

ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΜΟΥΣΙΚΗΣ ΚΑΙ ΟΠΤΟΑΚΟΥΣΤΙΚΩΝ ΤΕΧΝΟΛΟΓΙΩΝ
Τμήμα Μουσικής Τεχνολογίας και Ακουστικής



Πτυχιακή εργασία

**Εργαλεία αυτόματης ανάλυσης ηχητικού σήματος
για την υποβοήθηση της διαδικτυακής μουσικής εκπαίδευσης.**

Ιωάννης Βίγλης (ΤΑ2087)

Επιβλέπουσα: Χρυσούλα Αλεξανδράκη
Επικουρη Καθηγήτρια

Ρέθυμνο, Σεπτέμβριος 2022

Ευχαριστίες

Οφείλω πρώτιστα θερμές ευχαριστίες στην επιβλέπουσα καθηγήτριά μου, Χρυσούλα Αλεξανδράκη, όχι μόνο για την ευκαιρία που μου έδωσε να ασχοληθώ με το συγκεκριμένο θέμα αλλά σίγουρα και για την άψογη συνεργασία, την εμπιστοσύνη και την ενθάρρυνση που μου πρόσφερε όλο το χρονικό διάστημα της εκπόνησής της εργασίας μου. Οι ιδέες και οι συμβουλές της ήταν ιδιαίτερα χρήσιμες και πολύτιμες.

Επίσης, θα ήθελα να ευχαριστήσω τον Μιχάλη Καλοχριστιανάκη, μέλος του ερευνητικού προσωπικού του Πανεπιστημίου Κρήτης και της ερευνητικής ομάδας του έργου MusiCoLab, για τη βοήθειά του στην υλοποίηση της εφαρμογής. Εκτός από τη διαχείριση των εξυπηρετητών που φιλοξενούσαν τον κώδικα που ανέπτυσσα και τις διορθώσεις που πρότεινε, η συμβολή του ήταν αναγκαία και καθοριστική στην ασύγχρονη διασύνδεση μεταξύ τμημάτων της εφαρμογής που ήταν υλοποιημένα σε διαφορετικές γλώσσες προγραμματισμού.

Ακόμη, θα ήθελα να ευχαριστήσω τον πρόσφατα απόφοιτο του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΛ.ΜΕ.ΠΑ., Δημήτρη Μήλιο, για την απλόχερη προσφορά της γνώσης και της εμπειρίας του στον προγραμματισμό πίσω από την πλατφόρμα MusiCoLab, της οποίας επέκταση είναι η εφαρμογή που αναπτύχθηκε στο πλαίσιο της εργασίας αυτής. Πέρα από τη διαχείριση των διαφορετικών εκδόσεων της εφαρμογής κατά την ανάπτυξή της, σημαντική ήταν η συνεισφορά του και στην διόρθωση προβλημάτων που αφορούσαν τη συνολική πλατφόρμα σε σχέση με τις προσθήκες και τροποποιήσεις που απαιτούσε η εφαρμογή.

Τέλος, θα ήθελα να ευχαριστήσω τον συμφοιτητή μου στο τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Αλέξανδρο Νούσια, για τη συνεισφορά του στη διασύνδεση μεταξύ της εφαρμογής απεικόνισης της μουσικής σημειογραφίας και της εφαρμογής συγχρονισμού που αναπτύχθηκε στο πλαίσιο αυτής της εργασίας αλλά και για τη γενικότερη συνεργασία μας καθώς αναπτύσσαμε ταυτόχρονα διαφορετικές εφαρμογές στην ίδια πλατφόρμα και η εμπειρία του που μοιράστηκε μαζί μου ήταν ιδιαίτερα χρήσιμη.

Περίληψη

Αυτή η εργασία στοχεύει στην ανάπτυξη μιας διαδικτυακής εφαρμογής, η οποία αξιοποιεί πρόσφατα ερευνητικά επιτεύγματα στην αυτόματη ανάλυση μουσικού περιεχομένου, προκειμένου να υποβοηθήσει τη διαδικτυακή μουσική εκπαίδευση.

Συγκεκριμένα, η εργασία διερεύνησε λειτουργίες ανάκτησης μουσικής πληροφορίας από ψηφιακά σήματα ήχου, τα οποία καταγράφονται ή παρουσιάζονται κατά τη διάρκεια ενός μαθήματος μουσικής. Τέτοιες πληροφορίες, συνήθως αφορούν στην ανίχνευση μελωδίας, ρυθμού ή τονικότητας ενός μουσικού κομματιού, καθώς και σε εργαλεία που μπορούν να συγχρονίζουν μουσική σημειογραφία (πχ παρτιτούρα) με τη ζωντανή μουσική εκτέλεση.

Λειτουργίες που παρέχονται από προγραμματιστικές βιβλιοθήκες τεχνολογίας αιχμής, όπως η Librosa και η Essentia διερευνώνται σε σχέση με τη βελτίωση της αποτελεσματικότητας των μαθημάτων μουσικής.

Το αποτέλεσμα αυτής της εργασίας είναι μια καινοτόμος διαδικτυακή εφαρμογή για μουσική εκπαίδευση η οποία είναι προσβάσιμη μέσω ενός προγράμματος περιήγησης στο διαδίκτυο (Internet browser). Η ανεπτυγμένη εφαρμογή, μπορεί να χρησιμοποιηθεί τόσο σε σύγχρονα μαθήματα μουσικής (ταυτόχρονη παρουσία εκπαιδευτή και εκπαιδευόμενου), όσο και σε ασύγχρονη μάθηση και διδασκαλία (πχ εξάσκηση μαθητή).

Εκτός από το ότι δεν απαιτείται η προηγούμενη εγκατάσταση λογισμικού, η εφαρμογή αυτή μπορεί να ενσωματωθεί και να χρησιμοποιηθεί μέσω Συστημάτων Διαχείρισης Μάθησης, όπως το Open eClass ή το Moodle.

Abstract

This work aims at the development of a web application, which investigates the use of recent research achievements in automatic audio signal content analysis in the context of online music education.

Specifically, the work investigates functionalities for retrieving music information from digital audio signals, which are recorded or presented during an online music lesson. Such information commonly concerns the detection of melody, rhythm or the tonality of a music piece, as well as tools that can synchronize music notation artifacts (e.g. scores) with live music performance.

Functionalities provided by state-of-the-art programming libraries such as Librosa and Essentia are investigated with respect to improving the efficacy of music lessons. The developed application may be used in synchronous music lessons (simultaneous presence of instructor and trainee) as well as asynchronous learning and teaching (e.g. student practice).

The outcome of this work is a novel web application for music education which is accessible through an Internet browser. Besides not requiring any prior software installation, the developed application may be integrated and used through Learning Management Systems, such as Open eClass or Moodle.

Πίνακας Περιεχομένων

1 Εισαγωγή.....	9
1.1 Στόχος.....	9
1.2 Τεχνολογίες e-Learning.....	9
1.3 Μουσική Εκπαίδευση από Απόσταση.....	10
1.4 Δομή Εργασίας.....	12
2 Πλατφόρμες Μουσικής Εκπαίδευσης.....	13
2.1 MusiCoLab.....	13
2.2 Yousician.....	16
2.3 Soundslice.....	19
2.4 Σύνοψη.....	22
3 Ανάλυση Μουσικού Περιεχομένου.....	24
3.1 Σύγχρονες Δυνατότητες.....	24
3.2 Προγραμματιστικές βιβλιοθήκες.....	25
3.2.1 Librosa.....	25
3.2.2 Essentia.....	27
3.3 Ιδέες και Προοπτικές για τη Μουσική Εκπαίδευση.....	30
4 Υλοποίηση Εφαρμογής.....	33
4.1 Αναλυτική Περιγραφή Λειτουργίας.....	33
4.2 Αρχιτεκτονική.....	38
4.3 Προγραμματιστικά εργαλεία.....	41
4.4 Γραφική Διεπαφή Χρήστη.....	42
4.5 Επιδόσεις Συγχρονισμού σε πλαίσιο μαθήματος.....	46
5 Συμπεράσματα.....	51
5.1 Σύνοψη.....	51
5.2 Μελλοντικές επεκτάσεις.....	51
5.2.1 Εκτίμηση συνολικής απόδοσης.....	51
5.2.2 Επισημάνση λαθών στο χρώμα ή το tempo.....	52
5.2.3 Απεικόνιση τρέχουσας νότας από ηχογράφηση.....	52
5.2.4 Συγχρονισμός σε πραγματικό χρόνο.....	53

6 Αναφορές.....54

1 Εισαγωγή

1.1 Στόχος

Η εργασία αυτή, έχει ως στόχο να διερευνήσει τις δυνατότητες που δίνονται από βιβλιοθήκες ανοιχτού κώδικα για επεξεργασία και ανάλυση ηχητικών δεδομένων και για την υποβοήθηση της διαδικτυακής μουσικής εκπαίδευσης. Παράλληλα, στοχεύει στην ανάπτυξη μιας καινοτόμου διαδικτυακής εφαρμογή για τη μουσική εκμάθηση η οποία είναι προσβάσιμη μέσω ενός προγράμματος περιήγησης στο διαδίκτυο.

1.2 Τεχνολογίες e-Learning

Η τεχνολογία διαδραματίζει κρίσιμο ρόλο στην καθημερινότητά μας και εκπαιδευτικοί, επαγγελματίες και εκπαιδευόμενοι, καλούνται να αναλογιστούν τις βασικές τους θεωρήσεις όσον αφορά τη χρήση της, ακόμα και στον εκ νέου σχεδιασμό της εκπαίδευσης και κατάρτισης. Οι σύγχρονες συσκευές μπορούν να προσφέρουν σημαντικό όφελος σε μαθητές και δασκάλους, ενισχύοντας τα μαθησιακά αποτελέσματα (Dias & Victor, 2017). Η εξ αποστάσεως διδασκαλία προσφέρει ευελιξία, διευκολύνοντας την επικοινωνία μεταξύ διδασκόμενου και διδάσκοντα. Η ταυτόχρονη χρήση πολυμέσων επαυξάνει τη μαθησιακή εμπειρία και είναι ήδη ενταγμένη σε εκπαιδευτικά συστήματα (Malik & Agarwal, 2012). Η εκπαίδευση από απόσταση μπορεί να πραγματοποιείται είτε σύγχρονα με απευθείας αλληλεπίδραση των συμμετεχόντων είτε ασύγχρονα με χρήση μαθησιακού υλικού από τους διδασκόμενους. Τις περισσότερες φορές, η σύγχρονη διδασκαλία συμπληρώνεται από την ασύγχρονη, καθώς το υλικό της διδασκαλίας και επιπλέον πληροφορίες μπορούν να δοθούν στους μαθητές μετά, ή ακόμα και νωρίτερα από το σύγχρονο μάθημα (Worthington, 2013). Επιπλέον, ο συνδυασμός της δια ζώσης εκπαίδευσης με την ασύγχρονη διδασκαλία, ο οποίος εφαρμόζεται περισσότερο στην τριτοβάθμια εκπαίδευση εδώ και αρκετά χρόνια, έχει σημαντικά πλεονεκτήματα, καθώς συνδυάζει τα οφέλη της δια ζώσης διδασκαλίας με αυτά που μπορούν να προσφέρουν οι τεχνολογίες ηλεκτρονικής μάθησης (Freeman & Field, 2004). Τα τελευταία χρόνια, λόγω και της πρόσφατης πανδημίας, οι τεχνολογίες σύγχρονης

εκπαίδευσης από απόσταση χρησιμοποιήθηκαν επί περίπου δύο σχολικά έτη, κατά κόρον και παγκοσμίως, σε όλες τις βαθμίδες της εκπαίδευσης με τα αποτελέσματα, ωστόσο, αυτής της εκπαιδευτικής διαδικασίας, να μένει να ερευνηθούν.

1.3 Μουσική Εκπαίδευση από Απόσταση

Η μουσική εκπαίδευση είναι μια δημιουργική διαδικασία με πολύ υψηλές απαιτήσεις σε νοητικές και κινητικές δεξιότητες. Παραδοσιακά, ένα μάθημα μουσικής γίνεται σε κλειστό χώρο και με την αλληλεπίδραση δασκάλου και μαθητή να αφορά τόσο στα δομικά (τονικότητα, ρυθμός, tempo, αρμονία) όσο και στα εκφραστικά στοιχεία ενός μουσικού έργου. Τα τεχνουργήματα, που συνήθως διαμεσολαβούν τη διεκπεραίωση ενός μουσικού μαθήματος, ποικίλουν και περιλαμβάνουν είτε αυτά που σχετίζονται με το περιεχόμενο της μουσικής (πχ αναπαράσταση μελωδίας σε πεντάγραμμο), είτε τη μουσική έκφραση και απόδοση (πχ μουσικό όργανο), είτε τεχνουργήματα επικοινωνίας και συντονισμού (πχ εντολές, μηνύματα, χειρονομίες κλπ.) των εμπλεκομένων. Παρόλη την πολυπλοκότητα αυτής της αλληλεπίδρασης, στις μέρες μας παρουσιάζεται ένα ολοένα αυξανόμενο ενδιαφέρον στη διαδικτυακή μουσική εκπαίδευση καθώς και στην ανεύρεση νέου μαθησιακού υλικού ενδιαφέροντος από ιστοσελίδες οπτικοακουστικού περιεχομένου.

Η διαδικτυακή μουσική εκπαίδευση εφαρμόζεται ήδη στην πράξη και υπάρχουν πολλά εκπαιδευτικά ιδρύματα, διαδικτυακές εκπαιδευτικές πλατφόρμες αλλά και μεμονωμένοι δάσκαλοι, που προσφέρουν μαθήματα σε όλες τις βαθμίδες της μουσικής εκπαίδευσης.

Οι ενδιαφερόμενοι χρήστες, μπορούν να αναζητήσουν ολοκληρωμένες μουσικές εκτελέσεις σε πλατφόρμες γενικού σκοπού (πχ YouTube) ή υποστηρικτικά μουσικά κομμάτια για εξάσκηση σε παρόχους μουσικού περιεχομένου, (πχ Aebersold Spotify Backing Tracks, Hal Leonard Play-along series, Music Minus One). Επίσης, μπορούν να έχουν πρόσβαση σε παρτιτούρες από σχετικές ψηφιακές συλλογές και βάσεις δεδομένων όπως οι MusicBrainz, Mutopia, MuseOpen και άλλες. Υπάρχει, ακόμα, πληθώρα μαθημάτων μουσικής που προσφέρονται από πλατφόρμες ανοιχτής μάθησης (MOOC, Massive Open Online Courses) όπως οι Udemy και Coursera. Επιπλέον, οι πλατφόρμες μουσικής εκπαίδευσης που έχουν δημιουργηθεί από ακαδημαϊκές πρωτοβουλίες (πχ Intermusic) βασίζονται στη λεγόμενη

μικτή μάθηση, που συνδυάζει τη συμβατική διδασκαλία σε αίθουσα με τη μαθησιακή αλληλεπίδραση μέσω διαδικτύου. Στην τελευταία περίπτωση, η εκπαίδευση διευκολύνεται από Συστήματα Διαχείρισης Μάθησης (LMS, Learning Management Systems) όπως για παράδειγμα το Moodle¹ και το OpenClass², τα οποία βοηθούν στην οργάνωση του διδακτικού υλικού και ταυτόχρονα παρέχουν τη δυνατότητα για την πραγματοποίηση σύγχρονης διδασκαλίας πχ με το BigBlueButton³. Τέλος, υπάρχουν εμπορικές πλατφόρμες για εκμάθηση μουσικής, όπως για παράδειγμα το Yousician⁴ που εστιάζει στην εκμάθηση απόδοσης μουσικών οργάνων για αρχάριους μαθητές, το Soundslice⁵ που παρέχει μια διαδικτυακή εφαρμογή που προσφέρει μαγνητοσκοπημένες αποδόσεις συγχρονισμένες με παρτιτούρες οι οποίες μπορούν να προσαρμοστούν σε ταχύτητα (tempo), το Genius Jamtracks⁶ που παρέχει ρυθμιζόμενη ρυθμική και αρμονική πολυπλοκότητα κομματιών Jazz και ούτω καθεξής (Alexandraki et al, 2022).

Τέλος, υπάρχουν μάλλον λίγες ακαδημαϊκές πρωτοβουλίες για τη μουσική εκμάθηση, οι περισσότερες από τις οποίες βασίζονται στη λογική τους στο παράδειγμα της μικτής μάθησης, γνωστή και ως υβριδική μάθηση, η οποία συνδυάζει ευκαιρίες για διαδικτυακή αλληλεπίδραση με συμβατικές τάξεις που βασίζονται σε χώρους. Για παράδειγμα, το έργο Intermusic⁷, που χρηματοδοτείται από το πρόγραμμα Erasmus+ και στοχεύει στη δημιουργία ευρωπαϊκών συνεργασιών στη μουσική εκμάθηση, ενώ άλλες πρωτοβουλίες στοχεύουν στην κατάρτιση περιπτώσιολογικών μελετών, πειραμάτων και αξιολογήσεων μικτής και αυτόνομης μουσικής εκμάθησης.

Εν ολίγοις, σε μια προσπάθεια ταξινόμησης διαδικτυακών δυνατοτήτων και πόρων εστιασμένων στη μουσική εκπαίδευση, μπορεί κανείς να διακρίνει διαδικτυακά αποθετήρια μουσικής, μαθήματα μουσικής που βασίζονται σε πλατφόρμες ανοιχτής μάθησης, εμπορικές

1 <https://moodle.org>

2 <https://www.openececlass.org>

3 <https://bigbluebutton.org>

4 <https://yousician.com>

5 <https://www.soundslice.com>

6 <https://geniusjamtracks.com>

7 <http://intermusicproject.eu>

πλατφόρμες για μουσική εκμάθηση και τέλος πλατφόρμες που αναπτύχθηκαν στο πλαίσιο ερευνητικών πρωτοβουλιών και ακαδημαϊκών συνεργασιών.

1.4 Δομή Εργασίας

Η παρούσα πτυχιακή εργασία έχει την ακόλουθη δομή:

Το Κεφάλαιο 2 επικεντρώνεται στην περιγραφή τριών διαφορετικών συστημάτων λογισμικού που στοχεύουν στη μουσική εκπαίδευση. Θα αναλυθούν οι δυνατότητες που παρέχει το σύστημα MusiCoLab⁸, που αναπτύσσεται στο πλαίσιο ενός χρηματοδοτούμενου ερευνητικού έργου, καθώς και οι εμπορικές πλατφόρμες Yousician και SoundSlice.

Στο Κεφάλαιο 3 γίνεται αναλυτική περιγραφή των προγραμματιστικών βιβλιοθηκών ανοιχτού κώδικα Librosa⁹ και Essentia¹⁰. Οι βιβλιοθήκες αυτές, υλοποιούν αλγόριθμους που μπορούν να επεξεργαστούν ή και να εξάγουν πληροφορίες από ηχητικά σήματα. Το κεφάλαιο ασχολείται με τη διερεύνηση της χρήσης αυτών των βιβλιοθηκών, για την ενίσχυση της εμπλοκής των συμμετεχόντων σε μια συνεδρία διαδικτυακής μουσικής εκπαίδευσης.

Στο Κεφάλαιο 4 περιγράφεται η υλοποίηση εφαρμογής που ενσωματώθηκε στην πλατφόρμα μουσικής εκπαίδευσης από απόσταση MusiCoLab. Η εφαρμογή αυτή, συγχρονίζει την παρτιτούρα που καλείται να εκτελέσει ο μαθητής με την ηχογραφημένη εκτέλεσή του. Γίνεται περιγραφή της ιδιαιτερότητας της χρήσης αλγορίθμων συγχρονισμού στο δεδομένο πλαίσιο χρήσης, της αρχιτεκτονικής, των εργαλείων υλοποίησης και της γραφικής διεπαφής χρήστη της εφαρμογής. Επίσης, αποτιμάται η απόδοση του συγχρονισμού και πώς αυτή διαφοροποιείται από τις συνθήκες μιας συνεδρίας μουσικής εκμάθησης.

Τέλος, στο Κεφάλαιο 5, συνοψίζονται τα αποτελέσματα της εργασίας και καταγράφονται πιθανές μελλοντικές επεκτάσεις που θα μπορούσαν να βελτιώσουν περαιτέρω την εμπειρία χρήσης τέτοιων εφαρμογών κατά τη μουσική εκπαίδευση.

8 <https://musicolab.hmu.gr>

9 <https://librosa.org>

10 <https://essentia.upf.edu/index.html>

2 Πλατφόρμες Μουσικής Εκπαίδευσης

2.1 MusiCoLab

Το MusiCoLab¹¹ είναι ένα ερευνητικό έργο που στοχεύει στην ανάπτυξη μιας ολοκληρωμένης τεχνολογικής πλατφόρμας, για την υποστήριξη της απομακρυσμένης μουσικής εκπαίδευσης σε πλαίσιο σύγχρονης και ασύγχρονης συνεργασίας (Alexandraki et al, 2022).

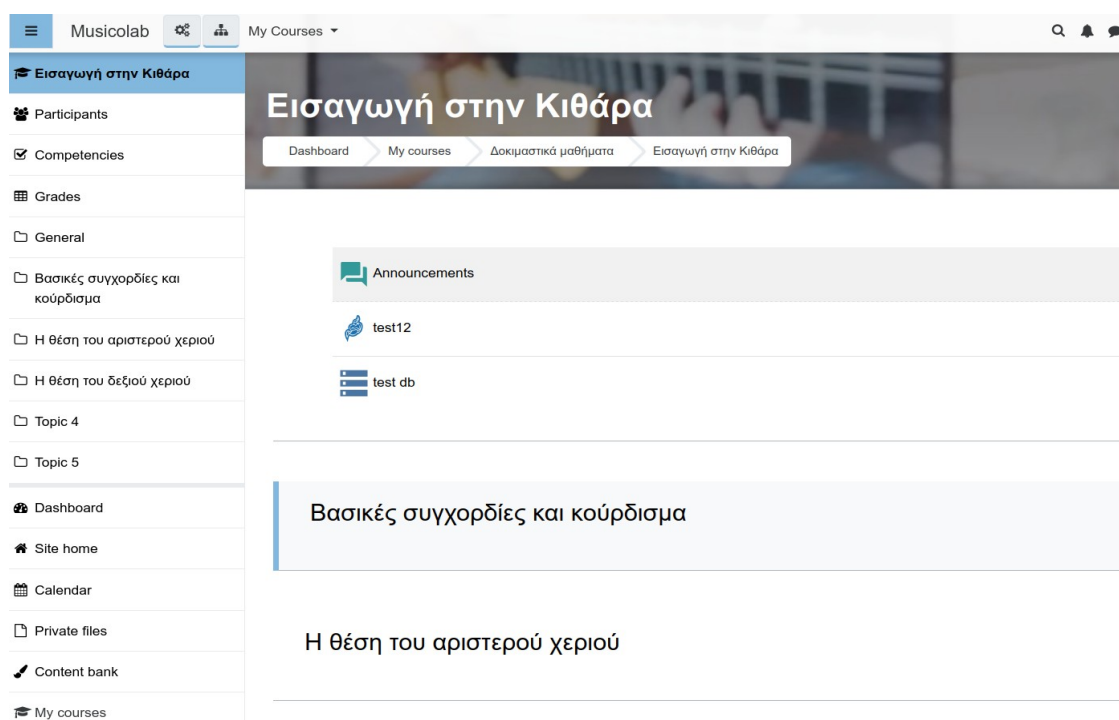
Η πλατφόρμα αυτή, περιλαμβάνει αποθετήριο ψηφιακών τεκμηρίων για τη μουσική που εμπλουτίζεται με πηγαίο υλικό και επισημειώσεις περιεχομένου, Σύστημα Διαχείρισης Μάθησης που εστιάζει στη μουσική εκπαίδευση, Σύστημα Τηλεδιάσκεψης που στοχεύει στην αποτελεσματική επικοινωνία με εικόνα και υψηλής πιστότητας ήχο, εργαλεία συνεργασίας μουσικών αναπαραστάσεων και μουσικολογικών δεδομένων και λογισμικό αυτόματης εναρμόνισης μελωδίας που θα μπορεί να αξιοποιηθεί σε ποικίλες εμπορικές εφαρμογές.

Η πλατφόρμα λειτουργεί σε περιβάλλον περιηγητή διαδικτύου και παρέχει το κατάλληλο περιβάλλον που ευνοεί τη συνεργασία των εμπλεκόμενων στην εκπαίδευση μερών (μαθητών και εκπαιδευτών). Η δομή της βασίζεται σε αρθρωτή αρχιτεκτονική όπου διάφορες ενότητες μπορούν να συνδυάζονται και να παρέχουν δυνατότητες είτε σύγχρονων συνεργατικών μαθημάτων είτε ασύγχρονης διδασκαλίας ή εξάσκησης.

Οι παιδαγωγικές μέθοδοι που υλοποιούνται, είναι αυτές του κοινωνικού κονστρουξιονισμού (social constructionism) και του κονστρουκτιβισμού (constructivism), όπου τα εμπλεκόμενα μέρη συνεργάζονται για να κατασκευάσουν αντικείμενα και η μάθηση προκύπτει από την αλληλεπίδραση με αυτά τα αντικείμενα (Knapp, 2019). Υποστηρίζεται η κατασκευή αντικειμένων, η αναζήτησή τους, ο διαμοιρασμός τους, η ανάλυσή τους κλπ, όπου τα αντικείμενα μπορούν να είναι είτε μουσικά αρχεία (αρχεία ήχου) είτε αρχεία μουσικής σημειογραφίας διαφόρων τύπων.

11 <https://musicolab.hmu.gr>

Μέσω της πλατφόρμας, οι εκπαιδευτές μπορούν ασύγχρονα να δομήσουν μαθήματα με τη βοήθεια ενός Συστήματος Διαχείρισης Μάθησης βασισμένο στο ελεύθερο λογισμικό Moodle¹² (Εικόνα 2.1), το οποίο παρέχει μηχανισμούς δημιουργίας μαθήματος, εγγραφής σε μάθημα, διαμοιρασμού αρχείων κλπ. Οι μαθητές μπορούν να αναζητήσουν τεχνουργήματα εκμάθησης μουσικής (MLAs, Music Learning Artifacts) που είναι σχετικά με ένα συγκεκριμένο μάθημα, τα οποία μπορεί να κωδικοποιούνται και αναπαρίστανται με μη-συμβατικές μορφές ψηφιακών εγγράφων (πχ παρτιτούρα, ταμπλατούρα, pianoroll, midi) ή να φτιάξουν δικά τους (πχ γράφοντας μια παρτιτούρα ή ηχογραφώντας ένα μουσικό κομμάτι).

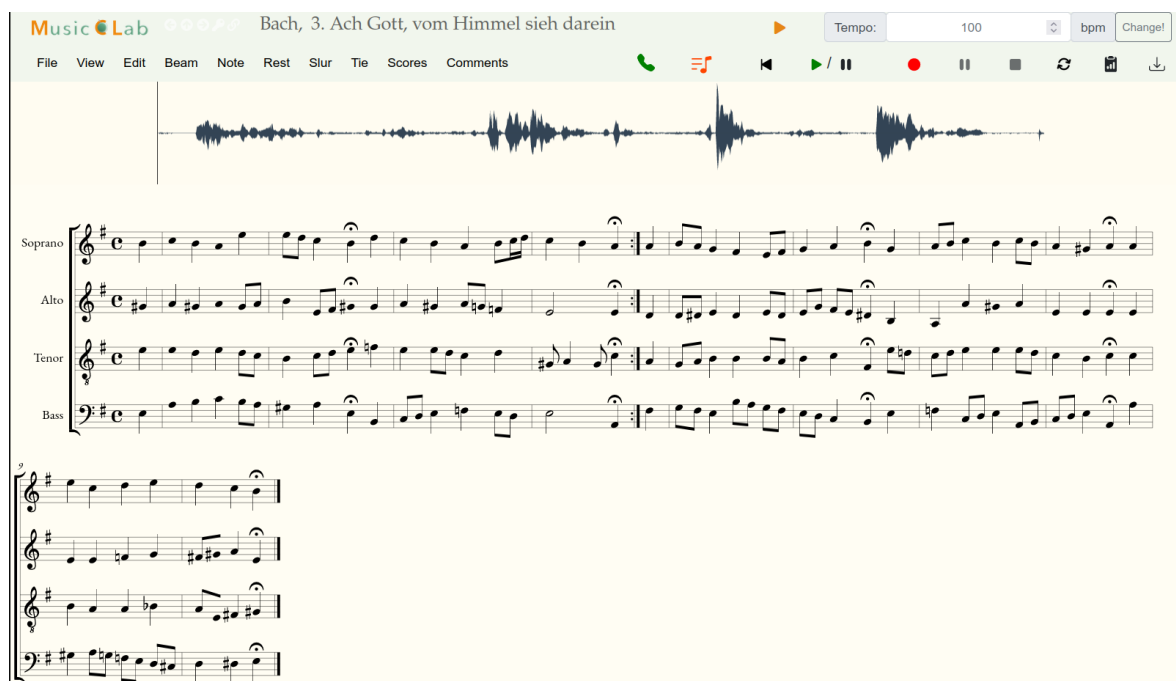


Εικόνα 2.1 - Περιβάλλον Συστήματος Διαχείρισης Μάθησης MusiCoLab (<https://musicolab.hmu.gr/moodle/>)

Οι μαθητές, με τη σειρά τους, μπορούν ασύγχρονα να ανακτούν αυτά τα MLAs, να ηχογραφούν δικές τους εκτελέσεις, να συγχρονίζουν αυτές τις ηχογραφήσεις με την αντίστοιχη παρτιτούρα και να κάνουν ανάλυση των ηχογραφήσεων για εξαγωγή συγχορδιών, ρυθμού, έκφρασης κλπ.

12 <https://moodle.org>

Στη σύγχρονη διδασκαλία, υπάρχει η δυνατότητα της δημιουργίας MLAs, της επαναδιαπραγμάτευσης και ανακατασκευής τους, του χειρισμού και της βελτίωσής τους. Μαθητές και καθηγητές μπορούν να επεξεργάζονται ταυτόχρονα παρτιτούρα, μπορούν να την αποδίδουν σε ήχο και να αλλάζουν συγχορδίες στην παρτιτούρα, χρησιμοποιώντας εργαλεία μηχανικής μάθησης για την αυτόματη παραγωγή (πολλών εναλλακτικών) εναρμονίσεων μιας δεδομένης μελωδίας. Η τελευταία λειτουργία είναι εστιασμένη σε μουσική Jazz. Μπορούν, επίσης, να ηχογραφούν τις δικές τους εκτελέσεις και να τις αναπαράγουν μαζί με τους δασκάλους τους, να συγχρονίζουν την ηχογράφιση με την παρτιτούρα ώστε να μπορούν να εντοπίζουν σημεία ενδιαφέροντος (Εικόνα 2.2), να αναλύουν τις ηχογραφήσεις ενώ ταυτόχρονα λειτουργεί σύστημα τηλεδιάσκεψης με εικόνα και ήχο βασισμένο στην ανοιχτού κώδικα πλατφόρμα Jitsi¹³ προσαρμοσμένο στις απαιτήσεις της μουσικής επικοινωνίας.

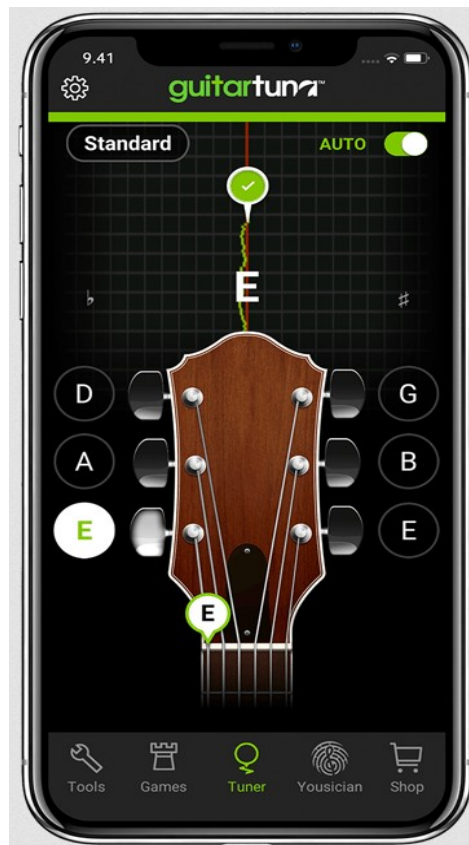


Εικόνα 2.2 - Περιβάλλον σύγχρονης διδασκαλίας MusiCoLab (<https://musicolab.hmu.gr/vhv/>)

13 <https://jitsi.org>

2.2 Yousician

Το Yousician¹⁴ είναι μια εμπορική πλατφόρμα διδασκαλίας της μουσικής, η οποία επίσης προσφέρει δυνατότητα σύγχρονης και ασύγχρονης διδασκαλίας. Οι εκπαιδευτικοί μπορούν να χρησιμοποιήσουν το Yousician για να παραδώσουν ιδιαίτερα μαθήματα σε μεμονωμένους μαθητές ή να πραγματοποιήσουν ομαδικά μαθήματα μουσικής. Μπορούν να χρησιμοποιήσουν ταυτόχρονα το Zoom, το Skype, το FaceTime ή παρόμοιες πλατφόρμες για τη δυνατότητα της τηλεδιάσκεψης. Τα μαθήματα είναι εστιασμένα σε συγκεκριμένα μουσικά όργανα (κιθάρα, μπάσο, πιάνο, γιουκαλίλι) και στο τραγούδι.

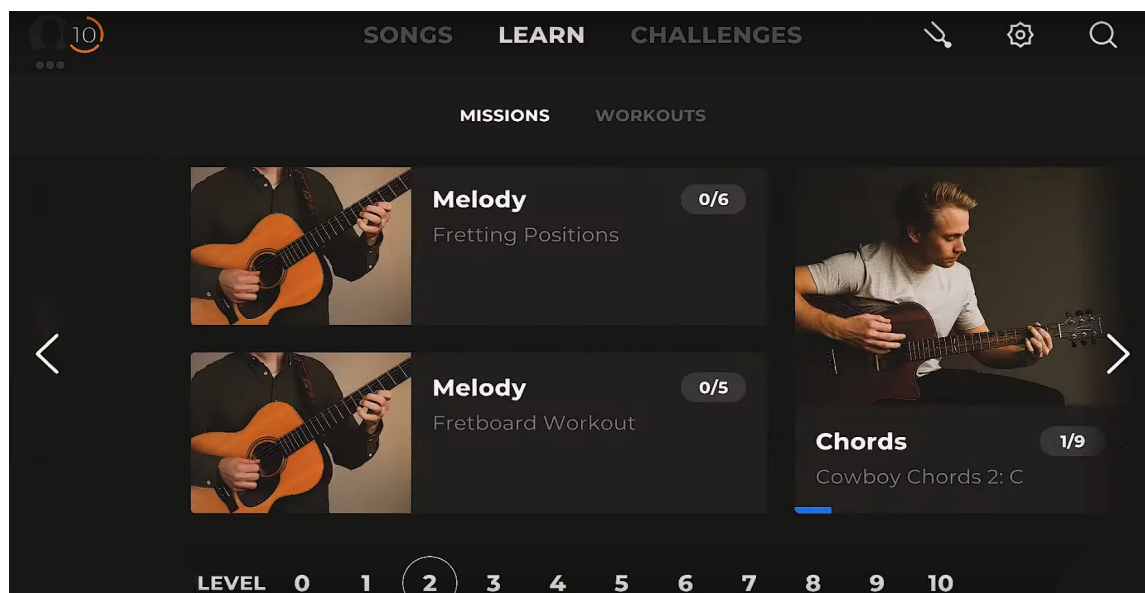


Εικόνα 2.3 - Χόρδισμα κιθάρας με την εφαρμογή Yousician (<https://yousician.com/guitar-tuner>)

14 <https://yousician.com>

Η χρήση της πλατφόρμας προϋποθέτει την εγκατάσταση εφαρμογής σε υπολογιστή με Windows ή MacOS ή σε κινητή συσκευή (tablet ή τηλέφωνο) με iOS ή Android. Μαθητές και εκπαιδευτικοί χρειάζεται να έχουν λογαριασμό στο Yousician. Αν πρόκειται για μαθήματα τραγουδιού, οι μαθητές μπορούν να βαθμονομήσουν το φωνητικό τους εύρος και να μεταφέρουν το τονικό ύψος ολόκληρου του κομματιού (transpose), προκειμένου να μπορούν να το ακολουθήσουν. Αν πρόκειται για μουσικό όργανο, μπορούν να το χορδίσουν με τη βοήθεια της εφαρμογής (Εικόνα 2.3).

Οι μαθητές επιλέγουν από τα διαθέσιμα μαθήματα, τα οποία έχουν κάποιο θέμα και αποτελούνται από δραστηριότητες - αποστολές, που καλούνται οι μαθητές να ολοκληρώσουν. Οι δραστηριότητες περιλαμβάνουν την εκμάθηση μελωδιών και solo, ενότητες που προσφέρουν γνώσεις μουσικής θεωρίας και γνώσεις πάνω στα μουσικά όργανα, ασκήσεις συγχορδιών, ρυθμού και συνοδείας κομματιού και άλλα (Εικόνα 2.4). Οι μαθητές μαθαίνουν όσο εκτελούν δραστηριότητες - αποστολές ενώ ταυτόχρονα “ξεκλειδώνουν” τις επόμενες. Μπορούν επίσης να παρακάμψουν θέματα και ασκήσεις που είναι χαμηλότερου επιπέδου από το δικό τους, πετυχαίνοντας σε τεστ δομημένα για αυτό το σκοπό, ξεκλειδώνοντας όλες τις προηγούμενες ενότητες.



Εικόνα 2.4 - Δραστηριότητες - Αποστολές στην εφαρμογή Yousician (<https://www.youtube.com/watch?v=EESFXkyn9Fs>)

Παράλληλα με τις παραπάνω δραστηριότητες - αποστολές, υπάρχουν δραστηριότητες εξάσκησης, κατηγοριοποιημένες σε διάφορα επίπεδα ικανοτήτων, που αφορούν για παράδειγμα την εξάσκηση σε συγχορδίες, την ακουστική αντίληψη, τις κλίμακες ή τη σημειογραφία και περιλαμβάνουν και σχετικά βίντεο εκμάθησης. Μια από τις δυνατότητες είναι ένα διαδραστικό παιχνίδι, στο οποίο καλείται ο μαθητής να ακολουθήσει με το μουσικό του όργανο, οπτικοποιημένο τμήμα ηχογραφημένης εκτέλεσης και βαθμολογείται η επίδοσή του.

Η πλατφόρμα, επίσης, περιλαμβάνει βάση δεδομένων με μουσικά κομμάτια κατηγοριοποιημένα ανά μουσικό είδος, ανά επίπεδο δυσκολίας και με το ποια ικανότητα του μαθητή θα μπορούσε να εξασκηθεί με το κάθε κομμάτι. Οι μαθητές μπορούν να αναζητήσουν κομμάτια και να επιλέξουν ανάμεσα σε εξάσκηση ή αναπαραγωγή του κομματιού. Στην εξάσκηση, δίνεται η δυνατότητα της προπόνησης σε συγκεκριμένο τμήμα του κομματιού, το οποίο για παράδειγμα σε κιθάρα, μπορεί να απεικονίζεται είτε με τη σωστή θέση των δαχτύλων στην ταστιέρα είτε με κλασική σημειογραφία (Εικόνα 2.5).



Εικόνα 2.5 - Περιβάλλον εξάσκησης με την εφαρμογή Yousician (<https://www.youtube.com/watch?v=AjcKdhFd8OI>)

Όσο περνάει ο χρόνος, η απεικόνιση κυλιέται ταυτόχρονα με την αναπαραγωγή του κομματιού. Τη στιγμή που ο μαθητής παίζει τη νότα, ανιχνεύεται η επιτυχία ή η αποτυχία του. Υπάρχει η δυνατότητα αλλαγής του tempo (ακόμα και το να αυξάνεται αυτόματα σταδιακά στο 100% αν ο μαθητής πετυχαίνει άριστο αποτέλεσμα στις προσπάθειές του). Υπάρχει, επίσης, η δυνατότητα μετρονόμου. Όταν ο μαθητής νιώσει έτοιμος, μπορεί να περάσει από την εξάσκηση στην πλήρη αναπαραγωγή του κομματιού, όπου πρέπει να ακολουθήσει σωστά όλο το κομμάτι και να βαθμολογηθεί για αυτό. Το διαδραστικό παιχνίδι του Yousician μοιάζει να κάνει την εξάσκηση αρκετά διασκεδαστική.

Αξίζει να σημειώσουμε, ότι στην πλατφόρμα παρέχονται βιντεοσκοπημένα μαθήματα (και ασκήσεις) από διάσημους καλλιτέχνες και μέλη συγκροτημάτων.

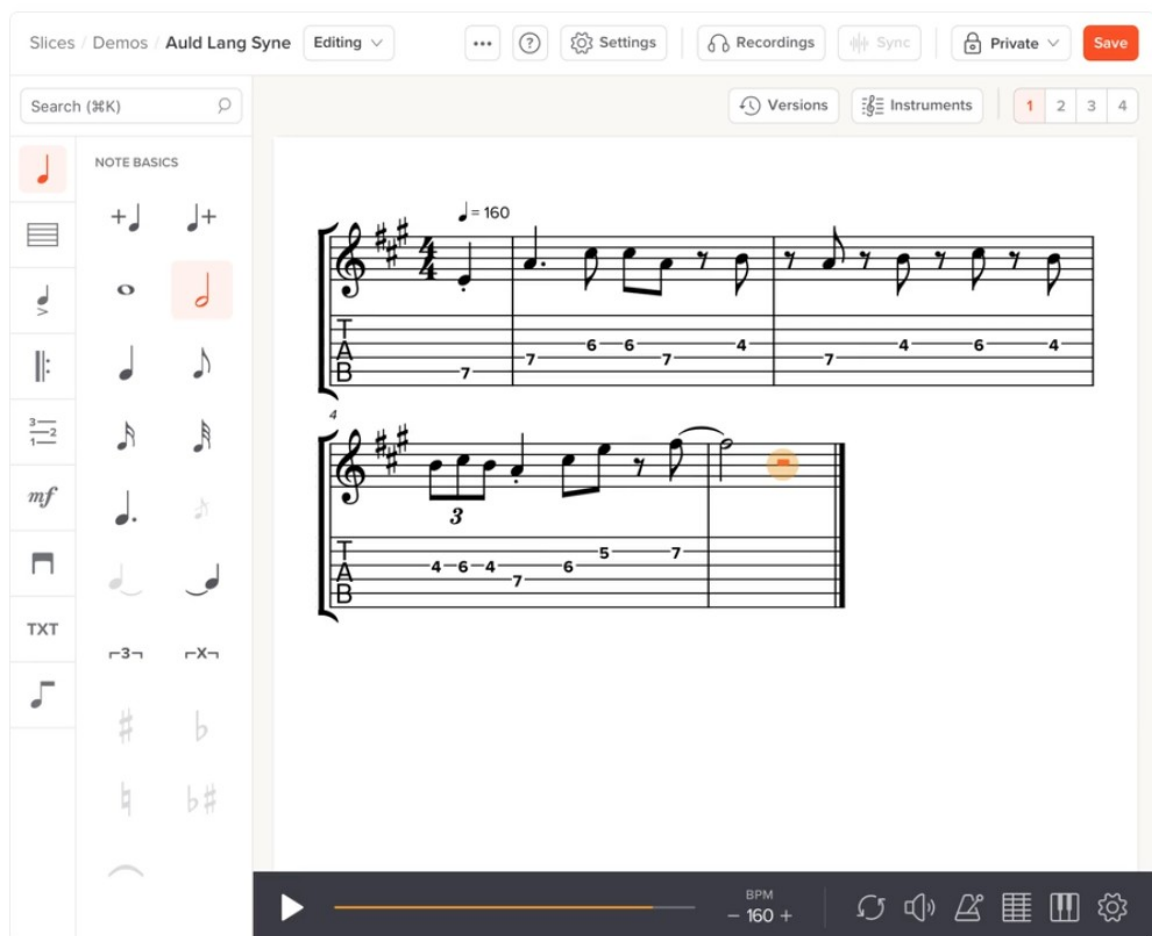
Η χρήση της πλατφόρμας Yousician μπορεί να γίνει και σε δια ζώσης μαθήματα. Οι καθηγητές μουσικής μπορούν να χρησιμοποιούν το Yousician κατά τη διάρκεια των μαθημάτων τους με μαθητές αλλά επίσης και για να καθοδηγήσουν την πρακτική των μαθητών στο σπίτι, μεταξύ των προγραμματισμένων μαθημάτων. Το μαθησιακό αποτέλεσμα προέρχεται από τον συνδυασμό αυτών των μεθόδων. Η ρύθμιση του Yousician για ένα μάθημα εξαρτάται από το μέγεθος της τάξης. Με έναν ή δύο μαθητές, οι δάσκαλοι μπορούν να έχουν το Yousician στον επιτραπέζιο υπολογιστή ή το tablet τους, έτσι ώστε οι μαθητές να μπορούν να δουν εύκολα την οθόνη. Για περισσότερους μαθητές στον ίδιο χώρο, υπάρχει το μειονέκτημα ότι δε μπορεί να εξατομικευτεί η ανάδραση του προγράμματος σε κάθε μαθητή.

2.3 Soundslice

Το Soundslice¹⁵ είναι μια επίσης εμπορική πλατφόρμα διδασκαλίας της μουσικής. Η δυνατότητα διδασκαλίας είναι ασύγχρονη και η χρήση της γίνεται σε περιβάλλον περιηγητή διαδικτύου χωρίς την απαίτηση συγκεκριμένου λειτουργικού συστήματος ή συσκευής.

15 <https://www.soundslice.com>

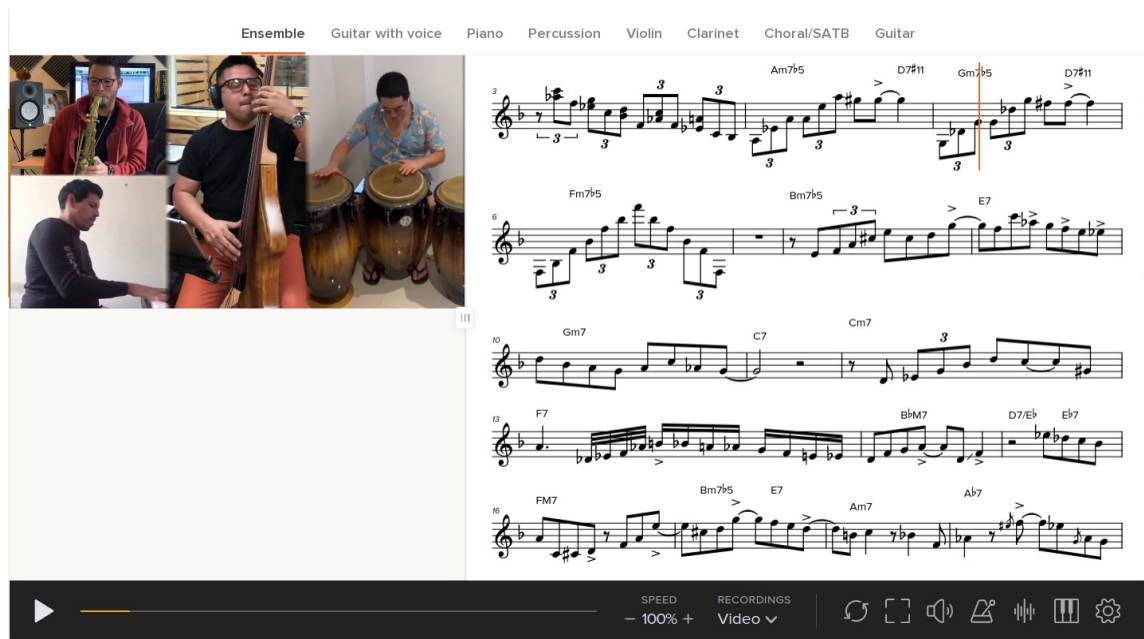
Οι καθηγητές δημιουργούν μαθήματα, τα οποία στην απλούστερη μορφή μπορεί να είναι μια παρτιτούρα. Στην πλατφόρμα υπάρχει επεξεργαστής μουσικής σημειογραφίας, με τον οποίο τόσο οι δάσκαλοι όσο και οι μαθητές μπορούν να επεξεργαστούν την παρτιτούρα (Εικόνα 2.6). Η παρτιτούρα μπορεί να απεικονιστεί με διάφορους τρόπους και μπορεί να γίνει εστίασή της ή και αλλαγή μεγέθους, ώστε να καταλάβει μικρότερο ή μεγαλύτερο τμήμα της οθόνης.



Εικόνα 2.6 - Επεξεργαστής Μουσικής Σημειογραφίας Soundslice (<https://www.soundslice.com>)

Ο καθηγητής μπορεί να συνοδέψει την παρτιτούρα με βίντεο ή ηχητικό απόδοσής της και να συγχρονίσει την παρτιτούρα με τον ήχο, με ακρίβεια νότας. Την ώρα που αναπαράγεται το

βίντεο ή ο ήχος, επισημαίνεται στην παρτιτούρα η τρέχουσα νότα (ή νότες) που ακούγεται (Εικόνα 2.7). Ο συνδυασμός ηχητικού και παρτιτούρας μπορεί να αποθηκευθεί ως μάθημα και, επιπλέον, οι δάσκαλοι μπορούν να ενοποιούν διάφορα μαθήματα σε ενότητες και κάθε μάθημα ή ενότητα να το κοινοποιούν σε όσους από τους μαθητές τους επιθυμούν.

The image shows a screenshot of the Soundslice website interface. At the top, there is a navigation menu with options: Ensemble, Guitar with voice, Piano, Percussion, Violin, Clarinet, Choral/SATB, and Guitar. Below the menu is a video player showing a live performance of a jazz ensemble. The ensemble consists of a saxophone player, a pianist, a bassist, and a drummer. To the right of the video player is a musical score for guitar. The score is written in treble clef and includes various chords and rhythmic patterns. The chords shown are Am7b5, D7#11, Gm7b5, D7#11, Fm7b5, Bm7b5, E7, Gm7, C7, Cm7, F7, BbM7, D7/Eb, Eb7, Fm7, Bm7b5, E7, Am7, and Ab7. The score also includes triplets and accents. At the bottom of the interface, there is a control bar with a play button, a speed control set to 100%, a recordings dropdown menu, and various playback controls like volume, full screen, and settings.

Εικόνα 2.7 - Περιβάλλον εξάσκησης με την εφαρμογή Soundslice (<https://www.soundslice.com>)

Οι μαθητές, με τη σειρά τους, χρειάζεται να έχουν έναν δωρεάν λογαριασμό και σύνδεση στο διαδίκτυο για να μπορούν να βλέπουν τα μαθήματα ή τις εργασίες που τους αναθέτουν οι εκπαιδευτές. Μπορούν να επιλέξουν τμήμα της παρτιτούρας και να το αναπαράξουν (μέσω της ηχογράφησης ή του βίντεο) μια φορά ή επαναλαμβανόμενα. Μπορούν να ρυθμίσουν το tempo (την ταχύτητα αναπαραγωγής) και πάνω σε αυτή την επανάληψη των βρόχων, να συνοδέψουν με το δικό τους μουσικό όργανο. Δεν υπάρχει περιορισμός στο μουσικό είδος ή το μουσικό όργανο στο οποίο θα εξασκηθεί ο μαθητής. Οι μαθητές μπορούν να σώζουν τις ηχογραφήσεις τους και να τις στέλνουν στον εκπαιδευτή για να πάρουν ανάδραση για την πρόοδό τους.

Το περιβάλλον είναι εύχρηστο και ευχάριστο, ωστόσο δεν υπάρχει η δυνατότητα της αυτόματης αξιολόγησης της απόδοσής της παρτιτούρας σε ήχο όπως στο Yousician.

2.4 Σύνοψη

Μια πρώτη παρατήρηση όσον αφορά τις παραπάνω πλατφόρμες, είναι ότι οι δύο τελευταίες (Yousician και Soundslice) είναι ολοκληρωμένες εμπορικές πλατφόρμες, ενώ η πρώτη (MusiCoLab) βασίζεται σε ελεύθερο λογισμικό και βρίσκεται ακόμα στο στάδιο έντονης ανάπτυξης.

Θα μπορούσαμε να ξεχωρίσουμε την εφαρμογή Yousician, όσον αφορά την ανάγκη εγκατάστασης συγκεκριμένου λογισμικού, ενώ οι άλλες δύο λειτουργούν σε περιβάλλον οποιουδήποτε περιηγητή διαδικτύου. Επίσης, η Yousician έχει οργάνωση των ασκήσεων με τρόπο που να μοιάζει περισσότερο με παιχνίδι (με αυτόματη αξιολόγηση της απόδοσης), ενώ οι MusiCoLab και Soundslice εστιάζουν στην εξάσκηση σε μουσικά κομμάτια χωρίς να γίνεται αξιολόγηση παρά μόνο από τον ίδιο τον μαθητή ή το δάσκαλό του. Επιπλέον, έχει περιορισμό στα μουσικά όργανα στα οποία μπορεί να εξασκηθούν οι μαθητές, ενώ οι άλλες δύο όχι.

Άλλη μια διαφορά, είναι ότι στη Yousician υποστηρίζεται η ταυτόχρονη επικοινωνία μαθητή - δασκάλου μόνο με χρήση εξωτερικής εφαρμογής, ενώ στη MusiCoLab αυτό γίνεται μέσω της ίδιας διεπαφής (με ενσωματωμένο σύστημα τηλεδιάσκεψης). Στο Soundslice δε μοιάζει να προβλέπεται αυτή η ανάγκη ή δυνατότητα.

Η εφαρμογή Soundslice ξεχωρίζει για τον συγχρονισμό ηχογραφημένης μουσικής (με βίντεο ή χωρίς) με την αντίστοιχη παρτιτούρα σε επίπεδο νότας, ενώ οι άλλες δύο σε επίπεδο νότας μπορούν μόνο να κάνουν αναπαραγωγή της παρτιτούρας με αυτόματο τρόπο (πχ με εφαρμογή αναπαραγωγής αρχείων midi). Επίσης μπορεί να γίνει ηχογράφηση της μουσικής εκτέλεσης του μαθητή και αποθήκευση αυτής της ηχογράφησης.

Στο πλαίσιο αυτής της εργασίας, αφού διαπιστώθηκε η ανάγκη, δόθηκε η δυνατότητα της ηχογράφησης του μαθητή και της αποθήκευσης της ηχογράφησης αυτής. Επιπλέον διαπιστώθηκε η ανάγκη συγχρονισμού της ηχογράφησης του μαθητή με την παρτιτούρα.

Μέσω αυτής, θα μπορεί ο δάσκαλος να υποδείξει στον μαθητή σημεία ενδιαφέροντος (πχ κάποιο λάθος στην εκτέλεση) στην παρτιτούρα και να εντοπιστεί άμεσα το αντίστοιχο σημείο της ηχογράφησης. Επίσης θα γίνεται πιο εύκολη η περιήγηση στα διάφορα σημεία της ηχογράφησης, καθώς η μετακίνηση του δείκτη στην κυματομορφή της θα μπορεί να γίνεται με βάση το περιεχόμενο και όχι κατά εκτίμηση της επιθυμητής χρονικής στιγμής.

Για την ανάπτυξη αυτής της εφαρμογής, που στόχο έχει να συνεισφέρει στη μουσική εκπαίδευση μέσω της επέκτασης της πλατφόρμας MusiCoLab, χρησιμοποιήθηκαν βιβλιοθήκες ελεύθερου λογισμικού, εστιασμένες στην ανάλυση μουσικού περιεχομένου.

3 Ανάλυση Μουσικού Περιεχομένου

3.1 Σύγχρονες Δυνατότητες

Οι χαμηλότερες ταχύτητες μεταφοράς δεδομένων των συνδέσεων του διαδικτύου στο παρελθόν, ιδίως στο μέρος που αφορούσε την αποστολή δεδομένων, αποτελούσαν εμπόδιο για τη ζωντανή μετάδοση υψηλής ποιότητας ήχου, πόσο μάλλον και συνδυασμό με ζωντανή εικόνα. Για παράδειγμα, ακόμα και για τη μετάδοση μονοφωνικού ήχου 44kHz με 16 bit πληροφορίας ανά δείγμα, χρειάζεται ταχύτητα μεταφόρτωσης 700kbps, που κάθε άλλο παρά δεδομένη ήταν στο παρελθόν. Ταυτόχρονα, η μη σταθερότητα των ταχυτήτων οδηγούσε σε αλλοιώσεις στο tempo κατά τη ζωντανή μετάδοση του ήχου (Driessen et al, 2022). Σήμερα, συνδέσεις VDSL ή ακόμα και οπτικών ινών μέχρι τους οικιακούς χρήστες, προσφέρουν τη δυνατότητα μεγάλων ταχυτήτων upload (πχ 10 ή 20Mbps) και παρέχουν υψηλή σταθερότητα αυτών των τιμών.

Δεν ήταν όμως μόνο αυτός ο περιορισμός για την ψηφιακή μετάδοση ήχου, καθώς υπήρχαν περιορισμοί ακόμα και από το ίδιο το υλικό που θα έκανε πχ τη μετατροπή του αναλογικού σήματος σε ψηφιακό. Οι σύγχρονες συσκευές, ακόμα και ένα κινητό τηλέφωνο, διαθέτουν σήμερα την απαιτούμενη επεξεργαστική ισχύ για τέτοιες εφαρμογές.

Άλλο ένα πρόβλημα του παρελθόντος ήταν το πρόβλημα της υπολογιστικής ισχύος που χρειαζόταν για να γίνει ζωντανή (online) επεξεργασία ήχου. Ο συνδυασμός της αύξησης της επεξεργαστικής ισχύος αλλά και οι ενσωματώσεις στους επεξεργαστές αλγορίθμων επεξεργασίας ψηφιακών σημάτων (Singh et al, 2014), λύνουν και αυτό το πρόβλημα.

Σε αντίθεση επομένως με το παρελθόν, σήμερα υπάρχουν τόσο οι αυξημένες υπολογιστικές δυνατότητες, όσο και οι εξειδικευμένες στην ανάλυση ηχητικών δεδομένων βιβλιοθήκες ανοιχτού λογισμικού. Ταυτόχρονα, υπάρχει μεγάλη αύξηση των ταχυτήτων μεταφοράς δεδομένων στο διαδίκτυο (ή και σε τοπικά δίκτυα), σταθερότητα της ταχύτητας αυτής και μεγάλη μείωση των χρόνων απόκρισης, λόγω και της περισσότερο εκτεταμένης χρήσης οπτικών ινών.

Ταυτόχρονα με την τεχνολογική πρόοδο που αφορά το υλικό, η εξέλιξη στους αλγορίθμους επεξεργασίας ηχητικών σημάτων έρχεται να χρησιμοποιήσει αυτό το υλικό και να δώσει νέες δυνατότητες. Όπως είδαμε στο προηγούμενο κεφάλαιο, ήδη στις πλατφόρμες μουσικής εκπαίδευσης υπάρχουν δυνατότητες ζωντανής αναγνώριση ηχητικών χαρακτηριστικών (πχ τονικού ύψους). Στην επόμενη παράγραφο, όπου αναλύονται οι προγραμματιστικές βιβλιοθήκες Librosa και Essentia, θα διερευνήσουμε το πλήθος των δυνατοτήτων που παρέχονται από αυτές.

3.2 Προγραμματιστικές βιβλιοθήκες

3.2.1 Librosa

Η Librosa είναι ένα πακέτο της Python για ανάλυση σημάτων ήχου και μουσικής (McFee et al., 2015). Στόχος της Librosa είναι να παρέχει τα δομικά συστατικά που επιτρέπουν σε προγραμματιστές την υλοποίηση εφαρμογών ανάκτησης μουσικής πληροφορίας. Παρακάτω παρατίθενται αλφαβητικά τα δομικά στοιχεία της Librosa και περιγράφονται περιληπτικά οι λειτουργίες τους.

librosa.beat: Περιλαμβάνει συναρτήσεις για την εκτίμηση του ρυθμού και τον εντοπισμό ρυθμικών παλμών (beats). (3 συναρτήσεις)

librosa.core: Περιλαμβάνει βασικές λειτουργίες για την φόρτωση αρχείων ήχου, τον υπολογισμό διάφορων αναπαραστάσεων ηχητικού φάσματος και διάφορα συχνά χρησιμοποιούμενα εργαλεία μουσικής ανάλυσης, όπως συναρτήσεις επεξεργασίας στο πεδίο του χρόνου, συναρτήσεις παραγωγής σημάτων και εύρους συχνοτήτων, συναρτήσεις υπολογισμού αρμονικών, συναρτήσεις μουσικής σημειογραφίας, συναρτήσεις ανάκτησης φάσης και άλλες. Για διευκόλυνση, όλη η λειτουργικότητα αυτής της υποενότητας είναι άμεσα προσβάσιμη από το ανώτερο επίπεδο namespace, librosa.*. (90 συναρτήσεις)

librosa.decompose: Συναρτήσεις για το διαχωρισμό αρμονικών - κρουστικών συνιστωσών ηχητικού σήματος (HPSS, Harmonic - Percussive Source Separation) και συναρτήσεις

γενικής αποσύνθεσης φασματογραφήματος, χρησιμοποιώντας μεθόδους αποσύνθεσης πινάκων που εφαρμόζονται στο `scikit-learn`¹⁶. (3 συναρτήσεις)

librosa.display: Ρουτίνες οπτικοποίησης και εμφάνισης με τη χρήση της `matplotlib`¹⁷. (11 συναρτήσεις)

librosa.effects: Επεξεργασία ήχου στο πεδίο του χρόνου, όπως μεταβολή τονικού ύψους και επιμήκυνση διάρκειας. Αυτή η υποενότητα περιλαμβάνει πακέτα που χρειάζονται στην `librosa.decompose` όσον αφορά την επεξεργασία στο πεδίο του χρόνου. (10 συναρτήσεις)

librosa.feature: Εξαγωγή και χειρισμός ηχητικών χαρακτηριστικών. Περιλαμβάνει χαμηλού επιπέδου εξαγωγή ηχητικών χαρακτηριστικών όπως χρωματογραφήματα, φασματογραφήματα στην κλίμακα Mel, Mel-frequency cepstral coefficients (MFCCs) και διάφορα άλλα φασματικά και ρυθμικά χαρακτηριστικά. Επίσης παρέχονται μέθοδοι χειρισμού χαρακτηριστικών. (22 συναρτήσεις)

librosa.filters: Δημιουργία Filter-banks (chroma, pseudo-CQT, CQT, κλπ.). Αυτές κυρίως είναι εσωτερικές συναρτήσεις που χρησιμοποιούνται από άλλες λειτουργίες της `librosa`.

librosa.onset: Εντοπισμός έναρξης ηχητικού γεγονότος (onset detection) και υπολογισμός της ισχύος της. (4 συναρτήσεις)

librosa.segment: Συναρτήσεις που χρησιμεύουν στη δομική τμηματοποίηση, όπως κατασκευή πινάκων επανάληψης (recurrence matrix construction), αναπαράσταση χρονικής υστέρησης (time-lag representation) και διαδοχικά περιορισμένη ομαδοποίηση (sequentially constrained clustering). (8 συναρτήσεις)

librosa.sequence: Συναρτήσεις για διαδοχική μοντελοποίηση (sequential modeling). Διάφορες μορφές αποκωδικοποίησης Viterbi και βοηθητικές συναρτήσεις για την κατασκευή πινάκων μετάβασης. (9 συναρτήσεις)

librosa.util: Βοηθητικές εφαρμογές όπως κανονικοποίηση (normalization), επένδυση (padding), κεντράρισμα (centering) κλπ. (33 συναρτήσεις)

16 <https://scikit-learn.org>

17 <https://matplotlib.org>

Το σύνολο των συναρτήσεων που υλοποιούν αλγορίθμους στη βιβλιοθήκη Librosa είναι, σήμερα (στην έκδοση 0.9.2), 193 συναρτήσεις.

3.2.2 Essentia

Η Essentia είναι μια βιβλιοθήκη ανοιχτού κώδικα, υλοποιημένη C++ και με συνδέσεις (bindings) στις γλώσσες Python και JavaScript, για την ανάλυση ηχητικών σημάτων με στόχο την ανάκτηση μουσικής πληροφορίας. Διατίθεται με την άδεια Affero GPLv3 και είναι επίσης διαθέσιμη με αποκλειστική άδεια κατόπιν αιτήματος. Η βιβλιοθήκη περιέχει μια εκτενή συλλογή αλγορίθμων, που υλοποιούν λειτουργίες εισόδου εξόδου για ήχο, τυπικές συναρτήσεις ψηφιακής επεξεργασίας σήματος, στατιστικό χαρακτηρισμό δεδομένων και μια μεγάλη ποικιλία φασματικών, χρονικών, τονικών και υψηλού επιπέδου περιγραφών μουσικής. Επιπρόσθετα, η Essentia μπορεί να συμπληρωθεί με ένα wrapper για συνεργασία με τα μοντέλα TensorFlow¹⁸ και τη βιβλιοθήκη Gaia¹⁹, μια βιβλιοθήκη υλοποιημένη σε C++ και με συνδέσεις Python, η οποία επιτρέπει την αναζήτηση ομοιότητας και την ταξινόμηση βάσει αποτελεσμάτων ανάλυσης ήχου (ισχύουν οι ίδιοι όροι άδειας χρήσης). Και οι δύο αυτές βιβλιοθήκες μπορούν να χρησιμοποιηθούν από την Essentia για υψηλού επιπέδου μουσική περιγραφή. Διατίθενται παραδείγματα προεκπαιδευμένων μοντέλων TensorFlow και μοντέλων ταξινόμησης SVM της Gaia, απευθείας με την εγκατάσταση.

Η Essentia δεν είναι ένα Framework, αλλά περισσότερο μια συλλογή αλγορίθμων (με κάποια επιπλέον υποδομή), ενταγμένων σε μια βιβλιοθήκη που είναι σχεδιασμένη με έμφαση στη σταθερότητα, την απόδοση και τη βελτιστοποίηση των παρεχόμενων αλγορίθμων, συμπεριλαμβανομένης της υπολογιστικής ταχύτητας και της χρήσης μνήμης, καθώς και στην ευκολία στη χρήση. Η ροή της ανάλυσης αποφασίζεται και υλοποιείται από τον χρήστη, ενώ η Essentia φροντίζει για τις λεπτομέρειες υλοποίησης των αλγορίθμων που χρησιμοποιούνται. Υπάρχει μια ειδική λειτουργία ροής, στην οποία είναι δυνατή η σύνδεση

18 <https://www.tensorflow.org>

19 <https://essentia.upf.edu/gaia>

των αλγορίθμων και η αυτόματη εκτέλεσή τους στην PureData²⁰ ή στη Max/MSP²¹. Μεγάλο μέρος της Essentia προορίζεται για εφαρμογές πραγματικού χρόνου.

Η Essentia έχει χρησιμοποιηθεί για ταξινόμηση μουσικής, σημασιολογική αυτόματη προσθήκη ετικετών, ομοιότητα και σύσταση μουσικής, οπτικοποίηση και αλληλεπίδραση με μουσική, ευρετηρίαση ήχου, ανίχνευση μουσικών οργάνων, ανίχνευση εξωφύλλου, ανίχνευση ρυθμικών παλμών (beats) και ακουστική ανάλυση ερεθισμάτων για μελέτες νευροαπεικόνισης.

Η Essentia και η Gaia έχουν χρησιμοποιηθεί εκτενώς σε πολλά ερευνητικά προγράμματα και βιομηχανικές εφαρμογές²².

Επί του παρόντος (Essentia 2.1-beta6), περιλαμβάνονται οι ακόλουθες κατηγορίες συναρτήσεων (μεταξύ άλλων):

Εισόδου/εξόδου αρχείων ήχου: δυνατότητα ανάγνωσης και εγγραφής σχεδόν όλων των τύπων αρχείων ήχου (wav, mp3, ogg, flac κλπ)

Τυπικές συναρτήσεις επεξεργασίας σήματος: FFT, DCT, frame cutter, windowing, envelope, smoothing

Φίλτρα (FIR & IIR): χαμηλο/υψηλο/ζωνοπερατά, απόρριψης ζώνης, αφαίρεσης DC, equal loudness

Στατιστικοί περιγραφείς (Statistical descriptors): median, mean, variance, power means, raw/central moments, spread, kurtosis, skewness, flatness

Περιγραφείς στο πεδίο του χρόνου: διάρκεια, ένταση, LARM, Leq, ένταση Vickers, zero-crossing-rate, log attack time και άλλοι signal envelope περιγραφείς

Φασματικοί περιγραφείς: Bark/Mel/ERB bands, MFCC, GFCC, LPC, spectral peaks, complexity, roll-off, contrast, HFC, inharmonicity και dissonance

Περιγραφείς τονικότητας: Pitch salience function, predominant melody και pitch, HPCP (chroma) related features, chords, key και scale, tuning frequency

20 <http://puredata.info>

21 <https://cycling74.com/products/max>

22 <https://essentia.upf.edu/applications.html>

Περιγραφείς ρυθμού: ανίχνευση ρυθμικών παλμών (beat detection), παλμοί ανά λεπτό (BPM), ανίχνευση έναρξης μουσικού γεγονότος (onset detection), αλλαγή ρυθμού (rhythm transform), ένταση ρυθμικών παλμών (beat loudness)

Άλλοι περιγραφείς υψηλού επιπέδου: danceability, δυναμική πολυπλοκότητα, audio segmentation, SVM classifier, TensorFlow wrapper for inference

Μοντέλα Μηχανικής Μάθησης: συνεργασία με ταξινομητές SVM και μοντέλα TensorFlow

Η βιβλιοθήκη λειτουργεί σε διάφορες πλατφόρμες, υποστηρίζοντας Linux και MacOS και εν μέρει συστήματα Windows, iOS και Android. Επίσης, μπορεί να μεταγλωττιστεί σε JavaScript (Correya et al., 2020), για χρήση σε εφαρμογές ιστού (Correya et al., 2021). Παρέχεται η dedicated wrapper βιβλιοθήκη `Essentia.js`²³.

Επιπλέον, η `Essentia` περιλαμβάνει συνδέσεις (bindings) στη γλώσσα Python (για Linux και OSX) και διάφορους προκαθορισμένους εξαγωγείς γραμμής εντολών για περιγραφείς μουσικής (για Linux, OSX και Windows), που διευκολύνουν τη χρήση της για γρήγορη δημιουργία πρωτοτύπων και επιτρέπουν τη δημιουργία ερευνητικών πειραμάτων πολύ γρήγορα. Επιπλέον, περιλαμβάνει ένα πρόσθετο `Vamp`²⁴ (για Linux και OSX) που μπορεί να χρησιμοποιηθεί με το `Sonic Visualiser`²⁵ για σκοπούς απεικόνισης. Υπάρχουν αρκετές επεκτάσεις τρίτων στην `Essentia` που επιτρέπουν τη χρήση της στα πλαίσια των `PureData` και `Max/MSP`, του `openFrameworks` και του `Matlab` (Bogdanov et al., 2013).

Όπως και στη `Librosa` έτσι και στην `Essentia`, υπάρχουν πολυάριθμες συναρτήσεις που υλοποιούν αλγόριθμους. Μέχρι σήμερα (στην έκδοση 2.1-beta6), στην πλήρη λίστα που παρατίθενται οι αλγόριθμοι²⁶, υπάρχουν 267 αλγόριθμοι από τους οποίους οι 253 εκτελούνται σε πραγματικό χρόνο.

Αξίζει σε αυτό το σημείο να αναφέρουμε τα μοντέλα μηχανικής μάθησης που έχουν υλοποιηθεί και εκπαιδευτεί για χρήση με την `Essentia`. Η λίστα με τα προεκπαιδευμένα

23 <https://mtg.github.io/essentia.js>

24 <https://www.vamp-plugins.org>

25 <https://sonicvisualiser.org>

26 https://essentia.upf.edu/algorithms_reference.html

μοντέλα TensorFlow που είναι διαθέσιμα στην Essentia, για διάφορες εργασίες ανάλυσης και ταξινόμησης ήχου (μουσικής), έχει ως εξής²⁷:

- Αναγνώριση συμβάντων ήχου
- Ταξινόμηση στυλ μουσικής
- Αυτόματη προσθήκη ετικετών μουσικής (είδος, διάθεση, εποχή, όργανα, κ.λπ.)
- Μεταφορά ταξινομητών μάθησης (είδος, διάθεση, χορευτικότητα, φωνή, ενορχήστρωση κ.λπ.)
- Εξαγωγείς χαρακτηριστικών
- Ανίχνευση τόνου
- Διαχωρισμός πηγών
- Εκτίμηση ρυθμού

Τα μοντέλα είναι σειριακά σε αρχεία Protocol Buffer κατάλληλα για συνεργασία με την Essentia και την TensorFlow. Επιπλέον, ορισμένα μοντέλα είναι επίσης διαθέσιμα σε μορφή TensorFlow.js. Κάθε αρχείο .pb συνδυάζεται με ένα αρχείο .json που περιέχει τα μεταδεδομένα του.

3.3 Ιδέες και Προοπτικές για τη Μουσική Εκπαίδευση

Οι παραπάνω βιβλιοθήκες μπορούν να χρησιμοποιηθούν σε πολλών ειδών εφαρμογές. Θα επικεντρωθούμε όμως σε διάφορες ιδέες για το πώς μπορούν να βοηθήσουν στη διδασκαλία και την εκμάθηση της μουσικής.

Καθώς αναφερόμαστε στη χρήση βιβλιοθηκών λογισμικού, η διδασκαλία ή η μελέτη/εξάσκηση θα είναι υποβοηθούμενη από κάποιου είδους ηλεκτρονική συσκευή. Η χρήση της, ωστόσο, μπορεί να είναι είτε μόνο κατά τη διδασκαλία ή την εξάσκηση, είτε ακόμα και στις δύο περιπτώσεις.

²⁷ <https://essentia.upf.edu/models.html>

Τόσο η Librosa όσο και η Essentia έχουν δυνατότητες ηχογράφησης, αποθήκευσης και αναπαραγωγής αρχείων ήχου σε διάφορες μορφές. Κατά την εξάσκηση, θα μπορούσε ο μαθητής να ηχογραφεί τις διαδοχικές εκτελέσεις του και να μπορεί να τις ακούσει ξανά.

Στη Librosa υπάρχουν συναρτήσεις που μπορούν να εξάγουν χαρακτηριστικά του ήχου από μια αποθηκευμένη ηχογράφηση. Μπορούμε να εξάγουμε πληροφορία για το χρώμα και τη χρονική στιγμή που παίχτηκε μια νότα με τις συναρτήσεις εντοπισμού έναρξης ηχητικού γεγονότος και εκτίμησης τονικού ύψους, χρώματος, ακόμα και ακολουθίας συγχορδιών. Επίσης, μπορούμε να εξάγουμε το ρυθμό της ηχογραφημένης μουσικής εκτέλεσης και τη θέση των ρυθμικών παλμών (beats) στη χρονική εξέλιξη. Η Essentia, επίσης παρέχει αυτές τις δυνατότητες και μάλιστα σε γλώσσα προγραμματισμού που ενδείκνυται και για εξαγωγή των χαρακτηριστικών αυτών σε πραγματικό χρόνο.

Η βιβλιοθήκη Librosa περιλαμβάνει ακόμα δυνατότητα μετρονόμου. Αυτή θα μπορούσε να χρησιμοποιηθεί για την τήρηση του tempo από το μαθητή κατά την εξάσκηση. Επίσης περιλαμβάνει γεννήτρια συχνοτήτων, που μπορεί να χρησιμοποιηθεί για το χόρδισμα μουσικών οργάνων.

Στην περίπτωση που οι μαθητές εξασκούνται σε σύνολο κρουστών και μελωδικών οργάνων, θα μπορούσε να χρησιμοποιηθεί συνάρτηση της Librosa που κάνει διαχωρισμό αρμονικής και κρουστικής συνιστώσας της αποθηκευμένης ηχογράφησης, προκειμένου να εντοπιστούν ευκολότερα τα σημεία ενδιαφέροντος από τους μαθητές, με τη βοήθεια των δασκάλων τους.

Με τη χρήση και των δύο βιβλιοθηκών, θα μπορούσαμε να εφαρμόσουμε φίλτρα, για παράδειγμα αυξάνοντας την ένταση ή μεταβάλλοντας το τονικό ύψος, προκειμένου να δείξουμε στο μαθητή το πώς θα έπρεπε να ακούγεται πιο σωστά η μουσική του εκτέλεση.

Ο μαθητής θα μπορούσε να κάνει περικοπή στις ηχογραφήσεις του ή διαχωρισμό μεταξύ συνεχόμενων ηχογραφήσεων, προκειμένου να αποστείλει στον δάσκαλό του το κομμάτι που θεωρεί πιο σωστό, στο πλαίσιο πχ μιας εργασίας.

Με τη βοήθεια συναρτήσεων της Librosa (όπως θα εφαρμόσουμε στη συνέχεια στην υλοποίηση της εφαρμογής μας), μπορούμε να συγχρονίσουμε αρχεία ήχου (ηχογραφήσεις) του ίδιου μουσικού κομματιού που μπορεί να έχουν παιχτεί σε άλλο tempo ή και με άλλο

μουσικό όργανο και με αυτό τον τρόπο να μπορούμε να περιηγηθούμε στις διαφορετικές χρονικές στιγμές που αντιπροσωπεύουν τα ίδια σημεία της παρτιτούρας.

Με τα μοντέλα μηχανικής μάθησης που έχουν υλοποιηθεί και εκπαιδευτεί για χρήση με την Essentia, μπορούμε με αυτοματοποιημένο τρόπο να οργανώσουμε βιβλιοθήκες με μουσικά κομμάτια και διαφορετικές εκτελέσεις τους, κατηγοριοποιημένες ανά μουσικό είδος, διάθεση, όργανα, εποχή κλπ.

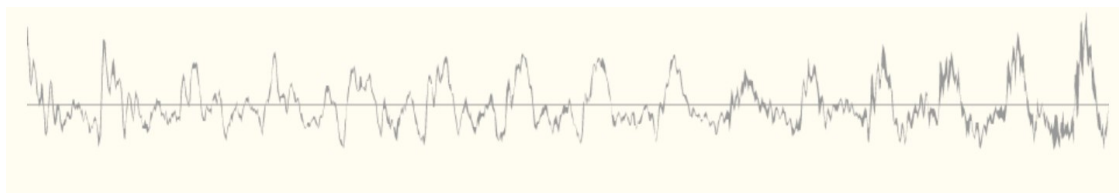
Επίσης, με τα μοντέλα αυτά, μπορούμε να κάνουμε ακόμα περισσότερα όσον αφορά το διαχωρισμό μουσικών οργάνων από μια ηχογράφιση. Πέρα από το διαχωρισμό δηλαδή των κρουστικών και αρμονικών συνιστωσών, μπορούμε να διαχωρίσουμε μια ηχογράφιση ενός μουσικού συνόλου σε 4 ή 5 μουσικά όργανα (φωνητικά, τύμπανα, μπάσο, πιάνο και κάποιο άλλο όργανο). Τα επιμέρους κομμάτια που δημιουργούνται μπορεί να χρειάζονται στους μαθητές που εκπαιδεύονται στο συγκεκριμένο μουσικό όργανο ή το τραγούδι.

Είναι βέβαιο ότι στα πλαίσια της παρούσας εργασίας δεν έχουμε εξαντλήσει το σύνολο των δυνατοτήτων που παρέχουν οι βιβλιοθήκες Librosa και Essentia. Η χρησιμότητά τους στη μουσική εκπαίδευση θα αναδύεται διαρκώς καθώς επεκτείνεται η χρήση τους. Νέες μέθοδοι και πρακτικές διδασκαλίας και εξάσκησης θα απαιτήσουν τη χρήση περισσότερων εργαλείων, που μπορεί να είναι ήδη διαθέσιμα ή να γίνουν διαθέσιμα μελλοντικά.

4 Υλοποίηση Εφαρμογής

4.1 Αναλυτική Περιγραφή Λειτουργίας

Στο πλαίσιο της εργασίας αυτής, υλοποιήθηκαν επεκτάσεις στην πλατφόρμα MusiCoLab προκειμένου να δοθεί η δυνατότητα ηχογράφησης από τους διδασκόμενους και να γίνουν διαθέσιμες η οπτική απεικόνιση της ηχογράφησης τους, η δυνατότητα αναπαραγωγής ή παύσης της από (και σε) οποιοδήποτε σημείο, η δυνατότητα λήψης της σε αρχείο ήχου μορφής .wav, και ο συγχρονισμός της ηχογράφησης με την αντίστοιχη παρτιτούρα που οι διδασκόμενοι απέδωσαν. Η παρτιτούρα προβάλλεται με τη χρήση του Verovio Humdrum Viewer²⁸ το οποίο έχει προσαρμοστεί στις ανάγκες του περιβάλλοντος του MusiCoLab. Όλη η λειτουργικότητα της πλατφόρμας και των προσαρμογών είναι διαθέσιμη σε περιβάλλον σύγχρονου περιηγητή διαδικτύου. Λειτουργεί, δηλαδή, ανεξάρτητα από την τερματική συσκευή και το λειτουργικό της σύστημα. Η εφαρμογή αυτή, ως μέρος του περιβάλλοντος MusiCoLab βρίσκεται σε μεμονωμένο σύνδεσμο²⁹.



Εικόνα 4.1 - Απεικόνιση κατά τη διάρκεια της ηχογράφησης

Για την απεικόνιση της κυματομορφής του ήχου κατά τη διάρκεια της εγγραφής (Εικόνα 4.1) αλλά και μετά την ολοκλήρωσή της, χρησιμοποιήθηκε η εφαρμογή wavesurfer.js που είναι υλοποιημένη σε JavaScript³⁰. Κατά τη διάρκεια της ηχογράφησης, χρησιμοποιείται η wavesurfer.microphone.js ενώ όταν ολοκληρώνεται η ηχογράφηση, χρησιμοποιείται η wavesurfer.cursor.js για την απεικόνιση της συνολικής κυματομορφής με ταυτόχρονη την

28 <https://verovio.humdrum.org>

29 <https://musicolab.hmu.gr/vhv>

30 <https://wavesurfer-js.org>

παρουσία κατακόρυφης γραμμής δείκτη χρόνου (cursor), που μπορεί να μετακινείται από το χρήστη και να δείχνει ταυτόχρονα τη χρονική στιγμή στην οποία βρίσκεται (Εικόνα 4.2).

Από την τρέχουσα θέση του δείκτη, μπορεί να γίνεται αναπαραγωγή και παύση της. Κάθε νέα ηχογράφηση σβήνει την προηγούμενη, οπότε μία μόνο ηχογράφηση μπορεί να εμφανίζεται και να είναι διαθέσιμη για λήψη. Οι βιβλιοθήκες αυτές καλούνται από κώδικα που γράφηκε σε JavaScript.



Εικόνα 4.2 - Απεικόνιση μετά το τέλος της ηχογράφησης

Η ηχογράφηση σε αρχείο .wav υλοποιείται με τη χρήση της βιβλιοθήκης recorder.js³¹. Η υλοποίηση βασίζεται σε κώδικα³² που χρησιμοποιεί τη recorder.js και με κατάλληλες προσαρμογές δημιουργεί ένα αντικείμενο τύπου Blob (Binary Large Object) που περιέχει την ηχογράφηση σε κωδικοποίηση .wav. Το αρχείο που δημιουργείται με το σταμάτημα της εγγραφής του ήχου, μπορεί να ληφθεί τοπικά στον υπολογιστή και το όνομα του αρχείου περιλαμβάνει την ημερομηνία και ώρα της ηχογράφησης. Και αυτή η βιβλιοθήκη καλείται από κώδικα που γράφηκε σε JavaScript.

Κατά τη διάρκεια της ηχογράφησης, δηλαδή, εκτελούνται παράλληλα δύο βιβλιοθήκες. Η πρώτη, με τη σειρά που αναφέρθηκαν, εμφανίζει την τρέχουσα τιμή της έντασης που λαμβάνει το μικρόφωνο και η δεύτερη αποθηκεύει τις τιμές αυτές σε αρχείο τύπου .wav.

Ο ηχογραφούμενος μαθητής, μπορεί ανά πάσα στιγμή και όσες φορές θέλει, να διακόψει την ηχογράφηση και να την συνεχίσει αργότερα. Η τελική ηχογράφηση θα περιλαμβάνει σειριακά τις ηχογραφήσεις για το σύνολο του χρονικού διαστήματος από το ξεκίνημα μέχρι

31 <https://github.com/mattdiamond/Recorderjs>

32 <https://github.com/addpipe/simple-recorderjs-demo>

το σταμάτημα της εγγραφής, εκτός από εκείνα τα χρονικά διαστήματα για τα οποία έγινε παύση.

Μετά το σταμάτημα της εγγραφής, ο μαθητής μπορεί να συγχρονίσει την ηχογράφησή του με τον ήχο που αναπαριστά η παρτιτούρα την οποία απέδιδε. Για να επιτευχθεί αυτό, αποστέλλονται από τον περιηγητή διαδικτύου που έγινε η εγγραφή, δύο αρχεία σε διακομιστή που τρέχει Linux και στον οποίο μπορούν να εκτελεστούν προγράμματα σε Python. Το ένα είναι το αρχείο .wav που ηχογραφήθηκε και το δεύτερο είναι το αρχείο τύπου midi (.mid) που εξάγεται από την παρτιτούρα που βλέπει και αποδίδει ο μαθητής. Τα δύο αυτά αρχεία γίνονται διαθέσιμα σε κώδικα υλοποιημένο σε γλώσσα προγραμματισμού Python. Το αρχείο .wav διαβάζεται από τη συνάρτηση librosa.load³³ και εκτός των ηχητικών δεδομένων, αποδίδεται σε μεταβλητή και η τιμή του ρυθμού δειγματοληψίας. Στη συνέχεια, με τη χρήση των συναρτήσεων Fluidsynth³⁴ και fs.midi_to_audio³⁵, μετατρέπεται το αρχείο midi σε αρχείο ήχου τύπου .wav με τον ίδιο ρυθμό δειγματοληψίας. Το αρχείο αυτό επίσης διαβάζεται από τη librosa.load.

Στο επόμενο βήμα, κατασκευάζονται κανονικοποιημένες χρωματικές αναπαραστάσεις των δύο αρχείων ήχου με τη χρήση της συνάρτησης librosa.feature.chroma_cens³⁶ ή της συνάρτησης librosa.feature.chroma_cqt³⁷. Μετά από δοκιμές, επιλέχθηκε η δεύτερη, καθώς τα αποτελέσματα του συγχρονισμού με την πρώτη είχαν κάποιες φορές λάθη. Οι αναπαραστάσεις αυτές έχουν στον ένα άξονα την αύξουσα αρίθμηση των τιμών των audio blocks και στον κατακόρυφο τα 12 ημιτόνια (ανεξάρτητα από τον αριθμό της οκτάβας) της ισοσυγκερασμένης μουσικής κλίμακας. Για κάθε ζεύγος αυτών των συντεταγμένων, ο πίνακας παίρνει κανονικοποιημένες τιμές, που εξαρτώνται από το ποσοστό της ενέργειας που αντιστοιχεί σε κάθε χρώμα (ημιτόνιο), για κάθε audio block (το μήκος του audio block έχει οριστεί στον κώδικα ως 1024 δείγματα).

33 <https://librosa.org/doc/latest/generated/librosa.load.html>

34 <https://www.fluidsynth.org>

35 <https://pypi.org/project/midi2audio>

36 https://librosa.org/doc/main/generated/librosa.feature.chroma_cens.html

37 https://librosa.org/doc/latest/generated/librosa.feature.chroma_cqt.html

Στη συνέχεια, αυτές οι αναπαραστάσεις εισάγονται ως ορίσματα σε μία ακόμα συνάρτηση της Librosa, την `librosa.sequence.dtw`³⁸, από την οποία εξάγονται δύο πίνακες, D και `wr`.

Πριν τη δημιουργία των D και `wr`, δημιουργείται πίνακας C διαστάσεων $M \times N$ όπου M, N οι αριθμός των block των δύο αρχείων ήχου. Αυτός ο πίνακας ονομάζεται πίνακας κόστους και για κάθε ζεύγος μπλοκ των δύο αρχείων ήχου περιέχει πληροφορίες σχετικά με το πόσο παρόμοιο -ή όχι- είναι το διάνυσμα χρώματος (οι τιμές του προηγούμενου πίνακα με τις χρωματικές αναπαραστάσεις) για τα αρχεία ήχου που πρόκειται να συγχρονιστούν. Η απόσταση συνημιτόνου (cosine distance) προτιμάται από την ευκλείδεια, επειδή δεν εξαρτάται από το μήκος των διανυσμάτων (την ενέργεια στην περίπτωση μας) αλλά μόνο από τη γωνία μεταξύ τους (δηλαδή την κατανομή της ενέργειας) η οποία και μας είναι χρήσιμη. Είναι επίσης πιο γρήγορος στον υπολογισμό, καθώς ο απαραίτητος πολλαπλασιασμός διανυσμάτων είναι αρκετά απλός για τις περισσότερες γλώσσες προγραμματισμού. Μετά τη δημιουργία του πίνακα κόστους, υπολογίζεται μια βέλτιστη διαδρομή μεταξύ του χρόνου έναρξης και λήξης και των δύο αξόνων. Αυτή η διαδρομή θα ακολουθεί τη βέλτιστη ακολουθία τιμών του πίνακα κόστους από την αρχή έως το τέλος και της διάρκειας του ήχου, έτσι ώστε το συνολικό κόστος να είναι το χαμηλότερο δυνατό, χρησιμοποιώντας μια πολλαπλής κλίμακας Dynamic Time Warping με backtracking διαδρομής σε δύο ακολουθίες (Müller, 2015). Έτσι δημιουργείται ο πίνακας D που λέγεται πίνακας συσσωρευμένου κόστους (accumulated cost matrix) και οι τιμές για κάθε στοιχείο του στη θέση (x,y) είναι το ελάχιστο άθροισμα των στοιχείων του πίνακα C που δημιουργούν διαδρομή από το στοιχείο (0,0) στο στοιχείο (x,y). Η βέλτιστη διαδρομή η οποία ενώνει τα σημεία (0,0) και (M,N) αποθηκεύεται στον πίνακα `wr` και αντιστοιχεί τα blocks των δύο αρχείων ήχου. Στον πίνακα `wr`, δηλαδή, αποτυπώνεται η ιδανική (ελάχιστου συνολικού κόστους) συσχέτιση μεταξύ των audio blocks των δύο αρχείων ήχου. Ο πίνακας αυτός, είναι διαστάσεων $2 \times N$, όπου N ο αριθμός των blocks του μεγαλύτερου σε διάρκεια αρχείου ήχου. Στην περίπτωση που θέλουμε μόνο το συγχρονισμό, μας χρειάζεται μόνο αυτός ο πίνακας.

Ωστόσο, στις περισσότερες περιπτώσεις που θα γίνει μια πραγματική ηχογράφιση, ο μαθητής θα ξεκινήσει την ηχογράφιση και θα ξεκινήσει μετά από κάποιο χρόνο και θα σταματήσει την ηχογράφιση αφού προφανώς τελειώσει τη μουσική του εκτέλεση. Αυτά τα

³⁸ <https://librosa.org/doc/latest/generated/librosa.sequence.dtw.html>

χρονικά διαστήματα - πριν και μετά - θα περιλαμβάνονται στην ηχογράφιση και δεν είναι απαραίτητο να πρόκειται για διαστήματα σιωπής. Με αλγόριθμο που αναπτύχθηκε για αυτές τις περιπτώσεις, εξετάζεται κάθε φορά ένας αριθμός από blocks για την ύπαρξη ή όχι επαναλήψεων του αύξοντα αριθμού των blocks στην πρώτη στήλη του πίνακα wr (που αφορά το προερχόμενο από το αρχείο midi αρχείο ήχου). Αν οι επαναλήψεις αυτές υπερβαίνουν κάποιο ποσοστό του συνολικού αριθμού των blocks που εξετάζονται, θεωρούμε με ασφάλεια ότι είμαστε εκτός της ηχογράφισης που αφορά την απόδοση της παρτιτούρας. Με audio block 1024 δειγμάτων, μετά από δοκιμές καταλήξαμε να εξετάζουμε κάθε φορά 150 blocks και σε αυτά να απαιτούμε να υπάρχουν λιγότερες από 30 επαναλήψεις. Σε αυτή την περίπτωση θα είμαστε εντός της ηχογράφισης που μας ενδιαφέρει να συγχρονίσουμε. Ο πίνακας wr περικόπτεται ανάλογα με το παραπάνω αποτέλεσμα στην αρχή και το τέλος του, μέρος του δηλαδή απορρίπτεται.

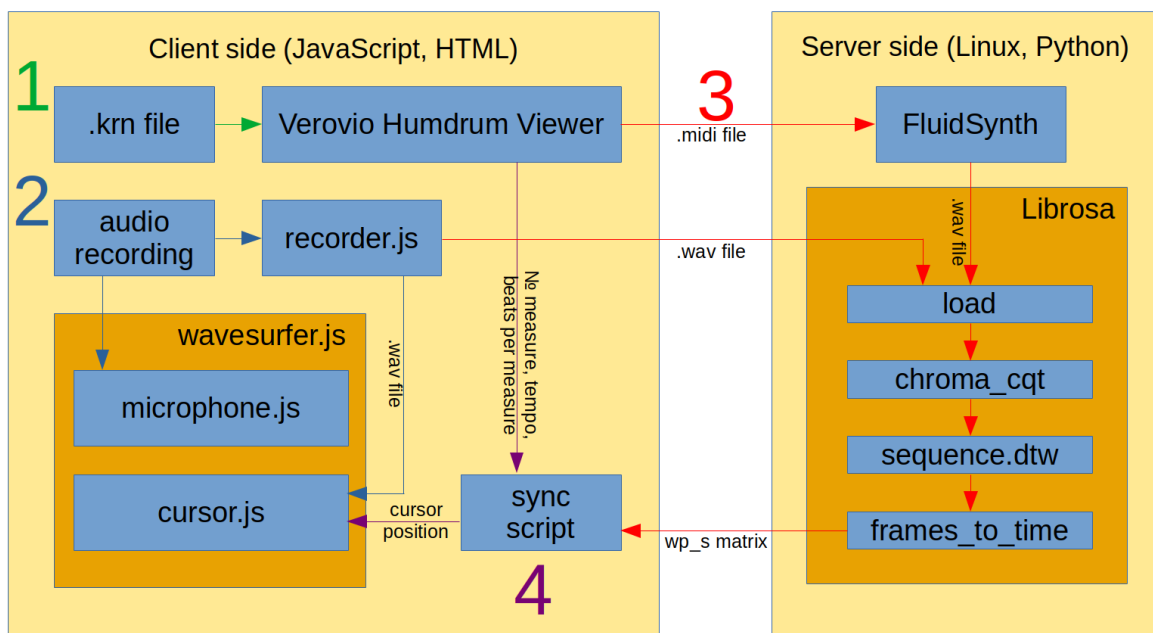
Στο τελευταίο βήμα της Python, ο πίνακας wr δίνεται ως όρισμα στη συνάρτηση librosa.frames_to_time³⁹ η οποία επιστρέφει πίνακα που αντί για αύξοντες αριθμούς audio blocks, έχει χρονικές στιγμές. Με αυτόν τον πίνακα, έστω wr_s, έχουμε πλέον τη χρονική συσχέτιση (συγχρονισμό) ανάμεσα στην ψηφιακή απόδοση της παρτιτούρας και την ηχογραφημένη από το μαθητή εκδοχή της.

Παράλληλα, στο περιβάλλον του περιηγητή, ο μαθητής επιλέγει αντικείμενο εντός του μέτρου στο οποίο θέλει να μεταβεί. Με κώδικα υλοποιημένο σε JavaScript, υπολογίζεται η χρονική στιγμή στην οποία αρχίζει το τρέχον μουσικό μέτρο. Στη συνέχεια ζητείται από τον διακομιστή το αρχείο wr_s και στον περιηγητή αναζητείται η τιμή του χρόνου που υπολογίστηκε νωρίτερα στην πρώτη στήλη του πίνακα και επιστρέφεται η αντίστοιχη τιμή της δεύτερης στήλης που είναι η αντίστοιχη τιμή του χρόνου στην ηχογράφιση. Στην τιμή αυτή του χρόνου, μετακινείται ο δείκτης στην κυματομορφή που αναπαριστά την ηχογράφιση.

39 https://librosa.org/doc/latest/generated/librosa.frames_to_time.html

4.2 Αρχιτεκτονική

Η εφαρμογή αποτελείται από διάφορα στοιχεία, τα οποία συνεργάζονται μεταξύ τους για την επίτευξη του αποτελέσματος. Ένα συνοπτικό διάγραμμα της αρχιτεκτονικής παρουσιάζεται στην Εικόνα 4.3.



Εικόνα 4.3 - Αρχιτεκτονική της εφαρμογής

Ένα μέρος της υλοποιείται σε εξυπηρετητή ο οποίος έχει τη δυνατότητα να εκτελεί κώδικα σε python. Εκεί βρίσκεται αποθηκευμένος ο κώδικας που εκτελεί το συγχρονισμό του αρχείου `.midi` που αφορά στην παρτιτούρα και του αρχείου `.wav` που αφορά στην ηχογράφιση.

Ένα δεύτερο μέρος της, το οποίο είναι υλοποιημένο σε HTML, ενσωματώθηκε στο αρχείο⁴⁰ του κώδικα της ιστοσελίδας που αφορά τη σύγχρονη διδασκαλία της πλατφόρμας μουσικής εκπαίδευσης Musicolab. Σε αυτό το μέρος διαμορφώνεται η διεπαφή των χρηστών. Με τη

⁴⁰ <https://musicolab.hmu.gr/vhv/index.html>

χρήση της HTML, δημιουργήθηκαν τα κουμπιά και έγινε η χωροταξία της εφαρμογής, σε συνδυασμό με τα ήδη υπάρχοντα στοιχεία της πλατφόρμας. Για τα εικονίδια των κουμπιών, χρησιμοποιείται η ανοιχτού κώδικα βιβλιοθήκη bootstrap-icons⁴¹. Επίσης, χωροθετούνται οι κυματομορφές που θα εμφανίζονται είτε κατά την εγγραφή είτε κατά τη συνολική απεικόνιση της ηχογράφησης.

Στον κώδικα του ίδιου αρχείου, καλούνται μια σειρά από αρχεία JavaScript τα οποία αποτελούν το τρίτο μέρος της εφαρμογής και βρίσκονται στον ίδιο εξυπηρετητή με την ιστοσελίδα της πλατφόρμας. Τα αρχεία wavesurfer.js, wavesurfer.microphone.js και wavesurfer.cursor.js δημιουργούν την κυματομορφή του ήχου, όποτε αυτό απαιτείται από την εφαρμογή. Πρόκειται για ανοιχτό κώδικα από τη βιβλιοθήκη wavesurfer.js⁴². Το πρώτο καλείται από τα άλλα δύο για να δημιουργείται η κυματομορφή (κατά την εγγραφή και τη συνολική της απεικόνιση αντίστοιχα). Τα δύο τελευταία καλούνται από ένα επιπλέον αρχείο JavaScript, το wave_one.js. Στον κώδικα αυτού του αρχείου, εκτός από την κλήση των προηγούμενων δύο (που γίνεται με τη χρήση των κουμπιών του περιβάλλοντος εργασίας), γίνεται και η αποστολή των αρχείων .midi (από την παρτιτούρα) και .wav από την ηχογράφηση, καθώς και η επιστροφή από τον server του πίνακα wr_s. Η αποστολή και η λήψη γίνονται με την τεχνική AJAX⁴³ (Asynchronous JavaScript And XML). Στο ίδιο script, δίνονται ως όρισμα ο αριθμός του τρέχοντος μουσικού μέτρου (εκεί που έχει γίνει επιλογή από το χρήστη), το tempo και ο αριθμός των beats ανά μέτρο. Για να δοθούν αυτά, έγιναν προσθήκες στον κώδικα αρχείων JavaScript που αφορούν την απεικόνιση σε παρτιτούρα. Με τα δεδομένα αυτά, στο wave_one.js, εφόσον έχει επιστραφεί από τον εξυπηρετητή και το αρχείο wr_s, υπολογίζεται η χρονική στιγμή που αντιστοιχεί στην αρχή του μέτρου και βρίσκεται η αντίστοιχη χρονική στιγμή της ηχογράφησης. Χρησιμοποιούνται δύο ακόμα αρχεία JavaScript, τα rec2wav.js και rec2wav_2.js, το πρώτο για τη δυνατότητα αποθήκευσης της ηχογράφησης σε αρχείο .wav και το δεύτερο για την κλήση του πρώτου και επιπλέον για την λειτουργικότητα των κουμπιών της διεπαφής χρήστη. Το πρώτο από αυτά τα αρχεία είναι από την ανοιχτού κώδικα βιβλιοθήκη recorder.js⁴⁴ και το δεύτερο είναι

41 <https://icons.getbootstrap.com>

42 <https://wavesurfer-js.org>

43 [https://en.wikipedia.org/wiki/Ajax_\(programming\)](https://en.wikipedia.org/wiki/Ajax_(programming))

44 <https://github.com/mattdiamond/Recorderjs>

τροποποιημένο και με προσθήκες από την ανοιχτού κώδικα βιβλιοθήκη `simple-recorderjs-demo`⁴⁵.

Η εφαρμογή, όσον αφορά την HTML και τη JavaScript, εκτελείται στον περιηγητή (browser) του χρήστη της πλατφόρμας. Όσον αφορά την Python, εκτελείται σε εξυπηρετητή ο οποίος δέχεται και επιστρέφει δεδομένα από και προς τον browser του χρήστη.

Κατά την εκτέλεση ενός τυπικού σεναρίου της εφαρμογής και ακολουθώντας την αρίθμηση του διαγράμματος της Εικόνας 4.3, έχουμε τα παρακάτω βήματα:

Βήμα 1^ο:

Φορτώνεται αρχείο `.kpn` (μουσικής σημειογραφίας). Το αρχείο απεικονίζεται ως παρτιτούρα από το `Verovio Humdrum Viewer`.

Βήμα 2^ο:

Με το πάτημα του κουμπιού της ηχογράφησης, ο μαθητής ερμηνεύει την παρτιτούρα και η ερμηνεία του ηχογραφείται. Κατά τη διάρκεια της ηχογράφησης, απεικονίζεται η τρέχουσα κυματομορφή με τη χρήση της βιβλιοθήκης `wavesurfer.microphone.js`. Ταυτόχρονα αποθηκεύονται δεδομένα ήχου από τη βιβλιοθήκη `recorder.js`, τα οποία με το σταμάτημα της ηχογράφησης, απεικονίζονται ως κυματομορφή με δείκτη χρόνου και είναι δυνατή η αναπαραγωγή/παύση τους με τη χρήση της βιβλιοθήκης `wavesurfer.cursor.js`.

Βήμα 3^ο:

Με το πάτημα του κουμπιού συγχρονισμού, αποστέλλονται σε εξυπηρετητή δύο αρχεία. Το ένα, είναι αρχείο μουσικής πληροφορίας τύπου `.midi` το οποίο παράγεται από το αρχείο `.kpn` από το `Verovio Humdrum Viewer`. Το δεύτερο, είναι αρχείο ήχου `.wav` το οποίο δημιουργείται από τα δεδομένα ήχου της ηχογράφησης από τη βιβλιοθήκη `recorder.js`. Το αρχείο `.midi` μετατρέπεται σε αρχείο ήχου `.wav` από τη βιβλιοθήκη `FluidSynth`. Τα δύο πλέον αρχεία `.wav`, φορτώνονται από τη συνάρτηση `librosa.load`, δημιουργούνται οι χρωματικές αναπαραστάσεις τους από τη συνάρτηση `librosa.chroma.cqt` και εφαρμόζεται αλγόριθμος DTW από τη συνάρτηση `librosa.sequence.dtw`. Ο παραγόμενος πίνακας αντιστοίχισης `audio blocks`, μετατρέπεται σε πίνακα αντιστοίχισης χρονικών στιγμών από τη συνάρτηση

45 <https://github.com/addpipe/simple-recorderjs-demo>

librosa.frames_to_time και ο πίνακας αυτός επιστρέφεται στο πρόγραμμα περιήγησης του μαθητή.

Βήμα 4^ο:

Ο μαθητής επιλέγει οποιοδήποτε στοιχείο της απεικόνισης της παρτιτούρας και πατάει το κουμπί μετάβασης στην επιλογή. Το Veronio Humdrum Viewer αποθηκεύει σε μεταβλητές τον άξοντα αριθμό του τρέχοντος μουσικού μέτρου (εκεί που έχει γίνει επιλογή από το χρήστη), το tempo και τον αριθμός των beats ανά μέτρο. Ο συνδυασμός αυτών των μεταβλητών δίνει τη χρονική στιγμή της αρχής του μουσικού μέτρου και με τον πίνακα αντιστοίχισης χρονικών στιγμών προσδιορίζεται η χρονική στιγμή της ηχογράφησης, η οποία χρησιμοποιείται από τη βιβλιοθήκη librosa.cursor.js για να μετακινηθεί κατάλληλα ο δείκτης στην κυματομορφή της ηχογράφησης.

4.3 Προγραμματιστικά εργαλεία

Στην πρώτη στήλη του Πίνακα 4.1 καταγράφονται όλες οι βιβλιοθήκες και οι επιμέρους συναρτήσεις τους που χρησιμοποιήθηκαν. Στη δεύτερη στήλη, για καθεμία από αυτές, καταγράφεται συνοπτικά ο ρόλος που είχε στην τελική εφαρμογή. Η σειρά που ακολουθείται είναι η αντίστροφη χρονικά από αυτήν της χρήσης τους κατά ένα τυπικό σενάριο χρήσης της εφαρμογής.

Βιβλιοθήκη/Συνάρτηση	Ρόλος της στην εφαρμογή
librosa.frames_to_time (Python)	Μετατροπή των αριθμών των audio blocks σε χρόνους (σε δευτερόλεπτα).
librosa.sequence.dtw (Python)	Αλγόριθμος DTW (Dynamic Time Warping) που εκτελεί το συγχρονισμό μεταξύ δύο πινάκων χρωματικών αναπαραστάσεων από δύο αρχεία ήχου
librosa.feature.chroma_cqt (Python)	Αλγόριθμος δημιουργίας χρωματικής αναπαραστάσης από αρχείο ήχου

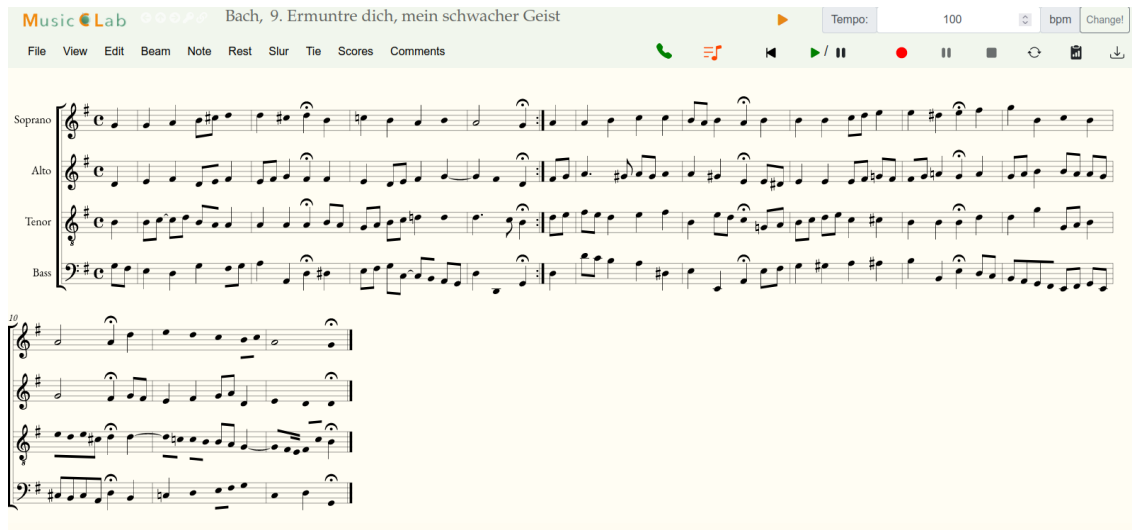
Βιβλιοθήκη/Συνάρτηση	Ρόλος της στην εφαρμογή
librosa.load (Python)	Φόρτωση στη μνήμη των αρχείων ήχου .wav
FluidSynth, midi2audio (Python, Linux)	Μετατροπή του αρχείου .midi σε αρχείο ήχου .wav
recorder.js, simple-recorder-demo.js (JavaScript)	Δημιουργία αρχείου .wav από την ηχογράφηση
wavesurfer.js, wavesurfer.cursor.js (JavaScript)	Απεικόνιση ηχογραφημένου ήχου με δείκτη χρόνου και δυνατότητα αναπαραγωγής παύσης και μετακίνησης του δείκτη.
wavesurfer.js, wavesurfer.microphone.js (JavaScript)	Απεικόνιση κυματομορφής ηχογράφησης
bootstrap-icons (SVG)	Εικονίδια κουμπιών διεπαφής χρήστη

Πίνακας 4.1 – Βιβλιοθήκες που χρησιμοποιήθηκαν και ο ρόλος τους στην τελική εφαρμογή

4.4 Γραφική Διεπαφή Χρήστη

Οι παραπάνω προσαρμογές ξεκίνησαν όταν στην ιστοσελίδα <https://musicolab.hmu.gr/vhv/> περιλαμβανόταν το Veronio Humdrum Viewer προσαρμοσμένο στις ανάγκες της μουσικής εκπαίδευσης. Οι προσαρμογές αφορούσαν επιπλέον λειτουργικότητα ως προς τη συνεργασία στην επεξεργασία της παρτιτούρας, με τη δυνατότητα αναπαραγωγής της καθώς και με τη δυνατότητα τηλεδιάσκεψης μαθητών-εκπαιδευτών.

Παράλληλα με την παρούσα εργασία, η πλατφόρμα ήταν υπό έντονη ανάπτυξη με συνεχείς βελτιώσεις και προσθήκες. Στη σημερινή μορφή της, η γραφική διεπαφή του χρήστη όταν φορτώνεται η πλατφόρμα και αφού ο χρήστης επιλέξει και ανοίξει ένα αρχείο παρτιτούρας, είναι όπως στην Εικόνα 4.4.



Εικόνα 4.4 - Γραφική διεπαφή χρήστη με εμφάνιση παρτιτούρας

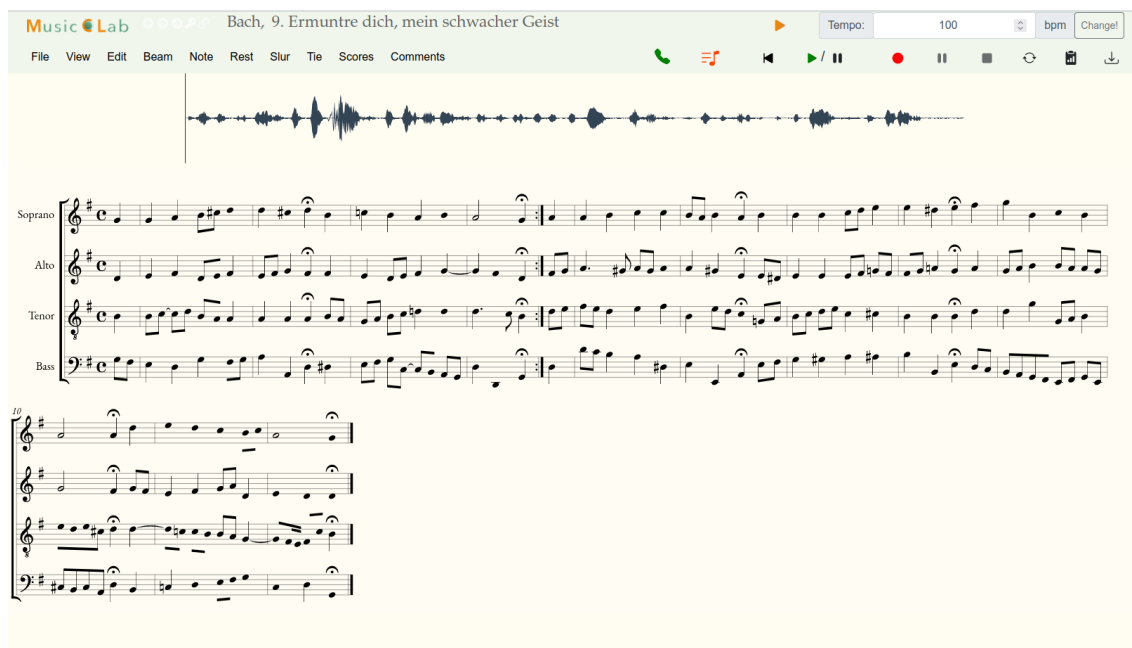
Στο πάνω δεξιό τμήμα της σελίδας και στη δεύτερη σειρά δημιουργήθηκαν μια σειρά από κουμπιά που υποστηρίζουν τη λειτουργικότητα που αναπτύχθηκε στα πλαίσια αυτής της εργασίας, τα οποία παρουσιάζονται στην Εικόνα 4.5.



Εικόνα 4.5 - Κουμπιά χειρισμού λειτουργιών

Το τέταρτο κουμπί (Record) ξεκινάει την ηχογράφιση. Όσο διαρκεί η ηχογράφιση, εμφανίζεται η κυματομορφή του εγγραφόμενου ήχου όπως στην Εικόνα 4.1. Το πέμπτο κουμπί (Pause Recording), το οποίο ενεργοποιείται κατά τη διάρκεια της ηχογράφισης, κάνει παύση και επανεκκίνηση της ηχογράφισης ενώ το έκτο κουμπί (Stop Recording) σταματά την εγγραφή και εμφανίζει την κυματομορφή της συνολικής ηχογράφισης όπως στην Εικόνα 4.2 και την Εικόνα 4.6.

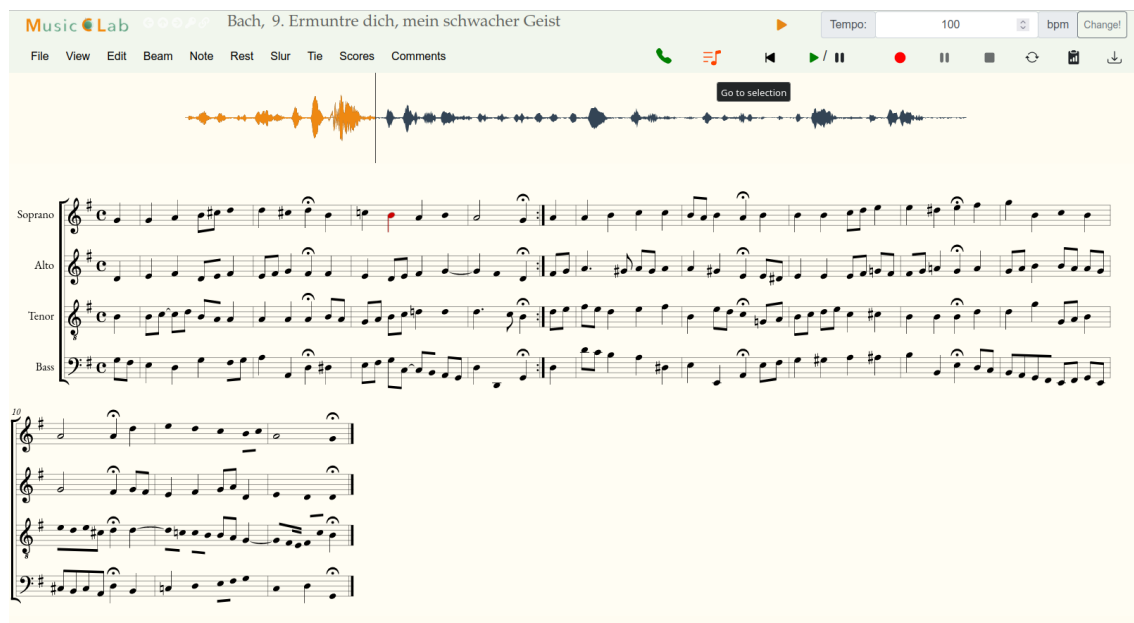
Το πάτημα του κουμπιού Stop Recording, επιπρόσθετα ενεργοποιεί τα τελευταία τρία κουμπιά. Το ένατο και τελευταίο κουμπί (Download) κάνει λήψη του αρχείου ήχου της ηχογράφησης στη συσκευή που βρίσκεται ο χρήστης. Το αμέσως προηγούμενο θα χρησιμοποιηθεί για την εξαγωγή χαρακτηριστικών της ηχογράφησης και είναι εκτός των πλαισίων αυτής της εργασίας. Το έβδομο κουμπί (Sync), είναι αυτό που ενεργοποιεί την αποστολή των δύο αρχείων στον διακομιστή για να γίνει ο συγχρονισμός.



Εικόνα 4.6 - Γραφική διεπαφή χρήση με εμφάνιση παρτιτούρας και ολοκληρωμένης ηχογράφησης

Η επόμενη κίνηση που μπορεί πλέον να κάνει ο χρήστης είναι να επιλέξει οποιοδήποτε στοιχείο από την παρτιτούρα. Η επιλογή, και μόνο, του στοιχείου επιστρέφει σε μεταβλητή της JavaScript τη χρονική στιγμή έναρξης του εκάστοτε μουσικού μέτρου. Καμία άλλη αλλαγή δε γίνεται όμως αν δεν πατηθεί το πρώτο κουμπί (Go to selection). Με το πάτημα του κουμπιού αυτού, ο δείκτης μετακινείται στην χρονική στιγμή της ηχογράφησης που αντιστοιχεί στην χρονική στιγμή της έναρξης του εκάστοτε μέτρου. Στην Εικόνα 4.7 έχει επιλεγεί η νότα με το κόκκινο χρώμα και έχει πατηθεί το κουμπί "Go to selection".

Το τρίτο κουμπί στην Εικόνα 4.5 (Play/Pause) ξεκινά ή κάνει παύση στην αναπαραγωγή του ηχογραφημένου αρχείου. Αν τη στιγμή που επιλέξουμε το κουμπί “Go to selection” γίνεται αναπαραγωγή του ήχου, τότε η αναπαραγωγή συνεχίζει από το σημείο στο οποίο μεταφέρθηκε ο δείκτης. Τέλος, το δεύτερο κουμπί (Stop) μετακινεί το δείκτη στην αρχή της ηχογράφησης και σταματάει την αναπαραγωγή, όποια και να ήταν η προηγούμενη κατάσταση.



Εικόνα 4.7 – Με επιλεγμένη τη νότα, το πάτημα του κουμπιού “Go to selection” μετακινεί το δείκτη

Το πλήρες σενάριο που αφορά την συγκεκριμένη εργασία, περιλαμβάνει ότι ο μαθητής φοράει ακουστικά για να ακούει και έχει ανοιχτό μικρόφωνο για να ηχογραφεί το δικό του μουσικό όργανο ή τη φωνή του. Επιλέγει την παρτιτούρα, πατάει το κουμπί της ηχογράφησης, πατάει το κουμπί της αναπαραγωγής της παρτιτούρας (πορτοκαλί κουμπί Play πάνω δεξιά), ακολουθεί παίζοντας ή τραγουδώντας αυτό που ακούει από τα ακουστικά και στη συνέχεια κάνει το συγχρονισμό προκειμένου να εντοπίσει και με τη βοήθεια του διδάσκοντα τα σημεία στα οποία τα πήγε καλά ή αυτά στα οποία χρειάζεται βελτίωση.

4.5 Επιδόσεις Συγχρονισμού σε πλαίσιο μαθήματος

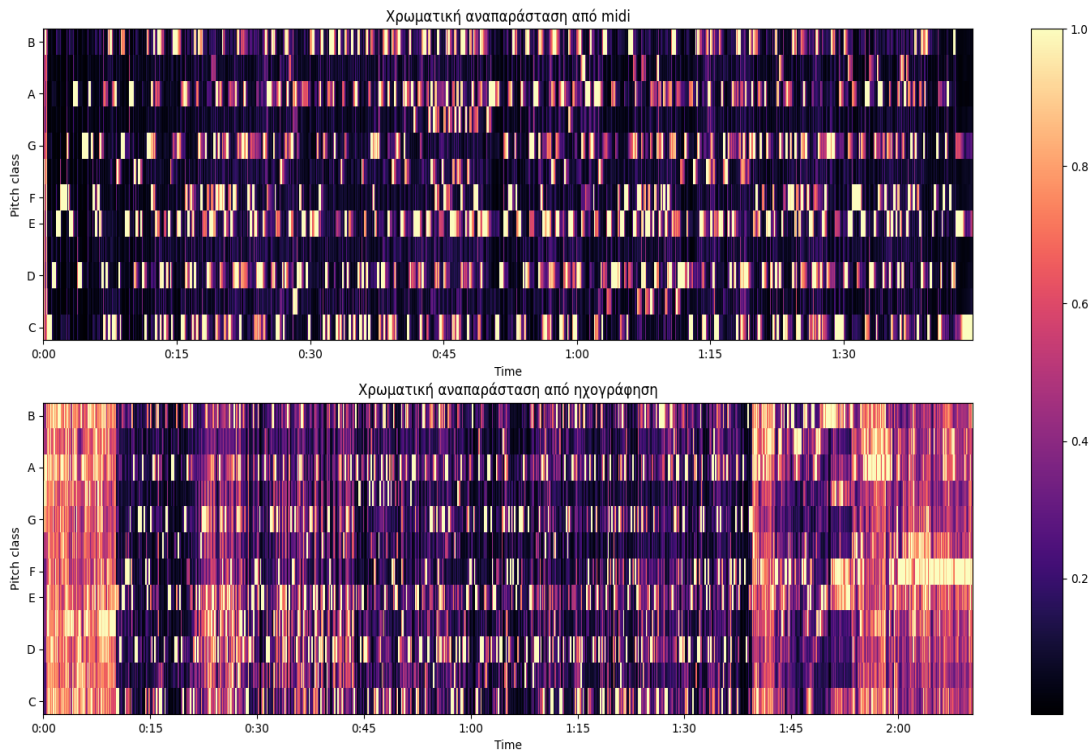
Υπάρχουν αρκετοί παράγοντες που παίζουν ρόλο στην απόδοση (ορθότητα) του συγχρονισμού. Ενδεικτικά, η ποιότητα της εγγραφής και το ποσοστό του περιβαλλοντικού ή όποιου άλλου θορύβου, η σωστή (ή όχι) απόδοση της παρτιτούρας όσον αφορά το χρώμα και πιθανές παραλείψεις ή επαναλήψεις, οι μεταβολές του tempo κλπ. Όσο περισσότερα είναι τα “λάθη” τόσο δυσκολότερος γίνεται ο συγχρονισμός. Στις περισσότερες ωστόσο περιπτώσεις, αν ο περιβαλλοντικός θόρυβος δεν επικαλύπτει την ηχογράφιση του μαθητή, ο συγχρονισμός γίνεται σωστά. Αξίζει να σημειώσουμε ότι η απόδοση της παρτιτούρας με διαφορετικό μουσικό όργανο από αυτό που περιγράφεται στην παρτιτούρα (και το αρχείο midi το οποίο μετατρέπεται σε ήχο για σύγκριση με την ηχογράφιση), δεν επηρεάζει το συγχρονισμό. Επίσης, δεν επηρεάζει το συγχρονισμό η μετατόπιση του ύψους κατά ακέραιο αριθμό οκτάβων.

Στα πλαίσια της αξιολόγησης της απόδοσης, πραγματοποιήθηκαν εγγραφές με κακής ποιότητας και τοποθέτησης μικρόφωνο, σε διαφορετικό tempo από αυτό που όριζε η παρτιτούρα, με κενά και αρκετό θόρυβο πριν και μετά τη μουσική απόδοση της παρτιτούρας, με μέτριο περιβαλλοντικό θόρυβο κατά τη διάρκεια των μουσικών εκτελέσεων, με παραλείψεις μεμονωμένων νότων ή ολόκληρων μέτρων σε διάφορα σημεία και με επανάληψη κάποιων μέτρων. Η χρωματική αναπαράσταση μιας τέτοιας περίπτωσης συγχρονισμού παρουσιάζεται στο διπλό διάγραμμα της Εικόνας 4.8. Το διάγραμμα αυτό, έχει παραχθεί από τη συνάρτηση `librosa.display.specshow`⁴⁶.

Το πάνω από τα διαγράμματα, αφορά το αρχείο ήχου που δημιουργήθηκε από το αρχείο της παρτιτούρας που μετατράπηκε αρχικά σε αρχείο midi. Το κάτω, αφορά το αρχείο ήχου από την ηχογράφιση του μαθητή. Στο υπόμνημα δεξιά, είναι ο χρωματικός κώδικας της ενέργειας που αφορά κάθε στιγμή και για κάθε χρώμα (ημιτόνιο) της ισοσυγκερασμένης μουσικής κλίμακας. Τα διαστήματα στην αρχή και το τέλος της ηχογράφησης που φαίνονται με περισσότερο ασαφές χρώμα (εκεί όπου είναι σε έλλειψη το μαύρο χρώμα, δηλαδή οι μηδενικές σχετικές τιμές ενέργειας), είναι είτε σιωπή είτε θόρυβος πριν και μετά την

⁴⁶ <https://librosa.org/doc/main/generated/librosa.display.specshow.html>

ηχογράφηση του μαθητή. Τα διαστήματα με ασαφές χρώμα κατά τη μουσική εκτέλεση (πχ γύρω από τη χρονική στιγμή 25'' της ηχογράφησης), είναι περιοχές όπου παράλληλα με τη μουσική εκτέλεση, καταγράφεται και θόρυβος.

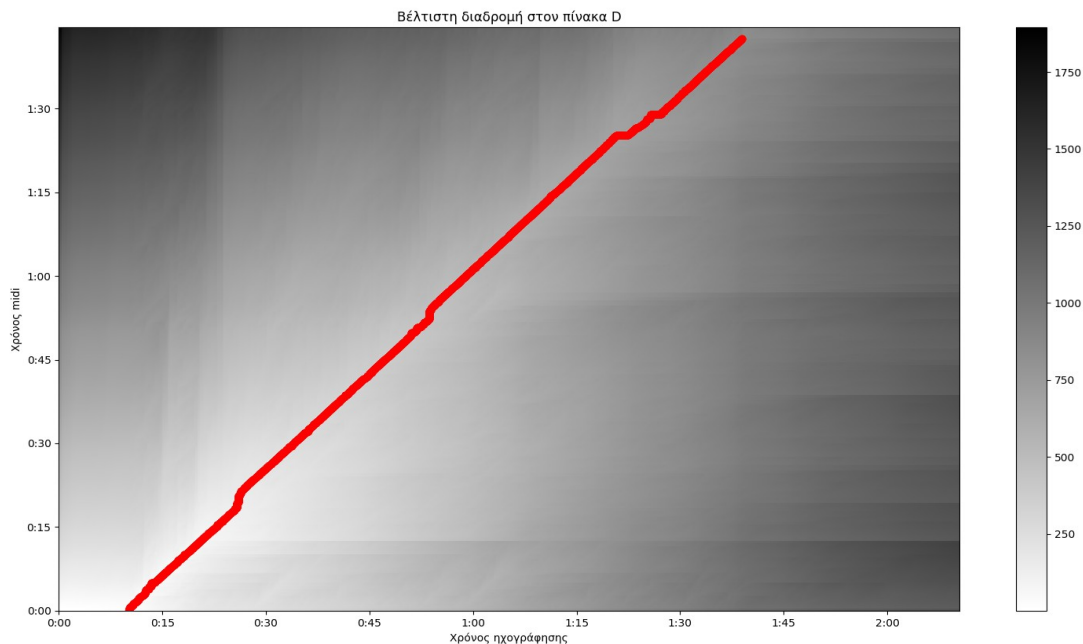


Εικόνα 4.8 – Χρωματικές αναπαραστάσεις αρχείου ήχου από midi και από ηχογράφηση

Ο συγχρονισμός των δύο αρχείων ήχου, φαίνεται στην κόκκινη διαδρομή πάνω στον πίνακα D (συσσωρευμένου κόστους) της Εικόνας 4.9. Η γραμμή αντιστοιχίζει τις χρονικές στιγμές των δύο αξόνων. Το διάγραμμα αυτό, όπως και το προηγούμενο, έχει δημιουργηθεί από τη συνάρτηση `librosa.display.specshow`.

Στον κατακόρυφο άξονα βρίσκονται οι χρονικές στιγμές των blocks του αρχείου ήχου που παράχθηκε από το αρχείο midi και στον οριζόντιο αυτές των blocks του αρχείου της ηχογράφησης. Τα τμήματα μέχρι περίπου τη χρονική στιγμή 10'' και από τη χρονική στιγμή 1'40'' και μετά, έχουν αναγνωριστεί ως μη αντιστοιχούντα στην παρτιτούρα και έχουν απορριφθεί. Στα δύο σημεία που η γραμμή έχει κατακόρυφη κλίση (περίπου στα 28'' και στα

55''), ο μαθητής παρέλειψε κάποιες νότες, ενώ στην περιοχή μεταξύ 1'20'' και 1'30'', ο μαθητής επανέλαβε μια σειρά από νότες. Καθώς έχουμε επανάληψη των ίδιων ακριβώς ηχητικών γεγονότων, είναι δύσκολο έως αδύνατο να καταφέρει ο αλγόριθμος μια απόλυτη αντιστοίχιση (πχ αποδοχή της πρώτης εκτέλεσης και απόρριψη της επανάληψης). Ωστόσο, το σφάλμα του μαθητή φαίνεται στο διάγραμμα, καθώς η γραμμή γίνεται οριζόντια σε δύο περιπτώσεις κατά τη διάρκεια της επανάληψης. Σε κάθε περίπτωση, η κόκκινη γραμμή πρέπει να έχει, σε αυτό το χρονικό διάστημα, συνολικά μικρότερη κλίση. Είτε δηλαδή κλιμακωτά (όπως στο διάγραμμα), είτε ακολουθώντας πχ την προηγούμενη κλίση κατά το πρώτο μισό διάστημα και συνεχίζοντας οριζόντια για το άλλο μισό, το σφάλμα στην εκτέλεση γίνεται αντιληπτό. Ο θόρυβος, επίσης, κατά τη διάρκεια της μουσικής εκτέλεσης φαίνεται πως δεν επηρέασε το συγχρονισμό.

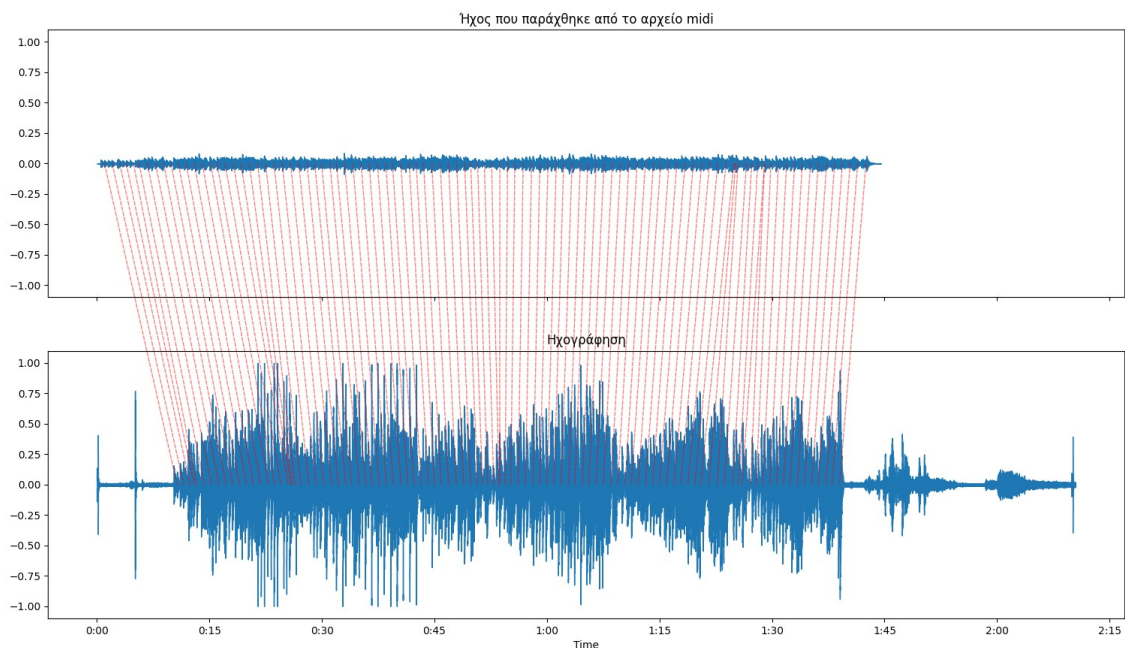


Εικόνα 4.9 – Βέλτιστη διαδρομή στον πίνακα συσσωρευμένου κόστους

Παρατηρούμε, επίσης, ότι το πάνω άκρο της γραμμής δε φτάνει στο μέγιστο ύψος του διαγράμματος (στο τέλος του χρόνου του αρχείου ήχου που παράχθηκε από το αρχείο midi). Αυτό αφορά ένα πολύ μικρό χρονικό διάστημα στο τέλος του αρχείου αυτού, στο οποίο η

ένταση της τελευταίας νότας φθίνει σταδιακά. Προκλήθηκε από την περικοπή που έγινε στο αρχείο wr και δεν επηρεάζει τη χρήση ή την απόδοση της εφαρμογής.

Μια άλλη αναπαράσταση του συγχρονισμού, αποτελεί το διάγραμμα που φαίνεται στην Εικόνα 4.10 όπου οι κόκκινες διακεκομμένες γραμμές ενώνουν τα αντιστοιχούντα σημεία στις κυματομορφές των δύο αρχείων ήχου. Το διάγραμμα αυτό έχει παραχθεί με τη συνάρτηση `librosa.display.waveshow`⁴⁷. Στο διάγραμμα αυτό, ο άξονας του χρόνου είναι κοινός, για να γίνει φανερή η διαφορετική διάρκεια των αρχείων ήχου. Επίσης, φαίνεται το διαφορετικό tempo, καθώς η διάρκεια του αντιστοιχιζόμενου μέρους, που αφορά την ηχογράφιση, δεν έχει το ίδιο μήκος. Ο μαθητής δηλαδή, απέδωσε την παρτιτούρα με μεγαλύτερη ταχύτητα από ό,τι θα έπρεπε.



Εικόνα 4.10 – Σύνδεση αντίστοιχων σημείων στις δύο κυματομορφές

Τα σημεία των παραλείψεων του μαθητή, που αναφέραμε νωρίτερα, είναι εκεί όπου οι διακεκομμένες γραμμές συγκλίνουν στην κάτω κυματομορφή. Η περιοχή της επανάληψης είναι εκεί που συγκλίνουν στην πάνω.

⁴⁷ <https://librosa.org/doc/main/generated/librosa.display.waveshow.html>

Επιπλέον, έχουμε εικόνα της έντασης του θορύβου πριν και μετά τη μουσική εκτέλεση, ο οποίος όπως φαίνεται δεν είναι ασήμαντος.

Ένα ακόμα στοιχείο που φανερώνεται στο ίδιο διάγραμμα, είναι το ότι η μεταβλητότητα της έντασης της ηχογράφησης δεν επηρεάζει το συγχρονισμό.

Με δεδομένο το γεγονός ότι ακόμα και στις ιδιαίτερα δυσμενείς συνθήκες που δοκιμάστηκε η εφαρμογή, ο συγχρονισμός γίνεται σωστά, μπορούμε με ασφάλεια να θεωρήσουμε ότι στις περισσότερες πραγματικές περιπτώσεις, το αποτέλεσμα θα είναι τουλάχιστο ικανοποιητικό.

Έγιναν επίσης δοκιμές με εγγραφές σε καλές συνθήκες (ελάχιστος θόρυβος και καλή ποιότητα ηχογράφησης) και αρκετά καλή απόδοση της παρτιτούρας. Σε αυτές τις περιπτώσεις ο συγχρονισμός ήταν αλάνθαστος.

5 Συμπεράσματα

5.1 Σύνοψη

Στο πλαίσιο της παρούσας εργασίας, αρχικά μελετήθηκαν οι δυνατότητες των συστημάτων μουσικής εκπαίδευσης Musicolab, Yousician και SoundSlice, ώστε να διερευνηθούν τρέχουσες ανάγκες των συστημάτων αυτών και να προσδιοριστούν οι λειτουργικές απαιτήσεις για την υπό ανάπτυξη διαδικτυακή εφαρμογή. Στη συνέχεια, διερευνήθηκαν οι δυνατότητες των βιβλιοθηκών Librosa και Essentia και παρουσιάστηκαν οι προοπτικές για την ανάπτυξη καινοτόμων εφαρμογών που να αξιοποιούν ερευνητικά επιτεύγματα αιχμής. Τέλος, αναπτύχθηκε μια διαδικτυακή εφαρμογή η οποία αφενός υποστηρίζει τη μουσική εκπαίδευση, ικανοποιώντας την ανάγκη του συγχρονισμού μιας μουσικής εκτέλεσης με τη συμβολική αναπαράσταση του μαθησιακού στόχου, αφετέρου αξιοποιεί σύγχρονα επιτεύγματα στην αυτόματη ανάλυση μουσικού περιεχομένου.

Η εφαρμογή αυτή ενσωματώθηκε στην πλατφόρμα μουσικής εκπαίδευσης Musicolab. Για την ανάπτυξή της, χρησιμοποιήθηκαν βιβλιοθήκες ελεύθερου λογισμικού και γράφηκε κώδικας σε HTML, JavaScript και Python.

Αντικείμενό της, είναι ο συγχρονισμός μουσικής σημειογραφίας (παρτιτούρας) με την ηχογραφημένη απόδοσή της από τον μαθητή. Οι δοκιμές στη χρήση της, έδωσαν ικανοποιητικά αποτελέσματα και επομένως αναμένεται να παρέχει ένα χρήσιμο διαδικτυακό εργαλείο, για διδάσκοντες και διδασκόμενους.

5.2 Μελλοντικές επεκτάσεις

Οι ιδέες και οι δυνατότητες με τις οποίες θα μπορούσε να επεκταθεί η εφαρμογή του συγχρονισμού είναι αρκετές. Ενδεικτικά θα παραθέσουμε κάποιες από αυτές:

5.2.1 Εκτίμηση συνολικής απόδοσης

Η τιμή του πίνακα D που αντιστοιχεί στην αντιστοίχιση των τελευταίων block των δύο αρχείων ήχου θα μπορούσε να είναι ένα μέτρο της συνολικής απόδοσης της παρτιτούρας σε

ήχο και θα μπορούσε να δίνεται υπό μορφή σκορ στο μαθητή, ώστε να έχει μια πρώτη εικόνα για το αν τα πήγε καλύτερα ή χειρότερα από την προηγούμενη φορά. Πέρα από αυτή την τιμή, που θα έδινε κυρίως τη συνολική εικόνα της χρωματικής απόδοσης, θα μπορούσε το “σκορ” να συνδυάζεται με την ομαλότητα ή μη του πίνακα `wr_s` και της κλίσης του (ιδανικά 45 μοίρες) ώστε να συνεκτιμηθεί και η απόδοση στο tempo.

5.2.2 Επισήμανση λαθών στο χρώμα ή το tempo

Ως επόμενο βήμα της παρούσας εφαρμογής, θα μπορούσε να γίνει εκτίμηση του πόσο πιστά (ή όχι) αποδόθηκε η παρτιτούρα στην ηχογράφιση του μαθητή. Δεδομένου ότι ο πίνακας `D` παράγεται ούτως ή άλλως κατά την εκτέλεση του κώδικα σε Python, θα μπορούσε σε συνδυασμό με τη βέλτιστη διαδρομή του πίνακα `wr_s` να παραχθεί πίνακας που να περιέχει τις τιμές του πίνακα κόστους πάνω στη βέλτιστη διαδρομή. Με αυτές τις τιμές, είναι δυνατόν να εντοπίσουμε τα σημεία στα οποία έχουν παιχτεί λάθος νότες από τον μαθητή και με γραφική αναπαράσταση αυτό να γίνεται εμφανές πάνω στην κυματομορφή της ηχογράφησης ή ακόμα και πάνω στην παρτιτούρα.

Παρόμοια, από τα δεδομένα του πίνακα `wr_s` είναι δυνατόν να έχουμε εικόνα για τα σημεία στα οποία το tempo παρουσίασε διακύμανση, δηλαδή ο μαθητής έπαιξε πιο αργά ή πιο γρήγορα από το μέσο tempo της εκτέλεσής του και πάλι αυτό να αποδίδεται γραφικά είτε στην παρτιτούρα είτε στην κυματομορφή.

5.2.3 Απεικόνιση τρέχουσας νότας από ηχογράφιση

Εύκολα θα μπορούσαμε να κάνουμε τον αντίστροφο συγχρονισμό, δηλαδή να μπορούμε να επιλέγουμε σημείο της ηχογράφησης και να επισημαίνεται το μέτρο στο οποίο αντιστοιχεί το σημείο αυτό. Επιπλέον όμως, θα μπορούσαμε να επεξεργαστούμε το αρχείο ήχου όσον αφορά την εύρεση της έναρξης ηχητικού γεγονότος (ατάκας) κάθε μιας νότας και του αντίστοιχου τονικού ύψους ώστε να μπορούμε κατά την αναπαραγωγή της ηχογράφησης να επισημαίνουμε την τρέχουσα νότα που ακούγεται και την ορθότητα ή μη αυτής σε επίπεδο χρόνου ή χρώματος.

5.2.4 Συγχρονισμός σε πραγματικό χρόνο

Σε μια εκτενέστερη επέκταση, θα μπορούσαμε να περάσουμε από την επεξεργασία αποθηκευμένων ηχητικών σημάτων σε επεξεργασία ζωντανών ηχητικών ροών καθώς αυτές λαμβάνονται κατά τη διάρκεια της μουσικής εκτέλεσης/ερμηνείας. Μια τέτοια υλοποίηση θα ήταν μάλλον αναγκαίο να γίνει τοπικά (πχ επεξεργασία μέσω του ίδιου του περιηγητή στον οποίο γίνεται η ηχογράφηση), ώστε να αποφευχθούν οι όποιες καθυστερήσεις οφείλονται στη μεταφόρτωση των ηχητικών δεδομένων σε κάποιο διαδικτυακό εξυπηρετητή. Ο κώδικας θα μπορούσε να γραφεί σε JavaScript και να γίνει και χρήση των βιβλιοθηκών ανοιχτού λογισμικού που έχουν γραφεί σε αυτή τη γλώσσα (πχ Essentia). Σε αυτή την περίπτωση, θα μπορούσαμε να έχουμε ζωντανή επισήμανση της κάθε νότας που παίζεται από το μαθητή, τη στιγμή που αυτή ακούγεται (Score Following), ενώ ακόμα θα μπορούσαν να επισημαίνονται και πιθανά λάθη κατά τη μουσική εκτέλεση.

6 Αναφορές

- Alexandraki, C., Akoumianakis, D., Kalochristianakis, M., Zervas, P., Kaliakatsos-Papakostas, M. & Cambouropoulos, E. (2022) MusiCoLab: Towards a Comprehensive Environment for Collaborative Music Learning, Web Audio Conference 2022, (WAC 2022), Cannes, France. <https://doi.org/10.5281/zenodo.6770559>.
- Bogdanov, D., Wack, N., Gómez Gutiérrez, E., Gulati, S., Boyer, H., Mayor, O., ... & Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. In Britto A, Gouyon F, Dixon S, editors. *14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil. ISMIR; 2013. p. 493-8.*. International Society for Music Information Retrieval (ISMIR).
- Correya, A. A., Bogdanov, D., Joglar-Ongay, L., & Serra, X. (2020). Essentia.js: A JavaScript library for music and audio analysis on the web. In Cumming J, Ha Lee J, McFee B, Schedl M, Devaney J, McKay C, Zagerle E, de Reuse T, editors. *Proceedings of the 21st International Society for Music Information Retrieval Conference; 2020 Oct 11-16; Montréal, Canada.[Canada]: ISMIR; 2020. p. 605-12.*. International Society for Music Information Retrieval (ISMIR).
- Correya, A. A., Marcos Fernández, J., Joglar-Ongay, L., Alonso Jiménez, P., Serra, X., & Bogdanov, D. (2021). Audio and Music Analysis on the Web using Essentia.js. *Transactions of the International Society for Music Information Retrieval. 2021; 4 (1): 167-81.*
- Dias, L., Victor, A. (2017). Teaching and learning with mobile devices in the 21st century digital world: Benefits and challenges. *European Journal of Multidisciplinary Studies, 2(5), 339-344.*
- Driessen, P. F., Darcie, T. E., & Pillay, B. (2011). The effects of network delay on tempo in musical performance. *Computer Music Journal, 35(1), 76-89.*
- Freeman, S. A., & Field, D. W. (2004). Student Perceptions of Web-Based Supplemental Instruction. *Journal of Technology Studies, 30(4), 25-31.*
- Knapp, N. F. (2019). The shape activity: Social constructivism in the psychology classroom. *Teaching of Psychology, 46(1), 87-91*
- Malik, S., & Agarwal, A. (2012). Use of multimedia as a new educational technology tool-A study. *International Journal of Information and Education Technology, 2(5), 468.*
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015, July). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference* (Vol. 8, pp. 18-25).

- Müller, M. (2015). *Fundamentals of music processing: Audio, analysis, algorithms, applications* (Vol. 5). Cham: Springer.
- Singh, M. P., & Jain, M. K. (2014). Evolution of processor architecture in mobile phones. *International Journal of Computer Applications*, 90(4), 34-39.
- Worthington, T. (2013, April). Synchronizing asynchronous learning-Combining synchronous and asynchronous techniques. In *2013 8th International Conference on Computer Science & Education* (pp. 618-621). IEEE.