραγωγό και τον τίτλο του κομματιού. Τα ηχητικά αποτυπώματα, σε συνδυασμό και με τις διάφορες προσεγγίσεις ταυτοποίησης μουσικού περιεχομένου που εξυπηρετούν, έχουν δημιουργηθεί όχι μόνο από την ανάγκη του κοινού να ανακαλύψει το τραγούδι που ακούει την ίδια στιγμή, αλλά και για να μπορούν να τηρούνται και να αποδίδονται τα πνευματικά δικαιώματα των τραγουδιών που παίζουν στο ραδιόφωνο ή στην τηλεόραση στους καλλιτέχνες.

Διαφορετικές τεχνικές έχουν δημιουργηθεί για τον εντοπισμό ομοιότητας μέσω ηχητικών αποτυπωμάτων. Αν και στην παρούσα πτυχιακή δεν θα αναλύσουμε σε βάθος το κομμάτι των ηχητικών αποτυπωμάτων, είναι αναγκαίο να αναφερθούμε σε αυτό, μιας και η τεχνική



αλλά και ο αλγόριθμος που χρησιμοποιήθηκε για τον συγχρονισμό των δικών μας ηχητικών αποσπασμάτων, λειτουργεί με παρόμοιο τρόπο.

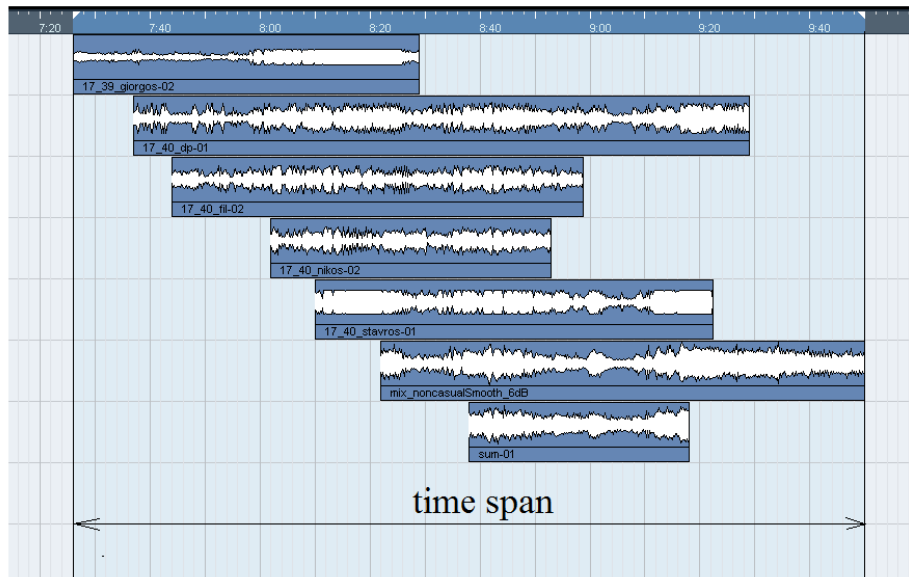
Τα ηχητικά αποτυπώματα κατά κανόνα βασίζονται σε τεχνικές ελάττωσης της διάστασης της πληροφορίας (dimension reduction), μετασχηματίζοντας την πληροφορία από το πεδίο του χρόνου στο πεδίο του χρόνου συχνότητας (ΧΣ). Έχουμε συνήθως έναν μετασχηματισμό της μορφής

$$x[n] \rightarrow X[\omega, \tau],$$

όπου  $x$  είναι το ηχητικό σήμα σε ψηφιακή μορφή,  $n$  είναι ο δείκτης του δείγματος,  $\omega$  ο δείκτης για τη ζώνη συχνότητας και  $\tau$  ο δείκτης χρονικού πλαισίου.

Σύμφωνα με μια από τις πιο διαδεδομένες προσεγγίσεις που έχουν προταθεί, από το πεδίο του ΧΣ δε χρειάζεται να κρατήσουμε όλη την πληροφορία, αλλά μόνο τα σημεία ΧΣ στα οποία έχουμε τοπικό μέγιστο (Wang 2003). Παίρνοντας συνδυασμούς γειτονικών τοπικών μεγίστων ανά δυο δημιουργούνται τα λεγόμενα hashes, από τα οποία η πληροφορία που διατηρείται είναι η σχετική απόσταση μεταξύ των δυο σημείων στο χρόνο και τη συχνότητα. Δύο ηχογραφήσεις από το ίδιο γεγονός αναμένεται να έχουν αρκετά κοινά hashes και μάλιστα, οι σχετικές διαφορές στους χρόνους εμφάνισής τους θα πρέπει να συμφωνούν. Αυτή είναι η φιλοσοφία πίσω από τη διαδικασία που περιγράφεται στο (Wang 2003) η οποία αν και έχει πολύ χαμηλή πολυπλοκότητα, επιτυγχάνει πολύ καλή επίδοση όταν πρόκειται για μουσικό περιεχόμενο.

Μια άλλη προσέγγιση για τη δημιουργία ηχητικών αποτυπωμάτων, η οποία είναι και αυτή που ακολουθήθηκε για τις ανάγκες αυτής της πτυχιακής, αξιοποιεί όλη τη διαθέσιμη πληροφορία στο πεδίου ΧΣ και όχι μόνο τα τοπικά μέγιστα (Haitsma 2002). Η φιλοσοφία εδώ είναι να κωδικοποιήσουμε τις διαφορές ανάμεσα σε γειτονικά σημεία στο ΧΣ, τόσο κατά τον άξονα του χρόνου όσο και τον άξονα της συχνότητας. Οι διαφορές αυτές κωδικοποιούνται ανάλογα με το πρόσημό τους, δηλαδή κωδικοποιούνται ως 1 αν είναι θετικές και ως 0 αν είναι αρνητικές. Κατά αυτόν τον τρόπο, επιτυγχάνεται μια σημαντική μείωση στον όγκο των δεδομένων για την αναπαράσταση της πληροφορίας και ταυτόχρονα, επιτυγχάνεται σημαντική ευρωστία απέναντι στο θόρυβο. Αν και η προσέγγιση αυτή οδηγεί σε ηχητικά αποτυπώματα μεγαλύτερου όγκου από ότι η προηγούμενη τεχνική, είναι πιο αποδοτική για εφαρμογές όπου το ηχητικό υλικό δεν είναι απαραίτητα μουσική (Stefanakis 2017).



**Σχήμα 2:** Αναπαράσταση της διαδικασίας του συγχρονισμού με βάση τις χρονικές διαφορές από την ηχογράφηση αναφοράς

## 2.2 Ταίριασμα και συγχρονισμός των αρχείων

Όπως αναφέρθηκε και πιο πάνω, ο σκοπός της χρήσης των ηχητικών αποτυπωμάτων (audio fingerprints) στην δικιά μας περίπτωση, είναι για να γίνει εύκολα το ταίριασμα των τραγουδιών μεταξύ τους έτσι ώστε, να είμαστε σίγουροι ότι έχουμε ένα ηχητικό απόσπασμα από την ίδια συναυλία και επιπρόσθετα, για να συγχρονίσουμε τα αποσπάσματα πάνω σε ένα κοινό χρονικό άξονα. Ο συγχρονισμός αυτός, αν και επιτυγχάνεται μέσω της ηχητικής πληροφορίας, είναι έγκυρος επίσης και για το οπτικό υλικό.

Για να εξάγουμε την απαραίτητη πληροφορία που θέλουμε, αξιοποιούμε τις ομοιότητες στα ηχητικά αποτυπώματα που λαμβάνουν υπόψιν όλους τους συνδυασμούς μεταξύ της ηχογράφησης αναφοράς και των άλλων ηχογραφήσεων ανά δύο (για  $N$  αρχεία έχουμε  $N-1$  τέτοιους συνδυασμούς). Αυτό που πρέπει να γίνει ουσιαστικά είναι ένας υπολογισμός όπου η συμφωνία μεταξύ των αποτυπωμάτων εξετάζεται για όλες τις πιθανές χρονικές μετατοπίσεις μεταξύ των δύο αρχείων. Ο υπολογισμός αυτός στην ουσία είναι η πράξη της ετεροσυσχέτισης και μαθηματικά μπορεί να αναπαρασταθεί ως

$$R_{i,j}(\tau) = \sum_{n=-\infty}^{+\infty} \mathbf{F}_i^T(n) \mathbf{F}_j(n + \tau)$$

όπου η  $\mathbf{F}_i(n)$  είναι ένα  $B \times 1$  διάνυσμα με τα ηχητικά αποτυπώματα για την ηχογράφιση  $i$  κατά μήκος του άξονα των συχνοτήτων για το χρονικό πλαίσιο με δείκτη  $n$ . Αν τα δύο ηχητικά αρχεία ( $i$  και  $j$ ) είναι όντως από το ίδιο γεγονός και επικαλύπτονται χρονικά, τότε η τιμή της καθυστέρησης  $\tau$  όπου μεγιστοποιείται το  $R_{i,j}(\tau)$  θα αντιστοιχεί στη χρονική διαφορά μεταξύ των δύο καταγραφών. Την ίδια στιγμή, η μέγιστη τιμή που παίρνει το  $R_{i,j}$  είναι μια ένδειξη του match strength, πόσο δηλαδή ταιριάζουν μεταξύ τους οι καταγραφές  $i$  και  $j$ . Για πολλές προσεγγίσεις στη βιβλιογραφία, αυτή η μέγιστη τιμή αποτελεί και το κριτήριο για την τελική απόφαση αν δύο ηχογραφήσεις είναι από το ίδιο γεγονός η όχι. Για τη δική μας προσέγγιση, δύο ηχογραφήσεις θεωρείται ότι ταιριάζουν αν το match strength είναι μεγαλύτερο από 10.

Για την περίπτωση που ένα πλήθος από ηχογραφήσεις ταιριάζει με την ηχογράφιση αναφοράς, τότε έχουμε απευθείας την πληροφορία που χρειάζεται για να συγχρονίσουμε όλο το σετ ηχογραφήσεων που είναι διαθέσιμο από το ίδιο ηχητικό γεγονός. Ας υποθέσουμε ότι δύο χρήστες καταγράφουν από το smartphone τους το ίδιο συναυλιακό γεγονός από δυο διαφορετικές οπτικές γωνίες. Ο θεατής νούμερο 1, ξεκινάει την καταγραφή από την αρχή του τραγουδιού. Ο θεατής νούμερο 2 όμως, ξεκινάει μετά από 2 λεπτά. Αν ορίσουμε το ηχητικό του θεατή νούμερο 1 για να το έχουμε ως αναφορά, τότε το ηχητικό νούμερο 2, θα μας δώσει την διάρκεια που απέχει από το ηχητικό 1, με θετικό πρόσημο. Σε διαφορετική περίπτωση, εάν πάρουμε ως αναφορά το ηχητικό νούμερο 2, τότε η διάρκεια που θα απέχει από το νούμερο 1, θα έχει αρνητικό πρόσημο. Για κάθε συναυλιακό γεγονός χρησιμοποιήθηκαν μέχρι και έξι ηχητικά αποσπάσματα. Με το ίδιο ακριβώς παράδειγμα, λειτουργεί ο αλγόριθμος και για τα έξι ηχητικά αποσπάσματα ταυτόχρονα.

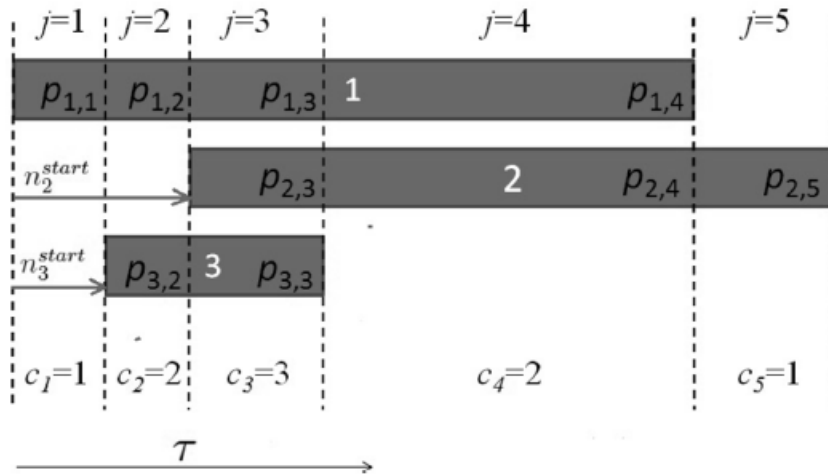
## 2.3 Κανονικοποίηση

Η κανονικοποίηση είναι μια σημαντική διαδικασία από την οποία πρέπει να περάσουν οι διαθέσιμες ηχογραφήσεις πριν την τελική μίξη. Απώτερος σκοπός της διαδικασίας είναι να προσδιορίσει έναν βάρος (gain) με το οποίο πρέπει να πολλαπλασιαστεί κάθε μία ηχογράφιση. Τα βάρη για τις ηχογραφήσεις υπολογίζονται συνδυαστικά, λαμβάνοντας

υπόψιν τις σχετικές διακυμάνσεις της ισχύος για τα χρονικά κομμάτια που οι ηχογραφήσεις επικαλύπτονται χρονικά. Λαμβάνοντας υπόψιν ότι, όλα τα προ ηχογραφημένα αρχεία που έχουν παραχθεί στις συναυλίες από τους χρήστες ήταν ηχογραφημένα από διαφορετικές θέσεις στον χώρο και με διαφορετικές φορητές συσκευές, η κανονικοποίηση θα φέρει όλα τα ηχητικά αρχεία στην ίδια ένταση ούτως ώστε κανένα ηχητικό αρχείο να μην επισκιάζει το άλλο όσον αφορά την ένταση και κανένα αρχείο να μην έχει υπερβολικά ασθενή στάθμη σήματος. Δε θα αναφερθούμε περαιτέρω στη διαδικασία και ο αναγνώστης παραπέμπεται στην εργασία (Stefanakis 2018).

## 2.4 Μίξη

Η μίξη, είναι η τελική επεξεργασία που χρειάζεται να γίνει, για την ολοκλήρωση του ακουστικού μέρους του τελικού βίντεο. Στόχος της είναι ένα μεγαλύτερο ενιαίο και αναβαθμισμένο ποιοτικά ακουστικό αρχείο, το οποίο συμπεριλαμβάνει όλες τις μεμονωμένες καταγραφές των χρηστών που παρακολούθησαν το συναυλιακό γεγονός. Στο σχήμα 3 απεικονίζεται μια περίπτωση όπου τρεις ηχογραφήσεις έχουν συγχρονιστεί και είναι έτοιμες για τη διαδικασία της μίξης. Για την κατανόηση της διαδικασίας που ακολουθείται, ας υποθέσουμε ότι κάθε ηχητικό απόσπασμα, αντιπροσωπεύει ένα fader σε μια κονσόλα ήχου. Το πρώτο μεμονωμένο ηχητικό, ξεκινάει να “παίζει” σε μέγιστη ένταση τη χρονική στιγμή 0. Το δεύτερο ηχητικό, ξεκινάει να “παίζει” τη χρονική στιγμή  $n_2^{start}$ . Στο πρώτο κανάλι θα πρέπει τώρα να κατέβει το fader, για να μειωθεί η ηχητική του ένταση, χωρίς να μειωθεί η συνολική ηχητική στάθμη από τη μίξη των δύο αρχείων στο master fader. Η ίδια λοιπόν διαδικασία θα πρέπει να επαναλαμβάνεται κάθε χρονική στιγμή όπου ένα νέο μεμονωμένο ηχητικό απόσπασμα θα ξεκινάει να συμμετέχει στη μίξη.



**Σχήμα 3: Τρεις μερικά επικαλυπτόμενες ηχογραφήσεις από τις οποίες πρέπει να παραχθεί μία ομοιόμορφη μίξη (Stefanakis 2019).**

Η διαδικασία της μίξης που ακολουθήθηκε σε αυτήν την πτυχιακή μπορεί να διατυπωθεί μαθηματικά ως

$$s^L[n] = \frac{1}{\sqrt{c_{j(n)}}} \sum_m g_m x_m^L[n - n_m^{start}]$$

για το αριστερό κανάλι (L) και ως

$$s^R[n] = \frac{1}{\sqrt{c_{j(n)}}} \sum_m g_m x_m^R[n - n_m^{start}]$$

για το δεξί (R) κανάλι. Εδώ  $s^L[n]$  είναι το παραγόμενο ακουστικό σήμα στο αριστερό κανάλι εξόδου και  $x_m^L[n]$  είναι αντίστοιχα το ακουστικό σήμα στο αριστερό κανάλι του UGR με δείκτη  $m$ . Τα βάρη του κάθε καναλιού που συμμετέχει στη μίξη, και τα οποία παράγονται από τη διεργασία της κανονικοποίησης συμβολίζονται ως  $g_m$ . Τέλος,  $c_{j(n)}$  είναι ένας θετικός ακέραιος ο οποίος ισούται με το πλήθος των UGRs που συμμετέχουν στη μίξη τη χρονική στιγμή  $n$ . Η διεργασία που περιγράφεται παραπάνω επεκτείνει ουσιαστικά την μονοφωνική προσέγγιση που παρουσιάζεται στην εργασία (Stefanakis 2019) θεωρώντας στερεοφωνικές ηχογραφήσεις και εν τέλει, στερεοφωνικό αποτέλεσμα στην έξοδο. Στην περίπτωση που κάποιο UGR είναι μονοφωνικό, η διεργασία εκτελείται χωρίς πρόβλημα θεωρώντας ίδιο περιεχόμενο στο left και right κανάλι.

## 3 ΥΛΟΠΟΙΗΣΗ

### 3.1 Εντοπισμός και κατέβασμα των βίντεο αρχείων

Για τον εντοπισμό των βίντεο αρχείων επιλέξαμε τον ιστότοπο του YouTube. Εκατοντάδες ώρες βίντεο “ανεβαίνουν” στο YouTube κάθε λεπτό ανά το παγκόσμιο. Εξαιτίας της εύκολης περιήγησης στην πλατφόρμα και της ακόμη πιο εύκολης αναζήτησης αρχείων, η συγκεκριμένη πλατφόρμα χρησιμοποιείται από όλους μας σχεδόν καθημερινά, οι επιλογές που υπάρχουν, στα διαφορετικά είδη βίντεο, είναι πάρα πολλές και για αυτόν τον λόγο η πλατφόρμα αυτή καλύπτει ένα τεράστιο εύρος ηλικιών και αυτό είναι που την κάνει τον πιο πολυσύχναστο ιστότοπο στο διαδίκτυο. Επιλέξαμε λοιπόν, να ασχοληθούμε με γνωστούς καλλιτέχνες ανά το παγκόσμιο αλλά και στον Ελλαδικό χώρο επιλέγοντας 4 διαφορετικές συναυλίες. Επιλέχθηκαν γνωστά μουσικά κομμάτια, έτσι ώστε το ηχητικό υλικό, να ήταν οικείο στους ερωτηθέντες. Με αυτόν τον τρόπο θα μπορούσαν πιο εύκολα να συγκρίνουν το τελικό αποτέλεσμα, σε σχέση με την ποιότητα που ήδη γνωρίζουν. Οπότε για αρχή, ξεκινήσαμε με το κομμάτι “Vittorio” από τον Έλληνα rapper Λέξ. Μια από τις τελευταίες του συναυλίες προσέκλισε σχεδόν δεκαπέντε χιλιάδες κοινό, κι έτσι μπορέσαμε να βρούμε οπτικοακουστικό υλικό από το συναυλιακό γεγονός. Στην συνέχεια, επιλέξαμε μια παγκόσμιου φήμης καλλιτέχνη, την Billie Eilish, με το κομμάτι “Happier Than Ever” (έχει πάνω από 500 εκατομμύρια streams στην μουσική πλατφόρμα Spotify και τεράστιο όγκο βίντεο αρχείων από συναυλίες της στην πλατφόρμα του YouTube). Το επόμενο κομμάτι που επεξεργάστηκε ήταν το “In The End” των θρυλικών Linkin Park, ένα από τα πιο δημοφιλή τους τραγούδια, η μπάντα κατέχει τον τίτλο, ανάμεσα σε άλλους, με τις μεγαλύτερες πωλήσεις του 21ου αιώνα. Το επόμενο και τελευταίο κομμάτι που επεξεργάστηκε, ήταν το “Dynamite” από το πρωτοεμφανιζόμενο συγκρότημα Νοτιοκορεατών BTS. Το κομμάτι αυτό έσπασε ρεκόρ στην πλατφόρμα του YouTube με τις πιο πολλές προβολές μέσα σε ένα 24ωρο. Σήμερα, κατέχει πάνω από 1,5 δισεκατομμύρια προβολές στο YouTube. Στον πίνακα 2 παρουσιάζονται οι ημερομηνίες και η τοποθεσίες στις οποίες πραγματοποιήθηκαν οι συναυλίες.

**Πίνακας 1: Τα μεμονωμένα βίντεο που πήραμε από τους χρήστες του YouTube.**

|                  |
|------------------|
| LEX – “Vittorio” |
|------------------|

|  |                        |                        |                        |                        |
|--|------------------------|------------------------|------------------------|------------------------|
| <a href="#">Video1</a>                     | <a href="#">Video2</a> | <a href="#">Video3</a> | <a href="#">Video4</a> | <a href="#">Video5</a> |
| <b>Billie Eilish – ‘Happier than ever’</b> |                        |                        |                        |                        |
| <a href="#">Video1</a>                     | <a href="#">Video2</a> | <a href="#">Video3</a> | <a href="#">Video4</a> | <a href="#">Video5</a> |
| <b>BTS – ‘Dynamite’</b>                    |                        |                        |                        |                        |
| <a href="#">Video1</a>                     | <a href="#">Video2</a> | <a href="#">Video3</a> | <a href="#">Video4</a> | <a href="#">Video5</a> |
| <b>Linkin Park – ‘In the end’</b>          |                        |                        |                        |                        |
| <a href="#">Video1</a>                     | <a href="#">Video2</a> | <a href="#">Video3</a> | <a href="#">Video4</a> | <a href="#">Video5</a> |

**Πίνακας 2: Η τοποθεσία και η ημερομηνία που διεξάχθηκαν οι συναυλίες.**

|                      |                                       |
|----------------------|---------------------------------------|
| <b>Billie Eilish</b> | State farm arena Atlanta, 05/02/2022  |
| <b>BTS</b>           | SOFI Stadium, Los Angeles, 27/11/2021 |
| <b>Lex</b>           | Θέατρο πέτρας Αθήνα, 14/06/2019       |
| <b>Linkin Park</b>   | The O2 London Arena, 03/07/2017       |

### 3.2 Ο αλγόριθμος Python

Στο προηγούμενο κεφάλαιο, αναλύσαμε τη διαδικασία που περνάει το ηχητικό αρχείο, που πήραμε από τον χρήστη, για να φτάσει στην τελική του ακουστική μορφή. Ένας σημαντικός συνεργάτης όμως στην διαδικασία αυτή είναι και ο αλγόριθμος που τρέξαμε στην γλώσσα προγραμματισμού Python. Παίρνοντας λοιπόν ένα έτοιμο εργαλείο που έχει προταθεί με βάση τον ήχο, θα ακολουθήσει μια συνοπτική περιγραφή και διαδικασία του εργαλείου αυτού.

Η συγκεκριμένη βιβλιοθήκη της Python έχει αναπτυχθεί από τον επίκουρο καθηγητή του τμήματος κύριο Νικόλαο Στεφανάκη, ακριβώς για τις ανάγκες του ταιριάσματος και συγχρονισμού ηχογραφήσεων. Η εκτέλεση του κώδικα έγινε στο command window σε περιβάλλον windows. Αφού έχουν απομονωθεί τα ηχητικά αποσπάσματα από τα βίντεο σε μορφή wav, δημιουργήθηκε ένας φάκελος με τα συγκεκριμένα ηχητικά αρχεία και αυτός ο

φάκελος δίνεται ως είσοδος στον αλγόριθμο. Ένα ηχητικό αρχείο από όλα τα διαθέσιμα επιλέγεται ως αρχείο αναφοράς και δίνεται και αυτό ως είσοδος. Ο αλγόριθμος θα εξάγει ηχητικά αποτυπώματα και θα προσπαθήσει να ταιριάζει όλα τα ηχητικά αρχεία με το αρχείο αναφοράς, δίνοντας μας στην έξοδο διάφορες πληροφορίες, όπως πχ τις χρονικές διαφορές ανάμεσα σε κάθε απόσπασμα και το απόσπασμα αναφοράς, τη διάρκεια της χρονικής επικάλυψης (overlap) και επίσης το match strength, το οποίο όσο πιο μεγάλο είναι, τόσο ισχυρότερη η ομοιότητα (ή συσχέτιση) ανάμεσα στο κάθε ένα ηχητικό αρχείο και στο αρχείο αναφοράς. Η χρονική διαφορά είναι μια θετική η αρνητική ποσότητα σε δευτερόλεπτα η οποία μας λέει πόσο πρέπει να καθυστερεί ή να προπορεύεται το κάθε ένα αρχείο ήχου σε σχέση με το αρχείο αναφοράς ώστε να έχουμε συγχρονισμό.

### 3.3 Μοντάζ

Το μοντάζ, είναι το τελικό στάδιο επεξεργασίας, με σκοπό να βελτιωθεί μια μεμονωμένη καταγραφή κυρίως όσον αφορά το οπτικό περιεχόμενο. Το πρόγραμμα που χρησιμοποιήθηκε για αυτόν τον σκοπό, ήταν το Adobe Premiere Pro 2020. Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο για τον συγχρονισμό των αρχείων, η πληροφορία που παίρνουμε από την Python για την χρονική διαφορά των ηχητικών, έρχεται και αξιοποιείται κατάλληλα για το μοντάζ. Επομένως, γνωρίζοντας στο παρών στάδιο την πληροφορία αυτή, τα βίντεο τοποθετούνται στον χρονικό άξονα του προγράμματος με βάση τις χρονικές διαφορές που παρέχονται από τον αλγόριθμο της Python.

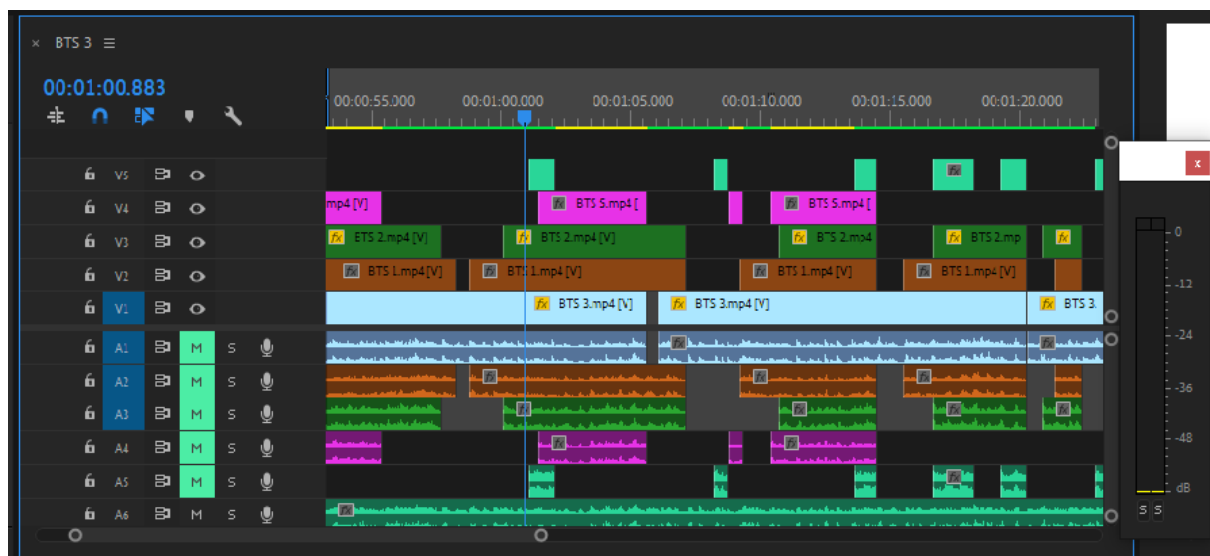
Από τη διαδικασία της μίξης παίρνουμε ένα ενιαίο ηχητικό αποτέλεσμα στο οποίο συμμετέχουν όλα τα UGVs και το οποίο θεωρητικά, έχει βελτιωμένα ποιοτικά χαρακτηριστικά. Αυτό το ηχητικό αρχείο, τοποθετείται στον χρονικό άξονα ως το μοναδικό ακουστικό αρχείο που θα αναπαράγεται στο τελικό βίντεο. Στη συνέχεια, αφού τοποθετηθούν όλα τα βίντεο στην σωστή χρονική στιγμή, γίνεται το μοντάζ, έχοντας πάντα υπόψιν, το ακουστικό μέρος, να αξιοποιείται συνδυαστικά με το οπτικό μέρος, επιλέγοντας να γίνουν εναλλαγές των βίντεο, σε “σημαντικές” χρονικές στιγμές όπου έχουν κάποιο ρυθμικό ενδιαφέρον, ή κάποτε να γίνονται εναλλαγές σε χρονικές στιγμές αναγκαστικά, με αφορμή, κάποιο όχι και τόσο καλής ποιότητας απόσπασμα βίντεο στην συγκεκριμένη χρονική στιγμή. Έτσι, εναλλάσσεται η οπτική πληροφορία ανάμεσα σε περίπου πέντε διαφορετικά UGRs βίντεο, χωρίς να χάνεται η χρονική συνοχή.



Για να έχουμε ένα μέτρο σύγκρισης σχετικά με το πόσο η συνεργατική επεξεργασία βελτιώνει την εμπειρία του θεατή σε σχέση με μεμονωμένες καταγραφές, δημιουργήθηκε και μία μεμονωμένη εκδοχή, η οποία καλύπτει το ίδιο ακριβώς χρονικό απόσπασμα με αυτό της συνεργατικής εκδοχής, αλλά «κλειδώνοντας» τη γωνία λήψης σε μία μόνο από τις διαθέσιμες κάμερες. Επίσης, για τις μεμονωμένες εκδοχές χρησιμοποιήθηκε ο αυθεντικός ήχος του βίντεο, χωρίς αυτός να αντικατασταθεί από το αποτέλεσμα της μίξης.

Όσον αφορά το βίντεο, χωρίς να συμπεριλαμβάνουμε τον ήχο, η οπτική πληροφορία έχει δεχθεί μια επεξεργασία για λειτουργικούς σκοπούς όπως, zoom out, zoom in και επεξεργασία του frame rate conversation. Δεν έγινε ωστόσο κάποια άλλη παρέμβαση για αισθητικούς λόγους.

Παρακάτω, στην εικόνα 1, βλέπουμε screenshot την στιγμή του edit για το τελικό συνεργατικό βίντεο στο πρόγραμμα Adobe premiere, ενώ στον πίνακα 3 παρατίθενται τα links από τα τελικά συνεργατικά βίντεο που δημιουργήθηκαν.



Εικόνα 1: Screenshot από την τη διαδικασία του τελικού μοντάζ.

Πίνακας 3: Links από τα τελικά συνεργατικά βίντεο.

|                      |                       |
|----------------------|-----------------------|
| <b>Lex</b>           | <a href="#">Video</a> |
| <b>Billie Eilish</b> | <a href="#">Video</a> |

|                    |                       |
|--------------------|-----------------------|
| <b>BTS</b>         | <a href="#">Video</a> |
| <b>Linkin Park</b> | <a href="#">Video</a> |

**Πίνακας 4: Links από τα μεμονωμένα βίντεο μικρής διάρκειας.**

|                      |                       |
|----------------------|-----------------------|
| <b>Lex</b>           | <a href="#">Video</a> |
| <b>Billie Eilish</b> | <a href="#">Video</a> |
| <b>BTS</b>           | <a href="#">Video</a> |
| <b>Linkin Park</b>   | <a href="#">Video</a> |

## 4 ΥΠΟΚΕΙΜΕΝΙΚΗ ΑΞΙΟΛΟΓΗΣΗ

Για την αξιολόγηση της μεθοδολογίας, σχεδιάσαμε ερωτηματολόγιο με πολυμεσικό υλικό, το οποίο διανεμήθηκε στο ευρύ κοινό. Παρακάτω αναλύονται οι παράγοντες που ελήφθησαν υπόψιν για τον σχεδιασμό του ερωτηματολογίου καθώς και τα αποτελέσματα που προέκυψαν από την επεξεργασία των απαντήσεων.

### 4.1 Υποθέσεις

Οι υποθέσεις που προσπαθούμε να επιβεβαιώσουμε ή να διαψεύσουμε μέσω της υποκειμενικής αξιολόγησης είναι οι εξής:

**Y1:** Η συνεργατική επεξεργασία των βίντεο από το ίδιο συναυλιακό γεγονός οδηγεί σε εμπειρία θέασης που είναι βελτιωμένη σε σχέση με τα μεμονωμένα βίντεο

**Y2:** Η συνεργατική επεξεργασία των βίντεο από το ίδιο συναυλιακό γεγονός οδηγεί σε εμπειρία ακρόασης που είναι βελτιωμένη σε σχέση με τα μεμονωμένα βίντεο

**Y3:** Η συνεργατική επεξεργασία των βίντεο από το ίδιο συναυλιακό γεγονός βελτιώνει την εμπειρία του χρήστη γενικά, σε σχέση με τα μεμονωμένα βίντεο.

Για να εξακριβωθεί κατά πόσο επιτυγχάνεται ο σκοπός αυτός, το ερωτηματολόγιο θα έπρεπε να επιτρέψει τη σύγκριση μεταξύ μεμονωμένων και συνεργατικών βίντεο, περιμένοντας τα αποτελέσματα να είναι αδιαμφισβήτητα υπέρ των συνεργατικών καταγραφών.

Έτσι, με σκοπό αυτό, περιμένουμε όχι μόνο η ποιότητα του ήχου να φθάνει σε ένα πολύ καλύτερο αποτέλεσμα (Y2), αλλά επίσης και η οπτική πληροφορία με τις διαφορετικές οπτικές γωνίες, να ξεπερνά σε εμπειρία το υπάρχον μεμονωμένο βίντεο (Y1). Τέλος, αναμένουμε ότι η συνολική εμπειρία του χρήστη (Y3) θα είναι βελτιωμένη κατά πολύ από το να έβλεπε μεμονωμένες καταγραφές στο διαδίκτυο.

### 4.2 Διατύπωση των ερωτήσεων

Ιδιαίτερα σημαντικό, οι ερωτηθέντες να υποβληθούν σε μια σειρά ερωτήσεων, οι οποίες δε θα προκαταβάλουν το χρήστη ως προς το να ευνοήσουν με τις απαντήσεις τους μεμονωμένες ή συνεργατικές καταγραφές. Πρέπει να σημειωθεί ότι πουθενά στο ερωτηματολόγιο δεν

αναφέρεται ο όρος «μεμονωμένος/-η» ή «συνεργατικός/ή» και πουθενά δεν γίνεται αναφορά στις τεχνικές/διαδικασίες που ακολουθήθηκαν για τη δημιουργία των συνεργατικών βίντεο. Επιπλέον, δεν γίνεται πουθενά αναφορά στην προέλευση των βίντεο ή στους όρους UGC ή UGV, κάτι που θα μπορούσε να προκαταλάβει το χρήστη αναφορικά με την ποιότητα του πολυμεσικού υλικού που καλείται να καταναλώσει. Έτσι, ο χρήστης ζητείται γενικά να πει τη γνώμη του για κάποια βίντεο που θα δει.

#### **4.2.1 Πρώτο μέρος ερωτήσεων**

Τα ερωτηματολόγια που δημιουργήσαμε και επεξεργαστήκαμε μέσω του Google Forms χωρίζονται σε δύο μέρη. Στο πρώτο μέρος, ο χρήστης καλείται να δει ένα απόσπασμα βίντεο που ανήκει σε μεμονωμένη καταγραφή και ένα που ανήκει σε συνεργατική επεξεργασία. Τα δύο αυτά βίντεο είναι από διαφορετικά γεγονότα. Ο χρήστης καλείται εδώ να αξιολογήσει το κάθε ένα βίντεο χωριστά από το άλλο. Η πρώτη ερώτηση εξυπηρετεί την υπόθεση Y1 και αφορά τον σχολιασμό της ποιότητας οπτικής πληροφορίας με τις επιλογές:

1. Κακή ερασιτεχνική.
2. Μέτρια ερασιτεχνική.
3. Πολύ καλή ερασιτεχνική.
4. Επαγγελματική.

Στη δεύτερη ερώτηση, εξυπηρετείται η υπόθεση Y2 και οι ερωτηθέντες καλούνταν να σχολιάσουν την ποιότητα του ήχου από το βίντεο, με τις ίδιες επιλογές όπως παραπάνω.

Στην τρίτη ερώτηση εξυπηρετείται η υπόθεση Y3 και ο χρήστης σχολιάζει αναφορικά με τη συνολική εμπειρία. Κρίθηκε σκόπιμο να δοθούν εδώ πέντε επιλογές, καθότι η αξιολόγηση της εμπειρίας χρήστη συνήθως γίνεται από μία κλίμακα από 1 έως 5. Οι επιλογές είναι:

1. Καθόλου ικανοποιητική
2. Λίγο ικανοποιητική
3. Μέτρια
4. Αρκετά ικανοποιητική
5. Πολύ ικανοποιητική.

Οι παραπάνω ερωτήσεις θέτονται τόσο για τη μεμονωμένη καταγραφή όσο και για το συνεργατικό βίντεο, όπου η σειρά εμφάνισης των βίντεο, αν θα είναι δηλαδή πρώτα η μεμονωμένη καταγραφή και μετά η συνεργατική ή το αντίθετο, ήταν διαφορετική σε κάθε μία από τις τέσσερις εκδοχές ερωτηματολογίου.

#### 4.2.2 Δεύτερο μέρος

Στο δεύτερο μέρος, είχαμε πάλι δύο πολυμεσικά αρχεία που έπρεπε να δει ο χρήστης, όμως το κάθε πολυμεσικό αρχείο αποτελείτο από δυο βίντεο από το ίδιο γεγονός, όπου το ένα ήταν συνεργατικό και το άλλο μεμονωμένο, με τυχαία όμως σειρά. Στην πρώτη ερώτηση, αξιολογείται η γενική προτίμηση (Y3) των ερωτηθέντων με 5 επιλογές ως εξής:

1. Η Β(συνεργατική) πολύ καλύτερη από την Α(μεμονωμένη).
2. Η Β(συνεργατική) λίγο καλύτερη από την Α(μεμονωμένη).
3. Και τα δύο το ίδιο.
4. Η Α(μεμονωμένη) λίγο καλύτερη από την Β(συνεργατική).
5. Η Α(μεμονωμένη) πολύ καλύτερη από την Β(συνεργατική).

Στην δεύτερη ερώτηση, καλούνταν οι ερωτηθέντες να επιλέξουν ποια εκδοχή από τις δυο(Α ή Β) είχε καλύτερο ήχο (Y2), έχοντας τις ίδιες πέντε πιθανές απαντήσεις με το προηγούμενο ερώτημα.

Για να κρατήσουμε αμείωτο το ενδιαφέρον, σε όλες τις περιπτώσεις του ερωτηματολογίου, επιλέξαμε μικρά αποσπάσματα από τις συναυλίες τα οποία δεν υπερέβαιναν το ένα λεπτό το καθένα. Για να αποφύγουμε την προκατάληψη των ερωτηθέντων, σχετικά με την επανάληψη των ίδιων διαδοχικών ερωτήσεων, δημιουργήσαμε 4 διαφορετικές εκδοχές ερωτηματολογίων. Βασικός λόγος, οι ερωτηθέντες να μην επηρεαστούν από τις προηγούμενες απαντήσεις τους. Ως εκ τούτου, κανένα από τα δύο συναυλιακά γεγονότα που παρουσιάζονται στο πρώτο μέρος του ερωτηματολογίου δεν εμφανιζόταν στο δεύτερο μέρος του ερωτηματολογίου. Με άλλα λόγια, κάθε βίντεο οντότητα που οι ερωτηθέντες καλούνταν να δουν εμφανιζόταν μόνο μία φορά. Στην αρχή αλλά και στα μέσα του ερωτηματολογίου, προτείναμε και υπενθυμίζαμε τους ερωτηθέντες να παρακολουθήσουν το οπτικοακουστικό υλικό με ακουστικά και να το παρακολουθήσουν επίσης στην καλύτερη δυνατή ποιότητα (1080p) και σε full screen. Στην εικόνα 2, φαίνεται ένα παράδειγμα για την μορφή των ερωτήσεων στο πρώτο μέρος, ενώ στην εικόνα 3 βλέπουμε την μορφή των ερωτήσεων στο δεύτερο μέρος.



Παρακολουθήστε το παραπάνω βίντεο και απαντήστε στις ακόλουθες ερωτήσεις:  
Σχολιάστε την ποιότητα της οπτικής πληροφορίας από το συγκεκριμένο βίντεο:

- Επαγγελματική
- Πολύ καλή ερασιτεχνική
- Μέτρια Ερασιτεχνική
- Κακή Ερασιτεχνική

**Εικόνα 2: Δείγμα δομής ερωτηματολογίου από το πρώτο μέρος.**



Παρακολουθήστε το παραπάνω βίντεο και απαντήστε στις ακόλουθες ερωτήσεις:  
Σχολιάστε την προτίμησή σας γενικά σχετικά με τις δύο εκδοχές A και B: \*

- Η B πολύ καλύτερη από την A
- Το B λίγο καλύτερη από την A
- Και οι δύο το ίδιο
- Η A λίγο καλύτερη από τη B
- Η A πολύ καλύτερη από τη B

**Εικόνα 3: Δείγμα δομής ερωτηματολογίου από το δεύτερο μέρος.**

Στον παρακάτω πίνακα 5, δίνονται μέσω links τα 4 διαφορετικά ερωτηματολόγια που χρησιμοποιήθηκαν.

Πίνακας 5: Όλα τα ερωτηματολόγια που διατέθηκαν στο κοινό.

| 1 <sup>ο</sup> Ερωτηματολόγιο | 2 <sup>ο</sup> Ερωτηματολόγιο | 3 <sup>ο</sup> Ερωτηματολόγιο | 4 <sup>ο</sup> Ερωτηματολόγιο |
|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| <a href="#">(1)</a>           | <a href="#">(2)</a>           | <a href="#">(3)</a>           | <a href="#">(4)</a>           |

### 4.3 Επεξεργασία των απαντήσεων και συμπεράσματα

Το ερωτηματολόγιο απαντήθηκε συνολικά από 37 άτομα. Παρακάτω θα δούμε αναλυτικά τις απαντήσεις που πήραμε από τα μεμονωμένα και τα συνεργατικά βίντεο σε μορφή διαγραμμάτων πίτας. Στη συνέχεια θα ακολουθήσουν τα συμπεράσματα.

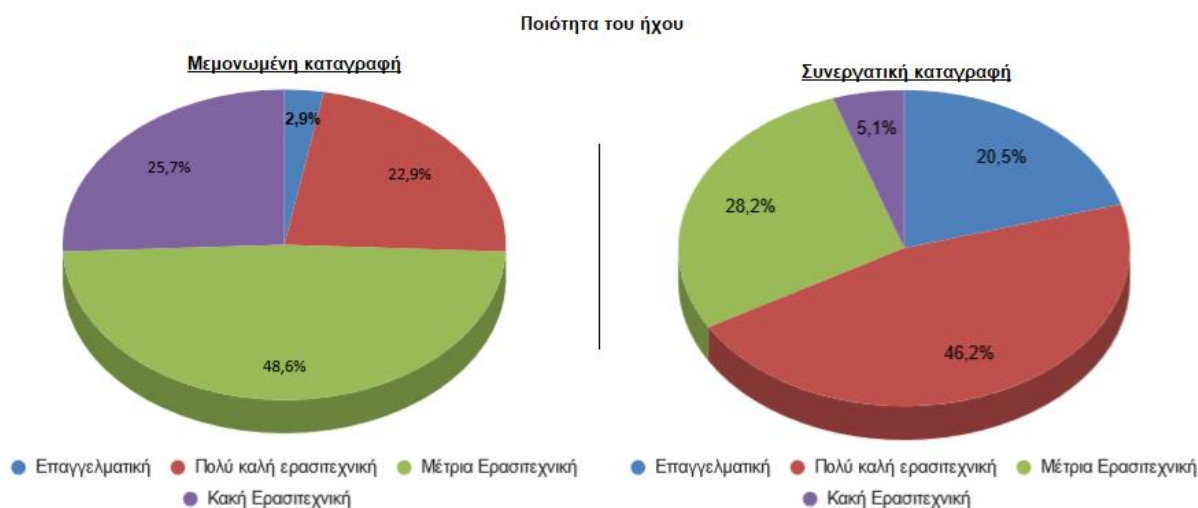
#### 4.3.1 Αποτελέσματα από το πρώτο μέρος

Η πρώτη ερώτηση λοιπόν, αφορούσε τα μονά βίντεο στο πρώτο μέρος και το κοινό καλούνταν να επιλέξει μια απάντηση όσον αφορά την οπτική πληροφορία όπως βλέπουμε στο σχήμα 4.



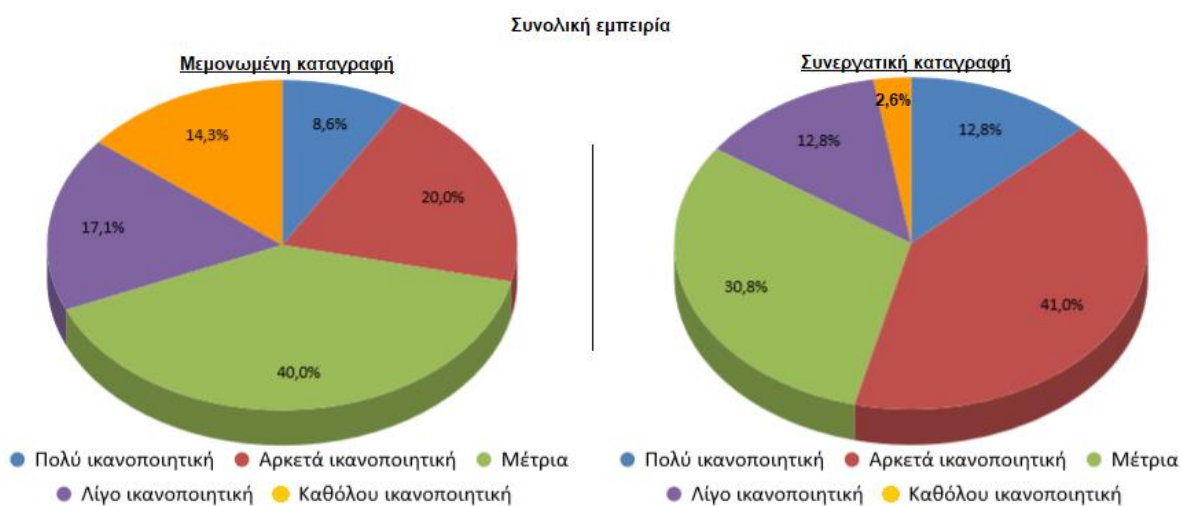
Σχήμα 4: Κατανομή των απαντήσεων όσον αφορά την οπτική πληροφορία για τις μεμονωμένες (αριστερά) και τις συνεργατικές (δεξιά) καταγραφές.

Στο επόμενο σχήμα 5, βρίσκονται οι απαντήσεις των ερωτηθέντων, όσον αφορά την ποιότητα του ήχου για τα μεμονωμένα και τα συνεργατικά βίντεο.



**Σχήμα 5:** Κατανομή των απαντήσεων όσον αφορά την ποιότητα του ήχου σε μεμονωμένες (αριστερά) και συνεργατικές (δεξιά) καταγραφές.

Τέλος, το Σχήμα 6 περιλαμβάνει τα αποτελέσματα του ερωτηματολογίου για την συνολική εμπειρία των ερωτηθέντων, για τα μεμονωμένα και τα συνεργατικά βίντεο.



**Σχήμα 6:** Κατανομή των απαντήσεων όσον αφορά την συνολική εμπειρία σε μεμονωμένες (αριστερά) και συνεργατικές (δεξιά) καταγραφές.

Όπως παρατηρήθηκε, ένα ποσοστό του 28,6%, θεωρεί την ποιότητα της οπτικής πληροφορίας στα μεμονωμένα βίντεο, από πολύ καλή ερασιτεχνική έως επαγγελματική, την

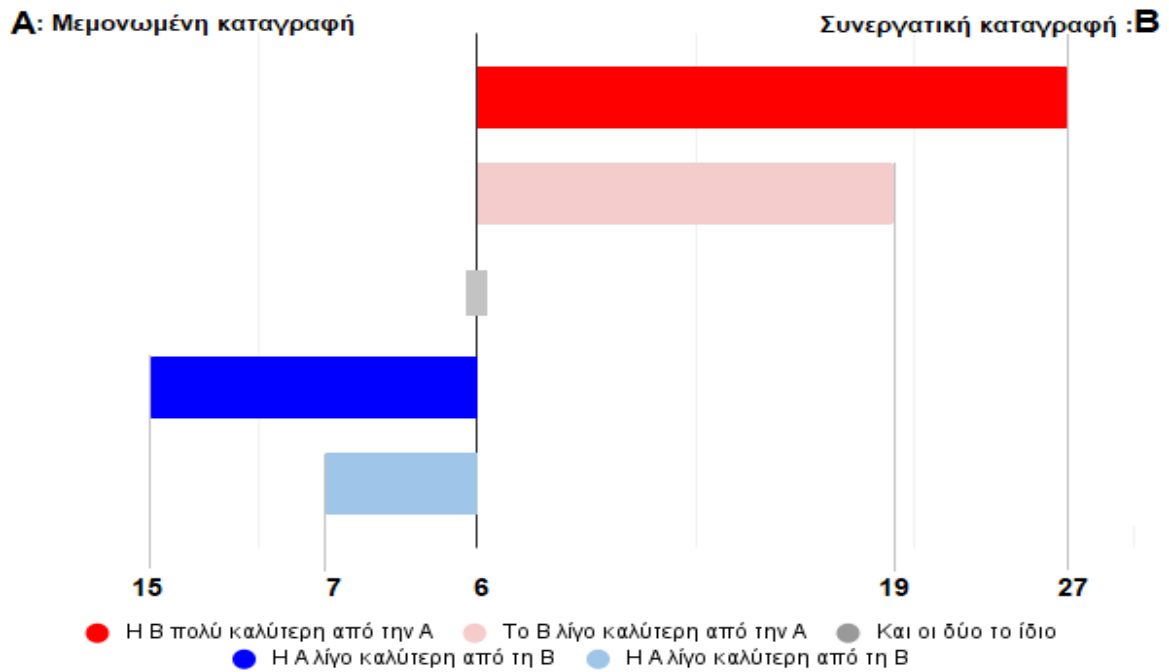


ίδια ώρα που στα συνεργατικά βίντεο, το ποσοστό αυτό ανέρχεται στο 64,1%. Επίσης, στις μεμονωμένες καταγραφές το ποσοστό που θεωρεί την ποιότητα του ήχου, κακή έως μέτρια ερασιτεχνική ανέρχεται στο 74,3%, τη στιγμή που το ίδιο ποσοστό στα συνεργατικά βίντεο είναι 33,3%.

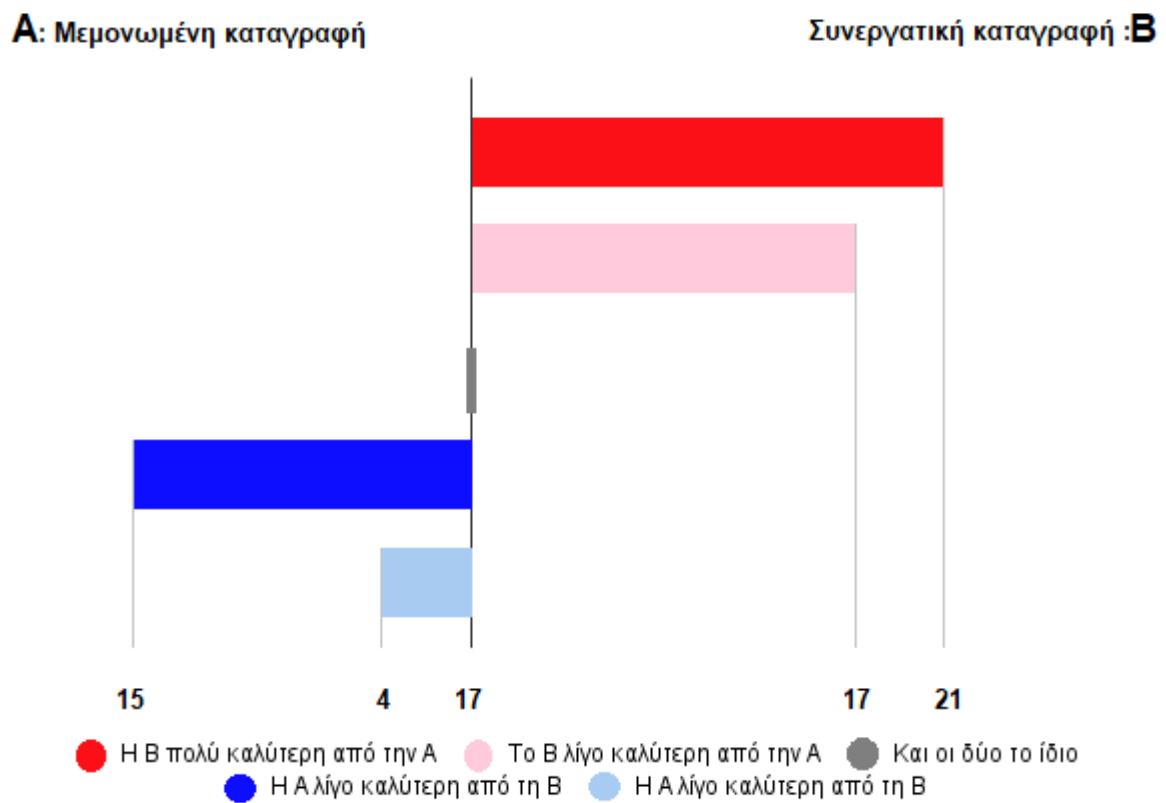
Ένα βίντεο λοιπόν, δεν μπορούμε να το κρίνουμε εξ ολοκλήρου από την οπτική του πλευρά, ή από την ηχητική. Η συνολική εμπειρία είναι αυτή που θα καθορίσει και το τελικό αποτέλεσμα, αφού έχουμε να κάνουμε με συνδυασμό πραγμάτων, ήχου και εικόνας. Στα μεμονωμένα βίντεο, το ποσοστό λοιπόν που σχολίασε την συνολική του εμπειρία αρκετά έως πολύ ικανοποιητική ήταν το 28,6%, ενώ για τα συνεργατικά βίντεο το ίδιο ποσοστό ήταν 53,8%. Βλέπουμε λοιπόν, ότι σε γενική προτίμηση, τα συνεργατικά βίντεο ανταποκρίνονται στον σκοπό δημιουργίας τους και προσφέρουν, βάση αποτελεσμάτων μια εμφανώς βελτιωμένη εμπειρία σε σχέση με τα μεμονωμένα. Το πρώτο μέρος των ερωτηματολογίων φαίνεται να επιβεβαιώνει επομένως και τις τρεις υποθέσεις, Y1, Y2 και Y3.

#### **4.3.2 Αποτελέσματα από το δεύτερο μέρος**

Ιδιαίτερα σημαντικά ήταν τα αποτελέσματα που πήραμε από τους ερωτηθέντες από τις ερωτήσεις σύγκρισης ανάμεσα στις μεμονωμένες και στις συνεργατικές καταγραφές. Τα αποτελέσματα φαίνονται στα Σχήματα 7 και 8 σε μορφή bar charts, όπου Α, θεωρούμε τις αυθεντικές μεμονωμένες καταγραφές και όπου Β, θεωρούμε τις συνεργατικές καταγραφές που δημιουργήσαμε. Για το Σχήμα 7 παρουσιάζονται οι απαντήσεις αναφορικά με την γενική εμπειρία (Y3) και προτίμηση ενώ στο Σχήμα 8, οι απαντήσεις αναφορικά με την εμπειρία ακρόασης (Y2)



**Σχήμα 7:** Κατανομή απαντήσεων όσον αφορά την γενική προτίμηση ανάμεσα στις περιπτώσεις Α (Μεμονωμένη) και Β (Συνεργατική).



**Σχήμα 8:** Κατανομή απαντήσεων όσον αφορά την προτίμηση αναφορικά με τον ήχο ανάμεσα στις περιπτώσεις Α (Μεμονωμένη) και Β (Συνεργατική).

Όσον αφορά τη συνολική εμπειρία, εδώ το 36,5% θεώρησε ότι η συνεργατική καταγραφή είναι πολύ καλύτερη από την μεμονωμένη. Ωστόσο, ένα ποσοστό 9,5% θεωρεί τις μεμονωμένες καταγραφές καλύτερες από τις συνεργατικές. Ένα μικρότερο ποσοστό (8,1%) επιλέγει ότι δεν αντιλαμβάνεται διαφορά ανάμεσα στις δύο εκδοχές. Αναφορικά με την εμπειρία ακρόασης, 28,4% των απαντήσεων που δόθηκαν θεωρεί την συνεργατική καταγραφή πολύ καλύτερη από την μεμονωμένη, ενώ ένα 5,4% θεωρεί την μεμονωμένη πολύ καλύτερη από την συνεργατική. Οι περιπτώσεις που θεωρούν ότι δεν αντιλαμβάνονται διαφορά στον ήχο είναι 23%.

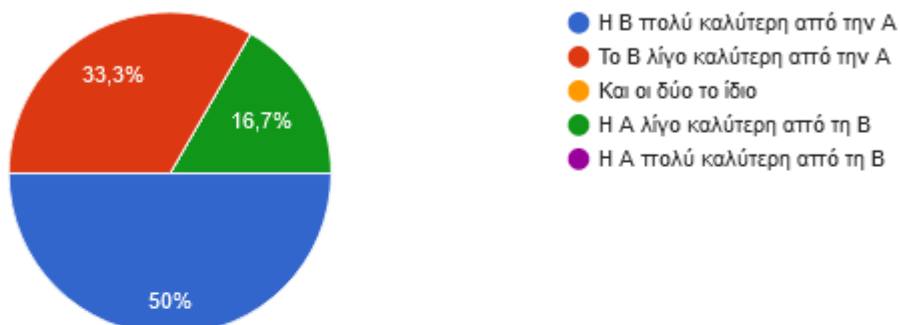
Εν τέλει, βλέπουμε στο δεύτερο μέρος η πλειοψηφία των απαντήσεων να είναι υπέρ των υποθέσεων Y2 και Y3. Ωστόσο, είναι επίσης φανερό ότι η προτίμηση των χρηστών δεν είναι πάντα υπέρ των συνεργατικών καταγραφών. Δυστυχώς δεν έχουμε στη διάθεσή μας στοιχεία όσον αφορά του λόγους που κάποιος προτιμάει τη μία ή την άλλη εκδοχή. Εκ των υστέρων, βλέπουμε ότι θα είχε μεγάλη αξία αν υπήρχε κάποια ερώτηση όπου ο χρήστης να μπορούσε αν ήθελε να αναφέρει τους λόγους που προτιμούσε τη μία ή την άλλη εκδοχή ή να αναφέρει πράγματα που τον ενόχλησαν.

### **4.3.3 Ιδιαίτερες περιπτώσεις**

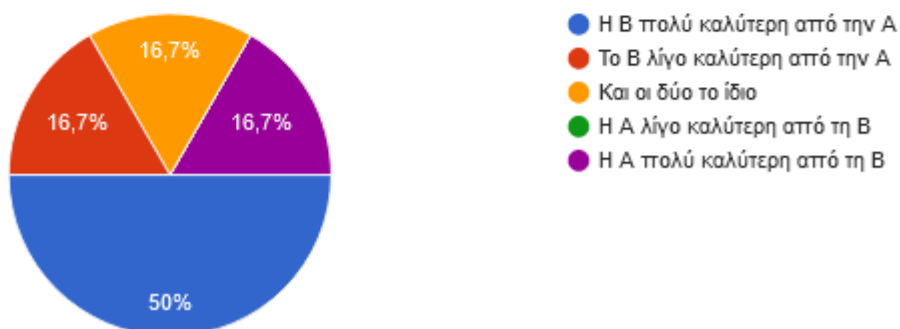
Μια περίπτωση που είχε τις περισσότερες ψήφους σε συμφωνία, ήταν στην συναυλία του Λεξ, με το 50% να συμφωνεί πως η συνεργατική καταγραφή είναι πολύ καλύτερη από την μεμονωμένη, όσον αφορά την γενική προτίμηση τους, αλλά και την ποιότητα στον ήχο. Τα αποτελέσματα των απαντήσεων εμφανίζονται παρακάτω στο Σχήμα 9.

Αφού τελειοποιήσαμε τα τελικά συνεργατικά βίντεο, συνειδητοποιήσαμε ότι ένα αρχείο βίντεο, από την συναυλία της Billie Eilish, δεν ήταν από την ίδια συναυλία, αλλά από μια συναυλία που είχε κάνει με διαφορετική ημερομηνία, διότι στο ίδιο τραγούδι (happier than ever) φορούσε διαφορετικά ρούχα. Παρόλο που ο συγχρονισμός του audio αρχείου με τα άλλα audio αρχεία πέτυχε, το match strength μας έδωσε μια τιμή λίγο πιο πάνω από το δέκα, συγκεκριμένα 10,6. Το κατώφλι συγχρονισμού είναι το 10, ο οποιοσδήποτε αριθμός κάτω από το 10 σημαίνει πολύ αδύναμο συγχρονισμό. Η εξήγηση που δίνουμε για τον συγχρονισμό πάνω από το 10, είναι ότι στην συγκεκριμένη συναυλία κάποια μουσικά όργανα ήταν προ ηχογραφημένα, με ένα ίδιο μοτίβο, με αποτέλεσμα οι διαφορετικές συναυλίες να συγχρονιστούν μεταξύ τους και ο αλγόριθμος να θεωρήσει ότι είναι από την ίδια συναυλία.

Σχολιάστε την προτίμησή σας γενικά σχετικά με τις δύο εκδοχές Α και Β:



Ποια εκδοχή κατά τη γνώμη σας έχει καλύτερο ήχο;



**Σχήμα 9: Οι απαντήσεις με την περισσότερη συμφωνία αναμεσα στις περιπτώσεις Β (Συνεργατική) και Α (Μεμονωμένη).**

## 5 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Η συνεργατική επεξεργασία των UGV αν και έχει οραματιστεί από αρκετούς ερευνητές, δεν έχει οδηγήσει μέχρι σήμερα σε κάποια ευρέως γνωστή υπηρεσία και οι δυνατότητες της δεν έχουν αξιολογηθεί λεπτομερώς. Αυτή η εργασία είναι μια πρώτη προσπάθεια πάνω σε αυτόν τον άξονα. Η διαδικασία που ακολουθήθηκε ήταν σχετικά απλή: ενιαίος ήχος από μίξη όλων των ηχητικών πηγών και μοντάζ στο οπτικό υλικό ώστε να έχουμε διαφορετικές οπτικές γωνίες. Ακόμα και σε αυτήν την απλή προσέγγιση η υποκειμενική αξιολόγηση κατέδειξε ότι κατορθώνεται σημαντική βελτίωση όσον αφορά την εμπειρία θέασης, την εμπειρία ακρόασης αλλά και τη συνολική εμπειρία.

Η συνεργατική αξιοποίηση των UGV από το ίδιο γεγονός θα μπορούσε να αξιοποιηθεί για να δημιουργηθούν υπηρεσίες που να εξυπηρετούν τόσο το απλό κοινό, όσο και τους επαγγελματίες καλλιτέχνες. Για παράδειγμα στην περίπτωση του απλού κοινού, κάποιος χρήστης θα μπορούσε να παρακολουθήσει μια συνεργατική καταγραφή σε περίπτωση που δεν κατάφερε να παρευρεθεί πραγματικά στην αγαπημένη του συναυλία. Με την συνεργατική καταγραφή λοιπόν, θα μπορεί να έχει μια πιο ολοκληρωμένη οπτικοακουστική εμπειρία, με πολλές γωνίες θέασης, που ακόμα και αν κατάφερνε να παρευρεθεί πραγματικά, δεν θα είχε την δυνατότητα να παρακολουθήσει το συναυλιακό γεγονός από τόσες διαφορετικές οπτικές γωνίες λόγω συνωστισμού.

Από την άλλη, όσων αφορά τους καλλιτέχνες, το διαδίκτυο έχει δώσει την ευκαιρία στον καθένα να αναδείξει την δουλειά του, χωρίς τις απαραίτητες γνωριμίες (όπως συνηθίζεται να συμβαίνει), αλλά και χωρίς την υποστήριξη δισκογραφικών εταιριών, τηλεοπτικών ή ραδιοφωνικών διαφημίσεων, τα οποία αν μη τι άλλο, φέρνουν τους καλλιτέχνες σε τεράστια έξοδα. Επομένως, η συνεργατική αξιοποίηση των UGV θα μπορούσε να παρέχει μια πιο οικονομική λύση για μουσικούς που θέλουν να προβληθούν αλλά ξεκινάνε τώρα την καριέρα τους και δεν έχουν την οικονομική άνεση να χρηματοδοτήσουν επαγγελματικές παραγωγές. Αυτό δεν αποκλείει βέβαια και το γεγονός ότι η αξιοποίηση των UGV έχει αξία και για τους καταξιωμένους καλλιτέχνες, δίνοντας τους τη δυνατότητα να ενσωματώσουν το επεξεργασμένο ερασιτεχνικό υλικό στο ήδη υπάρχων επαγγελματικής ποιότητας υλικό.

Μια ενδιαφέρουσα προοπτική θα ήταν να δημιουργηθεί μια εφαρμογή για κινητά και υπολογιστές, η οποία θα επιτρέπει στον χρήστη να περιηγηθεί σε έναν τρισδιάστατο χώρο, ο

οποίος θα αντικατοπτρίζει το κάθε συναυλιακό γεγονός ξεχωριστά. Με αυτόν τον τρόπο θα μπορούσε να παρέχεται η δυνατότητα σε άτομα τα οποία δεν μπόρεσαν να παρευρεθούν σε ένα μουσικό γεγονός, να παρακολουθήσουν την συναυλία από μια κινητή συσκευή και να νιώσουν, με τον καλύτερο δυνατό τρόπο, τον παλμό και το αίσθημα το ακροατηρίου. Για να επιτευχθεί αυτό, θα ήταν πολύ ενδιαφέρον ο χρήστης της εφαρμογής αυτής να μπορεί να περιηγηθεί εικονικά στον συναυλιακό χώρο και να μπορεί να επιλέξει την οπτική γωνία από την οποία θα ήθελε να παρακολουθήσει το γεγονός, ή να μπορεί εάν θέλει, να έχει και την επιλογή να ακούει τον ήχο από την συναυλία από διαφορετικές θέσεις. Για παράδειγμα, θα μπορεί να ακούει μόνο τον ήχο από ένα βίντεο στα μπροστά διαζώματα ή θα του παρέχεται η δυνατότητα να επιλέξει να ακούει μόνο τον ήχο από διαφορετική οπτική γωνία από τις πίσω θέσεις του ακροατηρίου. Αυτό θα μπορούσε να συναγωνιστεί την εμπειρία ακόμα και από φυσική παρουσία στη συναυλία, καθότι στην πράξη ένας ακροατής δεν μπορεί να αλλάξει θέση εύκολα αλλά και γρήγορα, ενώ επίσης, πολλές φορές οι μπροστινές θέσεις παρέχονται στους VIP καλεσμένους ή παρέχονται στο ευρύ κοινό με ιδιαίτερα αυξημένο κόστος εισιτηρίου.

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

- Bano S. and Cavallaro A. (2015), “Discovery and organization of multicamera user-generated videos of the same event,” *Journal of Information Sciences*, vol. 302, pp. 108–121.
- Cheong, H.J. and Morrison, M.A. (2008), “Consumers’ reliance on product information and recommendations found in UGC,” *Journal of interactive advertising*, 8(2), pp.38-49.
- Cotton C. and Ellis D. (2010) “Audio fingerprinting to identify multiple videos of an event,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 2386–2389.
- Haitsma J. and Kalker T. (2002), “A highly robust audio fingerprinting system,” in *International Society for Music Information Retrieval (ISMIR)*, pp. 107–115.
- J. Bryan, P. Smaragdis, and J. Mysore, (2012) “Clustering and synchronizing multi-camera video via landmark cross-correlation,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, pp. 2389 – 2392.
- Kennedy L. and Naaman M. (2009), “Less talk, more rock: Automated organization of community-contributed collections of concert videos,” in *Proceedings of the 18th international conference on World Wide Web*, pp. 311–320.
- M. Kim and P. Smaragdis, (2013) “Collaborative audio enhancement using probabilistic latent component sharing,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 896 – 900.
- M. Kim and P. Smaragdis, (2016) “Efficient neighborhood-based topic modelling for collaborative audio enhancement on massive crowdsourced recordings,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*.
- Shrestha P., Barbieri M., and Weda H. (2007), “Synchronization of multi-camera video recordings based on audio,” in *Proceedings of the 15th ACM international conference on Multimedia*, pp. 545–548.
- Stefanakis N, Mastorakis Y, Alexandridis A and Mouchtaris A, (2019) ‘‘Automatic Mixing of user-generated audio recordings from the same event’’, Undergraduate Thesis, Rethymno, Greece.

Stefanakis N. and Mouchtaris A. (2018), “Normalization of partly overlapping recordings from the same event based on relative signal powers,” in Proceeding of Acoustics, Speech and Signal Processing (ICASSP).

Stefanakis N., Chonianakis S. and Mouchtaris A., (2017) “Automatic matching and synchronization of user generated videos from a large scale sport event,” in Proceeding of Acoustics, Speech and Signal Processing (ICASSP).

Stefanakis N., Viskadouros M. and Mouchtaris A. (2017), “A subjective evaluation on mixtures of crowdsourced audio recordings”, in Proceeding of European Signal Processing Conference (EUSIPCO).

Valsamis Xristos (2018), “Αυτόματη μίξη ηχογραφήσεων παραγόμενες από χρήστες που παρακολουθούν το ίδιο δημόσιο γεγονός”, Rethymno, Greece July 2018.

Wang A. (2003), “An industrial-strength audio search algorithm,” in Proceedings of the 4th International Conference on Music Information Retrieval.