

**ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΣΧΟΛΗ ΚΑΛΩΝ ΤΕΧΝΩΝ
ΤΜΗΜΑ ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ**



**ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ**

Θέμα διπλωματικής εργασίας:

Ανάλυση ακουστικής σκηνής:

Διερεύνηση της διαδικασίας απομόνωσης ηχητικών stem στη μουσική μέσω του διαδικτυακού εργαλείου τεχνητής νοημοσύνης Spleeter.

Auditory scene analysis:

Exploring the process of stem separation in music using the online artificial intelligence tool Spleeter.

Φοιτητής:

Ιωάννης Στυμονιάρης AEM 2165

**Επιβλέπων καθηγητής:
Αιμίλιος Καμπουρόπουλος**

Θεσσαλονίκη, Σεπτέμβριος 2025

Πίνακας Περιεχομένων

Ευχαριστίες	4
Περίληψη.....	5
Abstract.....	6
Παραρτήματα	7
1. Εισαγωγή.....	8
1.1 Δομή της διπλωματικής εργασίας.....	9
2. Βιβλιογραφική επισκόπηση	11
2.1 Ανάλυση ακουστικής σκηής (ASA)	11
2.1.1 Εισαγωγή στην ASA	11
2.1.2 Albert Stanley Bregman	12
2.1.3 Η Ανάλυση Ακουστικής Σκηής (ASA) κατά τον Bregman.....	13
2.2.4 Ανάκτηση της Μουσικής Πληροφορίας (MIR)	16
2.2 Μοντέλα διαχωρισμού των μουσικών stem (Music stem separation models).....	16
2.2.1 Εισαγωγή στα μοντέλα διαχωρισμού	16
2.2.2 Συνελκτικά Νευρωνικά Δίκτυα (CNN)	17
2.2.3 Αρχιτεκτονική U-Net.....	17
2.2.4 Περιγραφή της λειτουργίας των μοντέλων διαχωρισμού	17
2.2.5 Βασικός και Μικρού Χρόνου Μετασχηματισμός Fourier (Short-Time Fourier Transform, STFT)	18
2.3 Περιγραφή και λειτουργία του Spleeter	19
2.3.1 Περιγραφή του Spleeter.....	19
2.3.2 Διαδικασία διαχωρισμού των stems και STFT	20
3. Μεθοδολογία	22
3.1 Ερευνητικός Σχεδιασμός.....	22
3.2 Στόχοι.....	22
3.3 Ερευνητικά Ερωτήματα.....	22
3.4 Περιγραφή των Εργαλείων.....	23
3.4.1 Spleeter	23
3.4.2 Sonic Visualiser	23
3.5 Επιλογή Μουσικών Έργων.....	24
3.6 Πρακτική Διαδικασία.....	26
3.7 Κλίμακα Καθαρότητας και Διαρροής (ΚΚΔ), Ποιότητα Διαχωρισμού.....	27

3.8 Υποθέσεις στο στάδιο της προεργασίας.....	28
3.9 Περιορισμοί της έρευνας.....	29
4. Παρουσίαση δεδομένων, Ανάλυση και Ευρήματα	31
4.1 Μουσικό παράδειγμα 1: Borislav Slavov: Baldur's Gate 3 – Down By The River.....	31
4.2 Μουσικό παράδειγμα 2: Camille Saint-Saëns: Carnival of the Animals, VII. Aquarium	37
4.3 Μουσικό παράδειγμα 3: Claude Bolling: Suite for Flute and Jazz Piano Trio.....	43
4.4 Μουσικό παράδειγμα 4: Dave Brubeck: Take Five.....	48
4.5 Μουσικό παράδειγμα 5: Kristen Anderson-Lopez & Robert Lopez: Show Yourself	54
4.6 Μουσικό παράδειγμα 6: Gareth Coker: Morgana’s Theme	59
4.7 Μουσικό παράδειγμα 7: Glenn Miller: Moonlight Serenade	64
4.8 Μουσικό παράδειγμα 8: G.F. Handel: Flute Sonata in F Major – III. Siciliana.....	69
4.9 Μουσικό παράδειγμα 9: Henry Mancini: Lujon	75
4.10 Μουσικό παράδειγμα 10: Christmas Oratorio, BWV 248: Part VI, 41. Arie "Ich will nur dir zu Ehren leben"	80
4.11 Μουσικό παράδειγμα 11: Kevin Penkin: Here, Tomorrow	85
4.12 Μουσικό παράδειγμα 12: Ola Gjeilo: Across The Vast, Eternal Sky	91
4.13 Μουσικό παράδειγμα 13: Ola Gjeilo: Dark & Luminous Night	96
4.14 Μουσικό παράδειγμα 14: S. Rachmaninoff: Symphonic Dances (first movement)	102
4.15 Μέση απόδοση του Spleeter	107
5. Συμπεράσματα και Συζήτηση.....	108
5.1 Βασικά Συμπεράσματα	108
5.2 Συζήτηση	109
5.3 Προτάσεις για Μελλοντική Έρευνα.....	109
6. Επίλογος	111
Βιβλιογραφία.....	112

Ευχαριστίες

Πριν από έξι χρόνια ξεκίνησα ένα μουσικό ταξίδι από το οποίο μου χάρισε πολλά περισσότερα από ένα απλό σύνολο γνώσεων. Πέρα από την κατάκτηση της μουσικής επιστημονικής μεθόδου και την εκμάθηση του τι ερωτήσεις πρέπει να θέτω, βρέθηκα σε πρωτόγνωρες για εμένα καταστάσεις που απαιτούσαν συνεχή προσαρμογή και ανάπτυξη νέων δεξιοτήτων. Αν και συχνά σε μια μοναχική πορεία, ωφελήθηκα βαθιά από την καλοσύνη και τη σοφία όλων των ανθρώπων που συνάντησα κατά τη διάρκεια αυτού του ταξιδιού. Είμαι ευγνώμων σε αυτούς για την καθοδήγηση, τη στήριξη, τη φιλία τους και τις στιγμές που δημιουργήσαμε και μοιραστήκαμε μαζί.

Ευχαριστώ τον επιβλέποντα καθηγητή μου, κύριο Αιμίλιο Καμπουρόπουλο, ο οποίος μου έδωσε την ευκαιρία να εκπονήσω μαζί του τη διπλωματική μου εργασία σε ένα νέο σχετικά αντικείμενο και με καθοδήγησε στη συγγραφή της. Μία διπλωματική είναι μια συνεχόμενη διαδικασία ανατροφοδότησης και προσεκτικού σχεδιασμού. Από τον καθηγητή μου επίσης έμαθα ότι η επιστήμη πρέπει να φτάνει πέρα από τους στενούς τοίχους του εργαστηρίου, να εμπνέει και να προσφέρει στην επιστημονική κοινότητα με κάθε δυνατό τρόπο.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου προς την οικογένειά μου, τους φίλους μου και την καθηγήτρια πιάνου μου, Αντωνέλα Πουλούλη, για τη συνεχή στήριξη και ενθάρρυνσή τους όλα αυτά τα χρόνια. Στάθηκαν στο πλευρό μου στις πιο δύσκολες στιγμές και με βοήθησαν να συνεχίσω την μουσική μου πορεία.

Περίληψη

Μέσω της διαδικασίας της Ανάλυσης Ακουστικής Σκηνής (Auditory Scene Analysis, ASA) πολύπλοκα ακουστικά περιβάλλοντα διαιρούνται σε επιμέρους διακριτές πηγές/ροές ήχου. Στη μουσική, ένα μέρος αυτής της διαδικασίας είναι και ο διαχωρισμός στελεχών (Music Stem Separation – MSS), κατά τον οποίο απομονώνονται ομάδες πηγών από το αρχικό σήμα, όπως για παράδειγμα όλα τα φωνητικά ή τα κρουστά. Λόγω της μεγάλης ταχύτητας και εξέλιξης της τεχνητής νοημοσύνης, η ανάπτυξη εργαλείων για αυτή τη διαδικασία είναι ευκολότερη διότι, ενσωματώνονται μοντέλα τα οποία βασίζονται στις διάφορες μεθόδους βαθιάς εκμάθησης (deep learning).

Στην παρούσα διπλωματική εργασία διερευνήθηκε η διαδικασία διαχωρισμού των stems μέσω της χρήσης του εργαλείου τεχνητής νοημοσύνης Spleeter. Αρχικά, παρουσιάζεται μια βιβλιογραφική ανασκόπηση σχετιζόμενη με το βασικό θέμα της εργασίας. Στη συνέχεια, γίνεται μία παρουσίαση και ανάλυση δεκατεσσάρων μουσικών παραδειγμάτων, ακολουθούμενα από μια συζήτηση και αξιολόγηση της απόδοσης του μοντέλου διαχωρισμού που επιλέχθηκε. Τελικός στόχος της εργασίας είναι να εντοπίσει τα πλεονεκτήματα και τους περιορισμούς του συγκεκριμένου εργαλείου διαχωρισμού.

Λέξεις κλειδιά: Μουσικός Διαχωρισμός, Ανάλυσης Ακουστικής Σκηνής, Διαχωρισμός Stem, Spleeter

Abstract

Through the process of Auditory Scene Analysis (ASA), complex acoustic environments are divided into distinct sound sources/streams. In music, part of this process is Music Stem Separation (MSS), during which groups of sources are isolated from the original signal, such as all vocals or drums. Rapid advances in artificial intelligence have simplified tool development by enabling the integration of models based on various deep learning methods.

In the present thesis, the stem separation process has been explored via the use of the artificial intelligence tool Spleeter. First, a bibliography overview related to the main subject of the study is presented. This is followed by the presentation and analysis of fourteen musical examples, along with a discussion and evaluation of the performance of the selected separation model. The final purpose of this work is to identify the advantages and limitations of this specific stem separation tool.

Key Words: Music separation, Auditory Scene Analysis, Stem Separation, Spleeter.

Παραρτήματα

Συντομογραφίες

ASA: Auditory Scene Analysis

ASS: Audio Source Separation

CNN: Convolutional Neural Network

CPU: Central Processing Unit

DAW: Digital Audio Workstation

FRSC: Fellowship of the Royal Society of Canada

FT: Fourier Transform

GPU: Graphic Processing Unit

LUFS: Loudness Unit related to Full Scale

MIR: Music Information Retrieval

MSS: Music Source Separation

SDR: Signal Distortion Ratio

STFT: Short-Term Fourier Transform

1. Εισαγωγή

Η μουσική αποτελεί μία από τις αρχαιότερες γλώσσες έκφρασης. Δεν είναι τυχαίο ότι η πρώτη μορφή επικοινωνίας μεταξύ των ανθρώπων δεν ήταν γραπτή αλλά πραγματοποιούνταν μέσω του ήχου (Γιάννου, 1995). Με το πέρασμα των αιώνων, η μουσική εξελίχθηκε σε τέχνη και επιστήμη. Στις μέρες μας, συνεχίζει να αποτελεί αναπόσπαστο μέρος της καθημερινότητας των ανθρώπων. Η ραγδαία τεχνολογική πρόοδος και ειδικότερα η ψηφιακή επανάσταση είναι αυτή που την έχει καταστήσει ευρέως προσβάσιμη από όλους, σχεδόν σε όλα τα πλάτη και μήκη του πλανήτη μας. (Volk et al., 2011) Διαδικτυακές πλατφόρμες (streaming), λογισμικά μουσικής παραγωγής και εφαρμογές τεχνητής νοημοσύνης έχουν μεταμορφώσει τόσο τη δημιουργία όσο και την «κατανάλωση» της ίδιας της μουσικής, δίνοντας δυνατότητες και εργαλεία στους καλλιτέχνες που προφανώς δεν προϋπήρχαν μέχρι σήμερα. Μέσα σε αυτό το ευρύτερο τεχνολογικό πλαίσιο εντάσσεται και το επιστημονικό πεδίο που ονομάζουμε **Διαχωρισμός Ακουστικών - Μουσικών ροών (Audio - Music Source Separation , ASS - MSS)** το οποίο αποτελεί το γενικό αντικείμενο της παρούσας διπλωματικής εργασίας.

Έπειτα, από σχεδόν μισό αιώνα ερευνών και σύνταξη βιβλιογραφίας, η ανάλυση και ο διαχωρισμός των ακουστικών ροών έχει φτάσει στο σημείο να έχει εφαρμογές εκτός από ερευνητικούς σκοπούς στα μουσικά πανεπιστήμια και σε διάφορα κέντρα μελέτης και έρευνας, στην καθημερινή ζωή των ανθρώπων λόγω και της χρήσης της από τις διάφορες εταιρίες για εμπορικούς και ψυχαγωγικούς σκοπούς (Pardo et al., 2018). Εκτός από τις πιο αληθοφανείς χρήσεις που μπορούμε να σκεφτούμε όπως είναι για παράδειγμα το караόκε (karaoke), η διαδικασία συνένωσης δειγμάτων (sampling) που χρησιμοποιούν οι DJs ή και ακόμη η δημιουργία συνοδευτικών οργανικών μερών σε κοντσέρτα, όπου δεν είναι δυνατή η εύρεση ορχήστρας, με σκοπό οι σολίστες να μάθουν καλύτερα το μέρος τους. Στη περίπτωση του караόκε αυτή η διαδικασία του διαχωρισμού των μουσικών ροών είναι αναγκαία διότι, είναι πιο εύκολο από ποτέ να δημιουργήσεις οργανικά μέρη για τη συνοδεία της φωνής αφαιρώντας τη βασική φωνή του αρχικού τραγουδιστή μέσα σε μερικά δευτερόλεπτα (Hennequin et al., 2020).

Στη συγκεκριμένη χρονική στιγμή που βρισκόμαστε, η ίδια η διαδικασία της απομόνωσης των ακουστικών ροών βασίζεται κυρίως σε αυτό που ονομάζουμε βαθιά μάθηση (deep learning). Η σύγχρονη έρευνα στην διαδικασία απομόνωσης ροών εστιάζεται σε μεθόδους του deep learning, χωρίς όμως να λαμβάνονται απαραίτητα υπόψιν συγκεκριμένα χαρακτηριστικά και ιδιότητες της ίδιας της μουσικής (Défossez et al., 2021). Οι αλγόριθμοι, αφού τους δοθεί μεγάλος αριθμός μουσικών έργων ως είσοδος, μπορούν να τα επεξεργαστούν και να «μαθαίνουν» από την ποικιλία των παραδειγμάτων. Με τον όρο «μαθαίνουν» εννοούμε ότι αποκτούν τη δυνατότητα να αναγνωρίζουν και να διακρίνουν διαφορετικούς ήχους και μουσικά όργανα μέσα σε αυτά.

Ένα μεγάλο πρόβλημα είναι ότι οι αλγόριθμοι, αν και ισχυροί, δεν αντιλαμβάνονται τη μουσική όπως την καταλαβαίνουμε εμείς οι άνθρωποι. (Ren et al., 2020; Lupker & Turkel, 2021; Schreiber et al., 2024) Δεν δίνουν σημασία σε μουσικά στοιχεία, όπως: **ρυθμός**: αν υπάρχει σταθερό ρυθμικό σχήμα ή χτύπημα (π.χ. τύμπανα σε συγκεκριμένο μοτίβο), **μελωδία**: πώς «κινείται» η μελωδική γραμμή μίας φωνής ή ενός

οργάνου (ανεβαίνει ή κατεβαίνει τη μουσική κλίμακα), **αρμονία**: ποιες νότες ενώ ηχούν μεταξύ τους ταυτόχρονα σχηματίζουν αυτό που λέμε συγχορδίες. Με άλλα λόγια, οι τεχνικές που χρησιμοποιούν μπορούμε να πούμε ότι είναι πολύ προχωρημένες, αλλά «βλέπουν» τον ήχο σαν ένα απλό σήμα και όχι σαν μουσική με νόημα, δομή και κατεύθυνση. Αποτέλεσμα αυτού είναι να περιορίζεται το πόσο «έξυπνα» και μουσικά σωστά γίνεται ο εκάστοτε διαχωρισμός.

Η παρούσα εργασία εστιάζει στο επιστημονικό αυτό πεδίο του Διαχωρισμού των Ακουστικών - Μουσικών ροών, μελετώντας πόσο ακριβής είναι τελικά ο διαχωρισμός. Μέσω της εισαγωγής δεκατεσσάρων (14) μουσικών παραδειγμάτων σε εργαλείο τεχνητής νοημοσύνης διαχωρισμού ακουστικών ροών, επιχειρείται μία εις βάθος ανάλυση, παρατήρηση και κατανόηση των δυνατοτήτων του διαχωρισμού από το αρχικό μουσικό έργο. Επιλέξαμε να επικεντρωθούμε στο **Spleeter**, ένα εργαλείο τεχνητής νοημοσύνης ανοιχτού κώδικα που αναπτύχθηκε από την Deezer και έχει εξελιχθεί σε ένα από τα πιο διαδεδομένα εργαλεία διαχωρισμού μουσικών ροών (Hennequin et al., 2020). Η επιτυχία του Spleeter οφείλεται στην ταχύτητα και την απλότητά του. Μπορεί να λειτουργήσει με μία μόνο εντολή και να αποδώσει καθαρά κομμάτια φωνητικών και συνοδείας μέσα σε λίγα δευτερόλεπτα. Επιπλέον, παρέχεται δωρεάν στο διαδίκτυο.

Η επιλογή του συγκεκριμένου θέματος προέρχεται από έναν συνδυασμό ερευνητικών κινήτρων. Από τη μία πλευρά, η φύση της ίδιας της μουσικής προσφέρει ένα πλούσιο πεδίο μελέτης. Από την άλλη, η ολοένα μεγαλύτερη ενσωμάτωση της τεχνητής νοημοσύνης στις ζωές των ανθρώπων εγείρει σημαντικά ερωτήματα για το πώς τελικά αναδιαμορφώνεται η ίδια η μουσική. Αναλύοντας τη διαδικασία του διαχωρισμού των stem μέσω του Spleeter, η εργασία αυτή επιχειρεί να συνδέσει τις θεωρητικές γνώσεις για την ανάλυση της ακουστικής σκηνής με τα τεχνητής νοημοσύνης εργαλεία διαχωρισμού.

1.1 Δομή της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία είναι δομημένη με τρόπο που εξυπηρετεί τη σταδιακή παρουσίαση του θέματος, προκειμένου να διασφαλιστεί η πλήρης κατανόηση τόσο των θεωρητικών όσο και των πρακτικών συστατικών του.

Στο Κεφάλαιο 1, παρατίθεται η εισαγωγή. Η εργασία ξεκινά με μια συνοπτική, αλλά περιεκτική γενική εισαγωγή, όπου εισάγονται βασικές έννοιες που είναι απαραίτητες για την κατανόηση του συγκεκριμένου ερευνητικού πεδίου. Οι έννοιες αυτές συνοδεύονται από ενδεικτικά παραδείγματα, ώστε να διευκολυνθεί η πρώτη επαφή του αναγνώστη με το αντικείμενο. Στη συνέχεια, θα επανέρθουμε σε αυτές τις έννοιες, τις οποίες θα αναλύσουμε σε μεγαλύτερο βάθος, δεδομένου ότι βοηθούν στην εξέλιξη της έρευνας στα παρακάτω κεφάλαια της διπλωματικής εργασίας.

Το Κεφάλαιο 2 είναι αφιερωμένο στη βιβλιογραφική επισκόπηση. Σε αυτό, παρουσιάζεται αναλυτικά η έννοια του Auditory Scene Analysis (ASA), με ιδιαίτερη έμφαση στη συμβολή του Albert Bregman, ο οποίος υπήρξε θεμελιωτής του συγκεκριμένου ερευνητικού πεδίου. Στη συνέχεια,

εμβαθύνουμε στην έννοια του Deep Learning, εξηγώντας τις βασικές αρχές της λειτουργίας του. Επίσης, θα αναλύσουμε τον τρόπο, την χρησιμότητα καθώς και αναγκαιότητα των αλγορίθμων που σχετίζονται με το source separation, Επιπλέον, γίνεται αναφορά στον μετασχηματισμό Fourier, καθώς αποτελεί βασικό εργαλείο στην ανάλυση των ακουστικών ροών. Τέλος, ιδιαίτερη έμφαση δίνεται και στην περιγραφή του τρόπου λειτουργίας και του εργαλείου τεχνητής νοημοσύνης Spleeter, το οποίο χρησιμοποιήθηκε στην παρούσα εργασία για τον διαχωρισμό των μουσικών παραδειγμάτων.

Στο Κεφάλαιο 3, παρουσιάζεται αναλυτικά η μεθοδολογία που ακολουθήθηκε. Εδώ περιγράφονται βήμα προς βήμα όλες οι ενέργειες και επιλογές που εφαρμόστηκαν προκειμένου να καταλήξουμε στα τελικά συμπεράσματα της έρευνας. Παράλληλα, γίνεται εκτενής αναφορά στα μουσικά έργα που χρησιμοποιήθηκαν ως παραδείγματα, εξετάζοντας τα χαρακτηριστικά τους, όπως ο αριθμός των μουσικών οργάνων, ο αναμενόμενος αριθμός των stems μετά την διαδικασία του διαχωρισμού, καθώς και τα κριτήρια βάσει των οποίων επελέγησαν. Στο τέλος του κεφαλαίου, καταγράφονται οι υποθέσεις οι οποίες στο στάδιο της προεργασίας καθώς και οι περιορισμοί, ώστε να αποσαφηνιστούν τα όρια και οι δυνατότητες της έρευνας..

Στο Κεφάλαιο 4, παρουσιάζονται τα αποτελέσματα της έρευνας. Γίνεται ταξινόμηση, περιγραφή και ερμηνεία των εξαγόμενων δεδομένων με σκοπό την πλήρη κατανόηση των ευρημάτων. Ειδικότερα, γίνεται αναλυτική παρουσίαση των δεδομένων που προέκυψαν μετά την εφαρμογή του εργαλείου Spleeter στον διαχωρισμό των μουσικών παραδειγμάτων. Στο τέλος κάθε παραδείγματος, περιλαμβάνεται ένας συνοπτικός πίνακας, ο οποίος συγκεντρώνει και οπτικοποιεί τα κύρια ευρήματα, διευκολύνοντας έτσι την συνοπτική ανασκόπησή τους.

Στο Κεφάλαιο 5, αναπτύσσονται τα συμπεράσματα. Σε αυτό το κεφάλαιο, συνοψίζονται τα βασικά συμπεράσματα της εργασίας, ενώ στη συνέχεια αναπτύσσεται η συζήτηση γύρω από αυτά, με στόχο την ερμηνεία και την αποτίμησή τους. Παράλληλα διατυπώνονται προτάσεις για μελλοντική έρευνα, η οποία μπορεί να επεκτείνει ή να εμβαθύνει περαιτέρω σε επιμέρους πτυχές του συγκεκριμένου αντικειμένου. Οι προτάσεις αυτές απευθύνονται σε μελλοντικούς ερευνητές και επαγγελματίες του χώρου, με στόχο την εξέλιξη του πεδίου.

Τέλος, στο Κεφάλαιο 6, αποτελεί και τον επίλογο της εργασίας. Ακολουθεί βιβλιογραφία στην οποία καταγράφονται όλες οι πηγές που χρησιμοποιήθηκαν κατά την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας.

2. Βιβλιογραφική επισκόπηση

2.1 Ανάλυση ακουστικής σκηνής (ASA)

2.1.1 Εισαγωγή στην ASA

Η ανάλυση ακουστικών ροών ή αλλιώς όπως αναφέρεται και στην βιβλιογραφία Auditory Stream Analysis είναι στην ουσία υποκατηγορία του γενικού όρου Ανάλυση Ακουστικής Σκηνής (Auditory Scene Analysis - ASA), η οποία αποτελεί έναν από τους σημαντικότερους τομείς της ακουστικής επιστήμης. Εστιάζει στη δυνατότητα του ανθρώπινου ακουστικού συστήματος να αποσπά, να ξεχωρίζει, να οργανώνει και να κατανοεί διαφόρων ειδών ηχητικές πηγές μέσα από ένα πολυσύνθετο ακουστικό περιβάλλον. (Magron et al., 2021; Kong et al., 2020) Σε αντίθεση με ό,τι συμβαίνει στην οπτική αντίληψη του ανθρώπου, όπου τα αντικείμενα συχνά είναι διακριτά (σε μία λογική για την ανθρώπινη όραση απόσταση), οι ήχοι από την άλλη πλευρά, οι οποίοι μπορεί να προέρχονται από διαφόρων ειδών ηχητικές πηγές π.χ. άλλους ανθρώπους, ήχους της φύσης, μουσικά όργανα, θόρυβος κτλ., παράγουν ηχητικά κύματα τα οποία φτάνουν στα αυτιά μας αναμειγμένα και στη συνέχεια συνενωμένα σε ένα ενιαίο ηχητικό κύμα. Οι βασικές μεταβλητές οι οποίες επηρεάζουν τη δημιουργία διακριτών ακουστικών ροών περιλαμβάνουν τη συχνότητα, το χρονικό διάστημα μεταξύ των ήχων, το ηχόχρωμα, και τη τοποθέτηση των ηχητικών πηγών μέσα στο χώρο.

Το παραπάνω φαινόμενο, βασίζεται σε μία λειτουργία του εγκεφάλου μας, η οποία ενεργοποιείται αυτόματα και επηρεάζεται από τη προσοχής μας. Δηλαδή, ο ακροατής μπορεί να επιλέξει συνειδητά να εστιάσει σε μία ροή, φιλτράροντας τις υπόλοιπες. Για παράδειγμα, όταν δύο μουσικά όργανα (π.χ. ένα φλάουτο και ένα βιολί) παίζουν εναλλάξ νότες με σαφώς διαφορετικές συχνότητες και ηχοχρώματα, ο ακροατής τείνει να τα διαχωρίσει σε δύο ξεχωριστές ροές, παρά το γεγονός ότι οι ήχοι αυτοί εναλλάσσονται στο ίδιο χρονικό πλαίσιο. Ένα επιπλέον χαρακτηριστικό μουσικό παράδειγμα, εντοπίζεται στα φαινόμενα αντίστιξης στην πολυφωνική μουσική όπως για παράδειγμα έργα του J.S. Bach. Εκεί ο εγκέφαλος καλείται να αναλύσει ταυτόχρονα πολλαπλές μελωδικές γραμμές, οργανώνοντας αυτές σε ξεχωριστές αντιληπτικές μονάδες. (Cambouropoulos, 2008; Huron, 2001; Tsougras & Cambouropoulos, 2025) Παρομοίως, στη σύγχρονη ηλεκτρονική μουσική, η χρήση διακριτών ρυθμικών μοτίβων (π.χ. διαφορετικά loops) οδηγεί στη δημιουργία ακουστικών ροών που συνυπάρχουν, άλλοτε συγχρονισμένα και άλλοτε με πολυρυθμικές αποκλίσεις. Η μελέτη αυτών των φαινομένων έχει άμεσες εφαρμογές όχι μόνο στη μουσικολογία, αλλά και στην τεχνητή νοημοσύνη, στην αναγνώριση φωνής και στην ανάπτυξη βοηθητικών ακουστικών συστημάτων, όπως τα κοχλιακά εμφυτεύματα ή τα προηγμένα ακουστικά βαρηκοΐας, τα οποία επιχειρούν να αναπαράγουν την ανθρώπινη ικανότητα ανάλυσης ακουστικών ροών. (Ding et al., 2024)

Μια πιο γενική αλλά ενδεικτική περίπτωση που μας βοηθά να κατανοήσουμε την ASA είναι όταν συνομιλούν τρία άτομα και όλοι αντιλαμβάνονται ότι κάποιος από αυτούς μιλάει ανάμεσά τους. Αναλυτικότερα: 1) Έχουμε το οπτικό ερέθισμα το οποίο περιλαμβάνει το κούνημα των χειλιών, τις εκφράσεις του προσώπου και των χειριών. 2) Το ηχητικό ερέθισμα δηλαδή την λήψη του ηχητικού κύματος της φωνής από τον καθένα ξεχωριστά. 3) Όλοι τους μπορούν να αντιληφθούν ποιος είναι ο άνθρωπος που ομιλεί κάθε φορά ακόμα και με κλειστά μάτια, λόγω του χαρακτηριστικού ηχοχρώματος του καθενός που έχει αποκωδικοποιηθεί προηγούμενος από τον εγκέφαλο. 4) Ταυτόχρονα με όλα τα παραπάνω ο ίδιος ο εγκέφαλος πάλι μπορεί να ομαδοποιεί τις λέξεις σε λεκτικά σύνολα και να βγάζει νόημα μέσω αυτών των ηχητικών κυμάτων που καταλήγουν να είναι οι λέξεις. Αυτή η διαδικασία, καθώς και το θεωρητικό υπόβαθρο που κρύβεται πίσω από αυτή, απασχόλησε ιδιαίτερα τον Καναδό ψυχολόγο Albert Bregman. (Bregman, 2001)

2.1.2 Albert Stanley Bregman

A: I was deciding between psychology and philosophy which both interested me, so I flipped a coin. It came up heads, meaning "philosophy."

K: So, you went to grad school in philosophy? How did you end up in the psych department then?

A: Oh no, I couldn't let a coin decide my life, so I went to study psychology.

Albert Stanley Bregman

A: Προσπαθούσα να διαλέξω ανάμεσα στην ψυχολογία και τη φιλοσοφία, που με ενδιέφεραν και τα δύο, οπότε έριξα ένα νόμισμα. Βγήκε η λέξη «φιλοσοφία».

K: Άρα, πήγες να φοιτήσεις στην φιλοσοφία; Πώς κατέληξες στο τμήμα ψυχολογίας τότε;

A: Ωχ όχι, δεν μπορούσα να αφήσω ένα νόμισμα να αποφασίσει τη ζωή μου, οπότε πήγα και σπούδασα ψυχολογία.

Albert Stanley Bregman

Σύμφωνα με την βιογραφία στον ιστότοπο του τμήματος ψυχολογίας του πανεπιστημίου McGill (McGill, 2025), ο Albert Stanley Bregman (Αλμπερτ Στάνλεϊ Μπρέγκμαν), γεννήθηκε στις 15 Σεπτεμβρίου 1936 και απεβίωσε στις 18 Μαΐου 2023. Ήταν ένας από τους σημαντικότερους και διακεκριμένους Καναδούς ακαδημαϊκούς και ήταν μέλος της FRSC (Fellowship of the Royal Society of Canada) Εταιρίας Βασιλικής Κοινωνίας του Καναδά. Η FRSC μπορεί να απονεμίσει τιμητική διάκριση σε άτομα τα οποία η ίδια κρίνει ότι έχουν προσφέρει ένα αξιοθαύμαστο έργο στις τέχνες, στις ανθρωπιστικές επιστήμες, στις φυσικές επιστήμες και στη δημόσια ζωή του Καναδά.

Η έρευνά του επικεντρώθηκε στο πώς οι ακροατές οργανώνουν νοητικά τα πολύπλοκα ακουστικά περιβάλλοντα σε διακριτές πηγές ήχου και άντλησε έμπνευση από την ψυχολογία της Gestalt (Γκεστάλτ), την επιστήμη των υπολογιστών και την τεχνητή νοημοσύνη. Το διαχρονικό του βιβλίο, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bregman, 2001) αποτελεί μέχρι και σήμερα ορόσημο για τον κλάδο αυτό. Οι πειραματικές μελέτες που ο ίδιος διεξήγαγε, μας αποκάλυψαν χρήσιμες γνώσεις για το πώς τελικά ακούει ο ανθρώπινος νους, πώς μπορεί να αντιληφθεί τη μελωδία και τον ρυθμό καθώς και πώς συμπληρώνει ο εγκέφαλος έναν ήχο ακόμη και όταν αυτός διακόπτεται από άλλους δυνατότερους ήχους του περιβάλλοντος. Σε αυτό το σημείο θα περιγράψουμε τι είναι η ASA.

2.1.3 Η Ανάλυση Ακουστικής Σκηής (ASA) κατά τον Bregman

Σύμφωνα με τον Bregman, η ASA (στον φυσικό κόσμο) αποτελεί μια διαδικασία διαχωρισμού και οργάνωσης των μεμονωμένων ήχων του περιβάλλοντος. (Bregman, 2001) Συνήθως αυτοί οι ήχοι αλληλεπικαλύπτονται και αλληλεπικαλύπτουν με τη σειρά τους άλλους, αναλόγως με τη συχνότητα, το χρόνο και την φάση. (Kunchur, 2023) Η ASA αποτελεί μια ιδιαίτερα δύσκολη και σύνθετη διαδικασία, καθώς στο ανθρώπινο αντί φτάνει, μέσω της μεταβολής της πίεσης, ένα ενιαίο ηχητικό κύμα που προκύπτει από το άθροισμα όλων των μεμονωμένων πηγών ήχου του περιβάλλοντος. Αυτές τις ξεχωριστές μονάδες ήχου τις αποκαλούμε πηγές (sources). (Kunchur, 2023) Αυτές οι πηγές μπορούν να ομαδοποιηθούν πολλές φορές σε ξεχωριστές ροές (streams). Σε αυτό το σημείο να αποσαφηνίσουμε τη διαφορά μεταξύ stream και stem καθώς και την σύνδεσή τους με τους όρους source και track.

Οι όροι stem και stream χρησιμοποιούνται πολλές φορές εναλλακτικά ο ένας με τον άλλο. Στην πραγματικότητα όμως, αναφέρονται σε τελείως διαφορετικές έννοιες που αφορούν την ερμηνεία και επεξεργασία του ίδιου του μουσικού ήχου. Η διάκριση μεταξύ τους γίνεται πιο κατανοητή αν συγκρίνουμε τον τρόπο με τον οποίο αντιλαμβανόμαστε τον ήχο στον φυσικό κόσμο σε αντιπαράθεση με τις διεργασίες οι οποίες εφαρμόζονται στα πλαίσια ανάκτησης της μουσικής πληροφορίας (MIR) και της ηχοληψίας.

Από την μία πλευρά, τα sources αντιστοιχούν σε επιμέρους ακουστικές πηγές στον φυσικό κόσμο. Δηλαδή, σε ήχους οι οποίοι προέρχονται για παράδειγμα από αντικείμενα, ανθρώπους, μουσικά όργανα (π.χ. μια κιθάρα, μια φωνή, κλπ.). (Bregman, 2001) Το stream, είναι το αποτέλεσμα της ακουστικής επεξεργασίας από το ανθρώπινο αντί και τον εγκέφαλο, δηλαδή η αντιληπτική ομαδοποίηση των ήχων, οι οποίοι προέρχονται από μία ή περισσότερες πηγές (sources). (Bregman, 2001) Ένα source μπορεί να είναι για παράδειγμα μια αντρική ή μια γυναικεία φωνή, ένα πιάνο ή ένα σαξόφωνο όμως η κατηγοριοποίησή τους σε streams είναι πιο πολύπλοκη. Για παράδειγμα, μια πηγή μπορεί να είναι μια αντρική ή μια γυναικεία φωνή, ένα πιάνο ή ένα σαξόφωνο. Ωστόσο, η ομαδοποίησή τους σε streams είναι πιο σύνθετη. Έστω ότι ο ακροατής ακούει ένα μουσικό έργο με δύο γυναικείες φωνές, μία αντρική φωνή, έγχορδα και πιάνο. Αρχικά, θα σχηματίσει τρία streams: φωνή, έγχορδα και πιάνο. Στη συνέχεια, μπορεί να ξεχωρίσει τις φωνές σε αντρικές και γυναικείες: έτσι τα streams γίνονται τέσσερα. Αν αντιληφθεί ότι

οι γυναικείες φωνές είναι δύο διαφορετικές, τότε τα streams γίνονται πέντε, και ούτω καθεξής. Στο σημείο που θα διαχωρίσει πλήρως όλες τις οντότητες, καταλήγει να αναγνωρίζει τις αρχικές πηγές.

Ιδιαίτερο ενδιαφέρον παρουσιάζει το φαινόμενο της «ψευδοπολυφωνίας» ή «λανθάνουσας πολυφωνίας», όπου ο εγκέφαλος έχει την αίσθηση πολλαπλών streams ενώ στην πραγματικότητα υπάρχει μία και μόνο πηγή. (Cambouropoulos, 2008) Αν για παράδειγμα ακούμε ένα έργο για σόλο βιολί, το ακουστικό μας σύστημα διαχωρίζει και ομαδοποιεί μοτίβα και τονικά ύψη ώστε να οργανώσει και να αντιληφθεί τον ήχο. (Cambouropoulos, 2008) Ένα χαρακτηριστικό παράδειγμα δίνει ο Αιμίλιος Καμπουρόπουλος στο άρθρο του «Διαχωρισμός φωνών: μουσικοαναλυτικές, γνωσιακές και υπολογιστικές προσεγγίσεις» με τη Chaconne του J.S. Bach (BWV 1004) σε ρε ελάσσονα. Ενώ το βιολί παίζει μια ενιαία γραμμή (a), ο ακροατής αντιλαμβάνεται ταυτόχρονα και ένα δεύτερο επίπεδο (b) λόγω της διαδικασίας της ομαδοποίησης, όπως την έχει περιγράψει ο Bregman (βλ. υποενότητα 2.1.3.1). Έτσι, παρότι υπάρχει μόνο μία πηγή, το αυτί αντιλαμβάνεται δύο διαφορετικές ροές.

a



b



Σχήμα 1: Bach's Chaconne for solo violin (mm. 33-36) from the D Minor Partita (BWV 1004) (Cambouropoulos, 2008)

Αντίθετα, στον υπολογιστικό κόσμο, τα stems αποτελούν σύνολα ηχητικών - οργανικών ομάδων, όπως αυτά καταγράφονται και χρησιμοποιούνται είτε μέσα στα πλαίσια ανάκτησης της μουσικής πληροφορίας (MIR, δεξ ενότητα 2.2) είτε της ηχοληψίας. Ένα stem περιλαμβάνει περισσότερες από μία πηγές που ανήκουν στην ίδια κατηγορία. Για παράδειγμα, ένα τζαζ έργο μπορεί να περιλαμβάνει αντρική και γυναικεία φωνή, ακουστικό μπάσο, μπάσο drum, κύμβαλα, πιάνο και σαξόφωνο. Εάν τα σκεφτούμε διαχωρισμένα σε sources (ή σε κανάλια, tracks με όρους ηχογράφησης) προφανώς έχουμε επτά (αντρική φωνή, γυναικεία φωνή, ακουστικό μπάσο, snare drum, κύμβαλα, πιάνο, σαξόφωνο), ενώ εάν τα ομαδοποιήσουμε σε stems προκύπτουν πέντε ομάδες (φωνητικά, κρουστά, μπάσο, πιάνο και λοιπά όργανα). Κάθε stem λοιπόν αποτελεί ένα σύνθετο ηχητικό σήμα, προερχόμενο από τον συνδυασμό επιμέρους ηχητικών σημάτων που έχουν ομαδοποιηθεί προγενέστερα ανάλογα με το είδος του οργάνου ή τη λειτουργία του. Με άλλα λόγια, τα streams περιγράφουν τον τρόπο που ο ήχος υπάρχει και γίνεται

αντιληπτός από τον άνθρωπο, ενώ τα stems είναι η τεχνική και υπολογιστική του αναπαράσταση. (Riou et al., 2025; Huang et al., 2021)

Με βάση όσα αναφέρθηκαν σχετικά με τις έννοιες source, stream και stem, καθίσταται σαφές ότι ακρόαση δεν αποτελεί κατά βάση μια παθητική λήψη ηχητικών πληροφοριών από το περιβάλλον. Αντίθετα, είναι μια σύνθετη ενεργή διαδικασία επεξεργασίας και ταυτόχρονης ερμηνείας η οποία ξεκινά από τη διάσπαση των ηχητικών πηγών και προχωρά στην ομαδοποίησή τους σε πιο σύνθετες ακουστικές δομές. (Bregman, 2001) Για να κατανοηθεί σε βάθος πώς το ακουστικό σύστημα επιτυγχάνει το παραπάνω, είναι απαραίτητο να εξεταστεί ο τρόπος με τον οποίο οι ήχοι ομαδοποιούνται. Αυτή η διαδικασία, η οποία αποτελεί το θεμέλιο της ASA, περιγράφεται από τον Bregman μέσα από συγκεκριμένους μηχανισμούς.

2.1.3.1 Διαδικασία ομαδοποίησης

Η φυσική διαδικασία ομαδοποίησης έχει δύο μορφές. Είτε είναι αυτόματη (primitive grouping), είτε βασίζεται σε προηγούμενη μάθηση και γνώση του ακροατή (schema-based grouping). Το ακουστικό σύστημα λειτουργεί δηλαδή με δύο τρόπους: από κάτω προς τα πάνω (bottom-up), όταν βασίζεται σε χαρακτηριστικά του ίδιου του ήχου, και από πάνω προς τα κάτω (top-down), όταν εμπλέκονται η εμπειρία, η προσοχή ή η πρόβλεψη του εκάστοτε ανθρώπου. (Bregman, 2001) Ένα από τα πρώτα πειράματα του Μπρέγκμαν περιελάμβανε ακολουθίες εναλλασσόμενων ήχων διαφορετικής συχνότητας (π.χ. υψηλά και χαμηλά). Όταν αυτοί οι ήχοι εναλλάσσονταν με γρήγορο ρυθμό, οι συμμετέχοντες του πειράματος αντιλαμβάνονταν δύο ξεχωριστές ακουστικές ροές, μία για τον υψηλό και μία για τον χαμηλό ήχο. Αντίθετα, όταν η εναλλαγή ήταν αργή, οι ίδιοι ήχοι γίνονταν αντιληπτοί ως μέρος μιας ενιαίας ακολουθίας. Αυτό το φαινόμενο, γνωστό ως ακουστικός διαχωρισμός των ροών (auditory stream segregation) αποτέλεσε τη βάση για τη μελέτη του πώς ο εγκέφαλος διαχωρίζει και ενοποιεί ηχητικές πληροφορίες. (Sussman, 2005)

2.1.3.2 Cocktail Party Effect

Η θεωρία της ASA έχει εφαρμογή σε πολλά πεδία, με χαρακτηριστικό παράδειγμα την αναγνώριση της φωνής μέσα σε θορυβώδη περιβάλλοντα, γνωστό και ως «cocktail party effect». Το πρόβλημα του cocktail party, όπως το όρισε ο Cherry το 1953, αναφέρεται στην εξαιρετική ικανότητα του ανθρώπινου ακουστικού συστήματος να αναγνωρίζει, να απομονώνει και να κατανοεί τον λόγο ενός ομιλητή, ακόμα και αν υπάρχουν κι άλλοι, οι οποίοι μιλούν ταυτόχρονα στον ίδιο χώρο. (Al-Hawawreh et al., 2024; Gutschalk & Dykstra, 2014) Παρά τη σημαντική πρόοδο της έρευνας, η διαδικασία αυτή εξακολουθεί μέχρι σήμερα να αποτελεί πρόκληση για τη μοντελοποίησή της, καθώς οι μηχανές υπολείπονται σε μεγάλο βαθμό σε σχέση με τις ανθρώπινες ικανότητες. Η μελέτη γύρω από το φαινόμενο αυτό μπορεί να συμβάλει ουσιαστικά στην καλύτερη κατανόηση των μηχανισμών επεξεργασίας

ακουστικών ερεθισμάτων από τον εγκέφαλο και να αποτελέσει μία βάση για την ανάπτυξη βελτιωμένων και αποδοτικότερων αλγορίθμων διαχωρισμού (Allan, 2011; Huron, 2001)

2.2.4 Ανάκτηση της Μουσικής Πληροφορίας (MIR)

Η σχέση μεταξύ MIR (Music Information Retrieval) και ASA βασίζεται στον κοινό τους στόχο, την εξαγωγή νοήματος από μία πολυσύνθετη ακουστική είσοδο. Ενώ η ASA επιχειρεί να εξηγήσει πώς οι άνθρωποι αντιλαμβάνονται της ηχητικές ροές, η MIR προσπαθεί να μιμηθεί αυτή την ικανότητα μέσω υπολογιστικών μοντέλων. (Cano et al., 2019) Η δεύτερη περιλαμβάνει ένα ευρύ φάσμα υπολογιστικών διεργασιών, οι οποίες κατηγοριοποιούνται σε χαμηλού, μεσαίου και υψηλού επιπέδου με βάση τα χαρακτηριστικά τους. Η χαμηλού επιπέδου ανάλυση περιλαμβάνει τεχνικές επεξεργασίας σήματος για εξαγωγή του τονικού ύψους, του ρυθμού και της χροιάς. Η μεσαίου επιπέδου ανάλυση αφορά στην αναγνώριση μουσικών δομών όπως ρυθμικά μοτίβα, αρμονικές ακολουθίες και μοτίβα. Τέλος η υψηλού επιπέδου ανάλυση σχετίζεται με το σημασιολογικό περιεχόμενο, όπως η αναγνώριση συναισθήματος, η ταξινόμηση σε είδος και η ανάκτηση με βάση τις ομοιότητες και διαφορές. (Casey et al., 2008; Kaminskas & Ricci, 2012) Μία απαραίτητη διαδικασία, η οποία συμπεριλαμβάνεται στην MIR είναι και η απομόνωση των stem, αντίστοιχα με εκείνη της ανθρώπινης ακουστικής αντίληψης. Γενικά, τα μοντέλα της MIR απαιτούν ακριβή εντοπισμό μουσικών γεγονότων (π.χ., ενάρξεις νοτών, χτύπους, αλλαγές συγχορδιών).

Παρά την τεχνολογική πρόοδο, τα συστήματα MIR συχνά υστερούν έναντι των ανθρώπινων ακουστικών δυνατοτήτων, ειδικά σε μουσικά περιβάλλοντα πολυφωνίας. Η καλύτερη ενσωμάτωση των διαδικασιών της ASA στην κατασκευή των μοντέλων της MIR προσφέρει μια μεγαλύτερη ακρίβεια στα συστήματα. Για παράδειγμα, η ενσωμάτωση της προσοχής των ακροατών και η δυναμική των έργων μπορεί να μοντελοποιήσει καλύτερα τον τρόπο με τον οποίο, οι ίδιοι εστιάζουν στα «σημαντικά» στοιχεία της μουσικής. (Cano et al., 2019)

2.2 Μοντέλα διαχωρισμού των μουσικών stem (Music stem separation models)

2.2.1 Εισαγωγή στα μοντέλα διαχωρισμού

Τα μοντέλα διαχωρισμού stem αποτελούν βασικό κομμάτι του πεδίου της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval - MIR), μέσω διάφορων τεχνικών βαθιάς εκμάθησης και επεξεργασίας σήματος, επιτρέπουν την απομόνωση μεμονωμένων μουσικών στοιχείων, όπως φωνή, κρουστά, πιάνο κτλ., από σύνθετες ηχογραφήσεις. Η δυσκολία προκύπτει από το γεγονός ότι οι πηγές αλληλεπικαλύπτονται στον χρόνο και τη συχνότητα, κάνοντας τον διαχωρισμό δύσκολο με παραδοσιακά φίλτρα. (Bittner et al., 2014; H. Li et al., 2022; Riou et al., 2025) Τα σύγχρονα μοντέλα διαχωρισμού

stems χρησιμοποιούν κυρίως συνελκτικά ή αναδρομικά νευρωνικά δίκτυα τα οποία εκπαιδεύονται σε μεγάλα σύνολα δεδομένων μεμονωμένων ηχογραφήσεων από όργανα.

2.2.2 Συνελκτικά Νευρωνικά Δίκτυα (CNN)

Τα Συνελκτικά Νευρωνικά (Convolutional Neural Network (CNN)) Δίκτυα αποτελούν μια κατηγορία μοντέλων βαθιάς εκμάθησης σχεδιασμένων για την επεξεργασία δεδομένων με δομή πλέγματος, όπως οι εικόνες. Λόγω της ικανότητάς τους να αποτυπώνουν ιεραρχικά τις χωρικές σχέσεις, τα CNN έχουν εφαρμογή σε πλήθος εργασιών, μεταξύ των οποίων η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων και η ανάλυση ιατρικών εικόνων. (H. Li et al., 2022; T. L. Li et al., 2010; Saleem et al., 2022; Zhang & Li, 2025).

2.2.3 Αρχιτεκτονική U-Net

Μία εξειδικευμένη μορφή CNN είναι και η αρχιτεκτονική U-Net, στην οποία βασίζεται και το Spleeter, αποτελεί μια εξειδικευμένη μορφή CNN που σχεδιάστηκε για ανάλυση κατά τμήματα σε εικόνες (image segmentation). Διακρίνεται για τη συμμετρική δομή encoder-decoder με skip connections, οι οποίες διατηρούν στη μνήμη τους σημαντικές πληροφορίες σε συνάρτηση του χώρου και επιτρέπουν ακριβείς προβλέψεις σε επίπεδο pixel. Κάποια από τα πλεονεκτήματα της αρχιτεκτονικής U-Net είναι πρώτα η υψηλή ακρίβεια στην κατά τμήματα ανάλυση. Ο συνδυασμός της δομής encoder-decoder με skip connections επιτρέπουν ακριβέστερο διαχωρισμό των αντικειμένων. Δεύτερον, είναι η αποτελεσματικότητα της με περιορισμένα δεδομένα, καθώς μπορεί να εκπαιδευτεί επιτυχώς με σχετικά λίγα δείγματα, καθιστώντας την ιδιαίτερα χρήσιμη σε ιατρικές εφαρμογές. Τέλος, έχει προσαρμοστεί για διδιάστατες (2D) και τρισδιάστατες (3D) εικόνες, καθώς και για διαφορετικές εργασίες segmentation. (Ronneberger et al., 2015)

2.2.4 Περιγραφή της λειτουργίας των μοντέλων διαχωρισμού

Μέχρι και σήμερα, τα περισσότερα μοντέλα διαχωρισμού βασίζονται στη χρήση συνελκτικών νευρωνικών δικτύων τύπου U-Net, εκμεταλλεζόμενα μοτίβα όπως αρμονικές δομές, ηχώχρωμα και ρυθμικές ενδείξεις για να διαχωρίσουν τα stems. Μέσω της τεχνικής «masking» σε κάθε stem χωριστά, το μοντέλο απομονώνει και διαχωρίζει την κυματομορφή και την μετατρέπει το σήμα σε φασματογράφημα. (Saleem et al., 2022; Hummerson et al., 2014) Αυτό συμβαίνει σε συνάρτηση του πεδίου χρόνου (T) και της συχνότητας (F) όπου τονίζονται οι περιοχές του φασματογραφήματος που ανήκουν σε συγκεκριμένο όργανο, ενώ καταστέλλουν τις υπόλοιπες. (Cano et al., 2019; Défossez et al., 2021)

Μια συνηθισμένη προσέγγιση είναι η εφαρμογή του Μικρού Χρόνου Μετασχηματισμού Fourier (STFT). Ωστόσο, η χρήση του σε φασματογραφήματα εγείρει σημαντικά ζητήματα, κυρίως λόγω της απότομης αύξησης (εκθετική άνοδος) του άξονα συχνοτήτων, η οποία δυσχεραίνει την όλη διαδικασία. (Engel et al., 2020) Αν και μετασχηματισμοί όπως οι constant-Q ή Mel-scale μπορούν να μειώσουν το πρόβλημα στις εργασίες ταξινόμησης, δεν προσφέρουν ακριβή αναπροσαρμογή στο πεδίο χρόνου, καθιστώντας τους ανεπαρκείς για διαχωρισμό ήχου υψηλής λεπτομέρειας. (Ewert & Muller, 2012; Holighaus et al., 2013; Schörkhuber & Klapuri, 2010)

Ένα ακόμα σοβαρό πρόβλημα των διάφορων μεθόδων «masking» στο πεδίου χρόνου και συχνοτήτων, είναι η απουσία πληροφοριών φάσης. Συνήθως, υποβαθμίζεται ο διαχωρισμός σε πολυκάναλες ηχογραφήσεις, καθώς προκαλείται θόρυβος στις διαχωρισμένες πηγές, ιδίως στις χαμηλές συχνότητες όπως το μπάσο. Σε ψηλότερες συχνότητες, όπως τα φωνητικά, επηρεάζονται λιγότερο, ενώ τα τύμπανα παρουσιάζουν προβλήματα στις «ατάκες» τους λόγω της ανεπαρκούς ευθυγράμμισης της φάσης. Αυτό έχει ως αποτέλεσμα μια «θολότητα» στον ήχο σε ορισμένα όργανα όταν χρησιμοποιείται αποκλειστικά η συγκεκριμένη μέθοδος. (Cano et al., 2019)

2.2.5 Βασικός και Μικρού Χρόνου Μετασχηματισμός Fourier (Short-Time Fourier Transform, STFT)

Ο Μετασχηματισμός Fourier (FT) αποτελεί μια από τις θεμελιώδεις τεχνικές στην ανάλυση ήχου και βρίσκεται στο επίκεντρο της διαδικασίας διαχωρισμού της μουσικής στα επιμέρους συστατικά της. Η τεχνική αυτή επιτρέπει τη διάσπαση μιας ολόκληρης ηχογράφησης, όπως ένα απλό τραγούδι ή ένα σύνθετο ορχηστρικό έργο σε ξεχωριστά στοιχεία όπως φωνητικά, τύμπανα, μπάσο, πιάνο ή άλλο. Είναι το θεμέλιο πάνω στο οποίο βασίζονται οι περισσότερες εφαρμογές τεχνητής νοημοσύνης, οι οποίες χρησιμοποιούνται σήμερα στον διαχωρισμό. Στην ουσία, ο βασικός μετασχηματισμός Fourier είναι η μετατροπή της συνάρτησης του χρόνου του ήχου σε νέες συναρτήσεις, οι οποία περιγράφουν την κατανομή των συχνοτήτων της αρχικής συνάρτησης. (Mitianoudis & Davies, 2004)

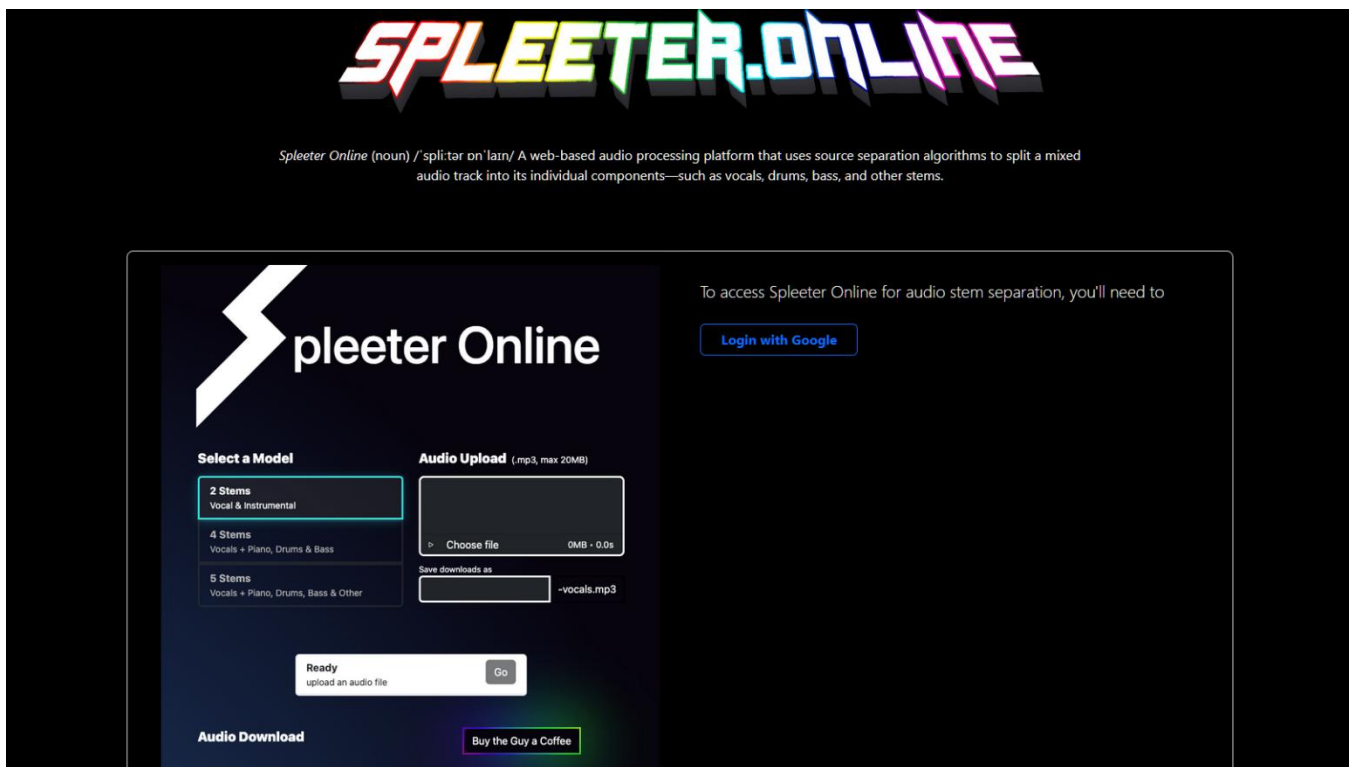
Για να κατανοήσουμε γιατί αυτή η μετατροπή είναι τόσο κρίσιμη, πρέπει να δούμε πώς λειτουργούν τα ηχητικά σήματα. Ένα σήμα ήχου είναι ουσιαστικά μια καταγραφή των μεταβολών της πίεσης του αέρα μέσα στον χρόνο. Ωστόσο, με αυτή τη μορφή είναι δύσκολο να αναγνωρίσουμε υπολογιστικά τα stems που ηχούν κάθε στιγμή. Ο μετασχηματισμός Fourier μας επιτρέπει να αναπαραστήσουμε το σήμα με βάση το φασματικό του περιεχόμενο. (Engel et al., 2020) Στο φασματογράφημα, είναι πολύ πιο εύκολο να διακρίνουμε, για παράδειγμα, μια μπάσα κιθάρα που παράγει σταθερούς, χαμηλούς τόνους από ένα κύμβαλα τα οποία παράγουν αιχμηρά και υψηλής συχνότητας ηχητικά σήματα. Αναλύοντας τα χαρακτηριστικά φασματικά μοτίβα, μπορούμε να εντοπίσουμε ποια μέρη του ήχου ανήκουν σε ποια πηγή. (Deville et al., 2018)

Τα περισσότερα συστήματα διαχωρισμού στηρίζονται όχι στον βασικό μετασχηματισμό Fourier, αλλά στην παραλλαγή του που ονομάζεται Μικρού Χρόνου μετασχηματισμός Fourier (STFT), ο οποίος επιτρέπει την ανάλυση του πώς εξελίσσονται οι συχνότητες στον χρόνο. Το αποτέλεσμα είναι ένα φασματογράφημα, μια δισδιάστατη δηλαδή αναπαράσταση του ήχου με τον χρόνο στον έναν άξονα, τη

συχνότητα στον άλλο και την ένταση αποτυπωμένη με τον συνδυασμό της φωτεινότητας και των διάφορων χρωμάτων. Αυτή η οπτική αναπαράσταση είναι συχνά το σημείο έναρξης για τα μοντέλα διαχωρισμού των stem, καθώς εκπαιδεύονται να αναγνωρίζουν τα τυπικά και χαρακτηριστικά φασματικά στοιχεία των διαφορετικών μουσικών οργάνων.

2.3 Περιγραφή και λειτουργία του Spleeter

2.3.1 Περιγραφή του Spleeter



Το Spleeter είναι ένα εργαλείο διαχωρισμού πηγών βασισμένο σε Python, το οποίο δημιουργήθηκε από την Deezer. Χρησιμοποιεί συνελκτικά νευρωνικά δίκτυα (CNN) τύπου U-Net και η λειτουργία του είναι βασισμένη στην απομόνωση και τον διαχωρισμό των φασματογραφημάτων από κάθε stem μέσω μιας διαδικασίας που ονομάζεται *masking*. (Hennequin et al., 2020) Τεράστιο μέρος της εκπαίδευσής του έχει γίνει στο TensorFlow (Abadi et al., 2016), ένα περιβάλλον όπου μπορείς να εκτελέσεις αλγόριθμους με σκοπό να εκπαιδεύσεις μία μηχανή. Είναι ένα προηγμένο εργαλείο που επιτρέπει στον χρήστη τον διαχωρισμό των επιμέρους ηχητικών στοιχείων από μουσικά κομμάτια με ταχύτητα και ακρίβεια. Άλλωστε, η ταχύτητα και η ευκολία στην χρήση ήταν οι κύριοι γνώμονες στους οποίους βασίστηκε και

η όλη ανάπτυξή του. (Hennequin et al., 2019) Προσφέρει τρία εκπαιδευμένα μοντέλα: 2 stem (φωνή και συνοδεία), 4 stem (φωνή, ντραμς, μπάσο, άλλα), 5 stem (φωνητικά, ντραμς, μπάσο, πιάνο, άλλα). Αυτή η ευελιξία που χαρακτηρίζει το συγκεκριμένο εργαλείο, το καθιστά κατάλληλο για πληθώρα εφαρμογών όπως ανάλυση μουσικού περιεχομένου, δημιουργία remix ή ακόμη και χρήση για εκπαιδευτικούς και ερευνητικούς σκοπούς. Όσον αφορά τους λόγους για τους οποίους το επιλέξαμε τους αναλύουμε στην υποενότητα 3.4.1. Σε αυτό το σημείο θα αναλύσουμε ολόκληρη τη τεχνική διαδικασία εισαγωγής των μουσικών αρχείων, την τελική εξαγωγή των stems και τον τρόπο με τον οποίο αξιοποιείται το STFT.

2.3.2 Διαδικασία διαχωρισμού των stems και STFT

Το εργαλείο δέχεται στην ουσία ως είσοδο φασματογράφημα μεικτών ηχητικών σημάτων και εξάγει περισσότερα φασματογραφήματα για τα επιμέρους stems, όπως τα φωνητικά, το πιάνο κτλ. Παρόλο που το νευρωνικό δίκτυο είναι υπεύθυνο για την αναγνώριση και απομόνωση των επιμέρους stem, ο μετασχηματισμός Fourier είναι η βάση στην οποία οποίο λειτουργεί όλο το δίκτυο. (Hennequin et al., 2020) Αναλυτικότερα, η διαδικασία περιλαμβάνει πρώτα τη μετατροπή του αρχείου ήχου σε φασματογράφημα μέσω STFT. Το φασματογράφημα αυτό εισέρχεται στο εκπαιδευμένο μοντέλο, το οποίο εξάγει ένα ή περισσότερα νέα φασματογραφήματα ανάλογα με την επιλογή ανάμεσα στα τρία μοντέλα διαχωρισμού. Τα φασματογραφήματα αυτά μετατρέπονται στη συνέχεια πίσω σε ηχητικά αρχεία μέσω του αντίστροφου μετασχηματισμού Fourier, δημιουργώντας ξεχωριστά αρχεία ήχου για κάθε stem. Για παράδειγμα, στο μοντέλο με τα 2 stem (φωνητικά και συνοδεία), το σύστημα μαθαίνει να καταστέλλει τις συχνότητες όπου τα φωνητικά απουσιάζουν και να ενισχύει εκείνες όπου είναι παρόντα. Το δίκτυο δεν «αναγνωρίζει» πως ακούγεται ένα φωνητικό μέρος όπως ένας άνθρωπος, αλλά μέσω της εκπαίδευσής του σε χιλιάδες παραδείγματα, μαθαίνει να ξεχωρίζει τα επιμέρους συστατικά τους, όπως για παράδειγμα το εύρος συχνοτήτων στη μεσαίες περιοχές. Ακόμα κι αν θαυμάζουμε τη «δουλεία» του μοντέλου διαχωρισμού, στην πραγματικότητα δεν θα λειτουργούσε τίποτα χωρίς τον μετασχηματισμό Fourier. Χωρίς αυτόν, δεν θα μπορούσαμε να δημιουργήσουμε τα φασματογραφήματα ούτε να επανασυνθέσουμε τα ηχητικά κύματα. (Hennequin et al., 2019)

Η χρήση του μετασχηματισμού Fourier στο παραπάνω πλαίσιο δεν είναι χωρίς προκλήσεις. Ένα βασικό πλεονέκτημα της φασματικής αναπαράστασης είναι ότι ταιριάζει περισσότερο με τον τρόπο κατά τον οποίο αντιλαμβάνεται τον ήχο το ανθρώπινο αυτί. (Balaban et al., 1992) Η ακοή μας είναι ιδιαίτερα ευαίσθητη σε συχνότητες και στις αλλαγές οι οποίες επέρχονται σε αυτές μέσα στον χρόνο. Για τον λόγο αυτό, ο διαχωρισμός stem φασματικά δίνει καλύτερα αποτελέσματα από ότι χρονικά, όπου τα σήματα των διαφόρων οργάνων αλληλεπικαλύπτονται. Παρόλα αυτά, ένα από τα βασικότερα προβλήματα είναι η ισορροπία μεταξύ του χρόνου και των συχνοτήτων. Δεν είναι δυνατό να έχουμε ταυτόχρονα υψηλή ακρίβεια στον χρόνο και στη συχνότητα. Αυτή είναι μια περιοριστική συνθήκη και αντιμετωπίζεται συνήθως μέσω συνδυασμού διαφορετικών παραμέτρων. (Hennequin et al., 2020; Huang et al., 2021; Klauri & Davy, 2006)

Ένα άλλο πρόβλημα είναι η ανακατασκευή της φάσης. Το φασματογράφημα περιέχει κυρίως το μέτρο της συχνότητας, αλλά για την ανακατασκευή του ήχου απαιτείται και η πληροφορία φάσης.

(Klapuri & Davy, 2006) Τα περισσότερα μοντέλα διαχωρισμού, όπως και το Spleeter, χρησιμοποιούν τη φάση του αρχικού σήματος κατά την ανασύνθεση, κάτι που μπορεί να προκαλέσει αλλοιώσεις. Ακόμη και σήμερα, τα αποτελέσματα είναι εντυπωσιακά και επαρκή για εμπορική και καλλιτεχνική χρήση, όμως οι ερευνητές προσπαθούν να βελτιώσουν την πρόβλεψη της φάσης.

Όπως ήδη έχουμε αναφέρει και στην εισαγωγή, η επιτυχία του Spleeter οφείλεται στην ταχύτητα και την απλότητά του. Μπορεί να λειτουργήσει με μία μόνο εντολή και να αποδώσει καθαρά κομμάτια φωνητικών και συνοδείας μέσα σε λίγα δευτερόλεπτα. Κάτω από αυτή την απλότητα όμως, κρύβεται μια πολύπλοκη κατασκευή, η οποία εξαρτάται απολύτως από τον μετασχηματισμό αυτό, για να μεταφράσει το ηχητικό σήμα, το οποίο στη συνέχεια θα διαχωριστεί. Κλείνοντας, θα συνειδητοποιήσουμε ότι, ο μετασχηματισμός Fourier παρέχει στην ουσία στις μηχανές έναν τρόπο να «ακούν» τη μουσική, ο οποίος προσεγγίζει τον ανθρώπινο τρόπο αντίληψης.

3. Μεθοδολογία

3.1 Ερευνητικός Σχεδιασμός

Η παρούσα διπλωματική εργασία εξετάζει τις δυνατότητες αλλά και τους περιορισμούς που συνεπάγονται από τη διαδικασία του stem separation. Τα παραπάνω προέκυψαν μέσω του εργαλείου τεχνητής νοημοσύνης Spleeter, το οποίο και χρησιμοποιήσαμε στα πλαίσια της εργασίας όπως έχουμε ήδη αναφέρει. Θα προσπαθήσουμε να προτείνουμε ένα ολοκληρωμένο μεθοδολογικό πλαίσιο για την όσο το δυνατότερο ακριβέστερη αξιολόγηση της απόδοσής του. Η προσέγγιση που θα ακολουθήσουμε είναι μικτή και συνδυάζει ακουστική και οπτική παρατήρηση. Αρχικά, η αξιολόγηση βασίζεται κυρίως στην ακρόαση, ενώ ταυτόχρονα πραγματοποιείται και σύγκριση των οπτικών δεδομένων που προκύπτουν από το φασματογράφημα (spectrogram) και την κυματομορφή (waveform). Επιπλέον, θα δοθεί ιδιαίτερη έμφαση στην καθαρότητα και στην πιθανή διαρροή μεταξύ των stems που μπορεί να προκύψει από το διαχωρισμό. Για παράδειγμα, κατάλοιπα από το stem της φωνής, μπορούν να εισβάλλουν στα stems των ντραμς ή του μπάσου αντίστοιχα.

Η ανάλυσή μας θα επικεντρωθεί στο Spleeter, όπου το συγκεκριμένο εργαλείο εστιάζει στο μοντέλο των πέντε (5) stems. Επιπλέον θα γίνει χρήση και του Sonic Visualiser ως βοηθητικό αναλυτικό εργαλείο για την εξαγωγή των κυματομορφών και φασματογραφήματων. Το σύνολο του μουσικού υλικού που επιλέχθηκε περιλαμβάνει δεκατέσσερα (14) μουσικά έργα ποικίλων στυλ και εποχής, για παράδειγμα από τη μπαρόκ και κλασική εποχή, τον ρομαντισμό, την τζαζ καθώς και από τον σύγχρονο κινηματογραφικό μουσικό ήχο. Με τον συνδυασμό της οπτικής παρατήρησης (Sonic Visualiser) και της ακουστικής επαλήθευσης (ανθρώπινο αυτί), θα επιχειρήσουμε να στοχεύσουμε στην παροχή συμπερασμάτων σχετικά με την απόδοση και τους περιορισμούς του stem separation που προκύπτει από το συγκεκριμένο εργαλείο τεχνητής νοημοσύνης.

3.2 Στόχοι

Η έρευνα οργανώθηκε γύρω από τους εξής στόχους:

- 1) Κριτική αξιολόγηση στην ακρίβεια του μοντέλου των πέντε (5) stem του Spleeter μέσα σε ένα ποικίλο μουσικά, εποχιακά και στυλιστικά σύνολο μουσικών έργων.
- 2) Ανίχνευση και ανάλυση τις καθαρότητας καθώς και των διαρροών στα stems που προκύπτουν από το separation.

3.3 Ερευνητικά Ερωτήματα

Σε αυτό το σημείο θα θέσουμε κάποια ερευνητικά ερωτήματα που προϋπήρχαν κατά τον αρχικό σχεδιασμό της έρευνας αλλά και άλλα που προέκυψαν κατά τη διάρκεια αυτής. Ξεκινώντας:

- 1) Πόσο αποτελεσματικά διαχωρίζει το Spleeter τα stems ιδιαίτερα σε μη-συμβατικά μουσικά παραδείγματα ;
- 2) Υπάρχουν κάποια βασικά χαρακτηριστικά στον τύπο των διαρροών που παρατηρούνται κατά τον διαχωρισμό; ;
- 3) Τί επιρροή έχουν παράγοντες όπως η ενορχήστρωση και ο αριθμός των οργάνων;
- 4) Πως η ίδια η ηχογράφηση επηρεάζει την ποιότητα του διαχωρισμού;

3.4 Περιγραφή των Εργαλείων

3.4.1 Spleeter

Το Spleeter είναι ένα δωρεάν και ανοιχτού κώδικα εργαλείο διαχωρισμού ηχητικών stem, το οποίο χρησιμοποιήθηκε στη μελέτη μας κυρίως λόγω της ταχύτητας, της ακρίβειας και της ευκολίας στη χρήση του (λειτουργεί με μία μόνο εντολή). Η αναλυτική του περιγραφή και του τρόπου λειτουργίας του περιγράφεται στο κεφάλαιο 2.3. Η δυνατότητά του να διαχωρίζει γρήγορα και αξιόπιστα τα επιμέρους στοιχεία ενός μουσικού κομματιού, όπως φωνητικά, ντραμς, μπάσο και άλλα όργανα, το καθιστά ιδανικό για εφαρμογές που απαιτούν επεξεργασία μεγάλου όγκου δεδομένων ήχου, ανάλυση μουσικού περιεχομένου, δημιουργία remix ή εκπαιδευτικές και ερευνητικές χρήσεις. Προσφέρει τρία εκπαιδευμένα μοντέλα: 2 stem (φωνή και συνοδεία), 4 stem (φωνή, ντραμς, μπάσο, άλλα), 5 stem (φωνητικά, ντραμς, μπάσο, πιάνο, άλλα).

Για τη συγκεκριμένη μελέτη χρησιμοποιήθηκε το μοντέλο των 5 stem, διότι παρείχε τον μεγαλύτερο αριθμό με stems που θα μπορούσαμε να χρησιμοποιήσουμε. Όπως αναφέρθηκε παραπάνω, ένα από τα πιο εντυπωσιακά χαρακτηριστικά, το οποίο μας ώθησε στη επιλογή του, είναι η υψηλή ταχύτητα. Είναι ικανό για παράδειγμα να διαχωρίσει σε 4 stems, ένα μουσικό έργο 100 δευτερολέπτων μέσα στο χρονικό όριο του 1 δευτερολέπτου με τη χρήση μίας και μόνο μίας κάρτας γραφικών (GPU). Με άλλα λόγια κάθε εκατό δευτερόλεπτα μουσικής ο διαχωρισμός γίνεται σε 1 δευτερόλεπτο. Έτσι λοιπόν, η ταχύτητά του το καθιστά εξαιρετικά αποτελεσματικό για την επεξεργασία μεγάλων όγκων δεδομένων ήχου. (Hennequin et al., 2020) Εδώ ίσως πρέπει να διευκρινίσουμε ότι γίνεται χρήση της κάρτας γραφικών διότι το ίδιο βασίζεται στην επεξεργασία των φασματογραφημάτων όπως ήδη ειπώθηκε παραπάνω. Ο επεξεργαστής (CPU) χρησιμοποιείται για την φόρτωση, μεταφόρτωση και εξαγωγή των αρχείων που απαιτούνται κάθε φορά. (Hennequin et al., 2019)

3.4.2 Sonic Visualiser

Το Sonic Visualiser, δημιουργήθηκε από το κέντρο ψηφιακής μουσικής (Centre for Digital Music) στο Queen Mary University of London και είναι ένα εργαλείο ανοικτής πρόσβασης. Μπορεί να συνδυάσει

λειτουργίες οπτικοποίησης και ανάλυσης του ηχητικού σήματος με στόχο να προσφέρει στους διάφορους ερευνητές (ειδικά σε εκείνους που ανήκουν στον τομέα της υπολογιστικής μουσικολογίας), εργαλεία ανάλυσης, οπτικής παρατήρησης και επεξεργασίας ηχητικού σήματος. Κύριος σκοπός του είναι η συμβολή του στην μουσική ανάλυση μέσω της εξαγωγής μουσικών δεδομένων που θα είναι εύκολα στη χρήση, για να μπορούν να χρησιμοποιηθούν για την τελική κατάληξη μουσικολογικών συμπερασμάτων από τους ερευνητές του πεδίου. (Cannam et al., 2010)

Το ίδιο το πρόγραμμα είναι βασισμένο στις κυματομορφές και τα φασματογραφήματα τα οποία συναντήσαμε και στο Spleeter. Με βάση τον χρόνο στον άξονα x του συστήματος γίνεται η απεικόνιση των κυματομορφών, φασματογραμμάτων, αναπαράσταση συχνοτήτων και άλλων μορφών δεδομένων. Αυτό επιτρέπει στους ερευνητές να παρατηρούν τις χρονικές και φασματικές διαφορές ή ομοιότητες ενός ηχητικού σήματος και πώς αυτό εξελίσσεται μέσα στο χρόνο. Μία πολύ χρήσιμη δυνατότητα της συγκεκριμένης εφαρμογής είναι ότι υποστηρίζει την «στοίβαξη» πολλαπλών μορφών απεικόνισης, επιτρέποντας την πολυεπίπεδη παρατήρηση και εξέταση των χαρακτηριστικών του μουσικού σήματος που εισάγεται σε αυτό κάθε φορά. Για αυτό το λόγο μπορεί να χρησιμοποιηθεί και για την ανίχνευση διαρροών των stems, μέσω ανάλυσης και συγκρίσεων που θα επιχειρήσουμε να πραγματοποιήσουμε σε αυτό με σκοπό να καταλήξουμε σε συμπεράσματα που αφορούν τη διαρροή και την καθαρότητα.

Συμπληρωματικά, ένα ιδιαίτερο χαρακτηριστικό του Sonic Visualiser είναι η δυνατότητα εισαγωγής επιπρόσθετων εργαλείων Vamp (Cannam et al., 2010). Αυτά τα εργαλεία είναι στην ουσία αλγόριθμοι οι οποίοι προσθέτουν επιπλέον δυνατότητες εξαγωγής διάφορων ακουστικών χαρακτηριστικών, όπως ρυθμό, τονικότητα, περιγραφή της δομής κτλ. Τα επιπλέον αυτά εργαλεία επεκτείνουν σημαντικά τις βασικές δυνατότητες ανάλυσης που προσφέρει το ίδιο το πρόγραμμα και επιτρέπουν στους χρήστες να προσαρμόζουν καθώς και να κατευθύνουν το αντικείμενο της μελέτης τους σε συγκεκριμένες πτυχές και ερευνητικά ερωτήματα. Στην δική μας έρευνα δεν θα γίνει χρήση αυτών των εργαλείων αλλά θεωρούμε ότι ήταν άξια αναφοράς.

3.5 Επιλογή Μουσικών Έργων

Για τις ανάγκες της παρούσας εργασίας επιλέχθηκαν 14 μουσικά έργα, τα οποία καλύπτουν ένα ευρύ φάσμα εποχών, οργάνων, ενορχήστρωσης, τεχνικών mixing και μουσικών ειδών (Μπαρόκ, Κλασική, Ρομαντική, Jazz, Κινηματογραφική μουσική). Η επιλογή τους βασίστηκε τόσο στην ποικιλομορφία τους όσο και στις προσωπικές προτιμήσεις του συγγραφέα της παρούσας εργασίας. Επιπλέον, στόχος είναι η δημιουργία ενός δείγματος παραδειγμάτων που θα βοηθήσει να αντιληφθούμε τις δυνατότητες και τους περιορισμούς του Spleeter μέσα σε διαφορετικές μουσικές συνθήκες από αυτές της εκπαίδευσής του.

Παρακάτω παρατίθενται ένας πίνακας με τους τίτλους των έργων και συνθετών, των βασικών συντελεστών και εκτελεστών καθώς και μία συνοπτική περιγραφή των οργάνων των μουσικών παραδειγμάτων που χρησιμοποιήθηκαν. Στην επιλογή των έργων περιλαμβάνονται: έργα για μικρά και μεγάλα σύνολα, με φωνή ή χωρίς, με φυσικά - ακουστικά ή ηλεκτρονικά όργανα, καθώς και έργα με

μεγάλο ή μικρό βαθμό επεξεργασίας στο στάδιο της μίξης. Επίσης, τα έργα παρουσιάζουν διαφορετικά χαρακτηριστικά στην ενορχήστρωση τους π.χ. συμφωνική ορχήστρα, κουαρτέτο εγχόρδων, jazz κουιντέτο κτλ., καθώς και προέρχονται από διαφορετικές τεχνικές ηχογράφησης όπως σε στούντιο με μικρόφωνα, εξολοκλήρου σε DAW περιβάλλοντα ή ζωντανές συναυλίες.

Τίτλος Έργου	Βασικοί Έκτελεστές	Βασική Περιγραφή των Οργάνων
1) <u>Borislav Slavov: Baldur's Gate 3, Down By The River</u>	Developed and Published by Larian Studios	Μικρή κινηματογραφική ορχήστρα, άρπα, τσελέστα, συνθετικό μπάσο (synth bass) και γυναικεία φωνή.
2) <u>Camille Saint-Saëns: Carnival of the Animals, VII. Aquarium</u>	St. Mark's Philharmonic Orchestra	Συμφωνική ορχήστρα και 2 πιάνο.
3) <u>Claude Bolling: Suite for Flute and Jazz Piano Trio</u>	Flute: Jean-Pierre Rampal Piano: Claude Bolling Double bass: Max Hédiguer Drums: Marcel Sabiani	Φλάουτο, πιάνο, κοντραμπάσο και τζαζ κρουστά.
4) <u>Dave Brubeck: Take Five</u>	The Dave Brubeck Quartet	Σαξόφωνο, πιάνο, ακουστικό μπάσο (κοντραμπάσο) και τζαζ κρουστά
5) <u>Kristen Anderson-Lopez & Robert Lopez: Show Yourself</u>	Disney orchestra, Idina Menzel (voice of Elsa) and Evan Rachel Wood (voice of Queen Iduna)	Μικρή κινηματογραφική ορχήστρα, ντραμς, πιάνο, σόλο γυναικεία φωνή και χορωδία.
6) <u>Gareth Coker: Morgana's Theme</u>	Ο συνθέτης	Κινηματογραφική ορχήστρα και σοπράνο φωνή.
7) <u>Glenn Miller: Moonlight Serenade</u>	Carly Simon	Τζαζ μπάσο και Άλτο φωνή.
8) <u>G.F. Handel: Flute Sonata in F Major – III. Siciliana</u>	Flute: Jean-Pierre Rampal Harpsichord: Robert Veyron-Lacroix	Φλάουτο και Τσέμπαλο.
9) <u>Henry Mancini: Lujon</u>	Δεν αναφέρονται	Τζαζ ορχήστρα.
10) <u>Christmas Oratorio, BWV 248: Part VI, 41. Arie "Ich will nur dir zu Ehren leben"</u>	Akademie für Alte Musik Berlin RIAS Kammerchor René Jacobs Werner Güra	Μπαρόκ ορχήστρα και Βαρύτονο.
11) <u>Kevin Penkin: Here, Tomorrow</u>	Voice: Lilas Lyrics: Christopher Keith O'Neill, Elspeth Bawden, Fabian Malabello	Ορχήστρα, κρουστά.
12) <u>Ola Gjeilo: Across The Vast, Eternal Sky</u>	The Choir of Royal Holloway · 12 ensemble	Ορχήστρα εγχόρδων, πιάνο και μικτή χορωδία SATB.
13) <u>Ola Gjeilo: Dark & Luminous Night</u>	Choir: Tenebrae choir Instruments: Thomas Gould, Ciaran McCabe, Jon Thorne, Matthew Sharp	Ορχήστρα εγχόρδων, πιάνο και μικτή χορωδία SATB.
14) <u>S. Rachmaninoff: Symphonic Dances (1η κίνηση)</u>	State Symphony Orchestra of Russian Federation conducted by Yevgeny Svetlanov	Πλήρης συμφωνική ορχήστρα, πιάνο (ελάχιστα σημεία) και άρπα.

3.6 Πρακτική Διαδικασία

Προεργασία:

Η διαδικασία ξεκίνησε με την επιλογή των δεκατεσσάρων μουσικών παραδειγμάτων, βάσει των κριτηρίων που παρατέθηκαν αναλυτικά στην ενότητα 3.5 (Επιλογή Μουσικών Έργων). Καταβλήθηκε μία μεγάλη προσπάθεια ώστε όλα τα επιλεγμένα πρωτογενή μουσικά έργα να είναι διαθέσιμα σε μορφή WAV, με χαρακτηριστικά 44.1 kHz και 16-bit, προκειμένου να εξασφαλιστεί η μέγιστη δυνατή ποιότητα και ομοιομορφία για την ανάλυση. Στη συνέχεια έπρεπε να σιγουρευτούμε ότι το αντικειμενικό ύψος έντασης (LUFUS, Loudness Unit related to Full Scale), το οποίο πρέπει να είναι ελεγχόμενο για όσο το δυνατόν μεγαλύτερη ακρίβεια στην ανάλυση του κάθε αρχείου ξεχωριστά, βρισκόταν στα ίδια επίπεδα. (Mason, 2011) Αυτό πραγματοποιήθηκε προγενέστερα σε μία αυτόματη διαδικασία από την πλατφόρμα όπου αντλήθηκαν τα έργα. Η πλατφόρμα αυτή χρησιμοποιώντας την κλίμακα μέτρησης LUFUS διασφαλίζει ότι όλα τα κομμάτια διατηρούνται στο ίδιο επίπεδο έντασης (Duggal, 2024). Ως εκ τούτου, δεν υπήρξε ανάγκη για επιπλέον παρέμβαση εκ μέρους μας όσον αφορά αυτόν τον παράγοντα.

Διαχωρισμός:

Ο διαχωρισμός των stems πραγματοποιήθηκε μέσω του Spleeter. Αφού έγινε η εισαγωγή του κάθε αρχείου ξεχωριστά στο εργαλείο τεχνητής νοημοσύνης, μετά από κάθε separation προέκυψαν 5 στερεοφωνικά αρχεία σύμφωνα με το μοντέλο των 5 stems που έχουμε επιλέξει.

Ανάλυση:

Αρχικά, από τη διαδικασία εισαγωγής όλων των stems, τα οποία προέκυψαν από κάθε μουσικό παράδειγμα, στο λογισμικό Sonic Visualiser, έχουμε στην διάθεσή μας συνολικά 140 εικόνες, εκ των οποίων 70 αντιστοιχούν σε waveforms και άλλες 70 σε spectrograms. Παράλληλα με την ακρόαση και την οπτική παρατήρηση των παραπάνω, έγινε μία προσπάθεια αξιολόγησης της καθαρότητας και ηχητικής διαρροής κάθε stem.

Αποτελέσματα:

Για την τεκμηρίωση των αποτελεσμάτων αξιοποιείται το σύνολο της προηγηθείσας ανάλυσης, ενώ στην παρουσίαση χρησιμοποιούνται επιπλέον η Κλίμακα Καθαρότητας και Διαρροής (ΚΚΔ) καθώς και η συνάρτηση Ποιότητας του Διαχωρισμού, τα οποία κατασκευάστηκαν ειδικά για τις ανάγκες της παρούσας μελέτης. Η λειτουργία και η χρήση τους θα αναλυθούν εκτενώς παρακάτω στην υποενότητα [3.7](#). Έτσι, με τη χρήση των παραπάνω καθίσταται δυνατή η εξαγωγή ερμηνευτικών δεδομένων και η διατύπωση συμπερασμάτων αναφορικά με την αποτελεσματικότητα του Spleeter στη διαδικασία διαχωρισμού των stems.

3.7 Κλίμακα Καθαρότητας και Διαρροής (ΚΚΔ), Ποιότητα Διαχωρισμού

Για την αξιολόγηση του φαινομένου της διαρροής μεταξύ των stems, αλλά και του βαθμού καθαρότητας του ηχητικού περιεχομένου, σχεδιάστηκε ειδικά για τις ανάγκες της παρούσας εργασίας μία υποκειμενική Κλίμακα Καθαρότητας και Διαρροής (ΚΚΔ) τεσσάρων βαθμών. Η κλίμακα αυτή επιτρέπει τη κατηγοριοποίηση των επιπέδων καθαρότητας και διαρροής, δίνοντας έμφαση στην ακρόαση, υποβοηθούμενη από την οπτική παρατήρηση της κυματομορφής και του φασματογραφήματος. Ο συνδυασμός αυτός, επιτρέπει την ταξινόμηση παρέχοντας ένα σαφές πλαίσιο για την αποτελεσματικότητα του εργαλείου διαχωρισμού. Επιπρόσθετα, θεωρούμε ότι έτσι διευκολύνεται η επαναληψιμότητα και σύγκριση της διαδικασίας ανάλυσης.

Η Κλίμακα Καθαρότητας και Διαρροής (ΚΚΔ) έχει ως εξής:

Κλίμακα Καθαρότητας και Διαρροής (ΚΚΔ)				
	0	1	2	3
Καθαρότητα (Κ)	Δεν ακούμε σχεδόν καθόλου την απομόνωση του stem	Είναι εντοπίσιμο το stem αλλά διακόπτεται αρκετά συχνά	Είναι διακριτό το stem στη μεγαλύτερη διάρκεια του	Το stem είναι σχεδόν πλήρως ή απόλυτα απομονωμένο και καθαρό
Διαρροή (Δ)	Από κανένα stem	Μόνο 1 stem	Από 2 - 3 stems	Από όλα τα 4 stems

Επιπλέον κατασκευάσαμε μία απλή συνάρτηση Π_{Δ} , η οποία αξιολογεί την συνολική Ποιότητα του Διαχωρισμού (Π_{Δ}) για κάθε stem, συνδυάζοντας τις προηγούμενες δύο μεταβλητές: Καθαρότητα (Κ) και Διαρροή (Δ).

Αναλυτικότερα:

Έστω ένα σύνολο

$$A = \{0, 1, 2, 3\}.$$

Ορίζουμε τη συνάρτηση Ποιότητας Διαχωρισμού:

$$\Pi_{\Delta}(K, \Delta) = \frac{K + (3 - \Delta)}{6}, \text{ όπου } K, \Delta \in A,$$

καθώς

$$D_{\Pi_{\Delta}} = A \times A = \{(K, \Delta) \mid K \in A, \Delta \in A\}$$

και

$$\Pi_{\Delta}(D_{\Pi_{\Delta}}) = \left\{0, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1\right\}.$$

Καθαρότητα (Κ)	3	1	0,83	0,66	0,5
	2	0,83	0,66	0,5	0,33
	1	0,66	0,5	0,33	0,16
	0	0,5	0,33	0,16	0
		0	1	2	3
		Διαρροή (Δ)			

Εδώ, τόσο η K όσο και η Δ εκφράζονται ως ακέραιες τιμές από το A . Ο τύπος της συνάρτησης είναι φτιαγμένος έτσι ώστε η $P\Delta$ να δίνει πάντοτε τιμές μεταξύ 0 και 1, όπου το 0 αντιστοιχεί στον χειρότερο διαχωρισμό (ελάχιστη καθαρότητα και μέγιστη διαρροή), ενώ το 1 αντιστοιχεί στον απόλυτο διαχωρισμό (μέγιστη καθαρότητα και μηδενική διαρροή). Με τον συνδυασμό καθαρότητας και διαρροής, η $P\Delta$ προσφέρει έναν σύντομο τρόπο σύγκρισης της αποδοτικότητας του διαχωρισμού.

3.8 Υποθέσεις στο στάδιο της προεργασίας

Κατά το προκαταρκτικό στάδιο της έρευνας, διατυπώθηκαν από τον συγγραφέα της συγκεκριμένης εργασίας και τον επιβλέποντα καθηγητή ορισμένες βασικές υποθέσεις. Αυτές στόχευαν σε μία αρχική πρόβλεψη της πιθανής απόδοσης του εργαλείου διαχωρισμού και στον προσδιορισμό των παραμέτρων οι οποίοι ενδέχεται να επηρεάσουν την αποτελεσματικότητά του.

Αναλυτικότερα:

- **Πολυφωνία και πυκνότητα της υφής:** Αναμένεται ότι έργα με υψηλό βαθμό πολυφωνίας, όπως των Saint-Saëns και Rachmaninoff, θα εμφανίσουν αυξημένα ποσοστά διαρροής. Το γεγονός αυτό αποδίδεται στην ταυτόχρονη παρουσία πολλαπλών stream, τα οποία ενδέχεται να δημιουργήσουν επικάλυψη στις συχνότητες. Το Spleeter, θα κληθεί να διαχειριστεί σύνθετα spectrograms, κάτι που ενδέχεται να υπερβαίνει τις δυνατότητές του.
- **Λιγότερο σύνθετα μουσικά περιβάλλοντα:** Αντιθέτως, σε έργα με απλούστερη δομή και μικρότερο αριθμό ηχητικών πηγών, όπως το Take Five του Dave Brubeck, αναμένεται πιο ξεκάθαρος και ακριβής διαχωρισμός των stems. Η μειωμένη πολυπλοκότητα στο stem separation επιτρέπει υψηλότερη πιστότητα στην απομόνωση
- **Ανίχνευση της φωνής:** Η φωνή, είναι από μόνη της ένα αρκετά έντονο ηχητικό στοιχείο με συγκεκριμένα και ξεκάθαρα χαρακτηριστικά φάσματος και ξεχωριστή δυναμική. Επιπλέον αποτελεί κεντρικό στοιχείο των δεδομένων εκπαίδευσης του μοντέλου (τραγούδια ποπ – ροκ). Αναμένεται λοιπόν, να αναγνωρίζεται με υψηλό ποσοστό επιτυχίας από το Spleeter. Ενδιαφέρον παρουσιάζει η αξιολόγηση της απόδοσης του εργαλείου σε χορωδιακά έργα, όπου η φωνή δεν παρουσιάζεται ως μονοφωνική αλλά μέσα σε ένα πολυφωνικό σύνολο ειδικά αν σκεφτούμε και το μέρος της ορχήστρας, αν αυτή υπάρχει, όπως στα δύο χορωδιακά του Ola Gjeilo.
- **Ειδικές περιπτώσεις οργάνων (πιάνο, μπάσο):** Η απόδοση του μοντέλου προβλέπεται να παρουσιάσει δυσκολίες στον διαχωρισμό ηχητικών πηγών όπως το πιάνο, λόγω της τεράστιας έκτασης του και της συνεχούς και ταυτόχρονης παρουσίας του σε πολλές περιοχές του φάσματος. Αντίστοιχα προβλήματα μη διαχωρισμού αναμένονται και στον

διαχωρισμό του μπάσου, ιδίως όταν πρόκειται για λειτουργία αρμονικού μπάσου όπως σε έργα εποχής μπαρόκ.

- **Επίδραση της αντήχησης (reverb):** Ένας παράγοντας που ενδέχεται να επιβαρύνει σημαντικά την ακρίβεια του διαχωρισμού είναι η πιθανή έντονη αντήχηση στην αρχική ηχογράφιση. Υποθέτουμε ότι όσο μεγαλύτερη είναι η αντήχηση, τόσο πιο «θολή» γίνεται η αναγνώριση και στη συνέχεια οριοθέτηση των πηγών, συχνά οδηγώντας σε τουλάχιστον μέτριου βαθμού διαρροή, σύμφωνα με την Κλίμακα Διαρροής (ΚΔ).
- **Γενική υπόθεση πολυπλοκότητας:** Ως γενική αρχή, διατυπώνεται η υπόθεση ότι: Όσο αυξάνεται η πολυπλοκότητα της μουσικής υφής και η επεξεργασία κατά τη μίξη (mixing), τόσο αυξάνεται και η πιθανότητα εμφάνισης διαρροής στα stems κατά τον διαχωρισμό. Αυτή η συσχέτιση θεωρούμε ότι είναι κρίσιμη για την ερμηνεία των αποτελεσμάτων και την αξιολόγηση του Spleeter σε διαφορετικά μουσικά παραδείγματα.

3.9 Περιορισμοί της έρευνας

Παρότι η μεθοδολογία της παρούσας μελέτης δομήθηκε με στόχο την μεγαλύτερη εγκυρότητα και αξιοπιστία των εξαγόμενων αποτελεσμάτων, είναι απαραίτητο να αναγνωριστούν καθώς και να αναφερθούν ορισμένοι περιορισμοί που επηρεάζουν τη γενίκευση των συμπερασμάτων.

Αναλυτικότερα:

- **Υποκειμενικότητα στην αξιολόγηση:** Η διαδικασία αξιολόγησης της διαρροής περιλαμβάνει, πέραν της ανάλυσης στο Sonic Visualiser, και την ακουστική επαλήθευση μέσω του ανθρώπινου αυτιού. Παρά την προσπάθεια εφαρμογής κοινών κριτηρίων μέσω της Κλίμακας Καθαρότητας και Διαρροής (ΚΚΔ), η ακουστική αντίληψη συμπεριλαμβάνει αναπόφευκτα ένα βαθμό υποκειμενικότητας. Η ανίχνευση της διαρροής μπορεί να διαφέρει ανάλογα με την εμπειρία, την προσοχή και την εκπαίδευση του εκάστοτε αξιολογητή.
- **Περιορισμοί του εργαλείου διαχωρισμού:** Το Spleeter έχει εκπαιδευτεί κυρίως σε μουσικά δεδομένα που ανήκουν στο πεδίο της δημοφιλούς μουσικής π.χ. pop, rock. (Hennequin et al., 2019) Ως εκ τούτου, η απόδοσή του παρουσιάζει μειωμένη αποτελεσματικότητα όταν εφαρμόζεται σε μουσικά έργα διαφορετικής αισθητικής, εποχής και μορφολογικής κατασκευής. Έτσι, τα μοντέλα διαχωρισμού δεν προσαρμόζονται επαρκώς στην πολυπλοκότητα που χαρακτηρίζει κάποια έργα τα οποία είναι μέρος των παραδειγμάτων που επιλέχτηκαν.

- **Τεχνικοί περιορισμοί της ανάλυσης:** Η χρήση του STFT, που αποτελεί τη βάση του διαχωρισμού της συγκεκριμένης προσέγγισης, μπορεί να προκαλέσει παραμορφώσεις ή ασάφειες στην αναπαράσταση ορισμένων ηχητικών περιοχών των παραδειγμάτων. Έτσι, ενδέχεται να μην αντιπροσωπεύεται η πραγματική ηχητική πληροφορία που εισήχθη, γεγονός που προσθέτει επιπλέον προβλήματα στην αξιολόγηση των stems.
- **Περιορισμένο μέγεθος δείγματος:** Η έρευνα βασίστηκε στην ανάλυση δεκατεσσάρων (14) μουσικών παραδειγμάτων. Αν και η επιλογή τους πραγματοποιήθηκε βάσει συγκεκριμένων κριτηρίων (βλ. ενότητα 3.5), το μέγεθος του δείγματος παραμένει περιορισμένο λόγω πόρων και χρόνου. Αυτό ενδέχεται να περιορίζει τη δυνατότητα γενίκευσης των ευρημάτων.

4. Παρουσίαση δεδομένων, Ανάλυση και Ευρήματα

(Ανάλυση, ευρήματα και παρατηρήσεις για κάθε stem)

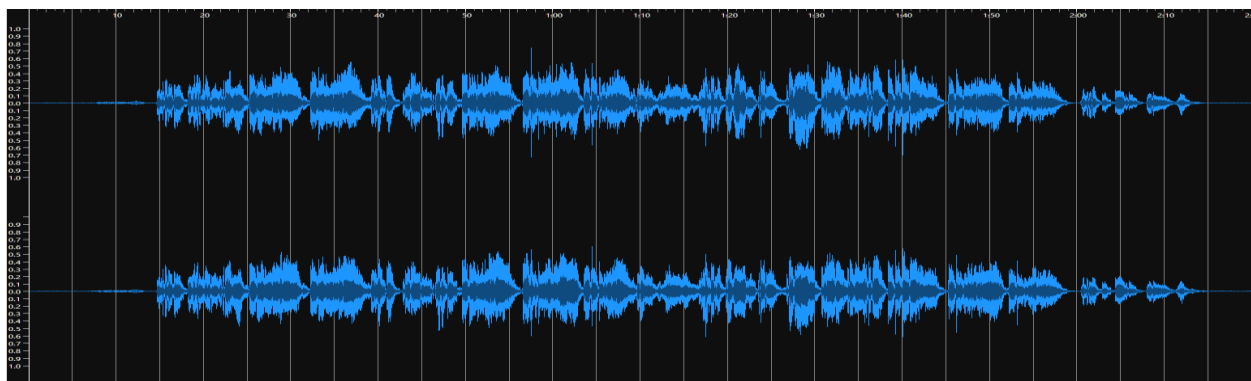
4.1 Μουσικό παράδειγμα 1: Borislav Slavov: Baldur's Gate 3 – Down By The River

Το πρώτο μουσικό παράδειγμα που αναλύθηκε είναι το Down By The River του Borislav Slavov, το οποίο αποτελεί μέρος από το soundtrack του βιντεοπαιχνιδιού Baldur's Gate 3. Πρόκειται για ένα αρκετά ατμοσφαιρικό και συναισθηματικά φορτισμένο έργο, ενώ το στοιχείο που κυριαρχεί είναι η γυναικεία φωνή. Οι στοίχοι που τραγουδά η φωνή πάνω σε ένα «χαλί» συνοδευτικών εγχόρδων, άρπας και κάποιον ηλεκτρονικών στοιχείων, σε συνδυασμό με την επεξεργασία του βάθους (depth) και της αντήχησης (reverb) δημιουργούν ως αποτέλεσμα ένα ηχητικό περιβάλλον «φαντασίας». Η επιλογή του έγινε λόγω της ιδιαίτερης υφής της φωνής αλλά και της αίσθησης ατμοσφαιρικότητας που αποτυπώνεται στα αυτιά του ακροατή.

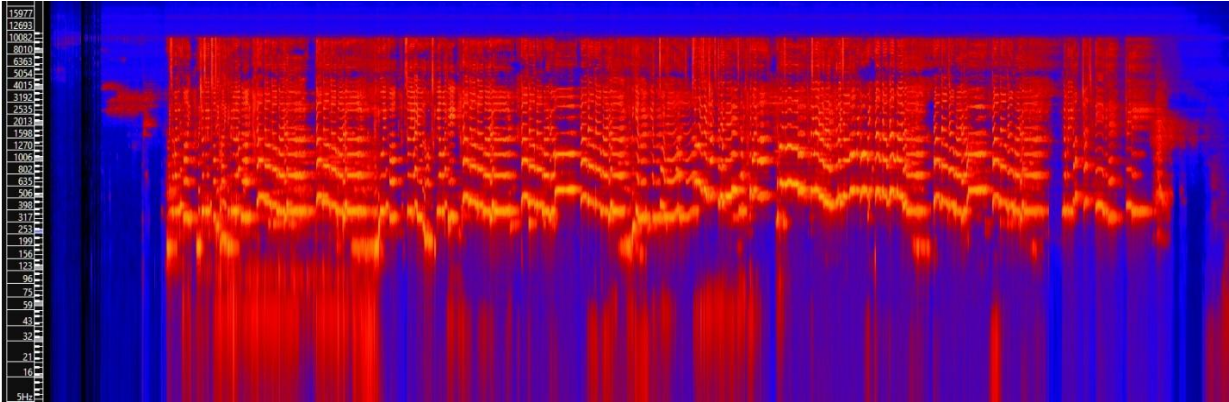
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems του Spleeter, το οποίο διαχωρίζει ένα ηχητικό αρχείο σε πέντε (5) stems, που ονομαστικά έχουν ως εξής: vocals, piano, drums, bass και other. Το αρχείο που εισήχθη στο Spleeter είχε διάρκεια περίπου 2 λεπτών και 11 δευτερολέπτων. Ο διαχωρισμός έγινε χωρίς καμία ανάγκη για περαιτέρω επεξεργασία ή μετατροπή στο πρωτογενές αρχείο. Η εξαγωγή ολοκληρώθηκε χωρίς τεχνικά προβλήματα (επαναλαμβανόμενες προσπάθειες, «crush» του εργαλείου κτλ.), και για τα πέντε παραγόμενα stems καθώς έγινε η αποθήκευσή τους στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 1). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 1: Waveform of vocal stem (Sonic Visualiser)



Εικόνα 2: Spectrogram of vocal stem (Sonic Visualiser)

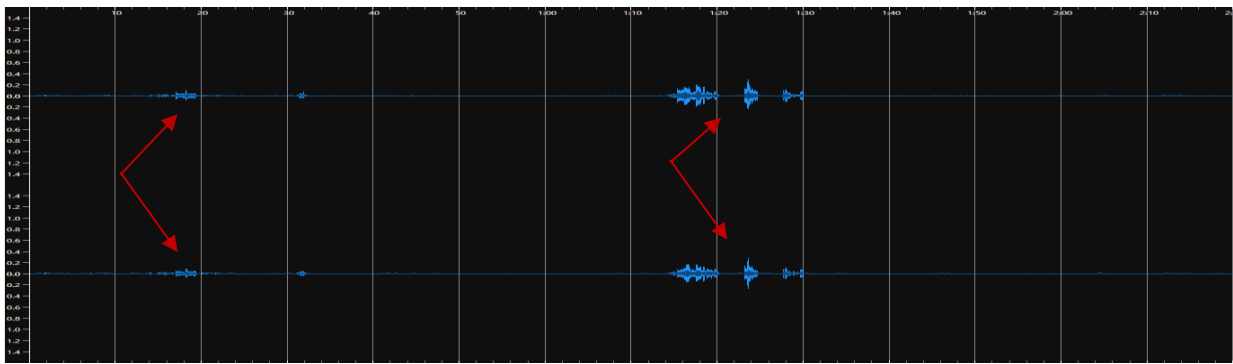
Λόγω της συνεχούς παρουσίας της φωνής σε όλη σχεδόν τη διάρκεια του έργου (με εξαίρεση τα αρχικά δευτερόλεπτα της εισαγωγής), το διαχωρισμένο stem αναμένεται να είναι ηχητικά πλήρες από τη φωνή μετά το 14ο δευτερόλεπτο. Η ακρόαση έδειξε ότι, παρά την ύπαρξη αντήχησης και των διαφόρων επιπρόσθετων εφέ, τα οποία είναι ενσωματωμένα στο αρχικό mixing, ο διαχωρισμός της φωνής επιτεύχθηκε σε μεγάλο βαθμό με επιτυχία. Αυτό θεωρούμε ότι είναι καθόλα σημαντικό, καθώς φαίνεται η ικανότητα του μοντέλου να απομονώνει με σχετική ευκολία τη φωνή από το αρχικό mixing.

Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα				X
Διαρροή		X		
Π.Δ.				0,83

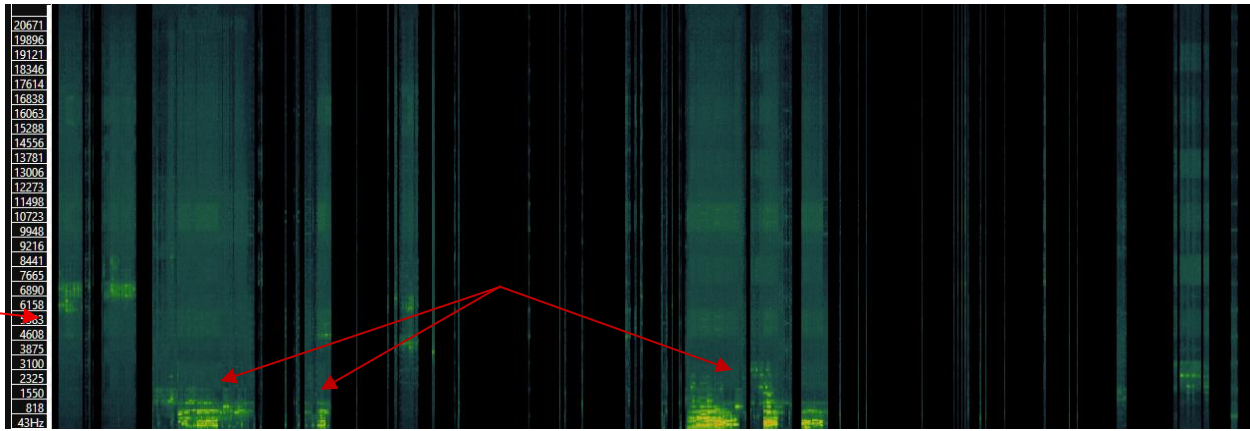
Ωστόσο, σε ορισμένα σημεία παρατηρείται διαρροή από το other stem, με αποτέλεσμα να ακούγονται ελάχιστα η άρπα και η τσελέστα. Η παρουσία αυτών των στοιχείων δεν επιδρά ουσιαστικά στην καθαρότητα του vocal stem, καθώς δεν δημιουργεί αισθητή παρεμβολή στο διαχωρισμό της φωνής. Παρόλα αυτά, εξαιτίας αυτής της διαρροής, ενώ ο διαχωρισμός της φωνής παραμένει καθαρός σχεδόν σε όλο το stem, η συνολική ποιότητα του διαχωρισμού συνεπάγεται με 0,83.

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 3: Waveform of piano stem (Sonic Visualiser)



Εικόνα 4: Spectrogram of piano stem (Sonic Visualiser)

Αρχικά, πρέπει να διασαφηνίσουμε ότι για την περίπτωση του πιάνου, αυτό δεν υπάρχει σαν όργανο στο συγκεκριμένο παράδειγμα. Παρ’ όλα αυτά, τόσο η κυματομορφή όσο και το φασματογράφημα υποδεικνύουν την παρουσία διαρροής σε σημεία στο συγκεκριμένο stem, το οποίο θεωρητικά θα έπρεπε να είναι κενό.

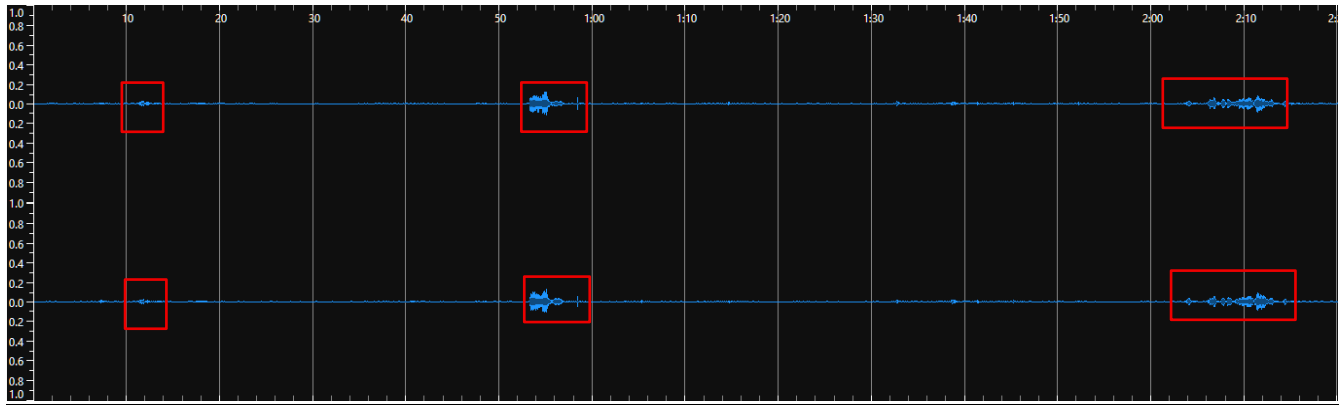
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα		X		
Διαρροή			X	
Π.Δ.	0,33			

Αρχικά, μέσω της ακρόασης διαπιστώνεται ότι το μοντέλο συγχέει ελαφρώς την άρπα με το πιάνο σε συγκεκριμένα σημεία μέσα στο έργο. Αυτό φυσικά θεωρούμε ότι είναι απόλυτα λογικό, διότι μέρος των συχνοτήτων και περιοχών των δύο οργάνων, ειδικά στις μεσαίες περιοχές τους, παρουσιάζουν σημαντική ομοιότητα (Fletcher & Rossing, 1991). Συνεπώς, είναι ορθό και αναμενόμενο το μοντέλο να παράξει το συγκεκριμένο αποτέλεσμα.

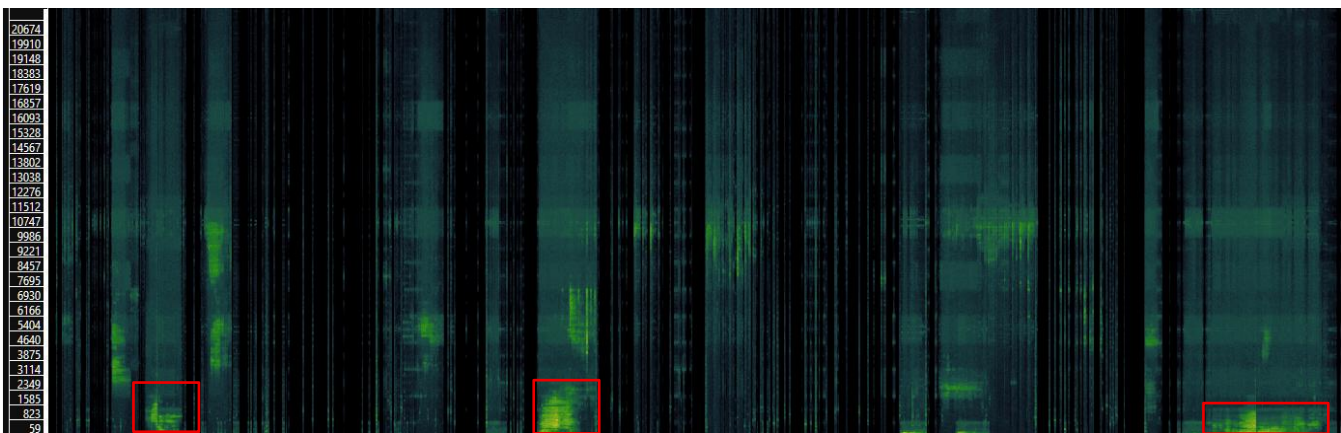
Στα σημεία του spectrogram όπου οι έντονες πράσινες περιοχές συχνοτήτων είναι ανάμεσα από 6500 και 7000 Hz (στην αρχή του αποσπάσματος) ακούμε ξεκάθαρα την εισαγωγή της άρπας, σε χαμηλή αλλά αισθητή ένταση και ψηλά συχνοτικά. Οπτικά επίσης την εντοπίζουμε και είναι απόλυτα εμφανής. Αυτός είναι και ο λόγος που στην ΚΚΔ αξιολογείται αρχικά με βαθμό διαρροής 1. Στη συνέχεια στα άλλα τρία σημεία ακούγεται μέρος της άρπας ξανά από το other αλλά αυτή τη φορά έχουμε και διαρροή από το bass stem. Αυτό συνεπάγεται τελικώς με διαρροή 2^{ου} βαθμού. Τέλος η συνολική ποιότητα του διαχωρισμού του stem σε συνάρτηση της διαρροής και καθαρότητας είναι 0,33.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 5: Waveform of drums stem (Sonic Visualiser)



Εικόνα 6: Spectrogram of drums stem (Sonic Visualiser)

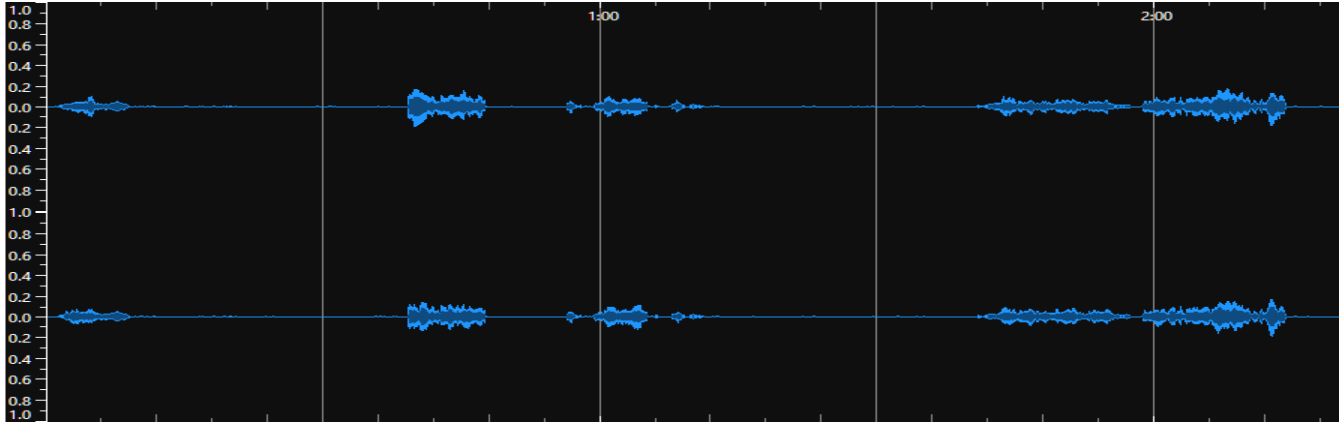
Αντιστοίχως με το πιάνο, πρέπει να τονίσουμε ότι και για την περίπτωση των drums, αυτά δεν υπάρχουν στο συγκεκριμένο παράδειγμα. Παρόλα αυτά, τόσο η κυματομορφή όσο και φασματογράφημα υποδεικνύουν την παρουσία διαρροής στο συγκεκριμένο stem, το οποίο θεωρητικά θα έπρεπε να είναι και αυτό εντελώς κενό.

Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

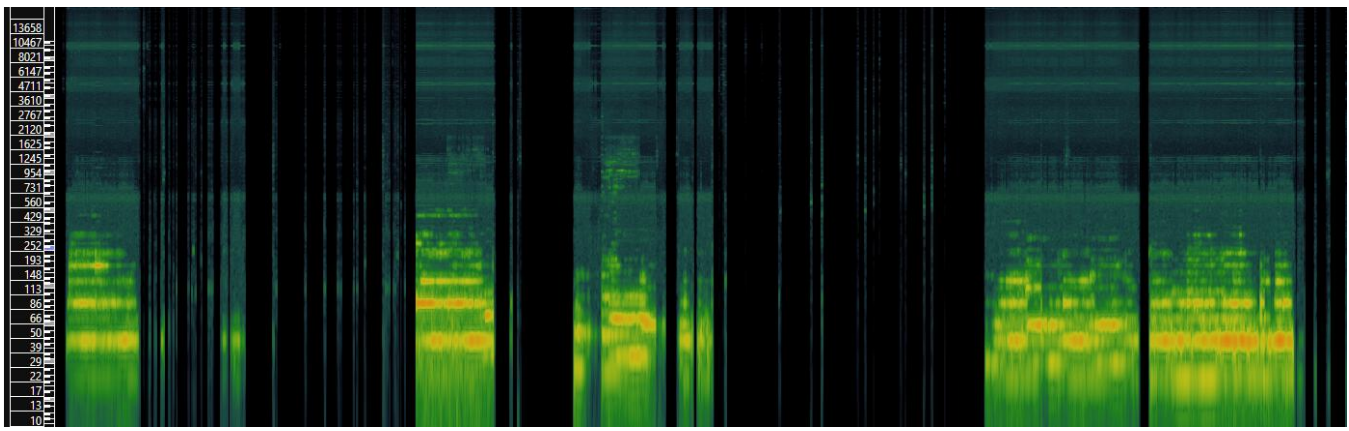
Σε αυτή την περίπτωση όμως, παρατηρείται διαρροή πάλι από τμήματα της άρπας κυρίως στα κόκκινα σημεία αλλά ταυτόχρονα και μεγάλη ακουστική αλλοίωση με περίεργους ήχους που δεν μπορούμε να τα ταυτίσουμε με κάποιο stem. Κατά την ακρόαση, περισσότερο αντιληπτά γίνονται τα τρία κόκκινα σημεία ενώ τα υπόλοιπα είναι λιγότερο αισθητά. Εκτός των τριών σημείων, στα οποία δεν είναι καθαρό το stem θα λέγαμε ότι η καθαρότητα είναι στο 2 ενώ η διαρροή στο 1 λόγω της άρπας.

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 7: Waveform of bass stem (Sonic Visualiser)



Εικόνα 8: Spectrogram of bass stem (Sonic Visualiser)

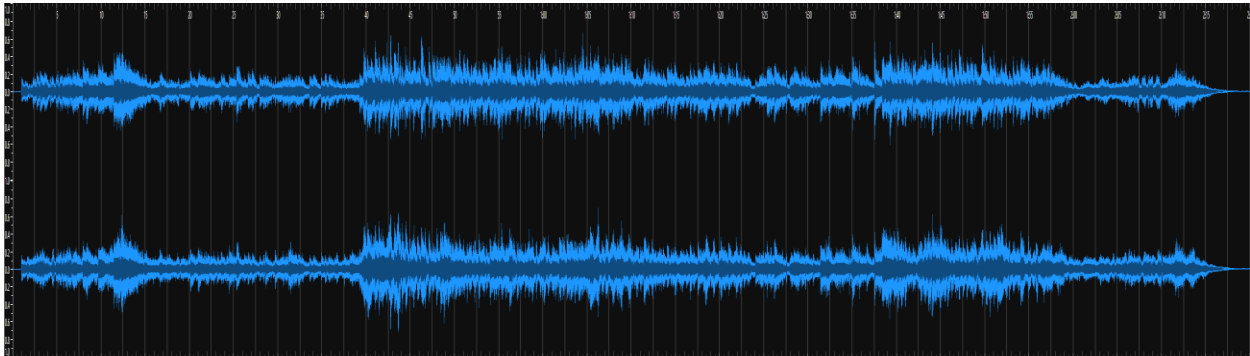
Στην περίπτωση με το stem του bass, αναμένουμε να ακούσουμε τις πολύ χαμηλές συχνότητες των εγχόρδων, οι οποίες επιτελούν τη λειτουργία του αρμονικού μπάσου. Κατά την ακρόαση, θα μπορούσε να ειπωθεί ότι πράγματι η απομόνωση είναι αρχικά χωρίς διαρροές από τα υπόλοιπα stems, διότι αυτό που ηχεί είναι ένας συνδυασμός ηλεκτρονικών μπάσων συχνοτήτων μαζί με τον ήχο των φυσικών εγχόρδων. Ωστόσο, ο διαχωρισμός δεν είναι απόλυτος, καθώς σε πολλά σημεία το stem του μπάσου είναι κενό ενώ γνωρίζουμε ότι το μπάσο ηχεί σε περισσότερα σημεία από εκεί που ανιχνεύτηκε.

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

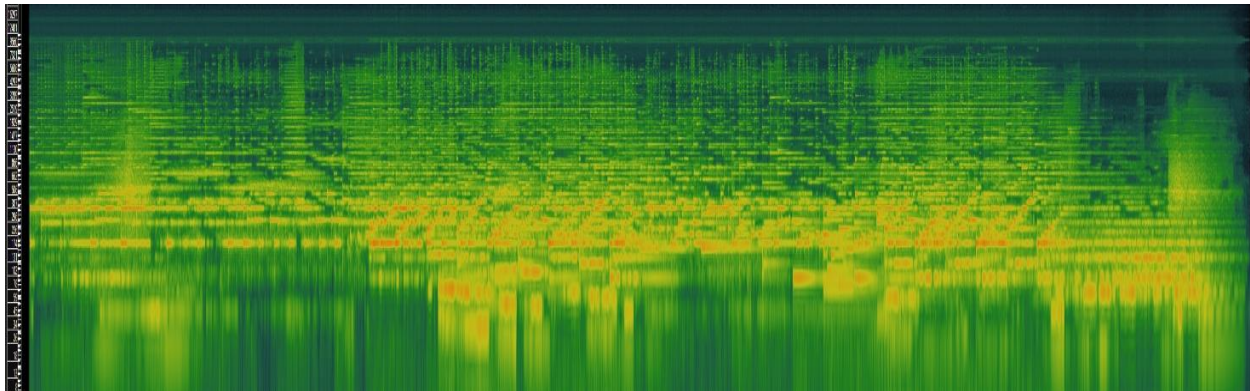
Συνεπώς, το μοντέλου διαχωρισμού δεν ανίχνευσε σωστά το συγκεκριμένο stem σε όλη τη διάρκειά του. Όπως ήδη αναφέρθηκε, οι συχνότητες που ακούγονται βρίσκονται εντός της περιοχής του μπάσου και δεν παρεμβαίνει εμφανώς κανένα άλλο stem. Για τον λόγο αυτό, η διαχωρισμός κρίνεται με διαρροή 0ου βαθμού ενώ η καθαρότητα στο 2 διότι δεν ανιχνεύτηκε σε όλη τη διάρκειά του.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 9: Waveform of other stem (Sonic Visualiser)



Εικόνα 10: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other περιέχεται η κύρια ενορχήστρωση, όπως έγχορδα, άρπα, celesta, πνευστά καθώς και διάφορα εφέ που βοηθούν στην συνθετική δημιουργία του ambient όπως reverb, echo κ.ά. Ακουστικά όμως, εντοπίζουμε επίσης κάποια ψήγματα και της φωνής, από τις χαμηλές κυρίως περιοχές της, το οποίο μας προκαλεί ιδιαίτερη εντύπωση και συνεπάγεται με διαρροή 1^{ου} βαθμού.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Πέρα από τη φωνή, το other περιλαμβάνει σε μεγάλο βαθμό την αναμενόμενη πληροφορία. Το πιο χαρακτηριστικό ζήτημα που παρατηρείται και είναι ακουστικά άμεσα εντοπίσιμο είναι η καθαρότητα της τσελέστας, η οποία ακούγεται θολά και λιγότερο ευδιάκριτα σε ορισμένα σημεία. Λόγω του παραπάνω η καθαρότητα συνεπάγεται σε 2^{ου} βαθμού .

Συμπεράσματα Παραδείγματος

Ο διαχωρισμός στο συγκεκριμένο έργο απέδωσε ικανοποιητικά αποτελέσματα, ιδιαίτερα όσον αφορά τον διαχωρισμό της φωνής. Ακουστικά, στο stem των vocals παρατηρήθηκαν πιο καθαρές περιοχές της φωνής και με σχετική ευκρίνεια. Η διαρροή από τα υπόλοιπα κανάλια ήταν ελάχιστη αλλά υπαρκτή.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα				X
	Διαρροή		X		
	Π.Δ.	0,83			
Piano	Καθαρότητα		X		
	Διαρροή			X	
	Π.Δ.	0,33			
Drums	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Bass	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Other	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			

Το εργαλείο κατόρθωσε να εντοπίσει σωστά το φωνητικό περιεχόμενο, και επιπλέον ήταν σε θέση να απομονώσει πλήρως τη φωνή από τα συνοδευόμενα εφέ, κάτι που είναι απόλυτα αναμενόμενο για το στάδιο εξέλιξης του συγκεκριμένου μοντέλου διαχωρισμού.

Όσον αφορά τα drums, το μοντέλο παρουσίασε μικρότερη αποτελεσματικότητα. Παρότι το stem αυτό δεν υπήρχε και θεωρητικά έπρεπε να ήταν καθαρό, διαπιστώθηκε διαρροή, η οποία ακουστικά και οπτικά ήταν αντιληπτή σε 3 σημεία. Το αντίθετο συνέβη στο bass, το οποίο δεν είχε εμφανή διαρροή, ωστόσο είμαστε βέβαιοι ότι το μοντέλο δεν κατάφερε να εντοπίσει πλήρως όλο το περιεχόμενο του μπάσου.

Τέλος, στο other περιέχεται η κύρια ενορχήστρωση, όπως έγχορδα, άρπα, celesta, πνευστά. Για το συγκεκριμένο παράδειγμα δεδομένου ότι περιλαμβάνει ότι «περίσσεψε» από τα υπόλοιπα stems, αυτό που μας κάνει ιδιαίτερη εντύπωση είναι η μικρή διαρροή της φωνής σε κάποια λιγοστά σημεία, κάτι που πιθανώς οφείλεται στην πυκνή ενορχήστρωση.

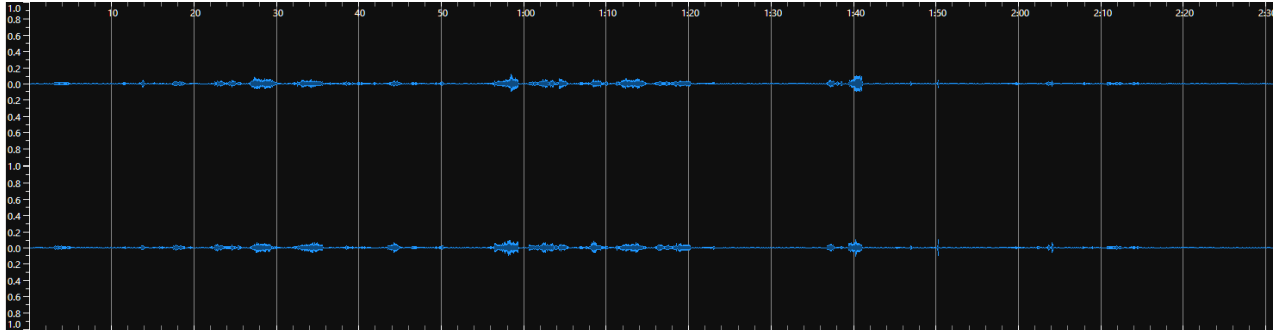
4.2 Μουσικό παράδειγμα 2: Camille Saint-Saëns: Carnival of the Animals, VII. Aquarium

Το δεύτερο μουσικό παράδειγμα που αναλύθηκε είναι το Carnival of the Animals, VII. Aquarium του Camille Saint-Saëns, ένα έργο στα τέλη της ρομαντικής περιόδου με επίσης έντονα ατμοσφαιρικά χαρακτηριστικά. Το κομμάτι χαρακτηρίζεται από τον ήχο των εγχόρδων, της άρπας, και κυρίως της τσελέστας. Ο συνδυασμός των παραπάνω με το πιάνο δημιουργούν μια ατμόσφαιρα που θυμίζει νερό και το τοπίο ενός ενυδρείου.

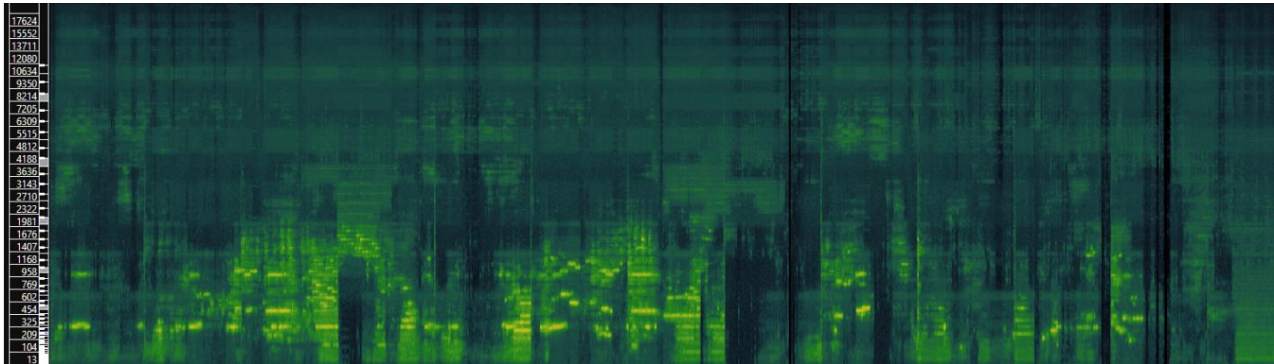
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems επίσης. Το αρχείο που εισήχθη στο Spleeter είχε διάρκεια περίπου 2 λεπτών και 30 δευτερολέπτων. Ο διαχωρισμός έγινε χωρίς καμία ανάγκη για περαιτέρω επεξεργασία ή μετατροπή στο πρωτογενές αρχείο. Η εξαγωγή ολοκληρώθηκε χωρίς τεχνικά προβλήματα (επαναλαμβανόμενες προσπάθειες, «crush» του εργαλείου κτλ.), και για τα πέντε παραγόμενα stems, τα οποία αποθηκεύτηκαν στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 2). Κατόπιν, εισήχθησαν στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της διαρροής και καθαρότητας του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 11: Waveform of vocal stem (Sonic Visualiser)



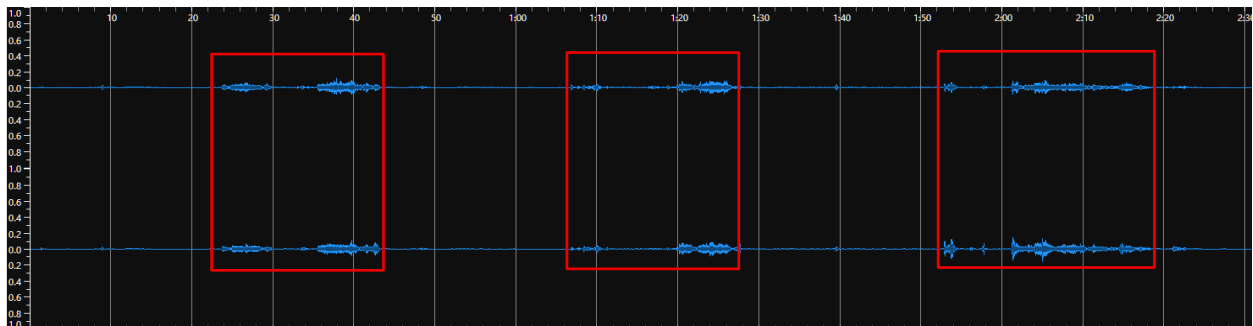
Εικόνα 12: Spectrogram of vocal stem (Sonic Visualiser)

Λόγω της απουσίας φωνής στο συγκεκριμένο έργο, το stem θα έπρεπε να είναι απολύτως καθαρό. Ωστόσο, σύμφωνα με την κυματομορφή και το φασματογράφημα εντοπίζονται αρκετές συχνότητες, οι οποίες μας προϊδεάζουν για διαρροή. Η παραπάνω παρατήρηση επιβεβαιώνεται και ακουστικά διότι στο stem των vocals ακούγονται σε όλη τη διάρκεια του τμήματα από τα stems των piano, bass και other γεγονός που οδηγεί στην διαρροή 2ου βαθμού.

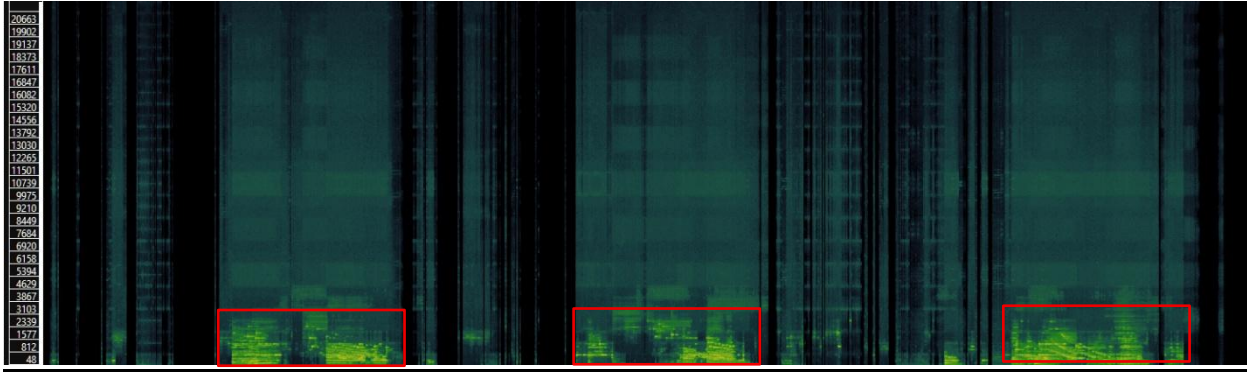
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα	X			
Διαρροή			X	
Π.Δ.	0,16			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 13: Waveform of piano stem (Sonic Visualiser)



Εικόνα 14: Spectrogram of piano stem (Sonic Visualiser)

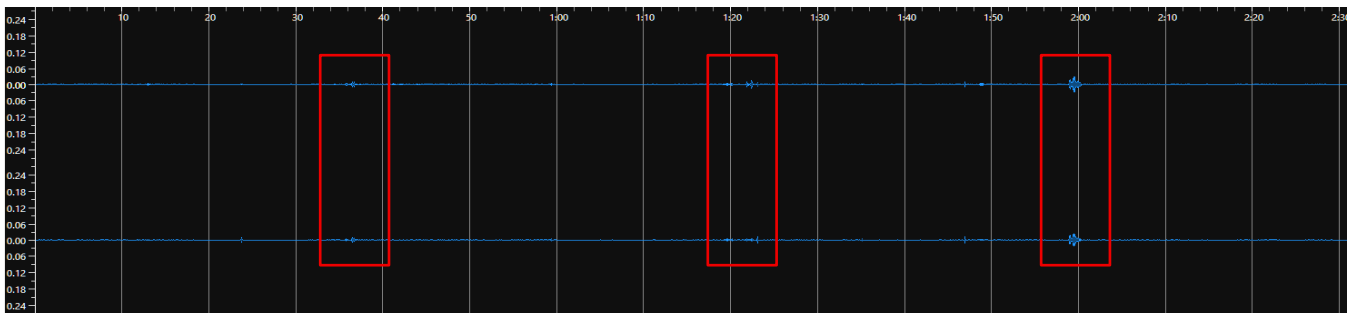
Ξεκινώντας με την ακρόαση ολόκληρου του έργου πριν το διαχωρισμό, διαπιστώνεται ότι το πιάνο είναι παρόν σε όλη τη διάρκειά του, εκτελώντας καταβάσεις από αρπές σε γρήγορο tempo. Μετά τον διαχωρισμό, το piano stem εμφανίζει σημαντικές αδυναμίες. Στα περισσότερα σημεία ακούγεται πολύ χαμηλά και θολά, αλλού με διακοπές, σε άλλα απουσιάζει εντελώς, ενώ μόνο σε 3 τμήματα διακρίνεται καθαρότερα. Το Spleeter δεν κατάφερε να το απομονώσει σε ικανοποιητικό βαθμό, με αποτέλεσμα να ακούγεται μόνο σε ελάχιστα σημεία. Όσο αναφορά τη διαρροή, ακουστικά εντοπίζονται στοιχεία από το other stem όπως η τσελέστα και τα έγχορδα. Συνεπώς, ακουστικά η διαρροή αξιολογείται ως 1^ο βαθμού, καθώς ανιχνεύεται και ξένο υλικό από άλλο stem πέρα από τον αναμενόμενο θόρυβο.

Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα		X		
Διαρροή		X		
Π.Δ.	0,5			

Τόσο η κυματομορφή όσο και το φασματογράφημα επαληθεύουν την ακρόαση και σε αυτή την περίπτωση. Με δεδομένο ότι γνωρίζουμε πως το πιάνο ηχεί σε όλη την διάρκεια του έργου, τα μεγάλα «κενά» που παρατηρούμε σε αυτά φανερώνουν ότι σημαντικά τμήματα του πιάνου δεν διαχωρίστηκαν καθόλου. Οπότε σε πολλά σημεία το stem παραμένει κενό. Συνεπώς η καθαρότητά του χαρακτηρίζεται ως 1^ο βαθμού γιατί εντοπίστηκε σε λιγοστά σημεία από το μοντέλο διαχωρισμού. συμπεραίνουμε περιορισμούς του μοντέλου στην απομόνωση του πιάνου για το συγκεκριμένο παράδειγμα.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 15: Waveform of drums stem (Sonic Visualiser)



Εικόνα 16: Spectrogram of drums stem (Sonic Visualiser)

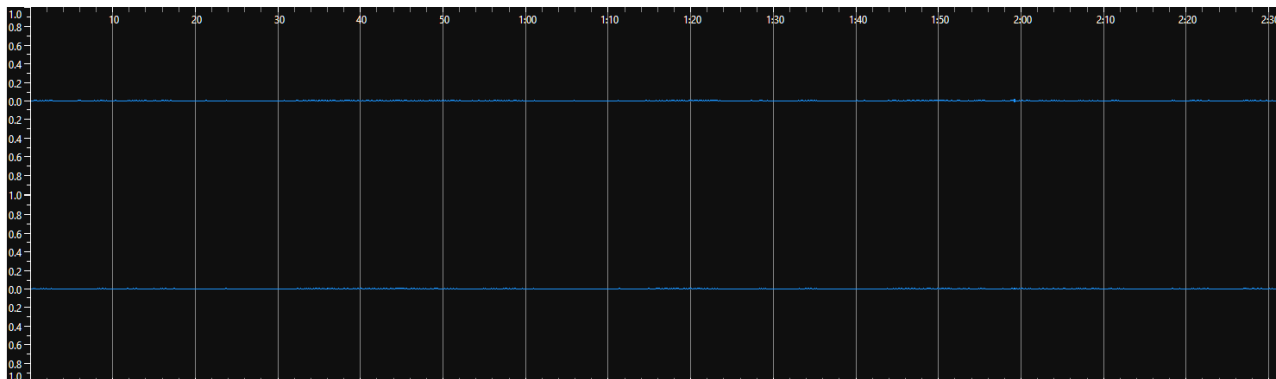
Για το συγκεκριμένο stem τα πράγματα είναι ιδιαίτερα απλά και σαφή. Σε αντίθεση με το πιάνο, το οποίο είναι παρόν σε όλη την διάρκεια του έργου και κατέχει τον βασικό ρόλο στο συγκεκριμένο μέρος, τα κρουστά απουσιάζουν πλήρως από την ενορχήστρωση. Θεωρητικά, η κυματομορφή και το φασματογράφημα θα έπρεπε να εμφανίζονται απολύτως κενά, χωρίς καμία ηχητική πληροφορία. Ωστόσο, η κυματομορφή υποδεικνύει την ύπαρξη ελάχιστης διαρροής.

Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα				X
Διαρροή			X	
Π.Δ.	0,66			

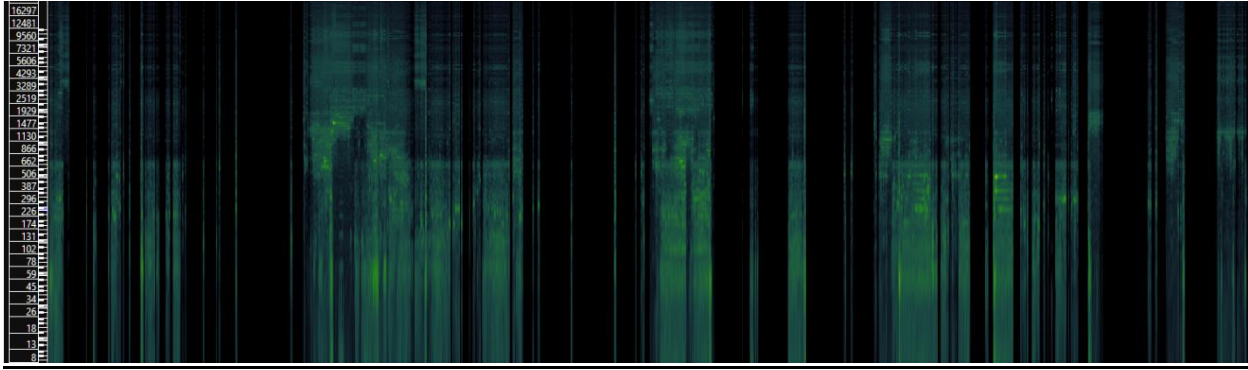
Ακουστικά, η διαρροή γίνεται πράγματι αντιληπτή σε τρία σημεία. Σε αυτά τα σημεία, είναι προερχόμενη από τα piano και other stems όπου η ένταση βρίσκεται σε τέτοιες τιμές, οι οποίες την καθιστούν αναγνωρίσιμη από τον ακροατή. Η διαρροή χαρακτηρίζεται ως 2^ο βαθμού, εύρημα το οποίο επιβεβαιώνεται και από το φασματογράφημα. Τέλος, η καθαρότητά του stem αξιολογείται ως 3^ο βαθμού, καθώς στη μεγαλύτερη διάρκεια του παραμένει ουσιαστικά κενό.

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 17: Waveform of bass stem (Sonic Visualiser)



Εικόνα 18: Spectrogram of bass stem (Sonic Visualiser)

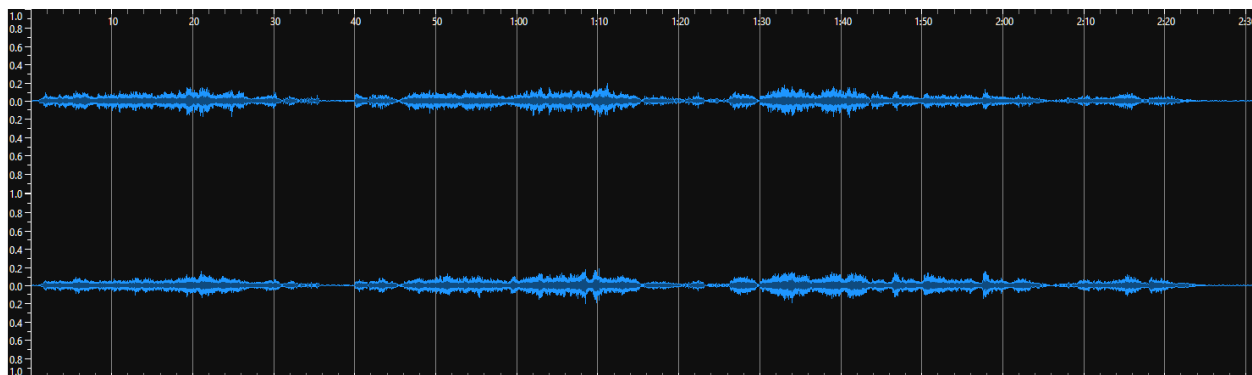
Στην περίπτωση του bass stem, θα ξεκινήσουμε με την ακρόαση. Στο συγκεκριμένο έργο, το ρόλο του μπάσου αναλαμβάνουν κυρίως οι χαμηλές συχνότητες του πιάνου και των εγχόρδων, οι οποίες έχουν την λειτουργία της αρμονικής θεμελίου του μπάσου. Ωστόσο, το εργαλείο διαχωρισμού δεν είναι εκπαιδευμένο να εντοπίσει αυτή τη λειτουργία και να τα απομονώσει. Για τον λόγο αυτό, αναμένουμε το stem να εμφανίζεται κενό.

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή		X		
Π.Δ.	0,83			

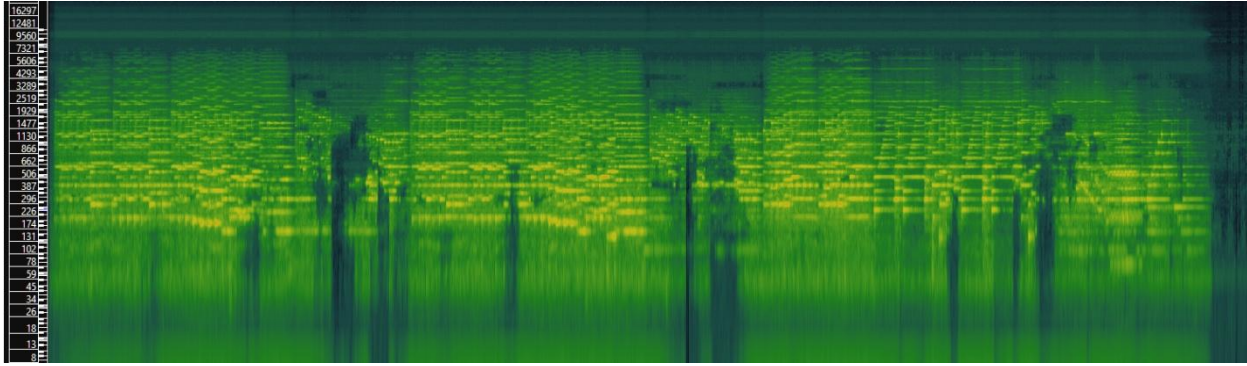
Ακουστικά, το αποτέλεσμα που ακούγεται περιορίζεται σε ελάχιστα αποσπάσματα εξαιρετικά χαμηλής έντασης, τα οποία προέρχονται πάλι από διαρροή του piano stem. Άρα συνεπάγεται με διαρροή 1^{ου} βαθμού. Όσο αναφορά την καθαρότητα, δεδομένου ότι αναμένουμε ένα κενό stem επειδή το μοντέλο δεν είναι ικανό να το εντοπίσει το μπάσο, όπως εξηγήσαμε παραπάνω, αυτή αξιολογείται ως 3^{ου}, καθώς στη μεγαλύτερη διάρκεια του παραμένει ουσιαστικά κενό. Παρόλο που η ΚΚΔ μας δίνει ποιότητα διαχωρισμού ίση με 0,83, η απομόνωση του μπάσου δεν πέτυχε σε καμία περίπτωση.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 19: Waveform of other stem (Sonic Visualiser)



Εικόνα 20: Spectrogram of other stem (Sonic Visualiser)

Θεωρητικά, το περιεχόμενό αυτού του stem θα έπρεπε να αποτελεί το «περίσσειμα» από την απομόνωση των υπολοίπων stems, δηλαδή των εγχόρδων, της άρπας και της τσελέστας. Ακουστικά όμως, παρατηρούμε μία πλήρη διαρροή.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα				
Διαρροή				
Π.Δ.				

Αναλυτικότερα, το φασματογράφημα ενώ εμφανίζεται «πλούσιο» συχνοτικά, δεν μας δίνει κάποια επιπλέον πληροφορία για την ύπαρξη διαρροής. Αυτό πράγματι επαληθεύεται κατά την ακρόαση, όπου το αποτέλεσμα του διαχωρισμού θυμίζει το έργο πριν από αυτόν με επιπλέον κάποια προσθήκη θορύβου. Συνεπώς, δεν υπάρχει πρακτική εφαρμογή της ΚΚΔ λόγω του μη διαχωρισμού.

Συμπεράσματα Παραδείγματος

Το παράδειγμα 2 ανέδειξε σημαντικούς περιορισμούς στην ακρίβεια διαχωρισμού του Spleeter. Στην περίπτωση του bass stem, το ρόλο του μπάσου αναλαμβάνουν κυρίως οι χαμηλές συχνότητες του πιάνου και των εγχόρδων, οι οποίες έχουν την λειτουργία της αρμονικής θεμελίου του μπάσου. Ωστόσο, το εργαλείο διαχωρισμού δεν είναι εκπαιδευμένο να εντοπίσει αυτή τη λειτουργία και να τα απομονώσει.

Το πιάνο, το οποίο ηχεί σε όλη τη διάρκεια του έργου, το μοντέλο δεν κατάφερε να το απομονώσει πλήρως, με αποτέλεσμα την ύπαρξη μεγάλων «κενών» στο stem. Η αξιολόγηση μας, χαρακτήρισε την διαρροή και την καθαρότητα ως 1^ο βαθμού καθώς αυτό διακόπτεται συνεχώς και εντοπίζεται σε λιγοστά σημεία από το μοντέλο διαχωρισμού. Επίσης, σε περιπτώσεις όπως των drums και των vocals παρατηρήθηκε διαρροή συχνοτήτων από το piano stem, ενώ αυτά θα έπρεπε να είναι κενά. Τέλος το other stem θύμιζε το έργο πριν τον

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα	X			
	Διαρροή			X	
	Π.Δ.	0,16			
Piano	Καθαρότητα		X		
	Διαρροή		X		
	Π.Δ.	0,5			
Drums	Καθαρότητα				X
	Διαρροή			X	
	Π.Δ.	0,66			
Bass	Καθαρότητα				X
	Διαρροή		X		
	Π.Δ.	0,33			
Other	Καθαρότητα				
	Διαρροή				
	Π.Δ.				

διαχωρισμό με επιπλέον κάποια προσθήκη θορύβου. Τα ευρήματα αυτά αναδεικνύουν τις δυσκολίες του διαχωρισμού των stems στο σύνθετο περιβάλλον της ορχήστρας.

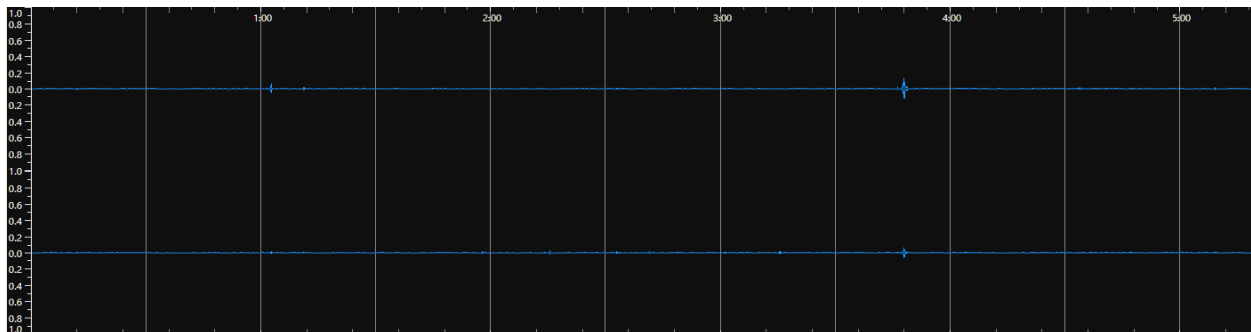
4.3 Μουσικό παράδειγμα 3: Claude Bolling: Suite for Flute and Jazz Piano Trio

Το τρίτο μουσικό παράδειγμα που αναλύθηκε είναι η Suite for Flute and Jazz Piano Trio του Claude Bolling, ένα έργο που συνδυάζει στοιχεία της κλασικής σουίτας και τζαζ μουσικής, με έντονα ρυθμικά μοτίβα και χαρακτηριστικό μουσικό διάλογο μεταξύ φλάουτου, πιάνου, κοντραμπάσου και ντραμς. Υπάρχουν συχνές και συνεχείς εναλλαγές μουσικών θεμάτων και αυτοσχεδιαστικών μερών που αναδεικνύουν τις δυνατότητες όλων των οργάνων.

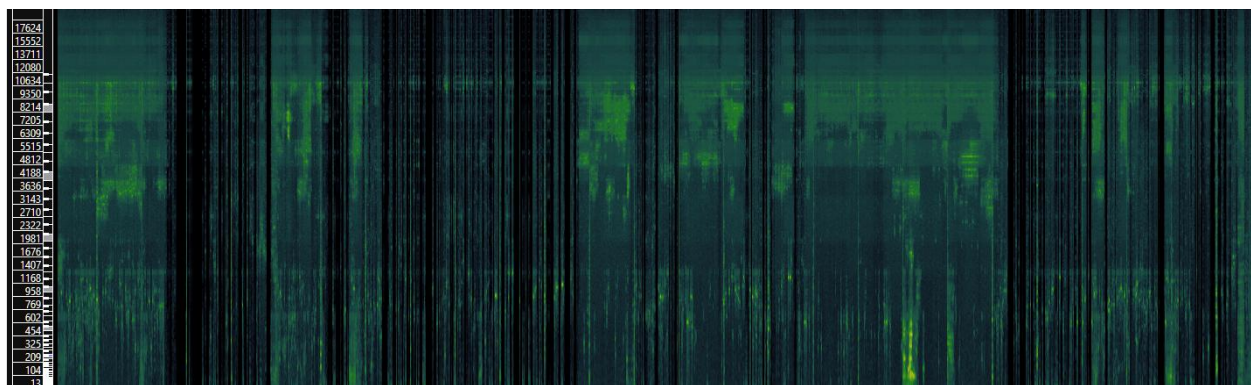
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη είχε διάρκεια περίπου 5 λεπτών και 19 δευτερολέπτων. Ο διαχωρισμός πραγματοποιήθηκε χωρίς να απαιτηθεί περαιτέρω μετατροπή ή επεξεργασία στο πρωτογενές αρχείο. Η διαδικασία εξαγωγής ολοκληρώθηκε ομαλά, χωρίς τεχνικά προβλήματα, και τα πέντε stems αποθηκεύτηκαν στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 3). Στη συνέχεια εισήχθησαν στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση ως προς τη διαρροή και την καθαρότητα του διαχωρισμού.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 21: Waveform of vocal stem (Sonic Visualiser)



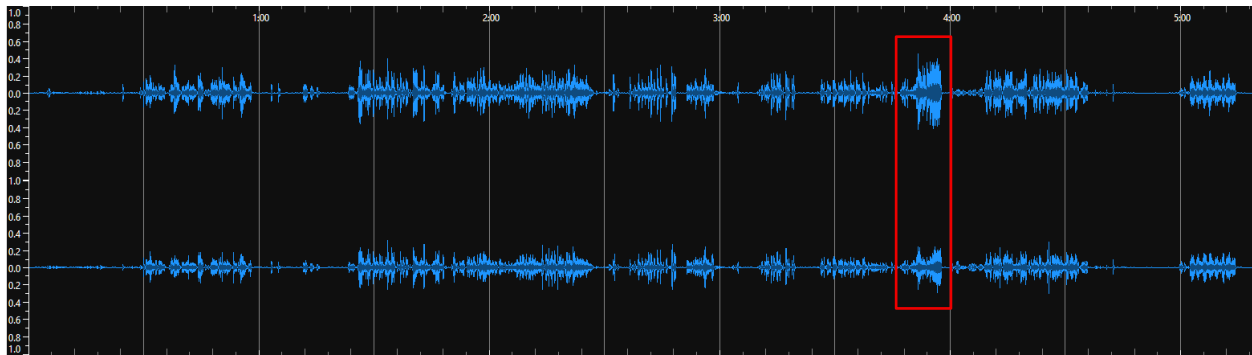
Εικόνα 22: Spectrogram of vocal stem (Sonic Visualiser)

Στο συγκεκριμένο έργο δεν υπάρχουν φωνητικά, συνεπώς το stem των vocals θα έπρεπε να είναι εντελώς κενό. Ωστόσο, ακουστικά σε όλη την διάρκεια του, διακρίνονται πολύ θολά και σιγανά τμήματα σε συνδυασμό με θόρυβο που μας παραπέμπουν σε αποσπάσματα από το πιάνο. Λόγω αυτού, συνεπάγεται σε 1^ο βαθμού διαρροή του piano stem. Επιπλέον έχουμε 2^ο βαθμού καθαρότητα καθώς η διαρροή δεν είναι τόσο έντονη όμως είναι συνεχής και σε χαμηλή σχετικά ένταση.

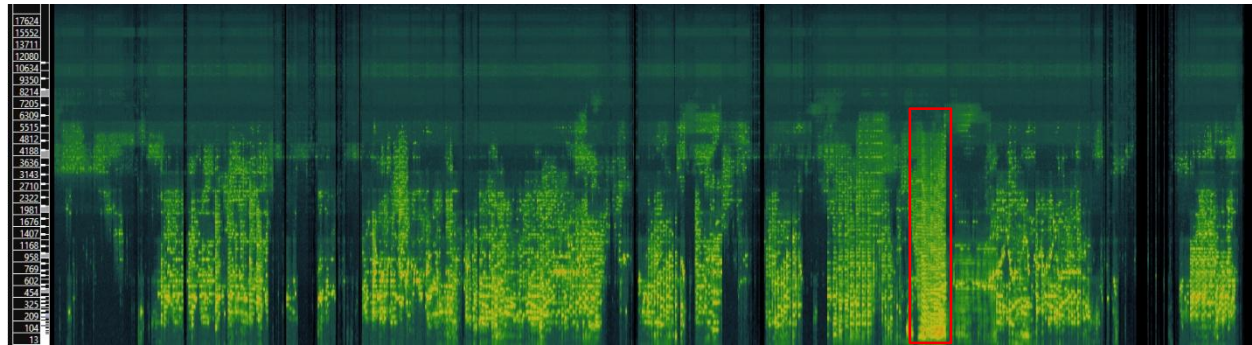
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 23: Waveform of piano stem (Sonic Visualiser)



Εικόνα 24: Spectrogram of piano stem (Sonic Visualiser)

Ξεκινώντας, το πιάνο είναι κυρίαρχο όργανο σε όλη τη διάρκεια του έργου, τόσο σε συνοδευτικό όσο και σε σολιστικό ρόλο. Ακουστικά, πράγματι ανιχνεύεται το πιάνο, αλλά με αισθητό θόρυβο και μειωμένη καθαρότητα. Σε αρκετά σημεία, παρόλο που είναι παρόν στο μεγαλύτερο μέρος του αρχικού έργου, δεν ανιχνεύεται πλήρως και φαίνεται να συγχέεται με το φλάουτο. Πιστεύουμε ότι αυτό συμβαίνει διότι κάθε φορά που ηχεί το φλάουτο μόνο του, το piano stem το περιλαμβάνει στο διαχωρισμό.

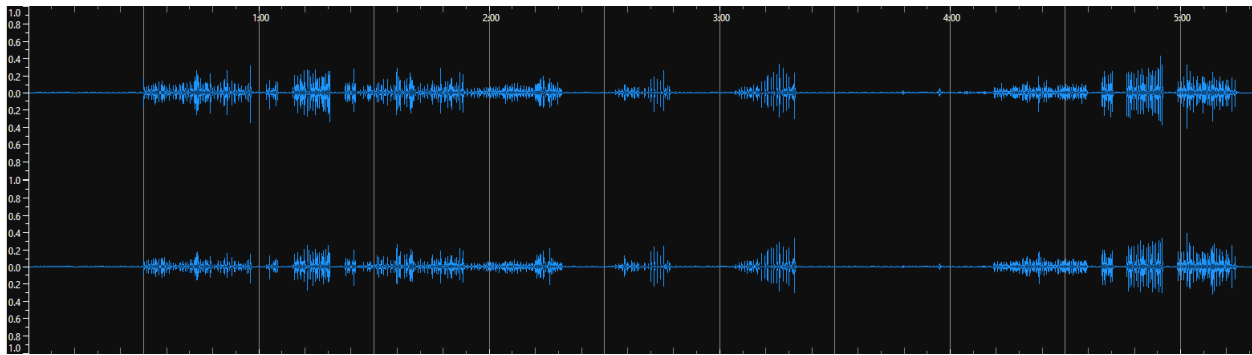
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα		X		
Διαρροή		X		
Π.Δ.	0,5			

Ένα επιπλέον χαρακτηριστικό σημείο εμφανίζεται λίγο πριν από το 4ο λεπτό, όπου το πιάνο παίζει μόνο του ένα μικρό σόλο μέρος. Το μοντέλο θα έπρεπε να το απομονώσει με απόλυτη καθαρότητα, όμως στην πράξη το αποτέλεσμα παραμένει θολό και σε χαμηλή ένταση. Το γεγονός αυτό αποκαλύπτει την αδυναμία του εργαλείου να διαχειριστεί σωστά ακόμη και μία ιδανική συνθήκη απομόνωσης όπως η συγκεκριμένη.

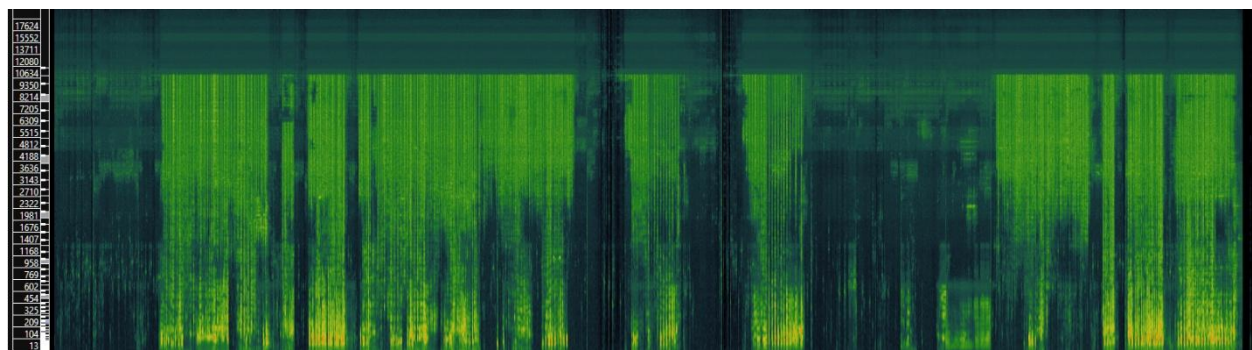
Συνολικά, ενώ το πιάνο δεν ανιχνεύεται ολόκληρο και σε αρκετά σημεία ακούγεται χαμηλό ή πολύ θολό, η διαρροή είναι 1 βαθμού, καθώς ακούγονται και στοιχεία που θυμίζουν το φλάουτο από το stem του other. Παράλληλα, στην κυματομορφή και το φασματογράφημα παρατηρούνται κενά σήματος σε τμήματα όπου το πιάνο θα έπρεπε να ηχεί. Αυτό υποδηλώνει, ότι σε σημεία που είναι πλήρως εκτεθειμένο το όργανο το μοντέλο παρουσιάζει σαφείς αδυναμίες για την πλήρη και σωστή απομόνωσή του. Εδώ καταλήγουμε σε αξιολόγηση της καθαρότητας ως 1^{ου} βαθμού και επιπλέον σε μη επαρκή ανίχνευση της πλήρους πληροφορίας του πιάνου.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 25: Waveform of drums stem (Sonic Visualiser)



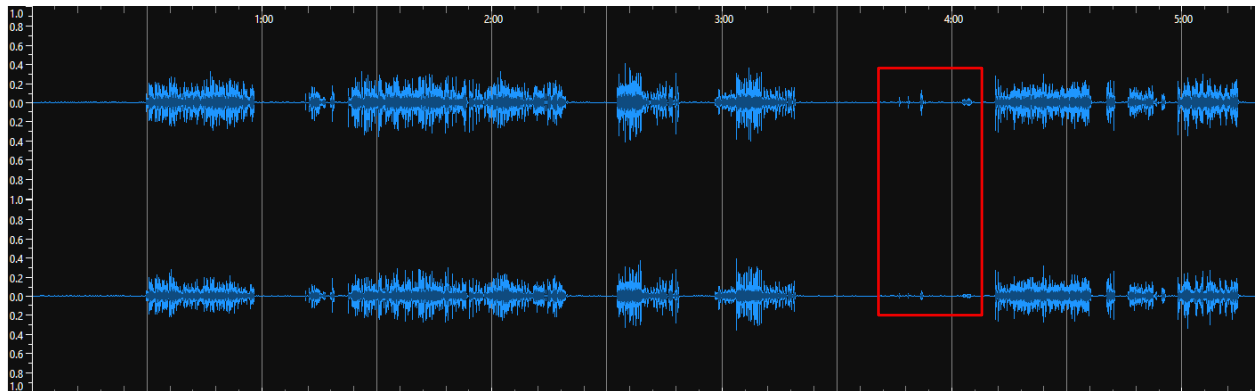
Εικόνα 26: Spectrogram of drums stem (Sonic Visualiser)

Τα ντραμς είναι παρόντα σχεδόν σε όλο το έργο με έντονο ρυθμικό χαρακτήρα. Το stem των drums παρουσιάζει ακουστικά αρκετά καθαρή παρουσία των κρουστών, αλλά με μικρή διαρροή άλλων stem σε χαμηλή ένταση, όπως είναι το bass και piano stem αντίστοιχα. Το γεγονός αυτό οδηγεί σε διαρροή 2^{ου} βαθμού. Όσον αναφορά την καθαρότητα του stem, αυτή αξιολογείται επίσης ως 2^{ου} βαθμού, καθώς τα κρουστά ακούγονται θολά σε ορισμένα σημεία και η διαρροή, παρότι υπαρκτή, δεν είναι παρεμβατική για να κατέβει περισσότερο η καθαρότητα.

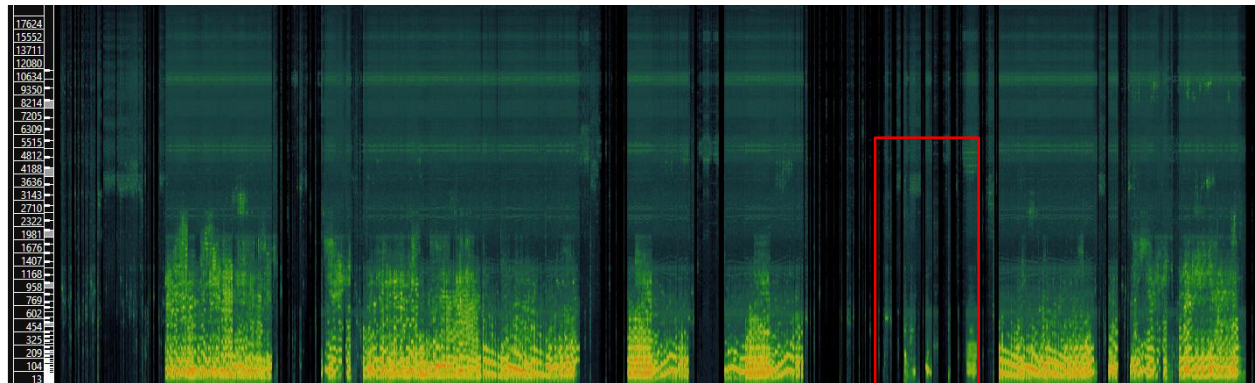
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή			X	
Π.Δ.	0,5			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 27: Waveform of bass stem (Sonic Visualiser)



Εικόνα 28: Spectrogram of bass stem (Sonic Visualiser)

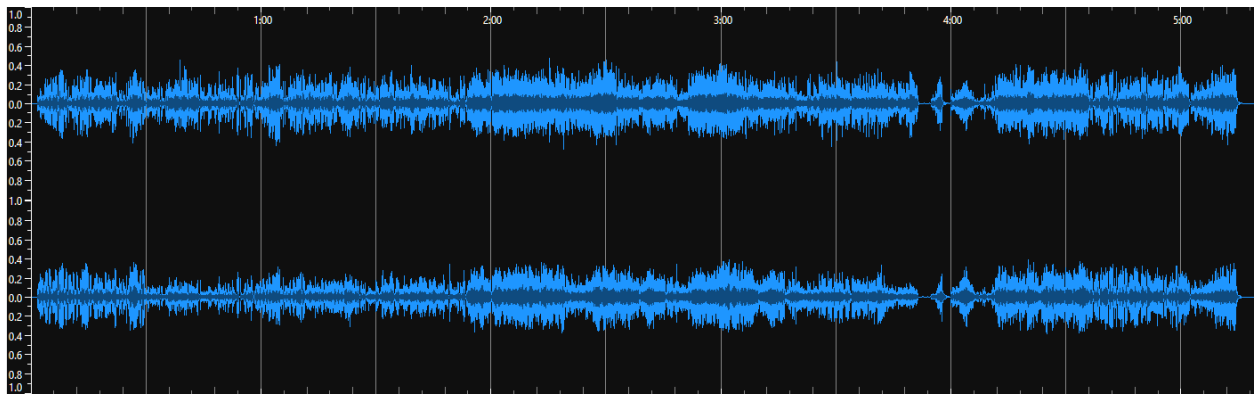
Το κοντραμπάσο αποτελεί βασικό στοιχείο της σύνθεσης, προσφέροντας τη «αρμονική βάση» για τα υπόλοιπα όργανα. Ακουστικά, το bass stem περιλαμβάνει καθαρά το κοντραμπάσο, με ελάχιστο αναμενόμενο θόρυβο σε πολύ χαμηλή ένταση. Ο διαχωρισμός του κρίνεται ιδιαίτερα επιτυχής, καθώς το όργανο έχει απομονωθεί στο σύνολό του. Ο χαρακτηριστικός ήχος του κοντραμπάσου στη τζαζ μουσική φαίνεται να διευκολύνει στη διαδικασία του διαχωρισμού για το μοντέλο.

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

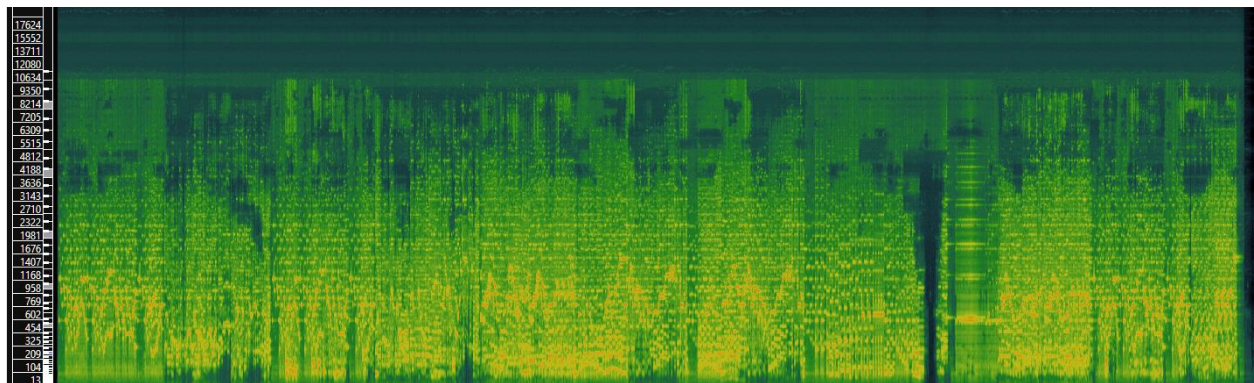
Στο φασματογράφημα, η ενέργεια συγκεντρώνεται κυρίως στις χαμηλές συχνότητες μεταξύ 90 και 150 Hz, εύρος που πράγματι αντιστοιχεί στην αναμενόμενη έκταση και στον χαρακτήρα του stem. (Fletcher & Rossing, 1991) Ωστόσο, σε τμήμα που το μπάσο απουσιάζει (κόκκινο σημείο), παρατηρείται μικρή παρουσία συχνοτήτων, οι οποίες δημιουργούν ελάχιστο θόρυβο χωρίς όμως την παραπομπή σε κάποιο άλλο stem. Με βάση τα παραπάνω, η διαρροή από τα υπόλοιπα stems είναι μηδενική και η καθαρότητα στο υψηλότερο βαθμό ίση με 3. Συνεπώς το bass stem αποτελεί το πρώτο παράδειγμα πλήρους και επιτυχούς διαχωρισμού από το μοντέλο.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 29: Waveform of other stem (Sonic Visualiser)



Εικόνα 30: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other αναμενόταν να απομονωθεί αποκλειστικά το φλάουτο, το οποίο στο συγκεκριμένο έργο έχει σολιστικό ρόλο. Ακουστικά, η παρουσία του φλάουτου είναι έντονη, αλλά μαζί του συμπεριλαμβάνονται σχεδόν και το μεγαλύτερο μέρος από τα υπόλοιπα όργανα, με αποτέλεσμα η διαρροή να εκτιμάται ως 3ου βαθμού. Αξίζει

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα	X			
Διαρροή				X
Π.Δ.	0			

να σημειωθεί ότι υπάρχουν σημεία στα οποία το φλάουτο δεν παίζει και ακούγονται τα υπόλοιπα όργανα. Εκεί, το stem θα έπρεπε να είναι κενό θεωρητικά.

Τέλος, βλέπουμε και ακούμε στην ακρόαση ότι δεν διακόπτεται το stem σε κανένα σημείο και συνεχίζει η παρουσία ήχου από τα υπόλοιπα stems. Αυτό φυσικά το αποτέλεσμα υποδηλώνει ότι ο διαχωρισμός απέτυχε, καθώς το stem θυμίζει μια πιο «θολή» εκδοχή του αρχικού έργου, χωρίς ουσιαστική απομόνωση του φλάουτου. Η συνολική εικόνα οδηγεί στο συμπέρασμα ότι ο διαχωρισμός απέτυχε πλήρως με μέγιστη διαρροή 3ου βαθμού, ενώ η καθαρότητα εκτιμάται στο μηδέν.

Συμπεράσματα Παραδείγματος

Το παράδειγμα 3 κατέδειξε ότι, σε έργα με περιορισμένο αριθμό οργάνων αλλά έντονη αλληλεπίδραση και συχνή εναλλαγή ρόλων, το Spleeter παρουσιάζει εμφανείς δυσκολίες στον διαχωρισμό. Στο stem των vocal, παρότι δεν υπάρχουν φωνητικά, ανιχνεύθηκε διαρροή από το piano stem. Ακουστικά το πιάνο φαίνεται να έχει διαρροή από το other καθώς ακούγεται το φλάουτο. Επιπλέον, παρουσιάζει σημαντικά κενά στο stem του με αποτέλεσμα τον ανεπαρκή διαχωρισμό.

Τα κρουστά είναι κατά το ήμισυ διαχωρισμένα καθώς ακούγονται «θολά» σε αρκετά σημεία και εμφανίζουν διαρροή 2^{ου} βαθμού. Στο other stem, όπου αναμενόταν η απομόνωση αποκλειστικά του φλάουτου, η διαρροή ήταν σχεδόν καθολική και εντοπίζεται και με τους δύο τρόπους αξιολόγησης. Αντίθετα, το stem του bass ήταν το μοναδικό το οποίο παρουσιάζεται καθαρό και είναι κοντά στον στόχο του διαχωρισμού.

Τα παραπάνω αποτελέσματα για το συγκεκριμένο παράδειγμα επιβεβαιώνουν ότι το μοντέλο έχει περιορισμούς στην απομόνωση των stems όταν τα παρουσιάζουν δεξιοτεχνική πολυπλοκότητα, επικάλυψη και συνεχή εναλλαγή ρόλων μεταξύ τους.

4.4 Μουσικό παράδειγμα 4: Dave Brubeck: Take Five

Το τέταρτο μουσικό παράδειγμα που αναλύθηκε είναι το Take Five του Dave Brubeck, ένα εμβληματικό έργο της τζαζ με χαρακτηριστικό ρυθμό 5/4 και όργανα όπως σαξόφωνο, πιάνο, κοντραμπάσο και ντραμς. Το κομμάτι βασίζεται σε ένα επαναλαμβανόμενο ρυθμικό και αρμονικό μοτίβο, το οποίο απαρτίζεται από τα drums, το πιάνο, το μπάσο, ενώ το σαξόφωνο αναλαμβάνει τον

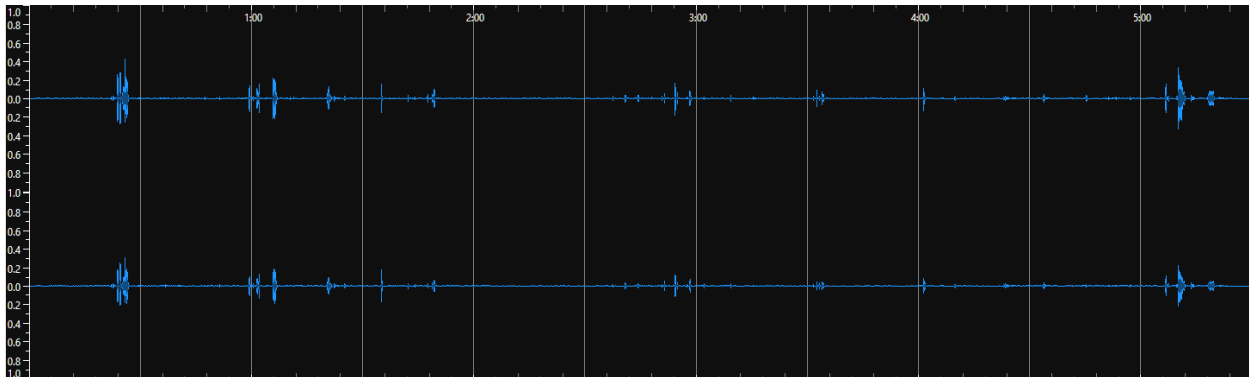
Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Piano	Καθαρότητα		X		
	Διαρροή		X		
	Π.Δ.	0,5			
Drums	Καθαρότητα			X	
	Διαρροή			X	
	Π.Δ.	0,5			
Bass	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Other	Καθαρότητα	X			
	Διαρροή				X
	Π.Δ.	0			

αυτοσχεδιασμό πάνω στην βασική μελωδία. Η αλληλεπίδραση μεταξύ των οργάνων είναι συνεχής, που αναδεικνύουν τη δεξιοτεχνία μεταξύ των εκτελεστών.

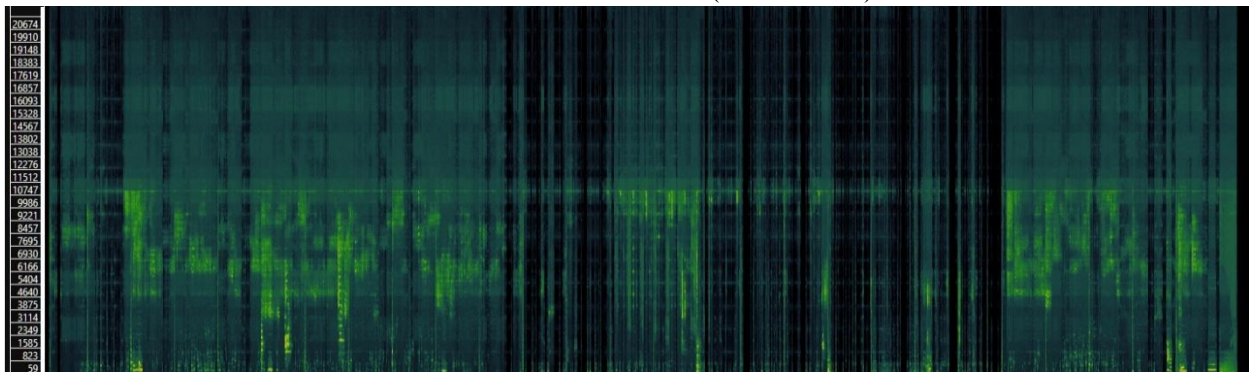
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη είχε διάρκεια περίπου 5 λεπτών και 28 δευτερολέπτων. Ο διαχωρισμός πραγματοποιήθηκε χωρίς να απαιτηθεί περαιτέρω μετατροπή ή επεξεργασία στο πρωτογενές αρχείο. Η διαδικασία εξαγωγής ολοκληρώθηκε ομαλά, χωρίς τεχνικά προβλήματα, και τα πέντε stems αποθηκεύτηκαν στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 4). Στη συνέχεια εισήχθησαν στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση ως προς τη διαρροή και την καθαρότητα του διαχωρισμού.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 31: Waveform of vocal stem (Sonic Visualiser)



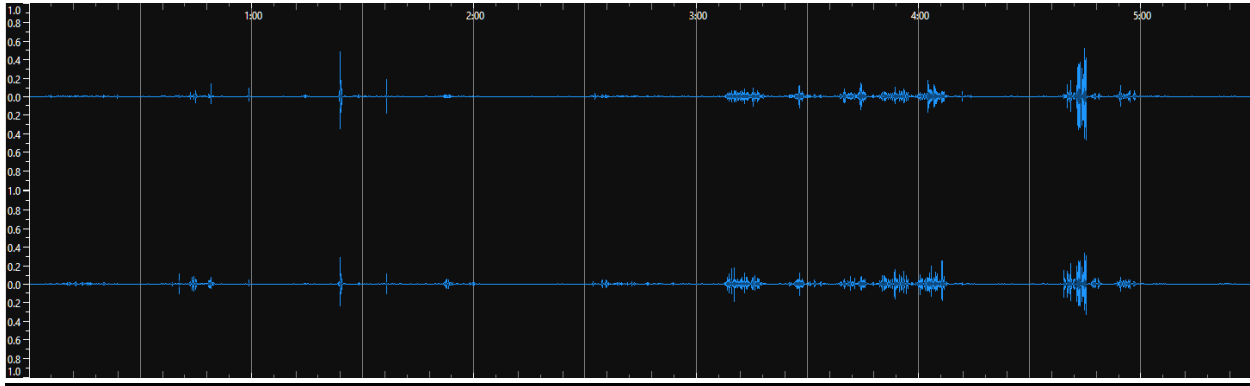
Εικόνα 32: Spectrogram of vocal stem (Sonic Visualiser)

Το κομμάτι δεν περιέχει φωνητικά, οπότε το stem των vocals θα έπρεπε θεωρητικά να είναι κενό. Ωστόσο, η κυματομορφή και το φασματογράφημα μας οδηγούν στο αποτέλεσμα της διαρροής. Ακουστικά, πράγματι ανιχνεύουμε το σαξόφωνο σε αρκετά σημεία αλλά σε χαμηλή σχετικά ένταση. Συνεπώς, οδηγούμαστε σε διαρροή 1ου βαθμού και 2^ο βαθμού καθαρότητα .

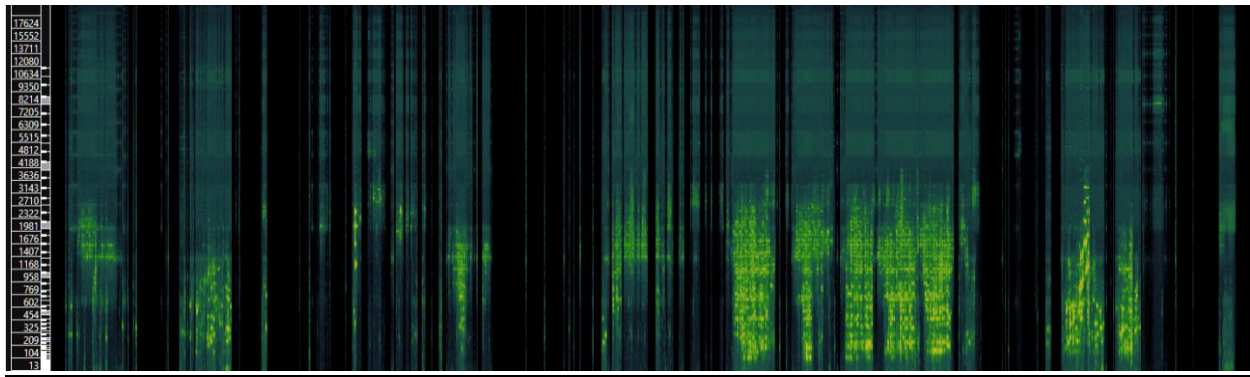
Κλίμακα ΚΚΔ			
Vocals	0	1	2 3
Καθαρότητα			X
Διαρροή		X	
Π.Δ.	0,66		

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 33: Waveform of piano stem (Sonic Visualiser)



Εικόνα 34: Spectrogram of piano stem (Sonic Visualiser)

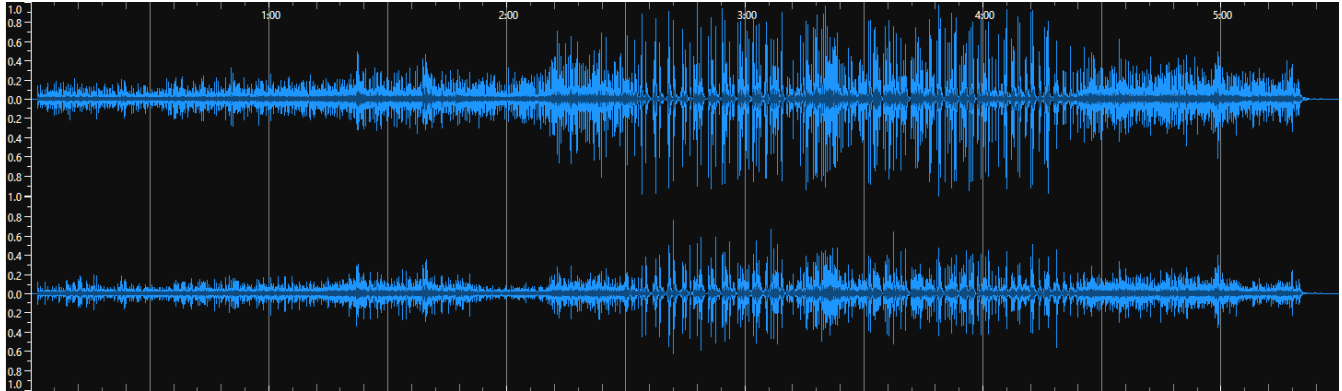
Το πιάνο συνοδεύει σταθερά σε όλη τη διάρκεια του έργου, παρέχοντας αρμονική βάση και ρυθμική υποστήριξη. Ακουστικά, το piano stem είναι ιδιαίτερα θολό και δεν μπορείς να το ξεχωρίσεις. Το παραπάνω καθιστά το συγκεκριμένο stem σε καθαρότητα μηδενικού βαθμού καθώς, η παρουσία του πιάνου είναι ελάχιστη έως ανύπαρκτη. Η ηχητική πληροφορία που απομένει και ακούμε παραπέμπει περισσότερο σε ένα πολύ «θολό» υπόβαθρο παρά ένα ξεκάθαρο και διακριτό αρμονικό μοτίβο συνοδευόμενο από ορισμένα μέρη του σαξοφώνου.

Στην κυματομορφή και το φασματογράφημα παρατηρούνται εκτεταμένα τμήματα με μειωμένη ένταση ή ακόμα και πλήρη απουσία συχνοτήτων, παρότι όπως αναφέραμε στο αρχικό έργο, το πιάνο παίζει ανελλιπώς. Η μη παρουσία σήματος είναι αποτέλεσμα του μη διαχωρισμού του, ειδικά σε περιοχές με έντονη παρουσία του σαξοφώνου, γεγονός που υποδηλώνει περιορισμούς του εργαλείου στην απομόνωση ακόμα και του σταθερού μοτίβου του πιάνου.

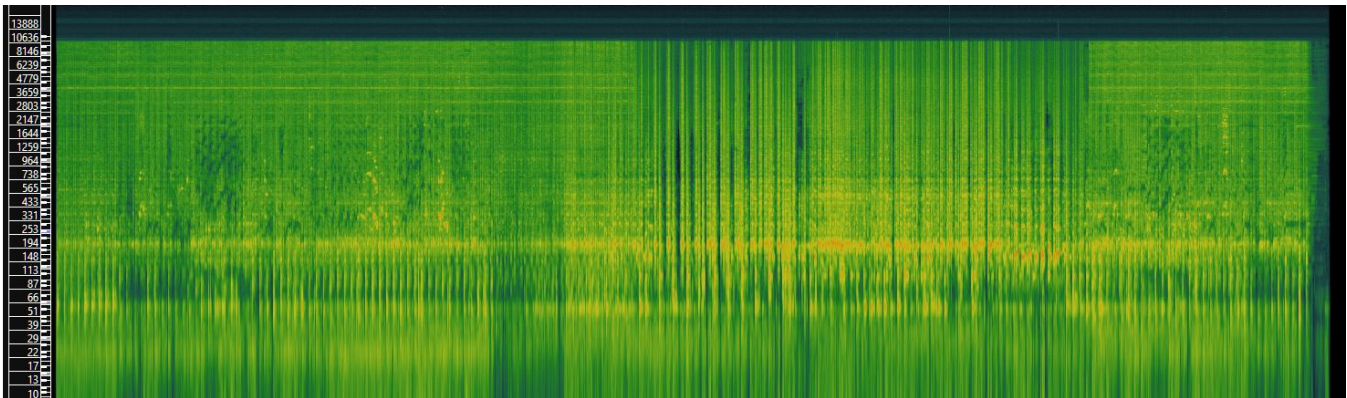
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα	X			
Διαρροή		X		
Π.Δ.	0,33			

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 35: Waveform of drums stem (Sonic Visualiser)



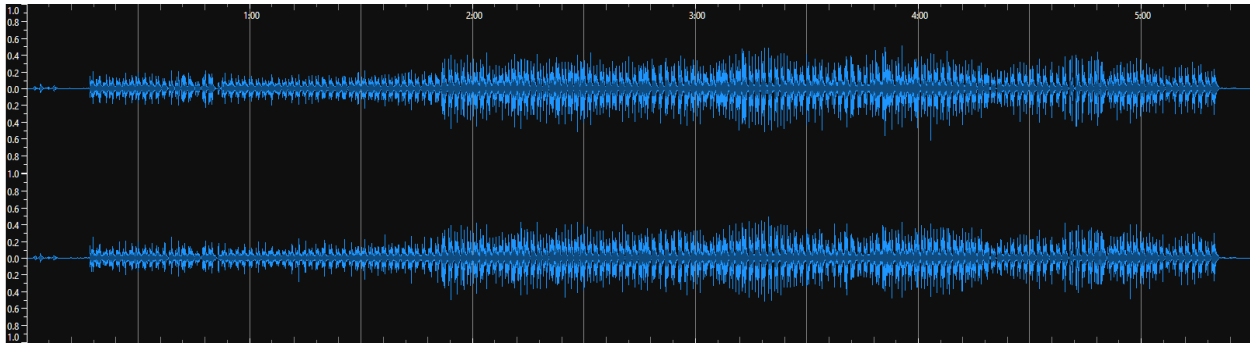
Εικόνα 36: Spectrogram of drums stem (Sonic Visualiser)

Τα ντραμς, με το έντονο ρυθμικό μοτίβο τους στα 5/4, είναι καθοριστικό στοιχείο του έργου. Ενώ αναμέναμε να είναι καθαρός ο διαχωρισμός ακουστικά, το stem των drums παρουσιάζει συνεχόμενη διαρροή 2^{ου} βαθμού από τα piano και other stems. Ακουστικά, τα κρουστά διακρίνονται με μεγάλη σαφήνεια και δυνατά, ωστόσο η διαρροή στο παρασκήνιο από το πιάνο και το σαξόφωνο είναι σταθερή και σε χαμηλή ένταση. Έτσι η καθαρότητα του stem αξιολογείται και αυτή ως 2^{ου} βαθμού. Η κυματομορφή και το φασματογράφημα μας παραπέμπουν σε κρουστά, χωρίς όμως να μπορούμε να αντλήσουμε επιπλέον πληροφορίες, οι οποίες θα μπορούσαν να συμβάλουν περαιτέρω στην αξιολόγησή μας.

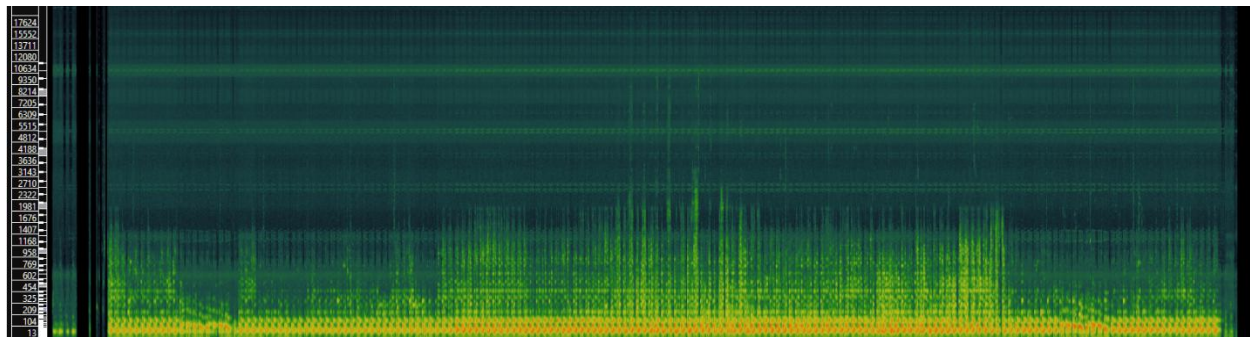
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή			X	
Π.Δ.	0,5			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 37: Waveform of bass stem (Sonic Visualiser)



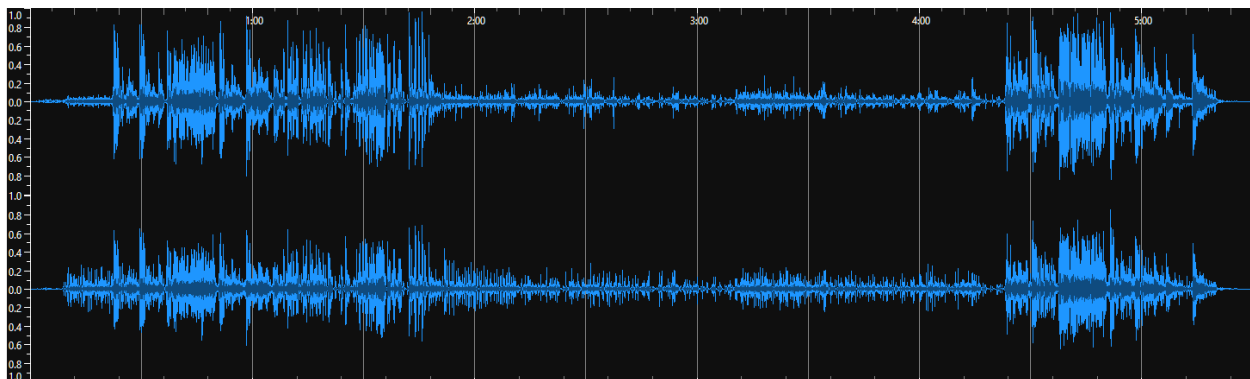
Εικόνα 38: Spectrogram of bass stem (Sonic Visualiser)

Το κοντραμπάσο ακολουθεί σταθερά το ρυθμικό μοτίβο, δημιουργώντας την αρμονική βάση πάνω στην οποία «κινούνται» το σαξόφωνο και το πιάνο. Στο bass stem, ο διαχωρισμός είναι σε μεγάλο βαθμό επιτυχής, με κύρια παρουσία χαμηλών συχνοτήτων της τάξεως κάτω των 104 Hz, όπως προκύπτει από το φασματογράφημα, εύρος το οποίο αντιστοιχεί στον χαρακτήρα του κοντραμπάσου. Σε ορισμένα μόνο σημεία, κυρίως προς στο τέλος, έχουμε μία μικρή διαρροή από το ρίαντο stem, χωρίς όμως αυτή να επηρεάζει ουσιαστικά στην καθαρότητα. Επομένως, το stem αξιολογείται με 3^ο βαθμού καθαρότητα και 1^ο βαθμού διαρροή.

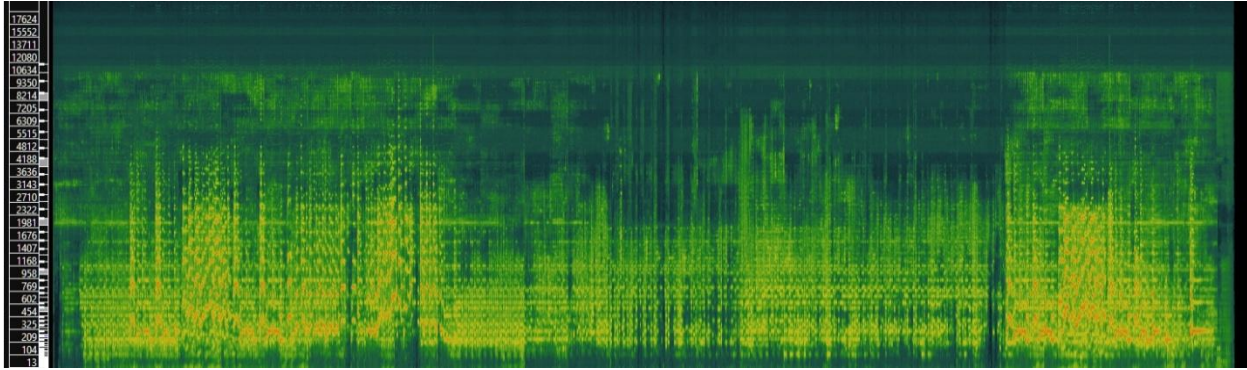
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή		X		
Π.Δ.	0,83			

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 39: Waveform of other stem (Sonic Visualiser)



Εικόνα 40: Spectrogram of other stem (Sonic Visualiser)

Στο other stem αναμενόταν η απομόνωση αποκλειστικά του σαξοφώνου, το οποίο αποτελεί το σολιστικό όργανο του έργου. Αν και το σαξόφωνο είναι ακουστικά κυρίαρχο, η διαρροή από τα υπόλοιπα stems είναι συστηματική, οδηγώντας σε διαρροή 3^{ου} βαθμού. Σε σημεία όπου το σαξόφωνο δεν παίζει, το stem συνεχίζει να περιλαμβάνει ήχους από άλλα stems, γεγονός που αποδεικνύει την πλήρη αποτυχία του διαχωρισμού.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα	X			
Διαρροή				X
Π.Δ.	0			

Συμπεράσματα Παραδείγματος

Το παράδειγμα 4 έδειξε ότι το Spleeter παρουσιάζει δυσκολίες στην απομόνωση των stem σε έργα με έντονη πολυρυθμία και διαρκή συγκοπτόμενο ρυθμό, όπως το Take Five. Στο stem των vocals, παρότι δεν υπάρχουν φωνητικά, ανιχνεύθηκε διαρροή από το other stem καθώς ακούγεται σε λιγοστά σημεία το σαξόφωνο. Το piano stem δεν ανιχνεύτηκε σωστά ενώ η ηχητική πληροφορία που απομένει στο stem παραπέμπει περισσότερο σε ένα πολύ «θολό» υπόβαθρο παρά ένα ξεκάθαρο και διακριτό αρμονικό μοτίβο. Επίσης, παρουσιάζει σημαντικά κενά που υποδηλώνουν τον ανεπαρκή διαχωρισμό.

Τα κρουστά διαχωρίστηκαν μερικώς, παρουσιάζοντας «θολά» σημεία και εμφανίζουν διαρροή 2^{ου} βαθμού. Στο other stem, όπου αναμενόταν η απομόνωση αποκλειστικά του σαξοφώνου, η διαρροή ήταν σχεδόν καθολική και εντοπίστηκε και με τους δύο τρόπους αξιολόγησης. Αντίθετα, το stem του bass αποτέλεσε την πιο επιτυχημένη περίπτωση διαχωρισμού, με σχετικά καθαρή απομόνωση σε σύγκριση με τα υπόλοιπα. Συνολικά, τα

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Piano	Καθαρότητα	X			
	Διαρροή		X		
	Π.Δ.	0,33			
Drums	Καθαρότητα			X	
	Διαρροή			X	
	Π.Δ.	0,5			
Bass	Καθαρότητα				X
	Διαρροή		X		
	Π.Δ.	0,83			
Other	Καθαρότητα	X			
	Διαρροή				X
	Π.Δ.	0			

παραπάνω αποτελέσματα επιβεβαιώνουν τους περιορισμούς του μοντέλου σε έργα όπου υπάρχει έντονη επικάλυψη και σταθερά επαναλαμβανόμενα μοτίβα στα όργανα σε όλη τη διάρκεια του μουσικού έργου.

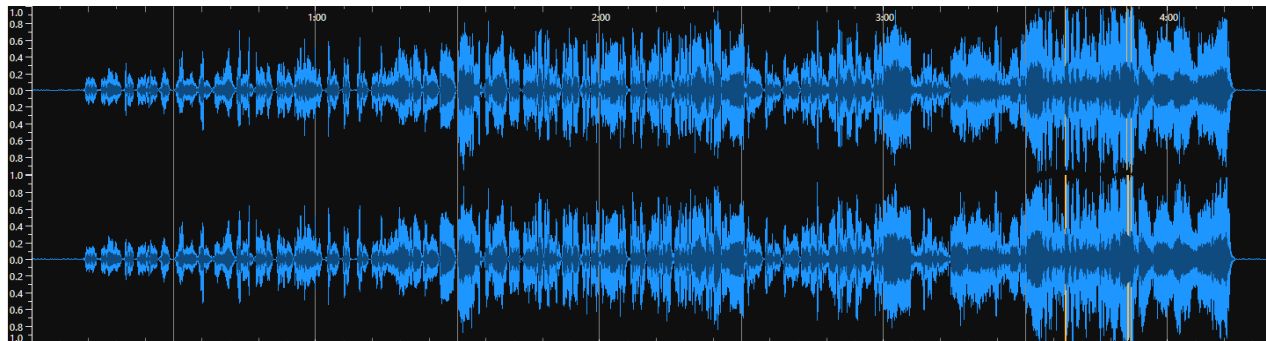
4.5 Μουσικό παράδειγμα 5: Kristen Anderson-Lopez & Robert Lopez: Show Yourself

Το πέμπτο μουσικό παράδειγμα το οποίο αναλύθηκε είναι το Show Yourself των Kristen Anderson-Lopez και Robert Lopez, από το soundtrack της ταινίας Frozen II. Πρόκειται για μια μπαλάντα με έντονα κινηματογραφικό χαρακτήρα, η οποία συνδυάζει τη φωνή από έναν από τους βασικούς χαρακτήρες της ταινίας, πιάνο και ορχηστρική συνοδεία. Το έργο βασίζεται στην εξέλιξη της πορείας της έντασης σε συνδυασμό με την ενορχήστρωση, ξεκινώντας ήρεμα και κορυφώνοντας με πλήρη ορχήστρα και χορωδιακά μέρη.

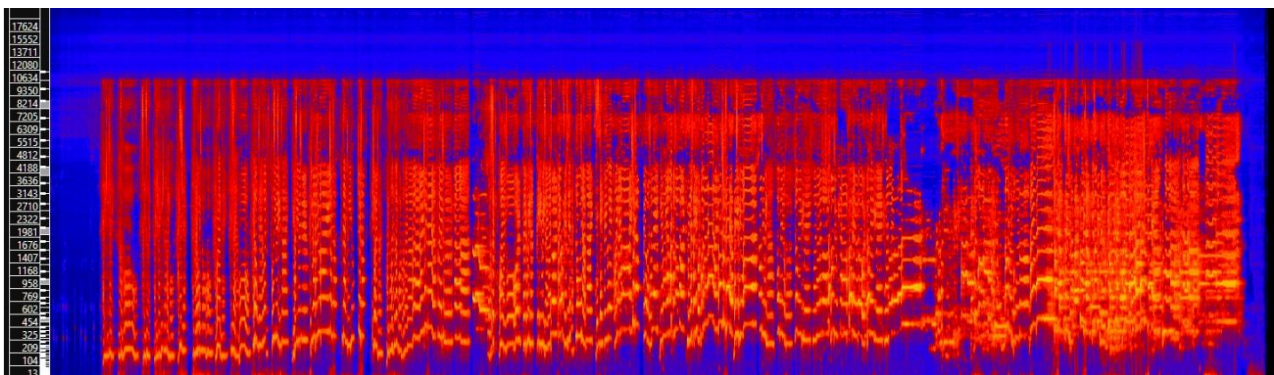
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη είχε διάρκεια περίπου 4 λεπτών και 20 δευτερολέπτων. Ο διαχωρισμός πραγματοποιήθηκε χωρίς περαιτέρω μετατροπή ή επεξεργασία στο πρωτογενές αρχείο. Η διαδικασία εξαγωγής ολοκληρώθηκε χωρίς τεχνικά προβλήματα και τα πέντε stems αποθηκεύτηκαν στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 5). Στη συνέχεια εισήχθησαν στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση ως προς τη διαρροή και την καθαρότητα του διαχωρισμού.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 41: Waveform of vocal stem (Sonic Visualiser)



Εικόνα 42: Spectrogram of vocal stem (Sonic Visualiser)

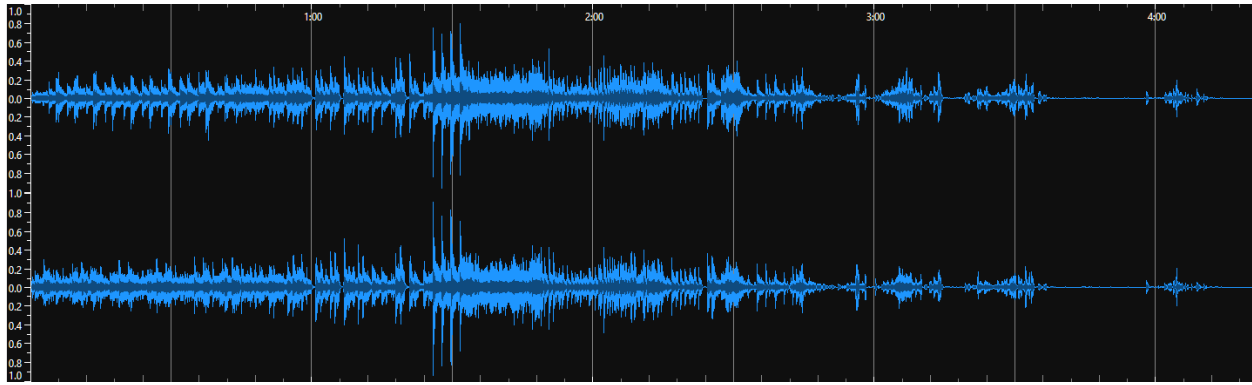
Στο συγκεκριμένο έργο, τα φωνητικά κατέχουν κυρίαρχο ρόλο, τόσο στα σόλο όσο και στα πολυφωνικά μέρη. Λόγω της συνεχούς παρουσίας της φωνής καθ’ όλη σχεδόν τη διάρκεια του έργου (με εξαίρεση τα αρχικά δευτερόλεπτα της εισαγωγής), το διαχωρισμένο stem αναμένεται να είναι ηχητικά πλήρες από τη φωνή μετά το 11^ο δευτερόλεπτο. Στην περίπτωση αυτού του stem, η αξιολόγηση βασίζεται κυρίως στην ακρόαση.

Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

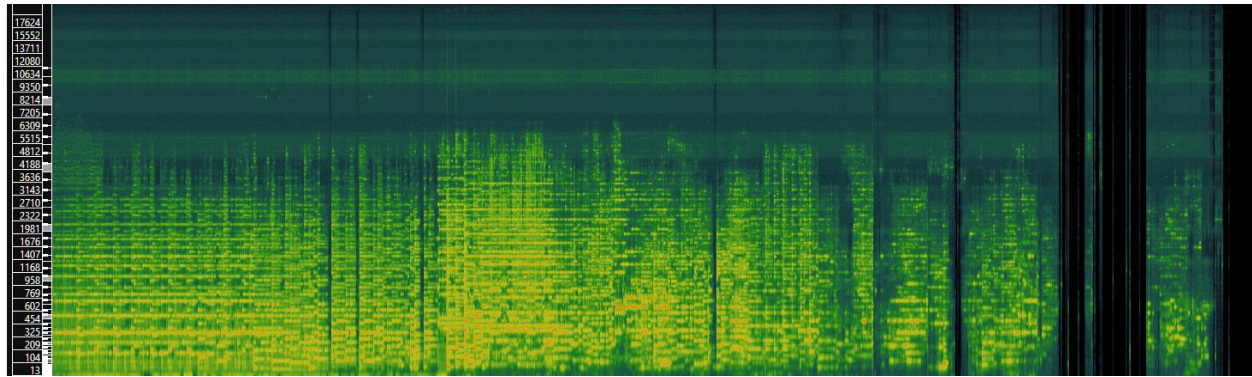
Η ακρόαση έδειξε ότι, ο διαχωρισμός της φωνής επιτεύχθηκε με επιτυχία. Το εύρημα αυτό είναι καθόλα σημαντικό, καθώς φαίνεται για άλλη μία φορά η ικανότητα του μοντέλου να απομονώνει με σχετική ευκολία τη φωνή από το αρχικό mixing. Σε λιγοστά σημεία ακούγεται ο αναμενόμενος θόρυβος ο οποίος προκύπτει από το διαχωρισμό χωρίς ωστόσο να παρατηρείται κάποια ουσιαστική διαρροή από τα υπόλοιπα stems. Αξίζει να αναφέρουμε ότι, εκτός από τη σόλο φωνή το μοντέλο κατάφερε να απομονώσει μαζί και η χορωδία. Αυτό είναι ένα θετικό χαρακτηριστικό το οποίο αποδεικνύει και την εκπαίδευση του Spleeter κυρίως σε μουσικό υλικό το οποίο περιλαμβάνει τραγούδι. Συνεπώς, η διαρροή αξιολογείται ως μηδενικού βαθμού.

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 43: Waveform of piano stem (Sonic Visualiser)



Εικόνα 44: Spectrogram of piano stem (Sonic Visualiser)

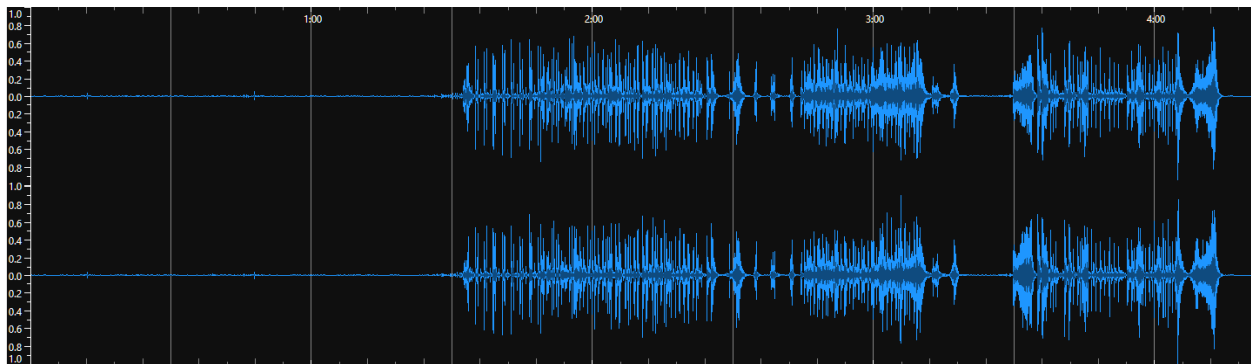
Το πιάνο είναι παρόν σχεδόν σε ολόκληρο το έργο, λειτουργώντας κυρίως ως συνοδευτικό της ορχήστρας και της φωνής. Ακουστικά, το piano stem περιλαμβάνει αποκλειστικά πιάνο, χωρίς εμφανή διαρροή από τα άλλα stem. Σχεδόν ένα λεπτό πριν το τέλος, το πιάνο σε ελάχιστα σημεία ακούγεται ελαφρώς θολό.

Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

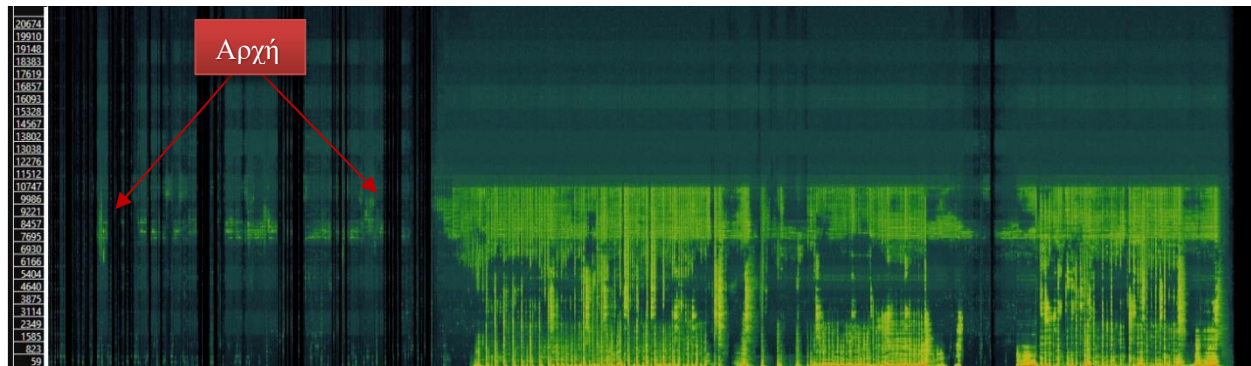
Στο φασματογράφημα εμφανίζονται κενά σε τμήματα όπου το πιάνο ακούγεται καθαρά στο πρωτότυπο έργο, κάτι που υποδεικνύει ότι το μοντέλο δεν ανίχνευσε πλήρως το stem σε όλη τη διάρκεια του. Όμως, αυτά τα σημεία είναι ελάχιστα και απολύτως λογικά, δεδομένου ότι το πιάνο διακατέχει συνοδευτικό ρόλο. Ως αποτέλεσμα των παραπάνω, η διαρροή είναι μηδενικού και η καθαρότητα 2^{ου} βαθμού.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 45: Waveform of drums stem (Sonic Visualiser)



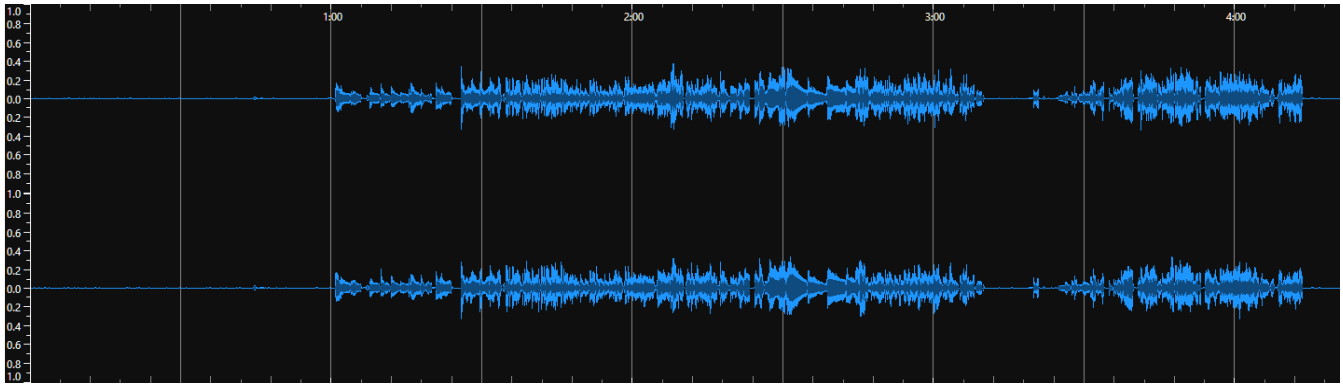
Εικόνα 46: Spectrogram of drums stem (Sonic Visualiser)

Τα κρουστά εμφανίζονται κυρίως στα πιο δυνατά και «tutti» τμήματα του έργου, μετά το 1' και 37'', εξυπηρετώντας ενορχηστρωτικά την ρυθμική συνοδεία και τέλος την κορύφωση. Στο drum stem, ακουστικά η παρουσία τους είναι ευδιάκριτη, αν και σε σημεία θολή, χωρίς να εντοπίζεται διαρροή. Η καθαρότητα αξιολογείται ως 2^{ου} βαθμού.

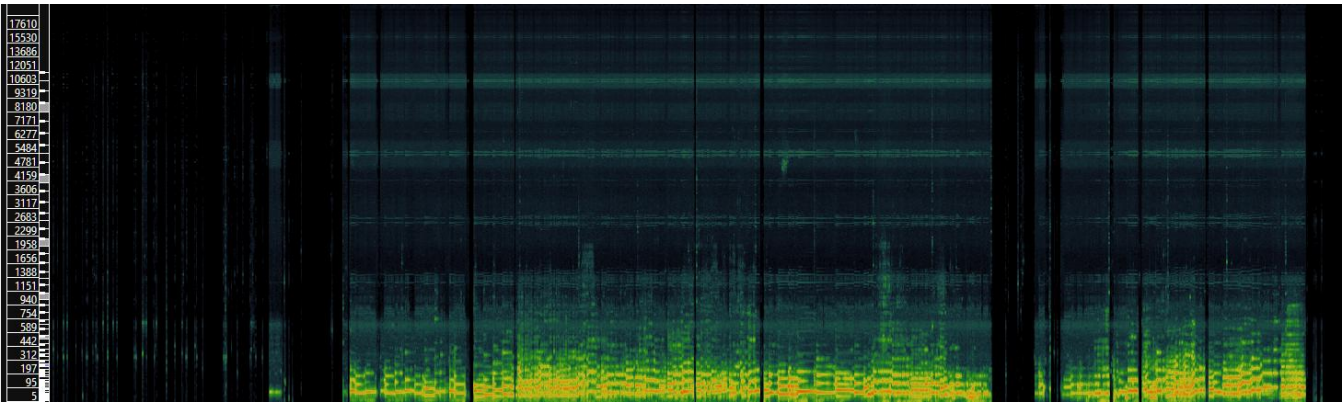
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 47: Waveform of bass stem (Sonic Visualiser)



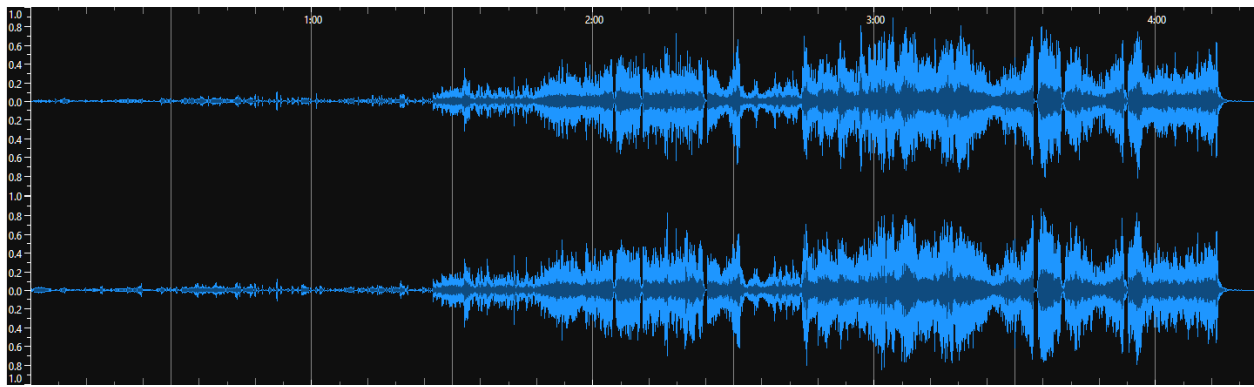
Εικόνα 48: Spectrogram of bass stem (Sonic Visualiser)

Στο συγκεκριμένο stem, το μπάσο (κυρίως ηλεκτρικό αλλά σε σημεία συνδυαστικά με κοντραμπάσο) είναι ευδιάκριτο, αν και ελαφρώς θολό. Δεν εντοπίζεται ουσιαστική διαρροή, εκτός του αναμενόμενου θορύβου ο οποίος προκύπτει από τη διαδικασία του διαχωρισμού. Αυτό συνεπάγεται σε μηδενική διαρροή και 2^ο βαθμού καθαρότητα.

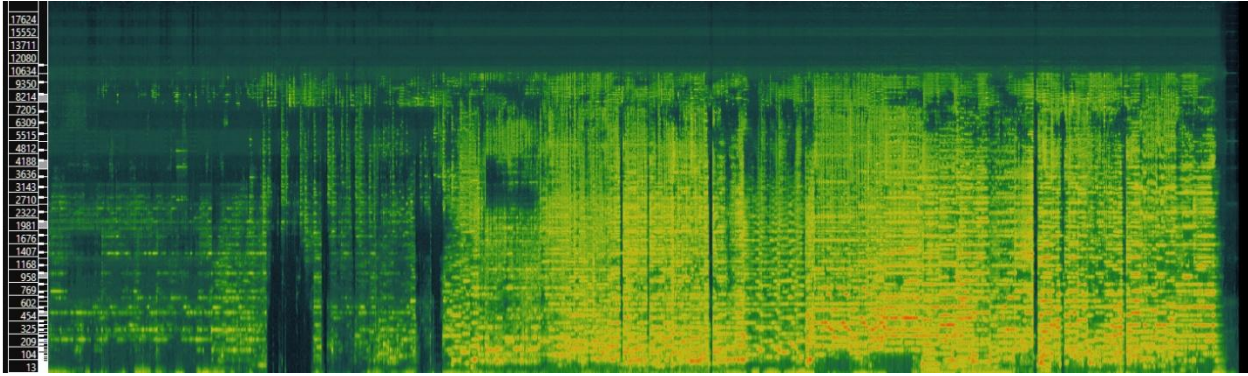
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 49: Waveform of other stem (Sonic Visualiser)



Εικόνα 50: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other δεν θα έπρεπε να περιλαμβάνονται η φωνή, το μπάσο, τα κρουστά και τα drums. Ωστόσο, ακουστικά η διαρροή είναι 2ου βαθμού, καθώς σε διάσπαρτα σημεία ακούγονται, το πιάνο και τα κρουστά. Αξίζει να σημειωθεί ότι παραδόξως δεν εντοπίζεται καθόλου η βασική φωνή, αλλά μόνο τμήμα της χορωδίας, γεγονός που αποκλίνει από το αναμενόμενο.

Κλίμακα ΚΚΑ				
Other	0	1	2	3
Καθαρότητα			X	
Διαρροή			X	
Π.Δ.	0,5			

Στο φασματογράφημα, ήδη από την αρχή του έργου, εμφανίζονται συχνότητες, οι οποίες μετά την ακρόαση που έχει προηγηθεί, μας οδηγούν στην αντιστοιχία τους με τη φωνή και το πιάνο, καθώς αυτά τα δύο ηχούν μόνα τους στο πρώτο λεπτό του έργου πριν την είσοδο της κιθάρας. Θεωρητικά, σε αυτό το σημείο το φασματογράφημα θα έπρεπε να είναι κενό, κάτι που όμως δεν ισχύει. Σε αντίθεση με ότι έχουμε συναντήσει σε άλλα παραδείγματα, το συγκεκριμένο stem δεν παρουσιάζει επικάλυψη με το αρχικό έργο, γεγονός που οδηγεί την καθαρότητα στο 2 και μας επιτρέπει να χαρακτηρίσουμε την ποιότητα του διαχωρισμού ως μεσαίου επιπέδου.

Συμπεράσματα Παραδείγματος

Το παράδειγμα 5 μας επιβεβαίωσε ότι, σε έργα τα οποία βασίζονται πρωτίστως στη φωνή, το Spleeter είναι ικανό να επιτύχει έναν σχετικά πλήρη διαχωρισμό με υψηλό βαθμό καθαρότητας. Το vocal stem διαχωρίστηκε με πλήρη επιτυχία, ενώ και τα stems των piano, drums και bass παρουσίασαν ιδιαίτερα ικανοποιητικά αποτελέσματα, καθώς δεν ανιχνεύτηκε ουσιαστική διαρροή. Ιδιαίτερη σημασία έχει το γεγονός ότι στο other stem, δεν παρουσιάστηκε επικάλυψη με το αρχικό έργο, σε αντίθεση με όσα διαπιστώσαμε σε προηγούμενα παραδείγματα. Το παραπάνω στοιχείο μας βοηθά να αντιληφθούμε ότι το μοντέλο αποδίδει σαφώς καλύτερα σε έργα με κυρίαρχο ρόλο την ίδια την φωνή.

Κλίμακα Διαρροής (ΚΚΑ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Piano	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Drums	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Bass	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Other	Καθαρότητα			X	
	Διαρροή			X	
	Π.Δ.	0,66			

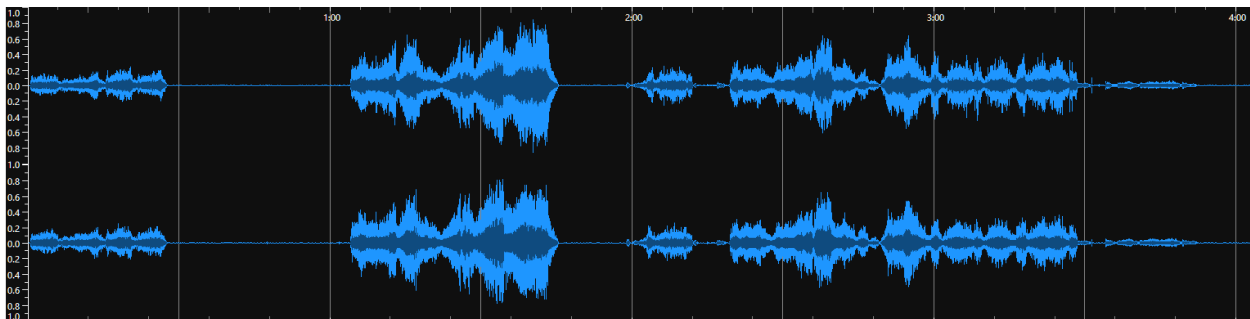
4.6 Μουσικό παράδειγμα 6: Gareth Coker: Morgana's Theme

Το έκτο μουσικό παράδειγμα που αναλύθηκε είναι το Morgana's Theme του Gareth Coker, το οποίο αποτελεί μέρος από το soundtrack του βιντεοπαιχνιδιού «The Mageseeker». Πρόκειται για μια σύνθεση φτιαγμένη για βιντεοπαιχνίδι με έντονα στοιχεία σκοτεινής ατμόσφαιρας, το οποίο έχει σκοπό να περιγράψει έναν σκοτεινό άγγελο με μωβ ενδυμασία. Η φωνή είναι το κυρίαρχο όργανο του έργου σε συνδυασμό με τις χορωδιακές υφές και την εναλλαγή ηχητικών εντάσεων και αρμονιών εξυπηρετείται η δραματική παρουσία καθώς και η περιγραφή του συγκεκριμένου χαρακτήρα. Η ενορχήστρωση περιλαμβάνει τη βασική σοπράνο φωνή, χορωδία, έγχορδα, πνευστά, κρουστά και ηλεκτρονικά στοιχεία.

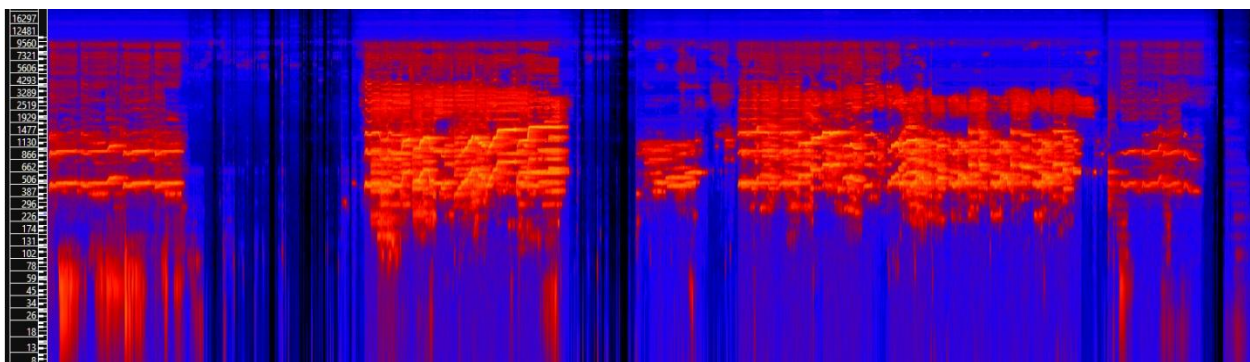
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη είχε διάρκεια περίπου 4 λεπτών και 9 δευτερολέπτων. Ο διαχωρισμός πραγματοποιήθηκε χωρίς να πραγματοποιηθεί κάποια μετατροπή ή πρόσθετη επεξεργασία του πρωτογενούς αρχείου και ολοκληρώθηκε ομαλά, με τα πέντε stems να αποθηκεύονται στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 6). Στη συνέχεια εισήχθησαν στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση ως προς τη διαρροή και την καθαρότητα του διαχωρισμού.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 51: Waveform of vocal stem (Sonic Visualiser)



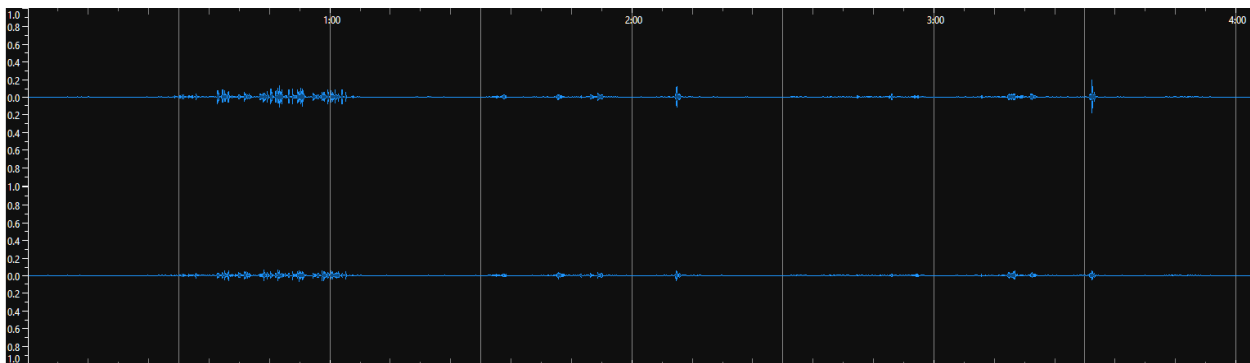
Εικόνα 52: Spectrogram of vocal stem (Sonic Visualiser)

Στο συγκεκριμένο έργο, το stem των vocals περιλαμβάνει τη βασική σοπράνο φωνή καθώς και τη χορωδία, τα οποία είναι τα κεντρικά στοιχεία της σύνθεσης. Ακουστικά, τόσο η γυναικεία σοπράνο φωνή όσο και η παρουσία της χορωδίας είναι καθαρή, με την ποιότητα του διαχωρισμό να είναι στο υψηλότερο επίπεδο ευκρίνειας. Αυτό συνεπάγεται με μηδενικό βαθμό διαρροής γιατί δεν ακούγονται στοιχεία από άλλα stems. Στο spectrogram, παρατηρείται ευρύ φάσμα συχνοτήτων αλλά κυρίως βρίσκονται μεταξύ 261 και 1046 Hz που είναι και το φυσικό εύρος μίας σοπράνο γυναικείας φωνής (Joliveau et al., 2004).

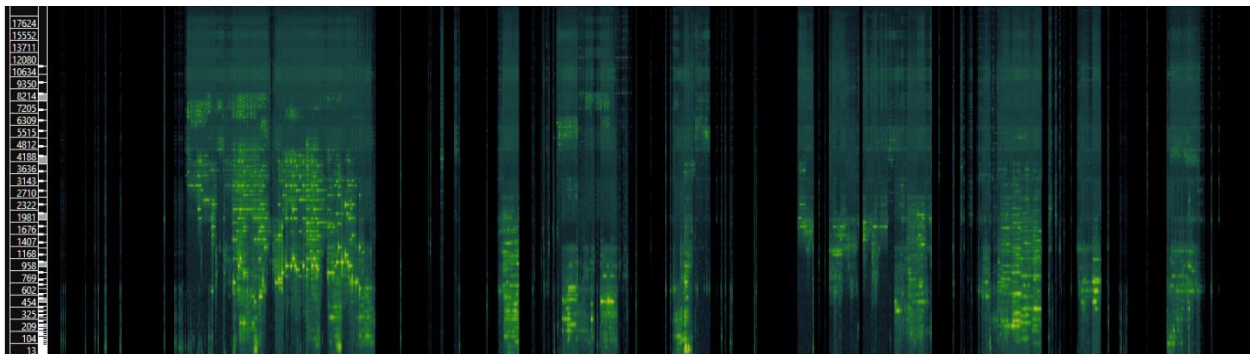
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα				X
Διαρροή		X		
Π.Δ.	1			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 53: Waveform of piano stem (Sonic Visualiser)



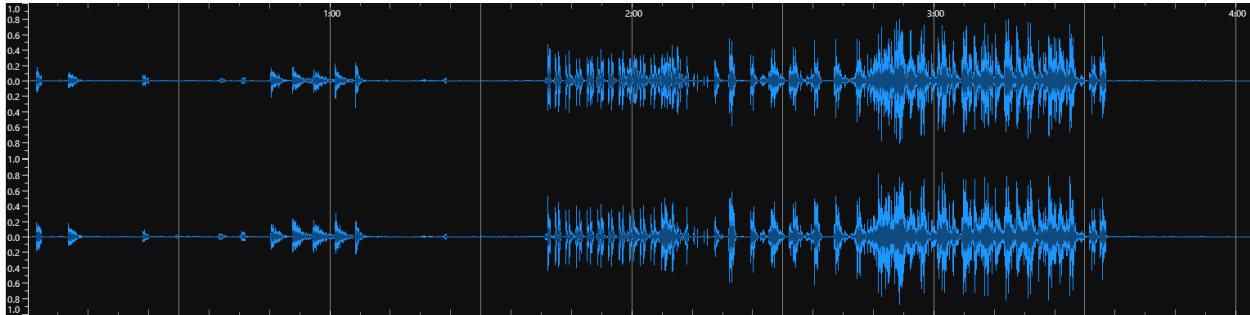
Εικόνα 54: Spectrogram of piano stem (Sonic Visualiser)

Αρχικά, πρέπει να επισημανθεί ότι στο συγκεκριμένο παράδειγμα το πιάνο δεν υπάρχει σαν όργανο μέσα στο έργο. Παρόλα αυτά, τόσο η κυματομορφή όσο και το φασματογράφημα υποδεικνύουν την παρουσία διαρροής στο συγκεκριμένο stem, το οποίο θεωρητικά θα έπρεπε να είναι κενό. Ακουστικά την εντοπίζουμε και είναι απόλυτα εμφανής καθώς εντοπίζονται ευδιάκριτα στοιχεία από το other stem. Για τον λόγο αυτό η διαρροή αξιολογείται ως 1^{ου} βαθμού. Η καθαρότητά παραμένει ως 2^{ου} βαθμού, αφού το stem δεν παραμένει πλήρως κενό, όπως θα αναμενόταν εξαρχής.

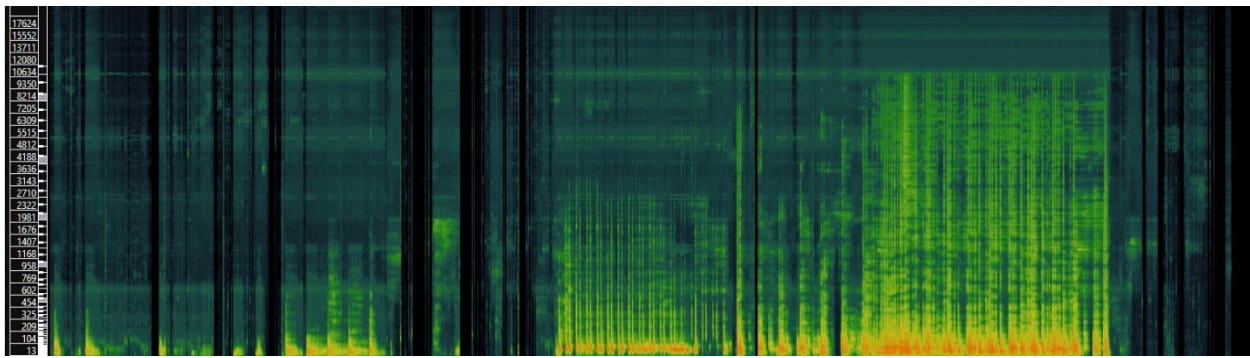
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 55: Waveform of drums stem (Sonic Visualiser)



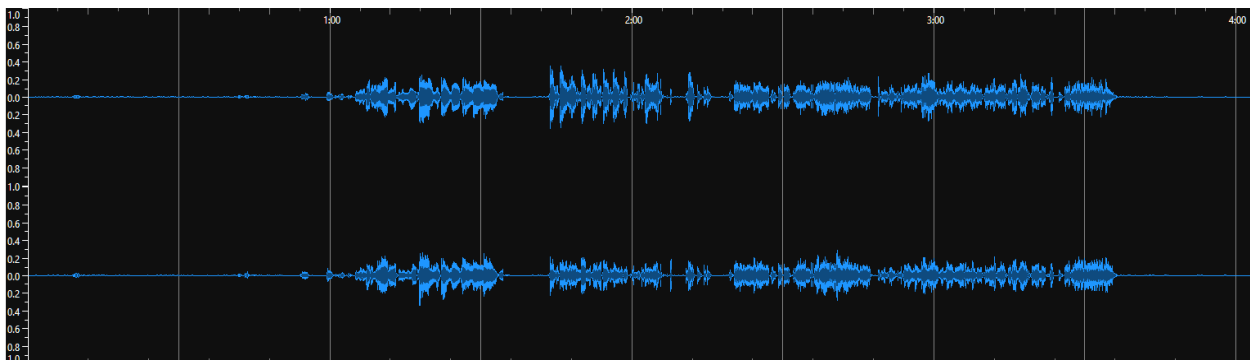
Εικόνα 56: Spectrogram of drums stem (Sonic Visualiser)

Τα κρουστά στο συγκεκριμένο παράδειγμα είναι ποικίλα και ηχούν σε όλη τη διάρκειά του έργου. Ακουστικά, το drum stem έχει διαχωριστεί με ιδιαίτερη επιτυχία, παρουσιάζοντας μηδενική διαρροή, καθώς δεν παρατηρείται να παρεμβαίνουν άλλα stems. Η καθαρότητα παραμένει υψηλή, με το φασματογράφημα να μας δίνει ξεκάθαρα την απεικόνιση που ανταποκρίνεται στον χαρακτήρα του φάσματος των κρουστών, χωρίς όμως να αντλήσουμε κάποια επιπλέον πληροφορία.

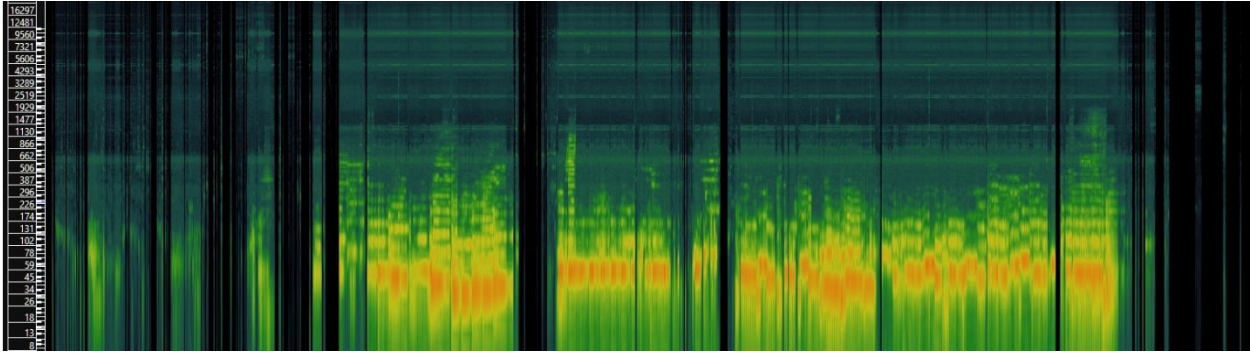
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 57: Waveform of bass stem (Sonic Visualiser)



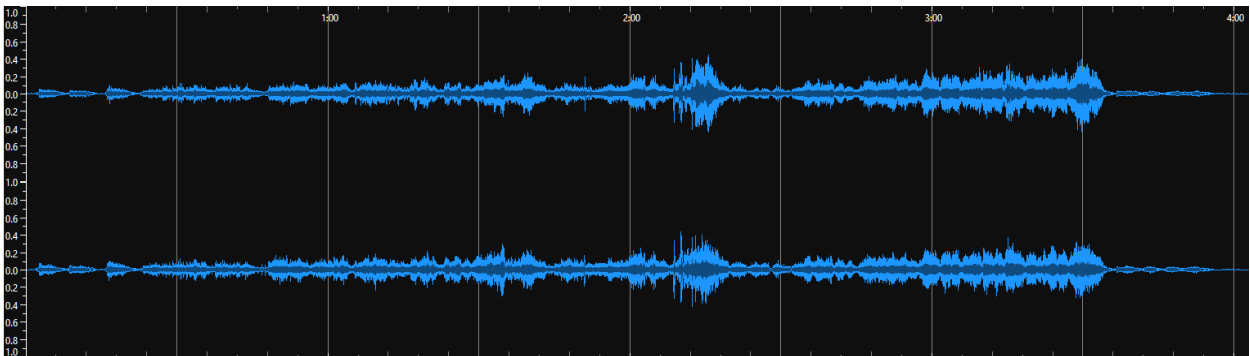
Εικόνα 58: Spectrogram of bass stem (Sonic Visualiser)

Στο συγκεκριμένο stem περιλαμβάνεται ένα κράμα από ακουστικό και ηλεκτρονικό μπάσο που χρησιμοποιούνται στο έργο. Η χαμηλή περιοχή του έργου κυριαρχείται από τα κοντραμπάσα, τα χαμηλά πνευστά και το ηλεκτρονικό μπάσο. Ακουστικά, το stem του bass έχει διαχωριστεί σωστά από τα υπόλοιπα stems με σχεδόν μηδενική διαρροή εκτός του θορύβου. Και σε αυτό το stem, το φασματογράφημα μας δίνει ξεκάθαρα την εικόνα που ανταποκρίνεται στην απεικόνιση των συχνοτήτων του μπάσου, χωρίς όμως να μπορούμε να αντλήσουμε κάποια επιπλέον πληροφορία όσο αναφορά τη διαρροή.

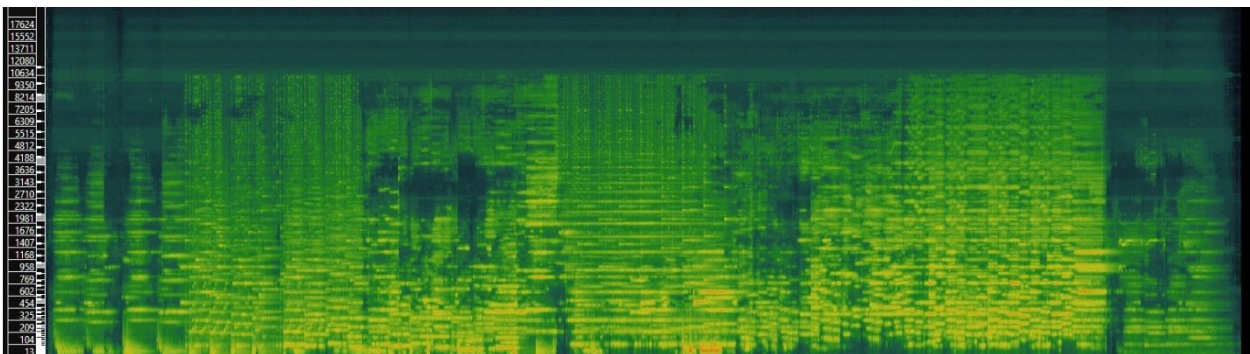
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 59: Waveform of other stem (Sonic Visualiser)



Εικόνα 60: Spectrogram of other stem (Sonic Visualiser)

Στο other stem αναμενόταν η συγκέντρωση των υπόλοιπων μελωδικών στοιχείων, κυρίως των εγχόρδων και ηλεκτρονικών ήχων. Ακουστικά, σε αντίθεση με ό,τι έχουμε παρατηρήσει μέχρι τώρα, δεν περιλαμβάνεται όλο το σύνολο των stems. Ένα πρώτο θετικό στοιχείο είναι η πλήρης απουσία της βασικής φωνής, με εξαίρεση ένα πολύ σύντομο σημείο στο τέλος. Εξίσου θετικό θεωρείται το γεγονός ότι δεν ακούγεται μεγάλο μέρος των κρουστών. Δυστυχώς όμως, παρατηρείται η παρουσία στοιχείων του μπάσο, αν και αυτά βρίσκονται σε πολύ χαμηλή ένταση και δεν είναι ιδιαίτερα αισθητά στα περισσότερα σημεία. Αυτό συνεπάγεται ακουστικά με διαρροή 1^{ου} βαθμού.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Το παραπάνω επαληθεύεται και οπτικά. Ήδη από την αρχή του έργου, εντοπίζονται πιο χαμηλές συχνότητες μεταξύ 11 και 104 Hz, οι οποίες μας παραπέμπουν στο μπάσο. Αυτό σαν στοιχείο είναι σημαντικό, αλλά δεν προκύπτουν επιπλέον ενδείξεις οι οποίες να υποδεικνύουν περαιτέρω παρεμβολές. Τέλος, παρατηρείται ότι το stem του other δεν περιλαμβάνει σχεδόν ολόκληρο το έργο σε «θολή» μορφή, όπως συνέβαινε σε κάποιες από προηγούμενες περιπτώσεις, αλλά περιορίζεται κυρίως στον ρόλο που πρέπει να έχει μετά τη διαδικασία του διαχωρισμού. γεγονός που οδηγεί την καθαρότητα στο 2 και μας επιτρέπει να χαρακτηρίσουμε την ποιότητα του διαχωρισμού ως μεσαίου επιπέδου.

Συμπεράσματα Παραδείγματος

Το παράδειγμα 6 επαλήθευσε αυτό που έχουμε παρατηρήσει στις αναλύσεις που παρουσιάσαμε μέχρι στιγμής. Στα έργα τα οποία βασίζονται στη φωνή, η διαδικασία του διαχωρισμού παρουσιάζει κατά κανόνα υψηλή καθαρότητα. Στην προκειμένη περίπτωση, είναι το πρώτο παράδειγμα στο οποίο το Spleeter κατάφερε με τεράστια επιτυχία να διαχωρίσει τρία από τα πέντε stems (vocals, drums και bass), επιτυγχάνοντας μηδενική διαρροή και καθαρότητα 3^{ου} βαθμού. Αξιοσημείωτο είναι ότι, στο stem του other δεν περιλαμβάνεται σχεδόν ολόκληρο το έργο σε «θολή» εκδοχή, όπως παρατηρήθηκε να συμβαίνει σε κάποιες από τις προηγούμενες περιπτώσεις. Αντιθέτως, προς έκπληξή μας, το stem του piano ήταν αυτό που παρουσίασε 1^{ου} βαθμού διαρροή αλλά πιο εκτεταμένη συγκριτικά με τα υπόλοιπα πέντε.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Piano	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Drums	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Bass	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Other	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			

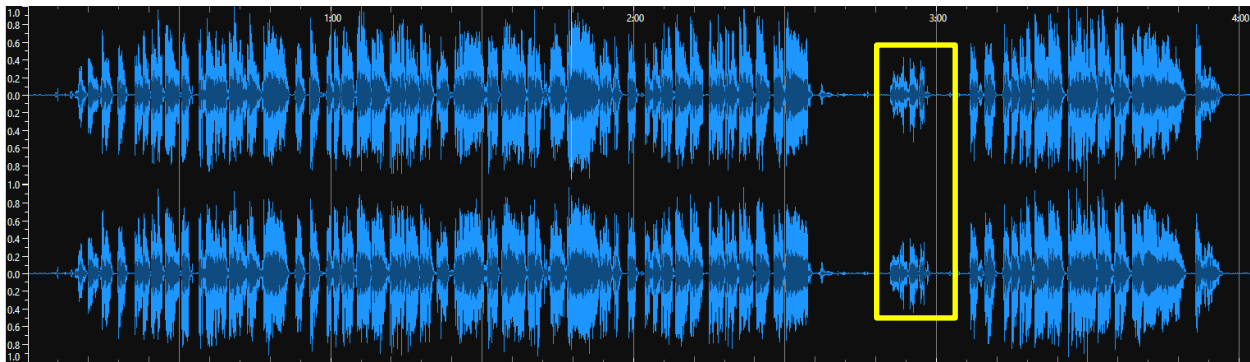
4.7 Μουσικό παράδειγμα 7: Glenn Miller: Moonlight Serenade

Το έβδομο μουσικό παράδειγμα που αναλύθηκε είναι το Moonlight Serenade του Glenn Miller, ένα εμβληματικό κομμάτι της big band swing εποχής. Η χαρακτηριστική του ενορχήστρωση με τον ξεχωριστό ήχο των πνευστών, την απαλή ρυθμική συνοδεία swing συντελούν στη σύνθεση της νοσταλγικής του ατμόσφαιρας. Εμείς σε αυτή την εργασία θα εξετάσουμε την εκδοχή που συμπεριλαμβάνει και τη φωνή και όχι μόνο το αρχικό ορχηστρικό. Η σύνθεση είναι στην ουσία της μια τζαζ μπαλάντα και στηρίζεται στον διάλογο της φωνής με τα πνευστά όπως σαξόφωνο τρομπόνι, τρομπέτα και κλαρινέτο πάνω σε ένα swing ρυθμικό «χαλί» από πιάνο, κοντραμπάσο και ντραμς.

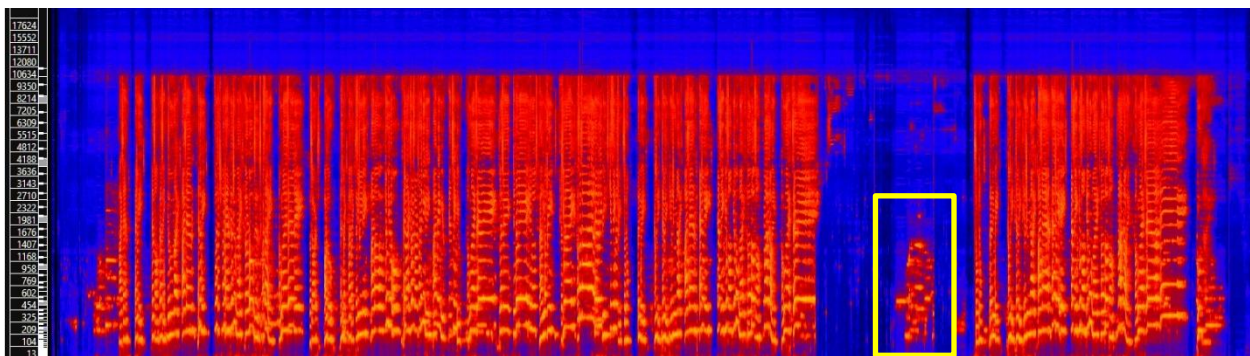
Η εξαγωγή ολοκληρώθηκε χωρίς τεχνικά προβλήματα (επαναλαμβανόμενες προσπάθειες, «crush» του εργαλείου κτλ.), και για τα πέντε παραγόμενα stems καθώς έγινε η αποθήκευσή τους στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 7). Στη συνέχεια εισήχθησαν στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση ως προς τη διαρροή και την καθαρότητα του διαχωρισμού.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 61: Waveform of vocal stem (Sonic Visualiser)



Εικόνα 62: Spectrogram of vocal stem (Sonic Visualiser)

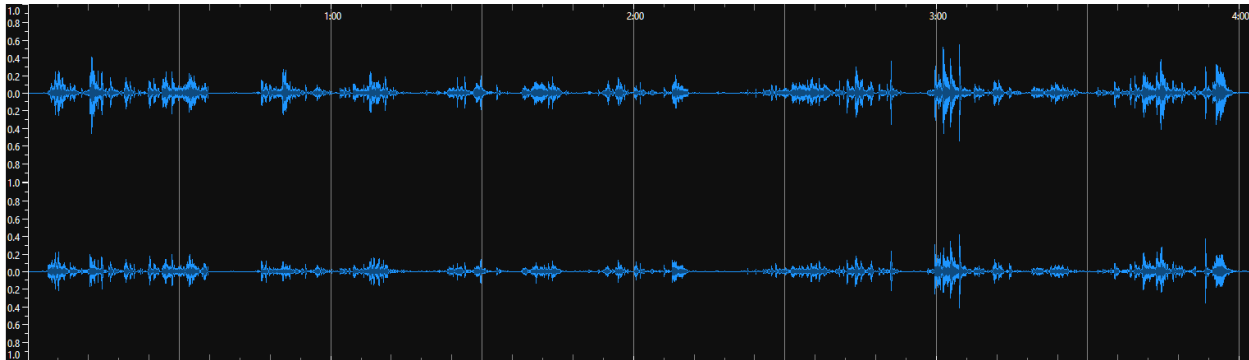
Λόγω της συνεχούς παρουσίας της φωνής σε όλη τη διάρκεια του έργου, το stem των φωνητικών αναμένεται να είναι και αυτό σχεδόν πλήρες. Η κυματομορφή και το φασματογράφημα παρέχουν χρήσιμες ενδείξεις για το τι μπορεί να αναμένει κανείς κατά την ακρόαση και επιτρέπουν την άμεση σύγκριση μεταξύ οπτικής και ακουστικής ανάλυσης.

Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα				X
Διαρροή			X	
Π.Δ.				0,66

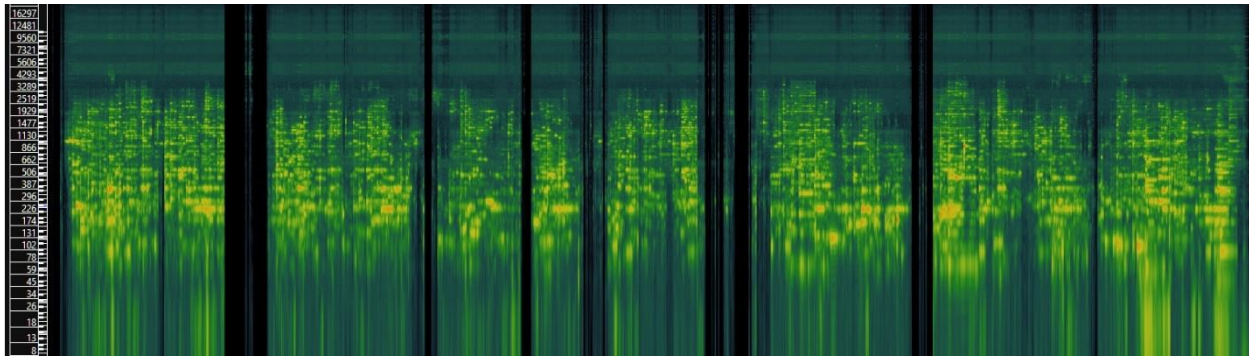
Ένα χαρακτηριστικό και εύκολα εντοπίσιμο στοιχείο είναι η διαρροή πριν το 3^ο λεπτό, όταν η φωνή δεν τραγουδά και ακούγονται τα σόλο μέρη του σαξοφώνου και του κλαρινέτου. Σε αυτό το σημείο, στο φασματογράφημα καταγράφονται συχνότητες, τις οποίες, μέσω της ακουστικής αξιολόγησής μας και της χρονικής τους τοποθέτησης, τις ταυτίζουμε με το σόλο μέρος του κλαρινέτου. Παρά την παραπάνω διαρροή, η συνολική απομόνωση των φωνητικών μπορεί να χαρακτηριστεί ιδιαίτερα επιτυχής. Η φωνή αποδίδεται καθαρά με βαθμό καθαρότητας 3, γεγονός που επιβεβαιώνει για άλλη μία φορά την αποτελεσματικότητα και συνέπεια του μοντέλου στον διαχωρισμό της φωνής. Ωστόσο, εντοπίζεται διαρροή από τα stems των drums και other η οποία είναι 2^{ου} βαθμού, και οδηγεί στη συνολική τιμή της ποιότητας του διαχωρισμού 0,5.

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 63: Waveform of piano stem (Sonic Visualiser)



Εικόνα 64: Spectrogram of piano stem (Sonic Visualiser)

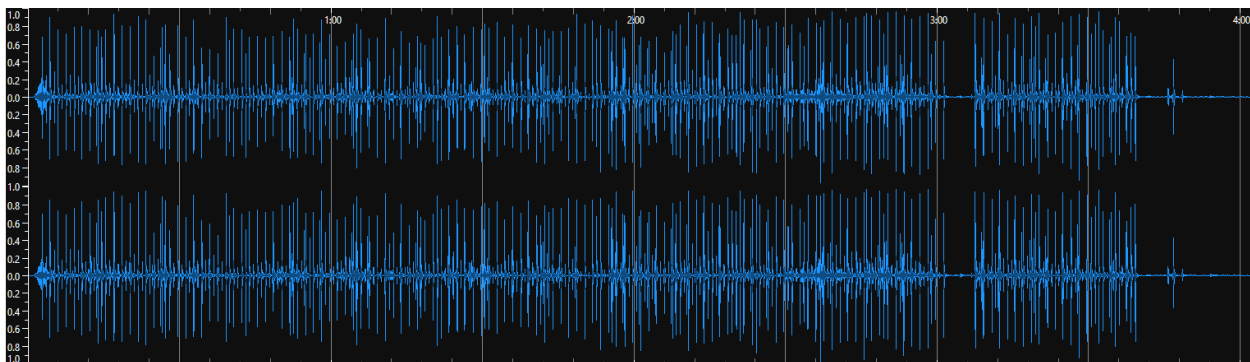
Το πιάνο στο συγκεκριμένο παράδειγμα ηγεί σταθερά, σχεδόν σε όλη τη διάρκεια του έργου. Ο ρόλος του είναι καθαρά συνοδευτικός και το ίδιο παρέχει αρμονική και ρυθμική υποστήριξη σε ρυθμό swing. Ακουστικά, το stem του piano είναι ελάχιστο ως και ανύπαρκτο καθώς καθίσταται ιδιαίτερα θολό και δεν μπορείς να το ξεχωρίσεις σχεδόν καθόλου. Το παραπάνω συνεπάγεται σε καθαρότητα μηδενικού βαθμού καθώς, δεν καθίσταται διακριτή η παρουσία του πιάνου, ενώ η ηχητική πληροφορία που απομένει και ακούμε θυμίζει περισσότερο ένα πολύ «θολό» υπόβαθρο παρά ένα ξεκάθαρο αρμονικά και ρυθμικά τζαζ πιάνο με επιπλέον στοιχεία διαρροής από τα πνευστά όργανα του other stem και των κρουστών.

Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα	X			
Διαρροή		X		
Π.Δ.	0,33			

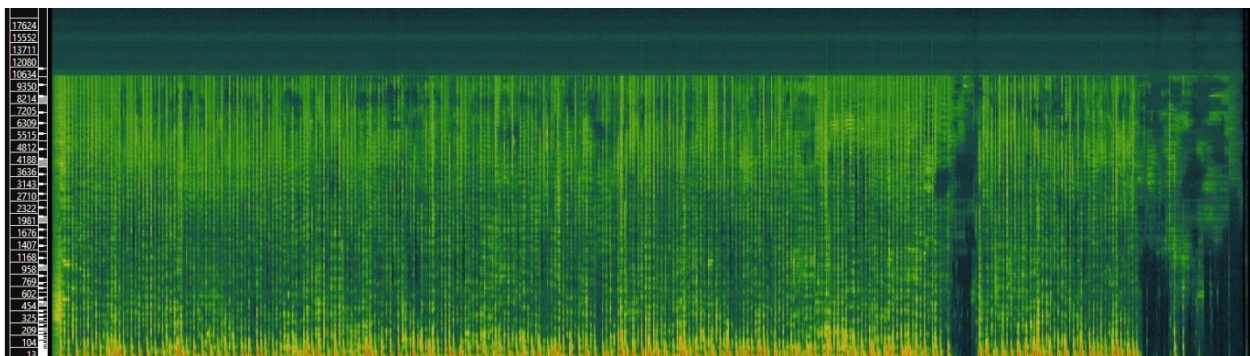
Στην κυματομορφή και το φασματογράφημα παρατηρούνται τμήματα με μερική ή και πλήρη απουσία συχνοτήτων, ακόμα και σε σημεία όπου το πιάνο παίζει ανελλιπώς στο αρχικό έργο. Η μη παρουσία σήματος είναι αποτέλεσμα της μη αναγνώρισής του από το μοντέλο και κατά επέκταση του μη διαχωρισμού του, ειδικά σε περιοχές με έντονη παρουσία της φωνής και των κρουστών. Συνεπώς, δεν μπορούμε να εξάγουμε ασφαλές συμπέρασμα οπτικά μέσω της παρατήρησης χωρίς να βασιστούμε εξολοκλήρου στην ακρόαση. Ωστόσο, μπορεί να διατυπωθεί η υπόθεση ότι από τη στιγμή που οι συχνότητες που διαχωρίζει το μοντέλο εγκύπτουν στην έκταση του συγκριμένου οργάνου, αναγνωρίζει μερικώς την παρουσία του, χωρίς όμως να επιτυγχάνει την ουσιαστική απομόνωσή του.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 65: Waveform of drums stem (Sonic Visualiser)



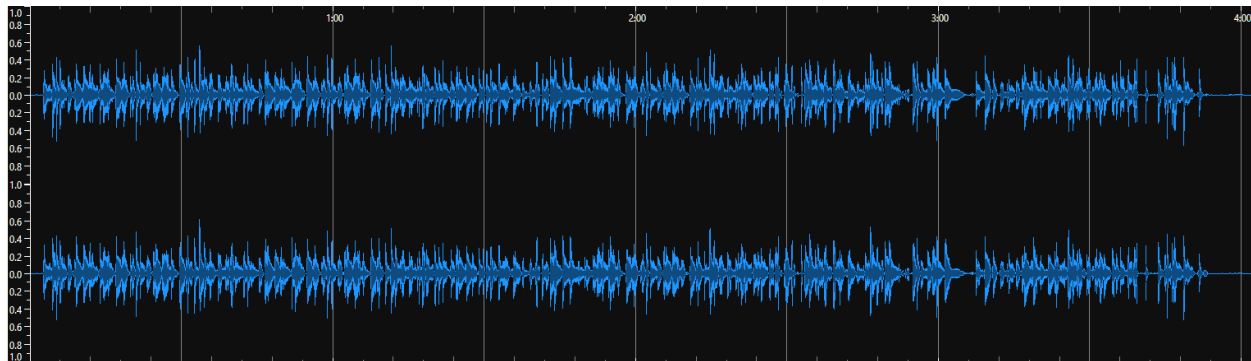
Εικόνα 66: Spectrogram of drums stem (Sonic Visualiser)

Τα ντραμς είναι παρόντα σε όλο το έργο με έντονο ρυθμικό χαρακτήρα swing. Ακουστικά, το stem των drums παρουσιάζει καθαρότητα ίση με 2 χωρίς να εντοπίζεται εμφανή διαρροή από άλλα stems. Υπάρχει, ωστόσο, χαμηλής έντασης θόρυβος που θολώνει ελαφρώς το stem σε μερικά σημεία, χωρίς όμως να μειώνει τη συνολική καθαρότητα κάτω από 2.

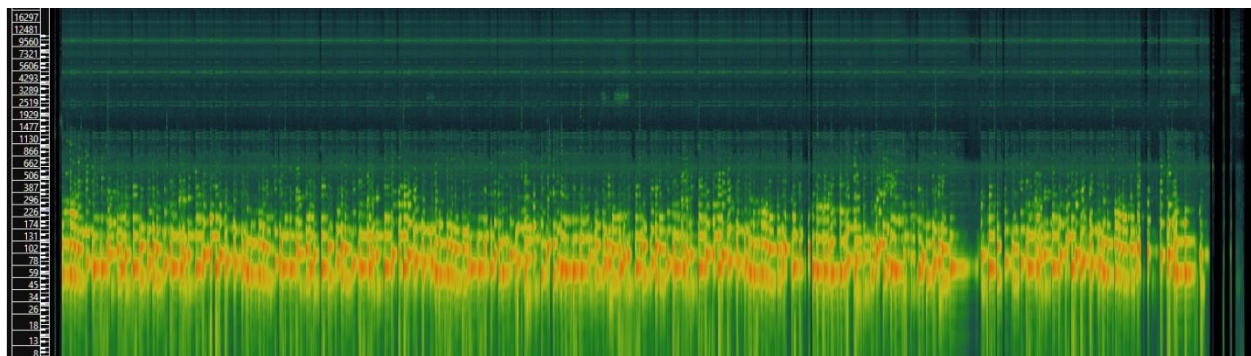
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 67: Waveform of bass stem (Sonic Visualiser)



Εικόνα 68: Spectrogram of bass stem (Sonic Visualiser)

Στην περίπτωση με του bass stem αναμένουμε να ακούσουμε τις πολύ χαμηλές συχνότητες των εγχόρδων, οι οποίες επιτελούν τη λειτουργία του μπάσου. Αξίζει να σημειωθεί ότι το μπάσο στη τζαζ και ειδικότερα το ακουστικό μπάσο, είναι από τα πλέον απαραίτητα για τη διαμόρφωση του ήχου αυτού του είδους μουσικής.

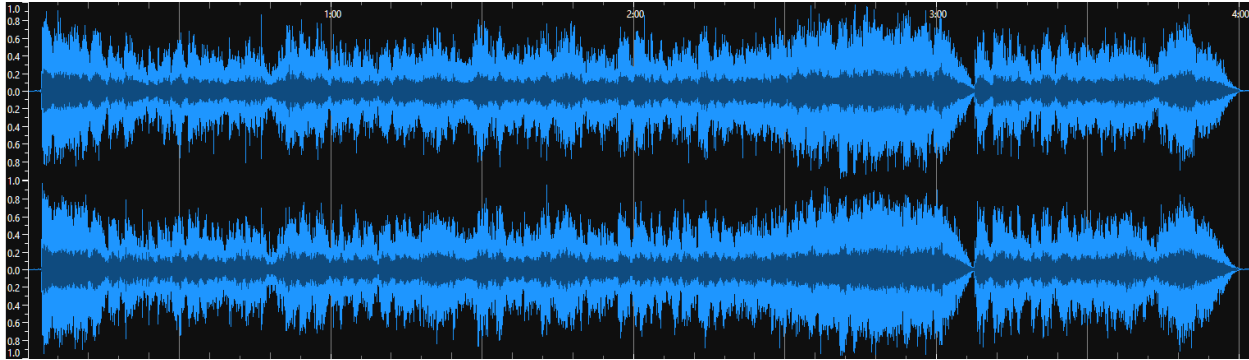
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Κατά την ακουστική αξιολόγηση λοιπόν, θα μπορούσε να ειπωθεί ότι πράγματι η απομόνωση είναι αρχικά χωρίς διαρροές από τα υπόλοιπα stems. Ο ήχος αποδίδεται ως ένας καθαρός και σαφής συνδυασμός χαμηλών συχνοτήτων, οι οποίες μας παραπέμπουν άμεσα το κοντραμπάσο των εγχόρδων, ενώ η παρεμβολή άλλων οργάνων είναι αμελητέα. Να σημειωθεί ότι παρόλο που ακούμε το συγκεκριμένο stem σε σχετικά χαμηλή ένταση ενώ είναι συνοδευόμενο από τον αναμενόμενο θόρυβο, ο οποίος προκύπτει λόγω της απομόνωσης, σε γενικές γραμμές ο διαχωρισμός μπορεί να χαρακτηριστεί επιτυχής. Η διαρροή αξιολογείται ως μηδενικού βαθμού, καθώς δεν είναι καθόλου αντιληπτά ψήγματα από τα

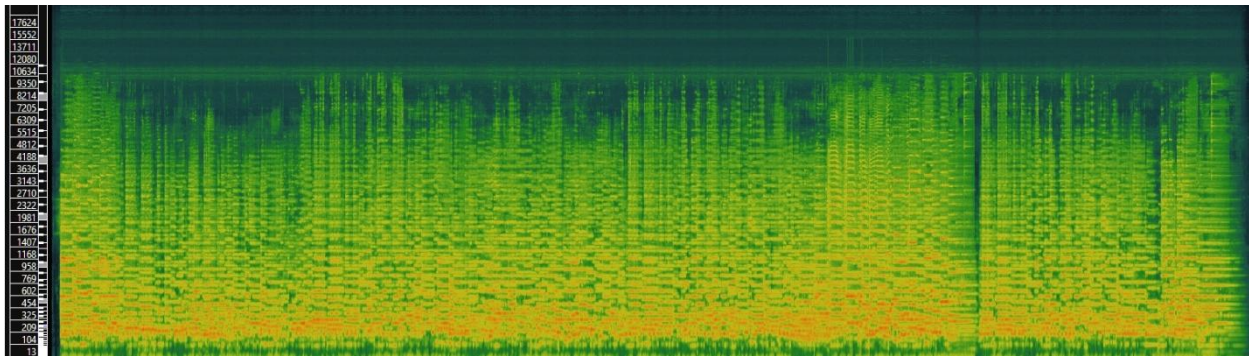
υπόλοιπα stems. Η παραπάνω εκτίμηση επιβεβαιώνεται και οπτικά μέσω του Sonic Visualiser. Οι θεμελιώδεις συχνότητες που καταγράφονται κυμαίνονται μεταξύ 45 και 131 Hz, εύρος που αντιστοιχεί ξεκάθαρα σε μπάσα όργανα. (Fletcher & Rossing, 1991)

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 69: Waveform of other stem (Sonic Visualiser)



Εικόνα 70: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other αναμενόταν να συγκεντρωθούν τα πνευστά (σαξόφωνα, τρομπόνα, τρομπέτες) και τα έγχορδα. Θεωρητικά, το περιεχόμενό του θα έπρεπε να αποτελεί το «υπόλοιπο» από την απομόνωση των υπολοίπων stems. Ακουστικά παρατηρείται διαρροή. Στο stem δεν περιλαμβάνεται καθόλου η φωνή, το οποίο από μόνο του είναι πολύ θετικό, αλλά εντοπίζεται μεγάλο μέρος της ρυθμικής συνοδείας των ντραμς, του μπάσου καθώς και αποσπάσματα από τα μέρη του πιάνου. Αυτό έχει ως αποτέλεσμα η διαρροή να αξιολογείται ως 2^ο βαθμού και όχι 3^ο, ακριβώς επειδή δεν έχουμε πλήρη διαρροή λόγω της απουσίας της φωνής. Η καθαρότητα κρίνεται και αυτή ως 2^ο βαθμού, καθώς ακούμε οτιδήποτε αναμένεται να συμπεριληφθεί στο συγκεκριμένο stem, αλλά με τη διαρροή να καθίσταται αισθητή όμως όχι παρεμβατική.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα			X	
Διαρροή			X	
Π.Δ.	0,5			

Συμπεράσματα Παραδείγματος

Στο έβδομο παράδειγμα, το εργαλείο κατόρθωσε να απομονώσει με επιτυχία το μπάσο, ενώ με ελαφρώς μικρότερη αποτελεσματικότητα τα ντραμς. Αντίθετα, απέτυχε πλήρως όσον αφορά το ριάνο stem. Στο συγκεκριμένο stem, το πιάνο καθίσταται από ελάχιστο ως και ανύπαρκτο καθώς αποδίδεται ιδιαίτερα θολό και δεν μπορείς να το ξεχωρίσεις - αναγνωρίσεις σχεδόν καθόλου.

Η ανάλυση του Moonlight Serenade επιβεβαιώνει για άλλη μια φορά ότι το Spleeter μπορεί να διαχωρίσει με σχετική ακρίβεια τη φωνή από τα υπόλοιπα stems. Παρόλο που το τελικό stem μπορεί να περιλαμβάνει και ελάχιστα σημεία από τα υπόλοιπα stems λόγω διαρροής, η φωνή ως επί το πλείστον είναι σαφώς διαχωρισμένη.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα				X
	Διαρροή			X	
	Π.Δ.	0,66			
Piano	Καθαρότητα	X			
	Διαρροή		X		
	Π.Δ.	0,33			
Drums	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Bass	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Other	Καθαρότητα			X	
	Διαρροή			X	
	Π.Δ.	0,5			

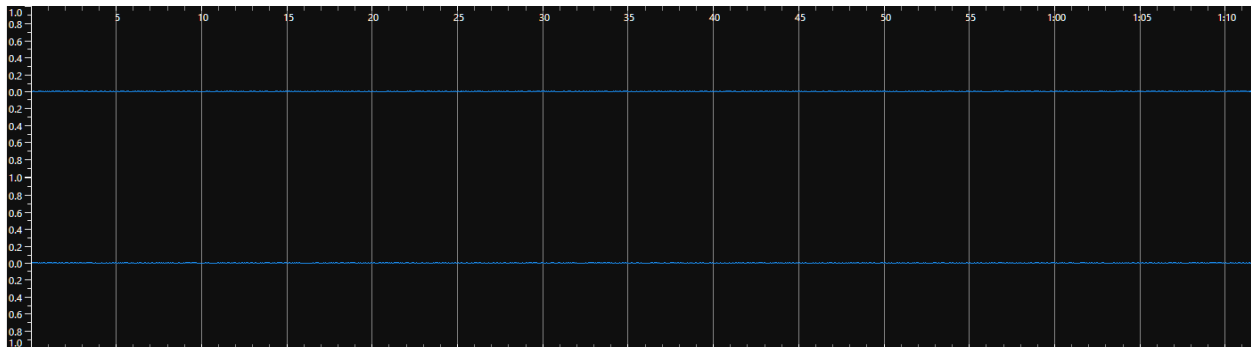
4.8 Μουσικό παράδειγμα 8: G.F. Handel: Flute Sonata in F Major – III. Siciliana

Το όγδοο μουσικό παράδειγμα που αναλύθηκε είναι η Flute Sonata in F Major – III. Siciliana του George Frideric Handel. Ένα έργο που ανήκει στην μπαρόκ μουσική, με έντονη την μελωδικότητα, λυρική και εκφραστικότητα του φλάουτου πάνω σε αρμονίες της εποχής. Το τσέμπαλο διατηρεί μία διακριτική και σταθερή συνοδεία ακολουθώντας το basso continuo όπως συνηθίζοταν στη μουσική του 18ου αιώνα.. Η επιλογή του συγκεκριμένου έργου δεν ήταν τυχαία. Αφενός, βασίζεται αποκλειστικά σε δύο μόνο όργανα τα οποία δεν ανήκουν σε κανένα stem. Αφετέρου, λόγω της πολυπλοκότητάς της γραφής του αναμένουμε στο διαχωρισμό να αποδοθεί ατόφιο στο stem του other, ενώ τα υπόλοιπα stems να παραμείνουν κενά. Αυτή η ιδιαιτερότητα το καθιστά μια πολύ ενδιαφέρουσα περίπτωση για τη μελέτη της αποτελεσματικότητας του διαχωρισμού του Spleeter.

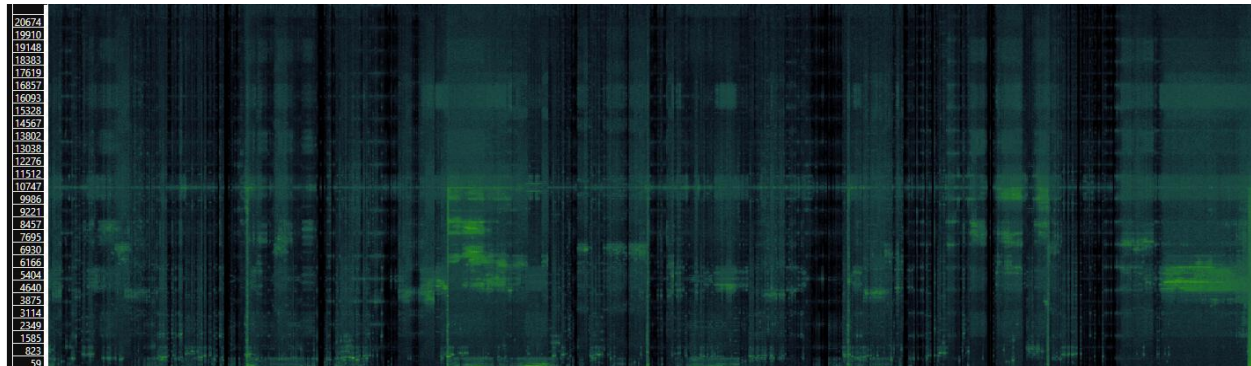
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο είχε διάρκεια περίπου 1 λεπτών και 11 δευτερολέπτων. Η εξαγωγή ολοκληρώθηκε χωρίς τεχνικά προβλήματα (επαναλαμβανόμενες προσπάθειες, «crush» του εργαλείου κτλ.), και για τα πέντε παραγόμενα stems καθώς έγινε η αποθήκευσή τους στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 8). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 71: Waveform of vocal stem (Sonic Visualiser)



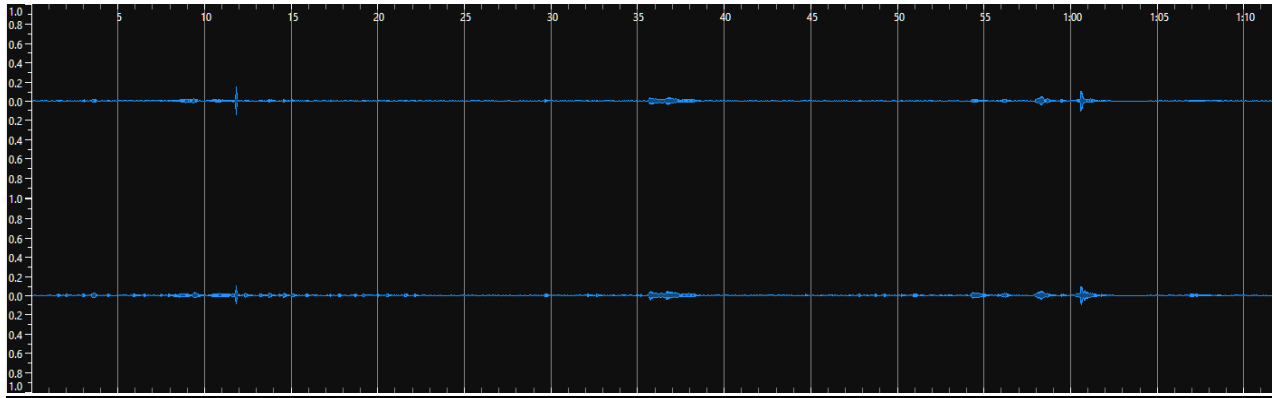
Εικόνα 72: Spectrogram of vocal stem (Sonic Visualiser)

Στο συγκεκριμένο έργο δεν υπάρχουν φωνητικά, συνεπώς το stem των vocals θα έπρεπε θεωρητικά να είναι απολύτως κενό. Πράγματι, στην κυματομορφή παρατηρείται ότι η ένταση είναι σχεδόν μηδενική. Ακουστικά, σε όλη την διάρκεια του stem δεν διακρίνονται τμήματα, τα οποία μπορούν να ταυτιστούν με κάποιο από τα υπόλοιπα stems. Υπάρχει μόνο ελάχιστος θόρυβος τον οποίο κάποιος μπορεί να τον εντοπίσει κάτω από ιδανικές συνθήκες ακρόασης: με ποιοτικά ακουστικά, σε υψηλή ένταση και σε απόλυτα ήσυχο περιβάλλον. Η διαρροή εκτιμάται ως μηδενικού βαθμού, ενώ η καθαρότητα στο 3, καθώς ο παρατηρούμενος θόρυβος είναι αμελητέος και καθόλου παρεμβατικός.

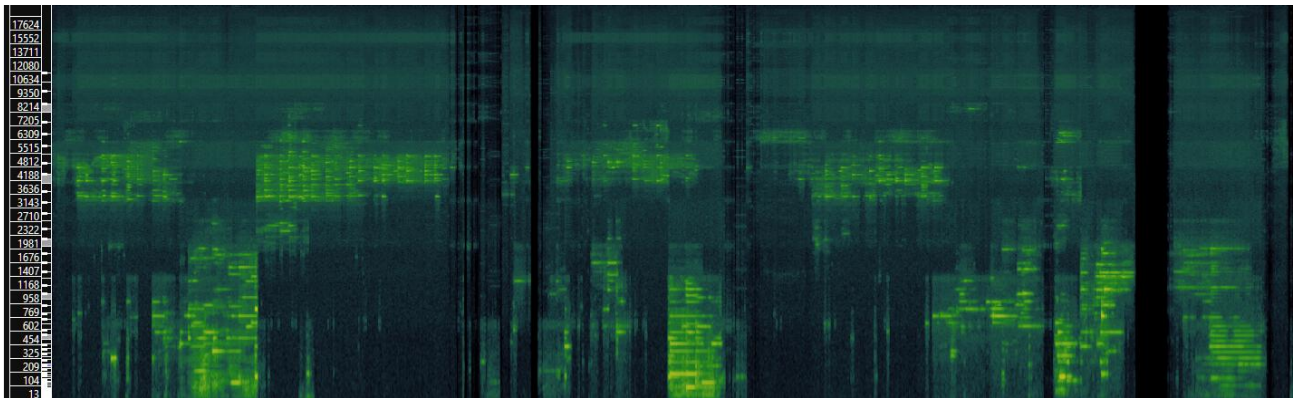
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 73: Waveform of piano stem (Sonic Visualiser)



Εικόνα 74: Spectrogram of piano stem (Sonic Visualiser)

Αρχικά, πρέπει να διασαφηνίσουμε ότι για την περίπτωση του πιάνου, αυτό δεν υπάρχει σαν όργανο στο συγκεκριμένο παράδειγμα. Παρόλα αυτά, τόσο η κυματομορφή όσο και το φασματογράφημα υποδεικνύουν μία μικρή παρουσία διαρροής στο συγκεκριμένο stem, το οποίο θεωρητικά θα έπρεπε να είναι απολύτως κενό.

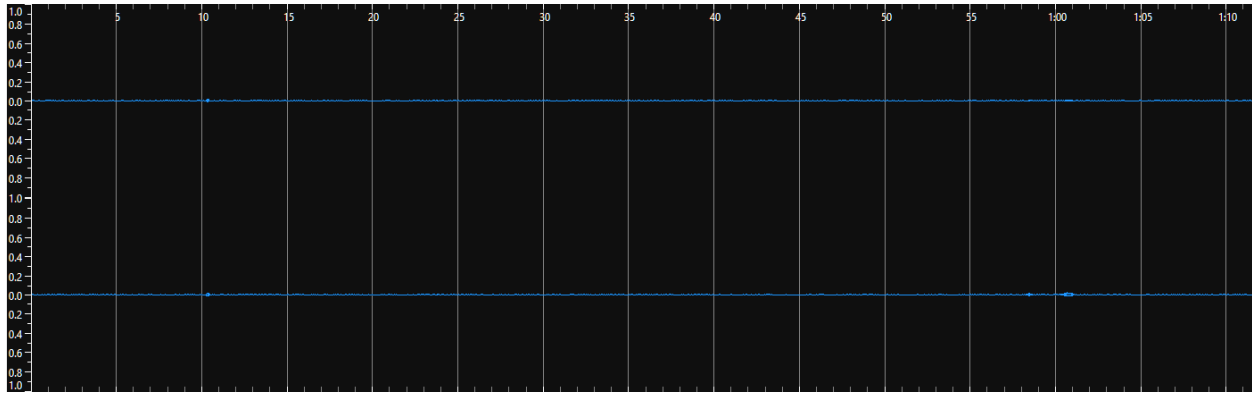
Κλίμακα ΚΚΑ				
Piano	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Αξίζει να σημειωθεί, ότι τόσο το πιάνο όσο και το τσέμπαλο είναι έγχορδα πληκτροφόρα όργανα, επομένως οι θεμελιώδεις συχνότητες που παράγουν επικαλύπτονται σχεδόν πλήρως. Το πιάνο έχει έκταση περίπου 27,5 και 4.186 Hz, (Fletcher & Rossing, 1991) ενώ και το τσέμπαλο με τη σειρά του βρίσκεται και αυτό ανάμεσα σε αυτό το διάστημα συχνοτήτων, με μικρότερο εύρος κατά 2 με 3 οκτάβες. Παρόλο που τα δύο όργανα συμπίπτουν συχνοτικά, το ηχόχρωμα και οι ατάκες του ήχου τους διαφέρουν έντονα. Ένα μοντέλο διαχωρισμού για να δουλέψει με μεγαλύτερη ακρίβεια, δεν βασίζεται μόνο στις θεμελιώδεις συχνότητες των οργάνων αλλά, χρησιμοποιεί και ηχοχρωματικά, αρμονικά, μοτιβικά και χρονικά χαρακτηριστικά του ήχου. (Johnson et al., 2023) Υποθέτουμε ότι αν το μοντέλο έχει εκπαιδευτεί να ξεχωρίζει τα όργανα μόνο με βάση το τονικό τους εύρος, τότε ναι, μπορεί να μπερδέψει το πιάνο με το τσέμπαλο. Αν όμως έχει εκπαιδευτεί ώστε να αναγνωρίζει το ηχόχρωμα και τα ειδικά χαρακτηριστικά της εκάστοτε ατάκας, τότε είναι σε θέση να τα ξεχωρίσει με σχετική ακρίβεια.

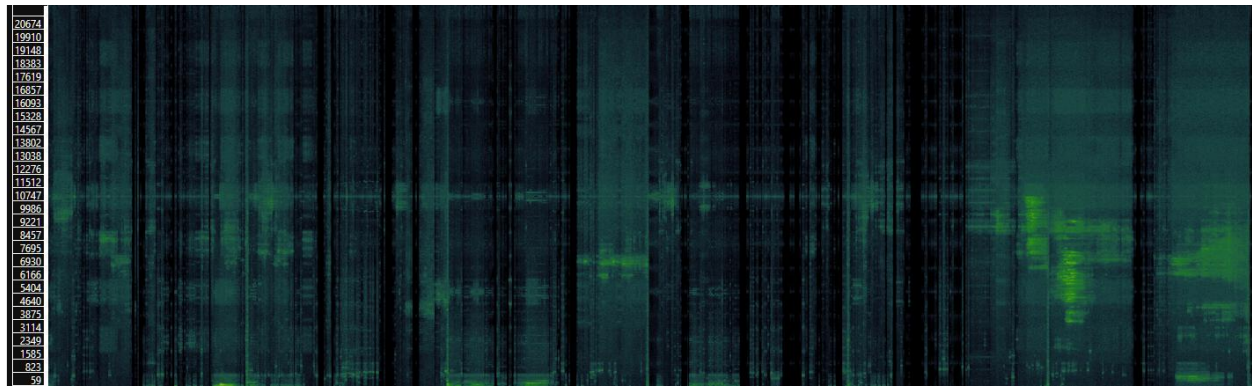
Σε συνδυασμό με την ακρόαση και την παρατήρηση της κυματομορφής και του φασματογραφήματος, διαπιστώνεται σε πολύ χαμηλή ένταση διαρροή από το τσέμπαλο από το other σε συνδυασμό με τον αναμενόμενο θόρυβο. Για τον λόγο αυτό η διαρροή είναι 1^{ου} βαθμού ενώ η καθαρότητα στο 2, καθώς η διαρροή είναι ελάχιστη.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 75: Waveform of drums stem (Sonic Visualiser)



Εικόνα 76: Spectrogram of drums stem (Sonic Visualiser)

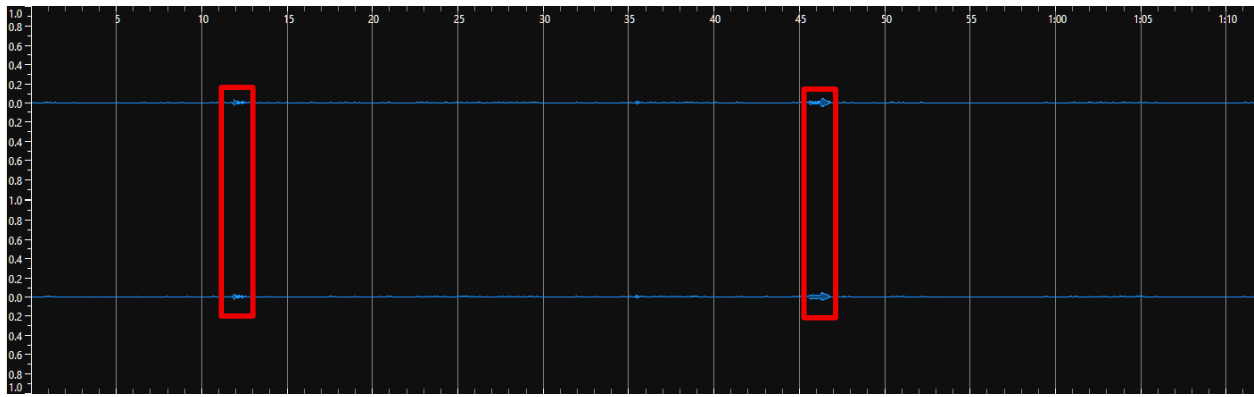
Αντιστοίχως με το πιάνο, πρέπει να τονίσουμε ότι και για την περίπτωση των drums, αυτά δεν υπάρχουν στο συγκεκριμένο παράδειγμα. Παρ' όλα αυτά, και σε αυτή την περίπτωση, παρατηρείται ελάχιστος θόρυβος αλλά καμία άλλη ηχητική πληροφορία η οποία να μας παραπέμπει σε κάποιο από τα υπόλοιπα stems. Κατά την ακρόαση, για να καταστεί αντιληπτός ο θόρυβος απαιτούνται και πάλι ιδανικές συνθήκες ακρόασης: με ποιοτικά ακουστικά, σε υψηλή ένταση και σε απόλυτα ήσυχο περιβάλλον. Σε συνδυασμό με την ακρόαση και την

Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

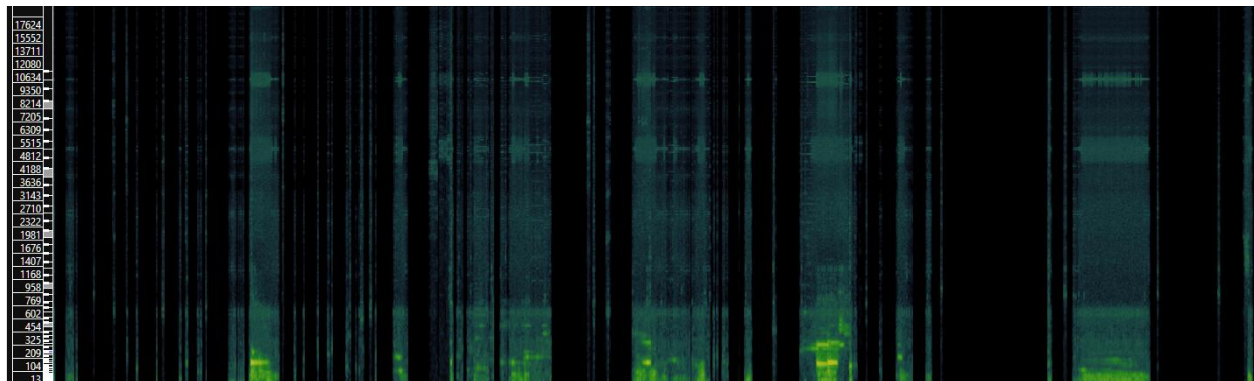
παρατήρηση της κυματομορφής και του φασματογραφήματος, διαπιστώνεται διαρροή μηδενικού βαθμού και καθαρότητα στο, καθώς ο αναμενόμενος θόρυβος είναι ελάχιστος και καθόλου παρεμβατικός.

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 77: Waveform of bass stem (Sonic Visualiser)



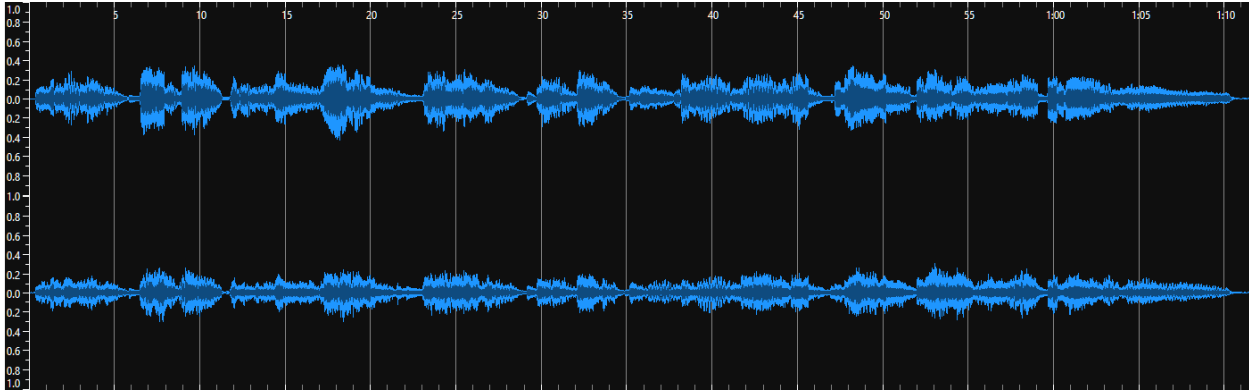
Εικόνα 78: Spectrogram of bass stem (Sonic Visualiser)

Παρομοίως με τα stems του piano και drums, το μπάσο δεν υφίσταται στο συγκεκριμένο παράδειγμα. Παρατηρείται μόνο ελάχιστος θόρυβος σε δύο σημεία, ο οποίος εμφανίζεται σε πολύ χαμηλή ένταση και δεν είναι καθόλου παρεμβατικός ως προς τη συνολική καθαρότητα του stem. Συνεπώς η διαρροή είναι μηδενικού βαθμού και η καθαρότητα 3^{ου}.

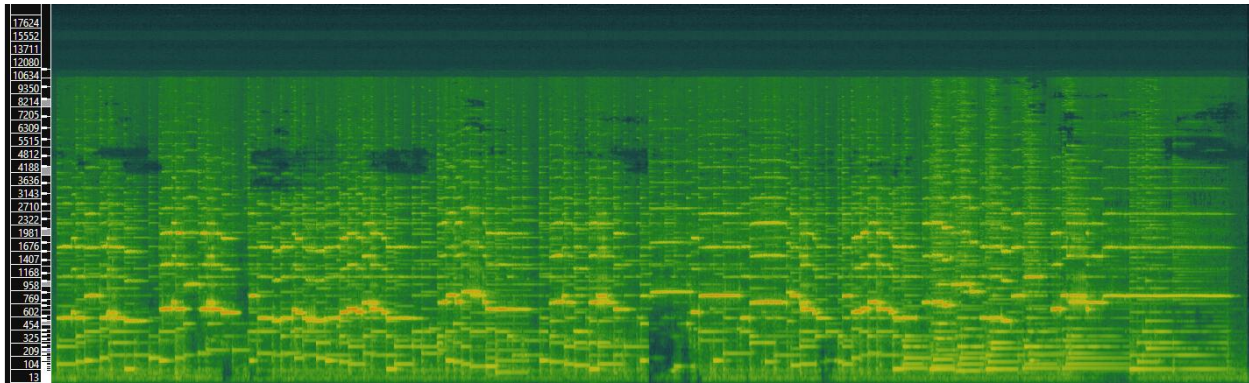
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 79: Waveform of other stem (Sonic Visualiser)



Εικόνα 80: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other αναμενόταν να απομονωθεί το φλάουτο και το τσέμπαλο, με το πρώτο να έχει τον κύριο σολιστικό ρόλο. Ακουστικά, το stem είναι σχεδόν πανομοιότυπο με το αρχικό έργο, το οποίο άλλωστε είναι και το ζητούμενο. Η μόνη εντοπίσιμη διαφορά είναι η προσθήκη επιπρόσθετου θορύβου μετά τον διαχωρισμό. Επιπλέον, σε ορισμένα σημεία το φλάουτο παρουσιάζεται ελαφρώς «θολό» και σε χαμηλότερη ένταση, στοιχείο το οποίο υποβαθμίζει την καθαρότητα ως 2^ο βαθμού. Το φαινόμενο αυτό υποθέτουμε ότι αποδίδεται στη λειτουργία του μοντέλου και αποτελεί παράμετρο προς βελτίωση.

Κλίμακα ΚΚΑ				
Other	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

Συμπεράσματα Παραδείγματος

Το παράδειγμα 8 έδειξε ότι σε έργα με περιορισμένο αριθμό οργάνων αλλά με πιο δεξιοτεχνική γραφή, το Spleeter καταφέρνει να αποδώσει με σχετική απομόνωση, αλλά εξακολουθούν να υπάρχουν μικρά προβλήματα διαρροής και θορύβου στα κενά stems. Τα stems των bass, drum και vocal ήταν τα πιο καθαρά σε σχέση με των piano και other, όπου η καθαρότητα ήταν μικρότερη. Ενώ το φλάουτο έπρεπε να αποδίδεται ατόφιο στο other, στην πραγματικότητα συνοδεύταν

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Piano	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			

από θόρυβο, ενώ επιπλέον σε ορισμένα σημεία ακουγόταν «θολό» και με μικρότερη ένταση. Το παραπάνω υποβαθμίζει την συνολική καθαρότητα του stem. Το φαινόμενο αυτό υποθέτουμε ότι αποδίδεται στη λειτουργία του μοντέλου και, κατά την άποψή μας, αποτελεί παράμετρο προς βελτίωση.

Drums	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.			1	
Bass	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.			1	
Other	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.			0,83	

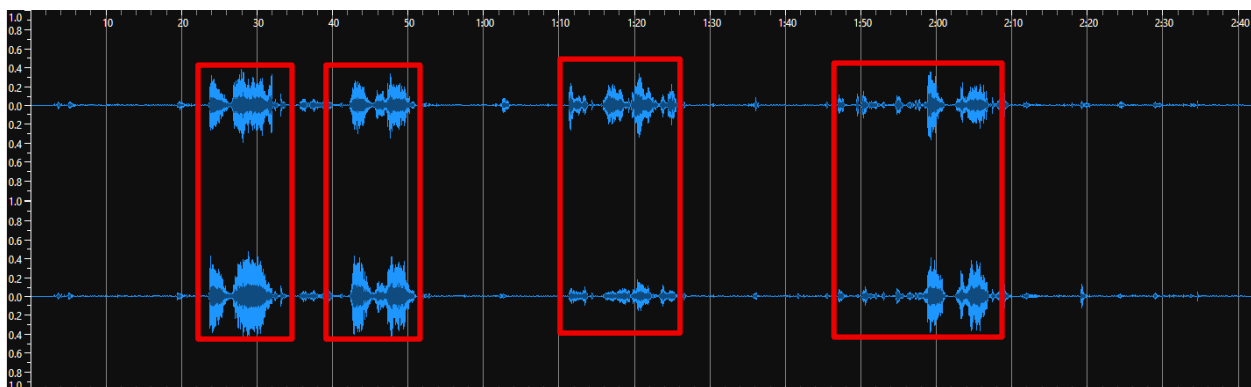
4.9 Μουσικό παράδειγμα 9: Henry Mancini: Lujon

Το ένατο μουσικό παράδειγμα που αναλύθηκε είναι το Lujon του Henry Mancini. Ένα χαρακτηριστικό έργο της τζαζ μουσικής με την ξεχωριστή και ιδιαίτερη ενορχήστρωση του συγκεκριμένου συνθέτη, ο οποίος αξιοποιεί με δεξιοτεχνία τις δυναμικές και τα ηχοχρώματα της ορχήστρας για να δημιουργήσει την αίσθηση χαλάρωσης και ατμοσφαιρικότητας στο έργο του. Χαρακτηριστική είναι η μεγάλη ποικιλία σε κρουστά καθώς και ο ίδιος ο ρυθμός, ο οποίος επαναλαμβάνεται από αυτά και το μπάσο. Πάνω στον ρυθμό αναπτύσσονται μελωδικά μοτίβα από τα έγχορδα, και το σαξόφωνο με τα πρώτα να έχουν και ρόλο αρμονικής υποστήριξης. Αυτή η πολυπλοκότητα με συνεχείς επικαλύψεις μεταξύ ρυθμομελωδικών και αρμονικών στοιχείων, προσθέτει μια ιδιαίτερη πρόκληση για το εργαλείο διαχωρισμού.

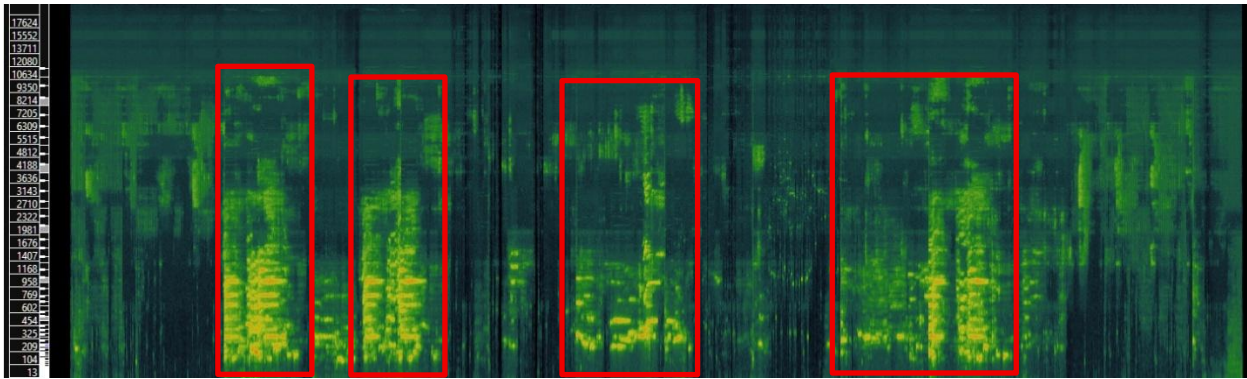
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο είχε διάρκεια περίπου 2 λεπτών και 42 δευτερολέπτων. Ο διαχωρισμός πραγματοποιήθηκε χωρίς πρόσθετη επεξεργασία και τα stems αποθηκεύτηκαν στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 9). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 81: Waveform of vocal stem (Sonic Visualiser)



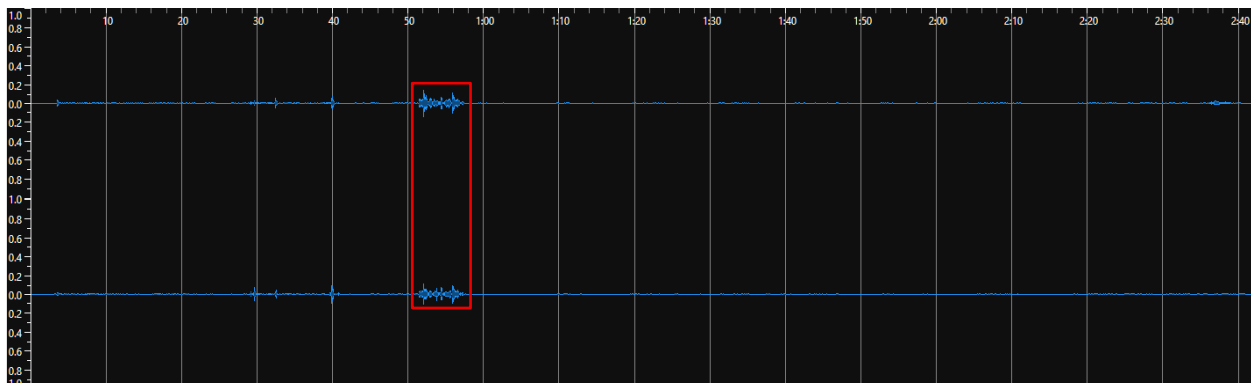
Εικόνα 82: Spectrogram of vocal stem (Sonic Visualiser)

Στο *Lujon* δεν υπάρχουν φωνητικά, συνεπώς το stem θα έπρεπε θεωρητικά να είναι κενό. Ωστόσο, ακουστικά διακρίνεται διαρροή 2^{ου} βαθμού, η οποία προέρχεται κυρίως από τις νότες των κρουστών, των εγχόρδων και του σαξοφώνου (drum και other stems). Η διαρροή, αν και σε γενικές γραμμές ακούγεται σε χαμηλή ένταση (εκτός των τεσσάρων σημείων όπου είναι πολύ έντονη ακουστικά), είναι συνεχής σε όλη τη διάρκεια του έργου συνοδευόμενη από τον αναμενόμενο θόρυβο. Στο φασματογράφημα αποτυπώνεται η συνεχής παρουσία συχνοτήτων που δεν δικαιολογείται ο λόγος ύπαρξής τους στο συγκεκριμένο stem, οδηγώντας σε καθαρότητα 1^{ου} βαθμού.

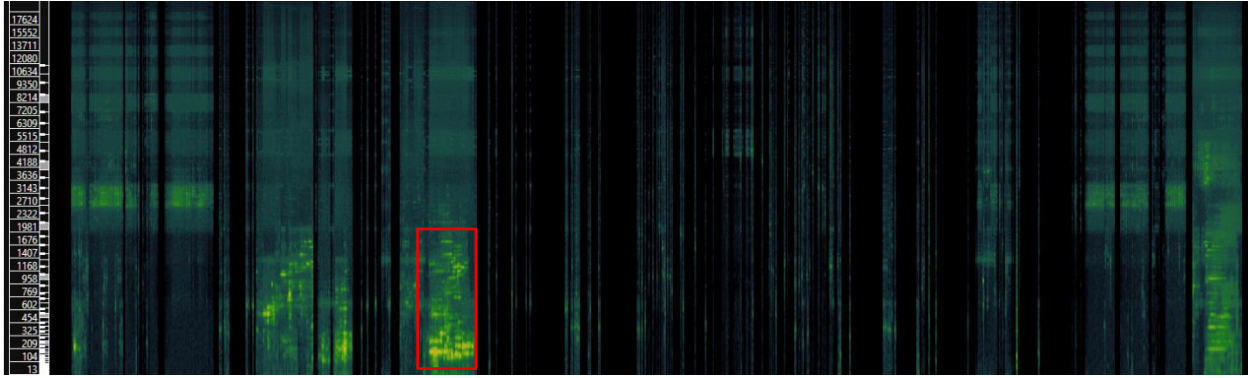
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα		X		
Διαρροή			X	
Π.Δ.	0,66			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 83: Waveform of piano stem (Sonic Visualiser)



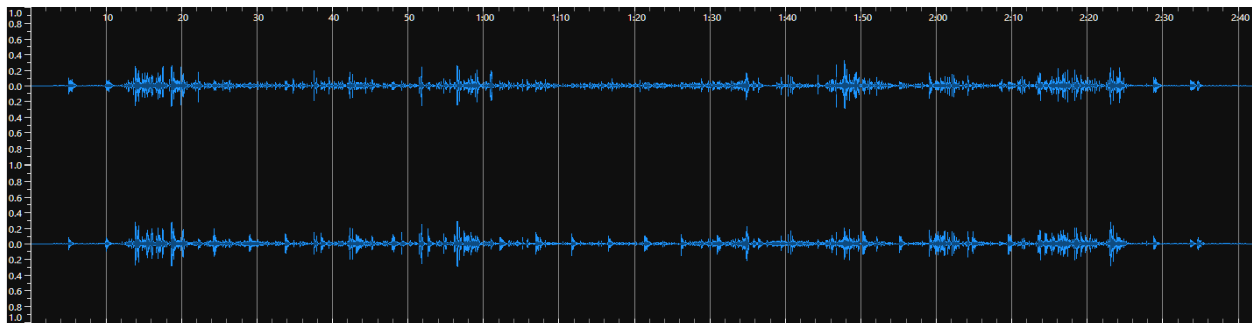
Εικόνα 84: Spectrogram of piano stem (Sonic Visualiser)

Στην περίπτωση του πιάνου, αυτό δεν υπάρχει σαν όργανο στο συγκεκριμένο παράδειγμα. Παρόλα αυτά, τόσο η κυματομορφή όσο και το φασματογράφημα υποδεικνύουν την παρουσία διαρροής στο stem, το οποίο θεωρητικά θα έπρεπε να είναι κενό. Ακουστικά, η διαρροή δεν καθίσταται αρκετά αισθητή, με εξαίρεση ενός σημείου στο οποίο εμφανίζεται πιο έντονη και σε μεσαία ένταση. Γενικά, η διαρροή χαρακτηρίζεται ως ελάχιστη, μη παρεμβατική και χωρίς σαφή ταύτιση με κάποιο άλλο stem. Για τον λόγο αυτό κρίνεται ως διαρροή 0^{ου} βαθμού, ενώ η καθαρότητα στο 2.

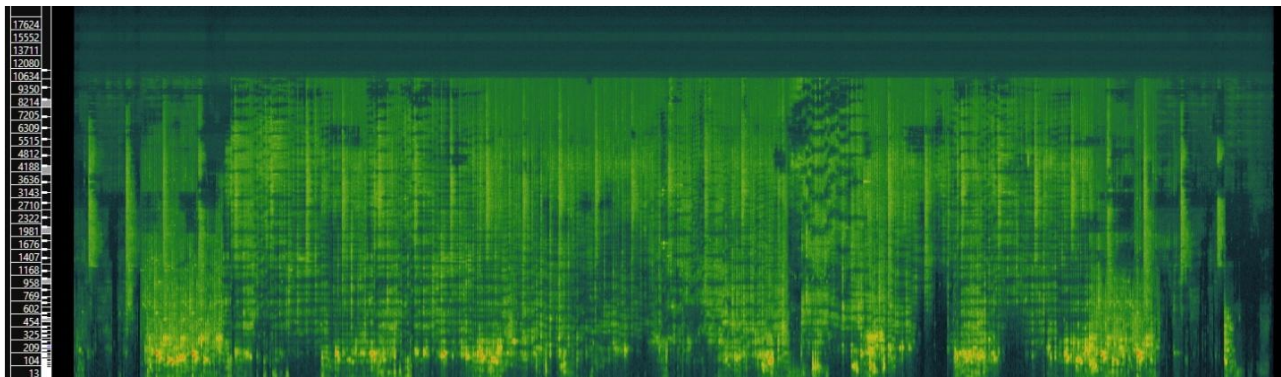
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 85: Waveform of drums stem (Sonic Visualiser)



Εικόνα 86: Spectrogram of drums stem (Sonic Visualiser)

Τα κρουστά παίζουν πρωταγωνιστικό ρόλο σε όλο το έργο με έντονο ρυθμικό χαρακτήρα. Μέσω του φασματογραφήματος και της κυματομορφής δεν προκύπτουν σαφείς ενδείξεις για τη διαρροή. Συνεπώς, η αξιολόγηση βασίζεται κυρίως στην ακουστική παρατήρηση.

Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή			X	
Π.Δ.	0,5			

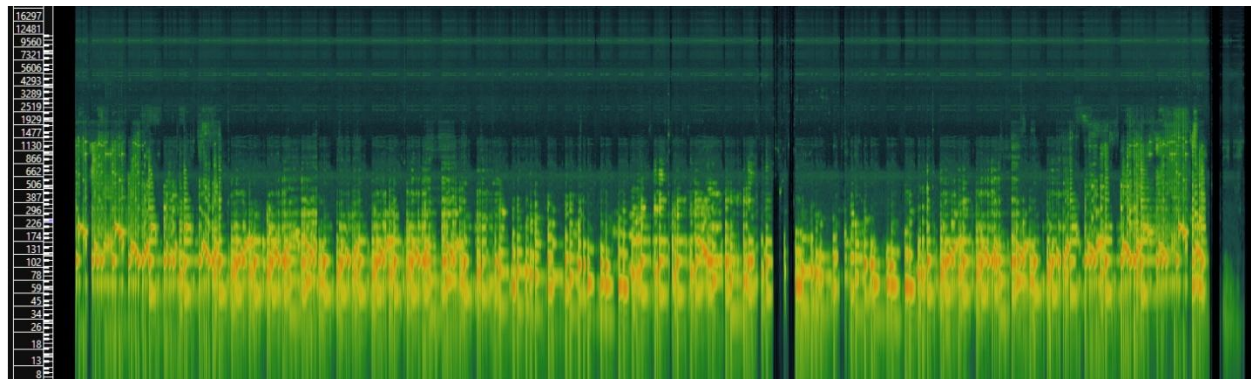
Το stem των drums παρουσιάζει ακουστικά μία όχι τόσο ευδιάκριτη παρουσία των κρουστών όσο αναμενόταν ενώ, παράλληλα ανιχνεύεται μια μικρή διαρροή από τα υπόλοιπα stem. Συγκεκριμένα, διακρίνεται σε χαμηλή ένταση η παρουσία του μπάσου και ορισμένων εγχόρδων, γεγονός που μας οδηγεί στην αξιολόγηση της διαρροής ως 2^ο βαθμού. Η διαρροή δεν είναι έντονα αισθητή και περιορίζεται σε ελάχιστα και συγκεκριμένα σημεία, χωρίς να επηρεάζει σημαντικά τη συνολική ποιότητα του διαχωρισμού. Συνεπώς η καθαρότητα παραμένει στο 2.

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 87: Waveform of bass stem (Sonic Visualiser)



Εικόνα 88: Spectrogram of bass stem (Sonic Visualiser)

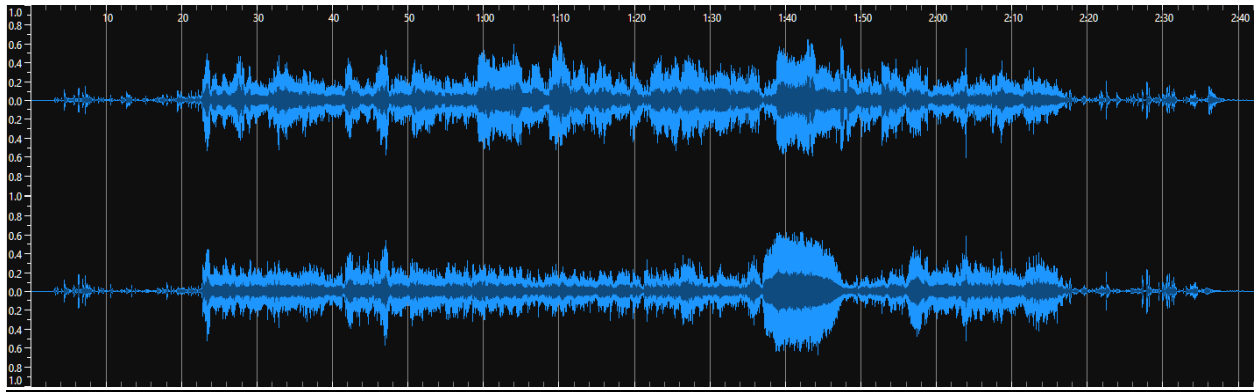
Στην περίπτωση με το stem του bass θα ξεκινήσουμε με την ακρόαση. Αναμένουμε να ακούσουμε τις πολύ χαμηλές συχνότητες των εγχόρδων και του βιμπραφώνου, τα οποία έχουν το σταθερό ρυθμομελωδικό μοτίβο που επιτελεί τη λειτουργία του μπάσου.

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

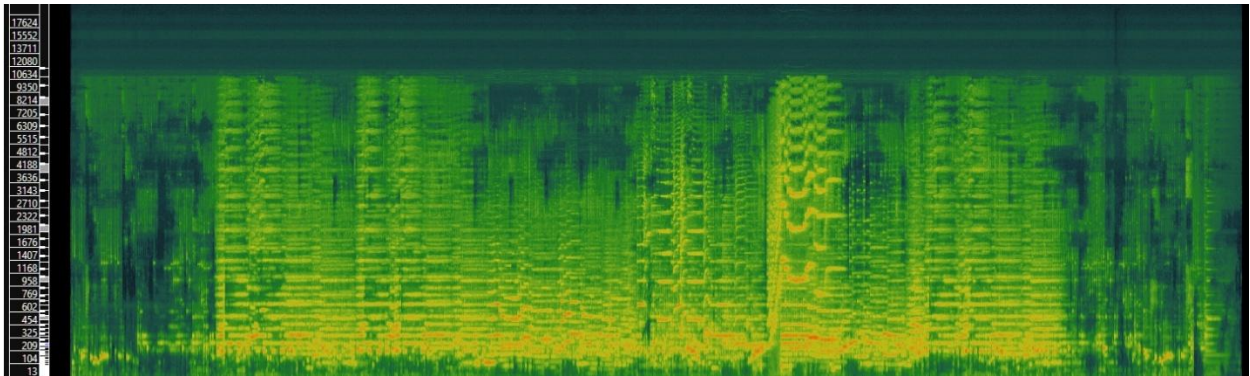
Κατά την ακρόαση, διαπιστώνεται ότι πράγματι η απομόνωση είναι αρχικά ιδιαίτερα επιτυχής, χωρίς εμφανείς διαρροές από τα υπόλοιπα stems. Το αποτέλεσμα που ακούμε είναι όντως ένας συνδυασμός των εγχόρδων και του βιμπραφώνου αλλά σε σημεία «θολό» και σε χαμηλή ένταση. Οι συχνότητες παραμένουν εντός της περιοχής του μπάσου και δεν παρεμβαίνει εμφανώς κανένα άλλο stem. Για τον λόγο αυτό, στην ΚΚΔ η διαρροή κρίνεται ως μηδενικού βαθμού ενώ η καθαρότητα στο 2. Τα παραπάνω επιβεβαιώνονται και οπτικά μέσω του Sonic Visualiser διότι, οι θεμέλιες συχνότητες οι οποίες παρατηρούμε να καταγράφονται κυμαίνονται μεταξύ 45 και 226 Hz.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 89: Waveform of other stem (Sonic Visualiser)



Εικόνα 90: Spectrogram of other stem (Sonic Visualiser)

Στο συγκεκριμένο stem, το περιεχόμενό του θα έπρεπε να αποτελεί το «υπόλοιπο» από την απομόνωση των υπολοίπων stems. Ακουστικά όμως, παρατηρείται μία πλήρη διαρροή. Αναλυτικότερα, το spectrogram εμφανίζεται «πλούσιο» συχνοτικά, το οποίο μας προϊδεάζει για την ύπαρξη διαρροής. Αυτό πράγματι επαληθεύεται κατά την ακρόαση, όπου το αποτέλεσμα του διαχωρισμού θυμίζει το έργο πριν από αυτόν, ακούγοντας τα drums και το bass στη

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα		X		
Διαρροή			X	
Π.Δ.	0,33			

μεγαλύτερη διάρκεια του stem. Συνεπώς, το επίπεδο διαρροής το εκτιμούμε ως 2^ο βαθμού και την καθαρότητα ως 1^ο βαθμού λόγω της συνεχόμενης διαρροής η οποία μειώνει σημαντικά την ποιότητα του διαχωρισμού.

Συμπεράσματα Παραδείγματος

Στο ένατο παράδειγμα, το μοντέλο διαχωρισμού κατόρθωσε να απομονώσει με επιτυχία το μπάσο, με μέτρια αποτελεσματικότητα τα κρουστά. Αντίθετα, δυσκολεύτηκε ιδιαίτερα όσον αφορά το stem των vocals και του other. Στα φωνητικά παρόλο που δεν υπάρχουν στο έργο και θα έπρεπε το stem να είναι κενό. Ακούγεται διαρροή σε χαμηλή ένταση η οποία είναι συνεχής σε όλη τη διάρκεια του συνοδευόμενη από τον αναμενόμενο θόρυβο. Στο other stem, η διαρροή ήταν σχεδόν καθολική, ακούγοντας έντονα τα drums και το bass.

Η ανάλυση του Lujon (ως τελευταίο τζαζ παράδειγμα) έδειξε για άλλη μία φορά ότι το Spleeter δυσκολεύεται με ηχογραφήσεις τζαζ μουσικής όπως εντοπίσαμε και στα παραδείγματα 4.3, 4.4, 4.6. Τα ευρήματα επιβεβαιώνουν ότι η σταθερή αλληλοεπικάλυψη, ο μεγάλος αριθμός των οργάνων καθώς και οι ιδιαιτερότητές τους, δυσκολεύουν και περιορίζουν σημαντικά την ακρίβεια του διαχωρισμού, καταλήγοντας σε θολά και ασαφή stems ενώ πάλι το bass stem είναι αυτό με τη μεγαλύτερη Π.Δ.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα		X		
	Διαρροή			X	
	Π.Δ.	0,33			
Piano	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Drums	Καθαρότητα			X	
	Διαρροή			X	
	Π.Δ.	0,5			
Bass	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Other	Καθαρότητα		X		
	Διαρροή			X	
	Π.Δ.	0,33			

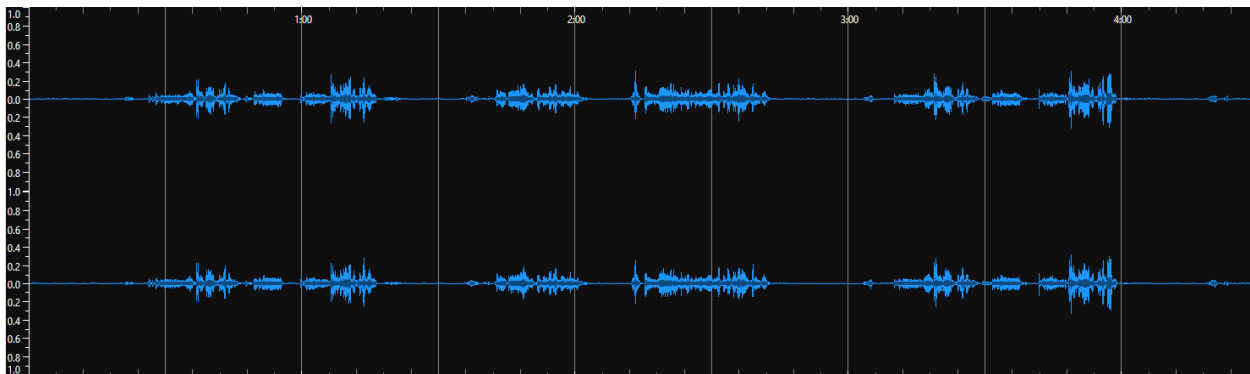
4.10 Μουσικό παράδειγμα 10: Christmas Oratorio, BWV 248: Part VI, 41. Arie "Ich will nur dir zu Ehren leben"

Το δέκατο μουσικό παράδειγμα που αναλύθηκε είναι το Christmas Oratorio, BWV 248: Part VI, 41. Arie "Ich will nur dir zu Ehren leben" του Johann Sebastian Bach. Το Χριστουγεννιάτικο ορατόριο είναι ένα έργο της ύστερης περιόδου του μπαρόκ με κύριο στοιχείο την πολυφωνία, την αυστηρή δομή και την πλούσια ενορχήστρωση. Το έκτο μέρος είναι μία άρια γραμμένη για βαρύτονο και ορχήστρα. Η κύρια μελωδία χαρακτηρίζεται από διαρκή κίνηση, ενώ τα έγχορδα και το basso continuo στο εκκλησιαστικό όργανο συμπληρώνουν άλλες φορές με τη μελωδία και άλλες ως αρμονικό υπόβαθρο.

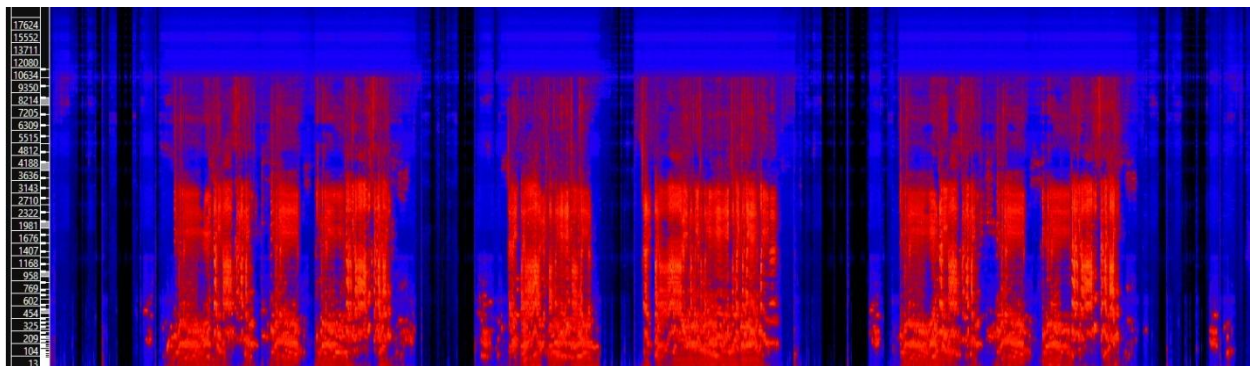
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη στο Spleeter είχε διάρκεια περίπου 4 λεπτών και 28 δευτερολέπτων. Ο διαχωρισμός έγινε χωρίς καμία ανάγκη για περαιτέρω επεξεργασία ή μετατροπή στο πρωτογενές αρχείο. Η εξαγωγή ολοκληρώθηκε έπειτα από δύο «crash» του εργαλείου και τα stems αποθηκεύτηκαν στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 10). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 91: Waveform of vocal stem (Sonic Visualiser)



Εικόνα 92: Spectrogram of vocal stem (Sonic Visualiser)

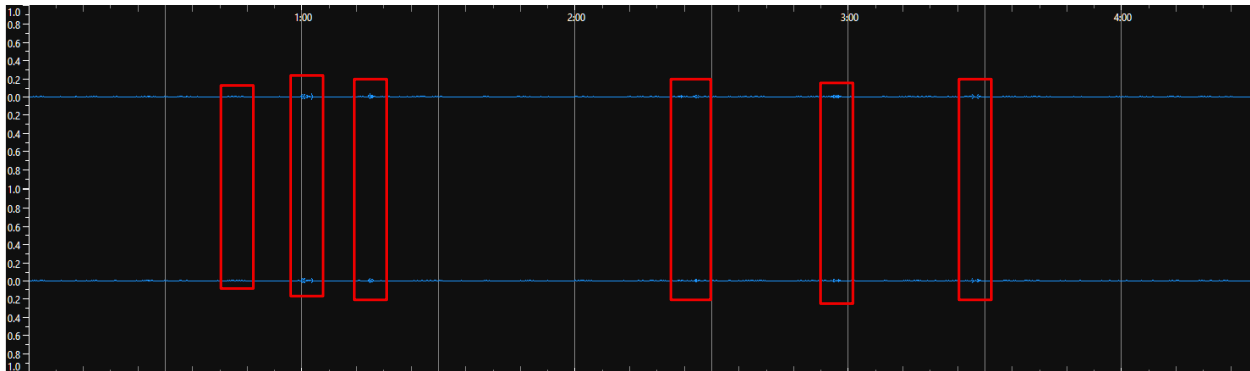
Λόγω της συνεχούς παρουσίας της φωνής καθ' όλη σχεδόν τη διάρκεια του έργου, για το stem των φωνητικών δεν μπορεί να συμβάλει ιδιαίτερα το Sonic Visualiser. Παρόλα αυτά, μπορούμε να αντλήσουμε πληροφορίες σχετικά με το τι μπορεί να αναμένει κανείς κατά την ακρόαση. Στην περίπτωση αυτού του stem, η αξιολόγηση βασίζεται κυρίως στην ακρόαση. Σε πολλά σημεία, το έργο ακούγεται σχεδόν απόφιο σε μειωμένη ένταση, γεγονός που υποδηλώνει χαμηλή καθαρότητα. Εξίσου σημαντικό είναι ότι εντοπίζεται διαρροή από τα έγχορδα και το εκκλησιαστικό από το other stem, ιδιαίτερα σε σημεία όπου η συνοδεία είναι έντονη μελωδικά, με

Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα		X		
Διαρροή		X		
Π.Δ.				0,5

αισθητή την παρουσία θολότητας και θορύβου κάτι που αξιολογείται ως 1ου βαθμού διαρροή και καθαρότητα.

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 93: Waveform of piano stem (Sonic Visualiser)



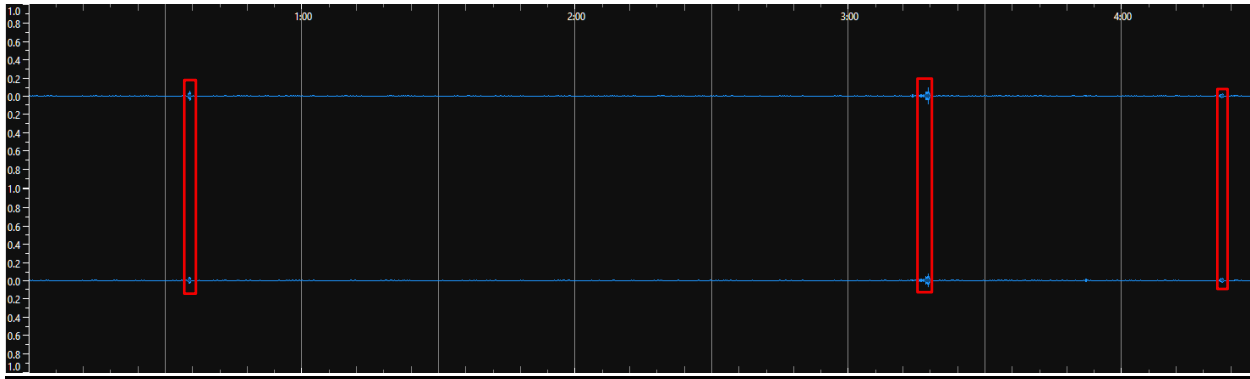
Εικόνα 94: Spectrogram of piano stem (Sonic Visualiser)

Δεδομένου ότι στο έργο δεν υπάρχει πιάνο, το stem του piano θα έπρεπε να είναι κενό. Ακουστικά, εμφανίζεται διαρροή η οποία είναι αντιληπτή κυρίως στα κόκκινα σημεία και παραπέμπει στο other stem με σε χαμηλή ένταση. Αυτό συνεπάγεται σε διαρροή 1^{ου} βαθμού . Στη συνέχεια παρόλο που εντοπίζουμε αυτά τα έξι σημεία όπου η διαρροή είναι αισθητή σε γενικές γραμμές το stem είναι καθαρό, γεγονός που καθιστά την καθαρότητα ως 2^{ου} βαθμού.

Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 95: Waveform of drums stem (Sonic Visualiser)



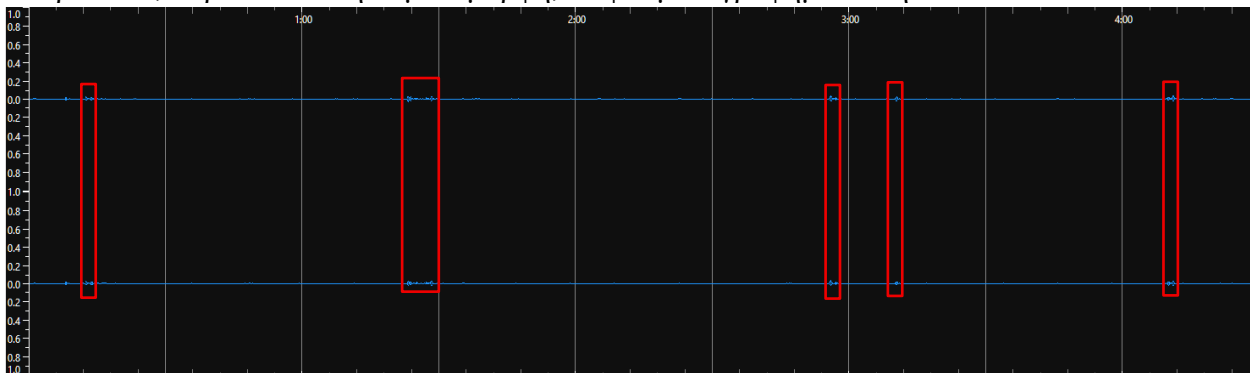
Εικόνα 96: Spectrogram of drums stem (Sonic Visualiser)

Στην μπαρόκ μουσική και ειδικότερα στο συγκεκριμένο μέρος δεν υπάρχουν κρουστά. Συνεπώς, το stem θα έπρεπε να είναι κενό. Ακουστικά και οπτικά το stem είναι καθαρό στο μεγαλύτερο μέρος της διάρκειάς του. Παρατηρείται μόνο σε ελάχιστα σημεία και σε πολύ χαμηλή ένταση η παρουσία θορύβου και συνεπάγεται σε ήπια μη παρεμβατική διαρροή. Επιπλέον, αυτή δεν ταυτίζεται με κάποιο από τα υπόλοιπα stems. Συνεπώς, η διαρροή είναι μηδενικού βαθμού και η καθαρότητα 3^{ου}.

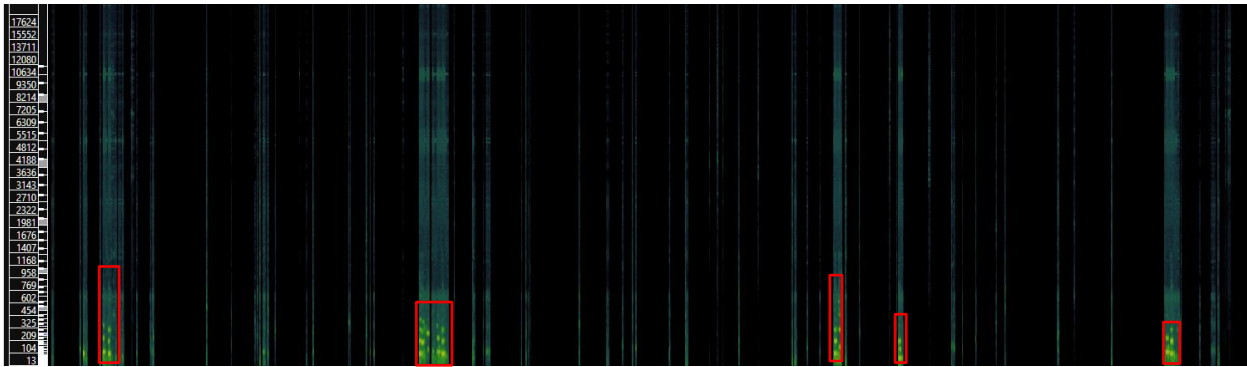
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 97: Waveform of bass stem (Sonic Visualiser)



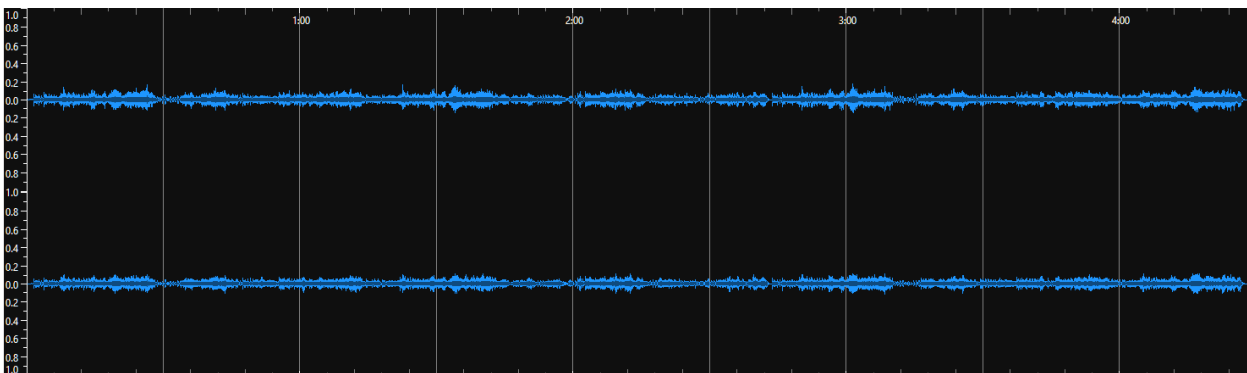
Εικόνα 98: Spectrogram of bass stem (Sonic Visualiser)

Στην περίπτωση με το stem του bass θα ξεκινήσουμε πάλι με την ακρόαση. Σε αυτό το stem αναμένουμε να ακούσουμε τις πολύ χαμηλές συχνότητες του εκκλησιαστικού οργάνου και του φαγκότου που έχουν την λειτουργία του basso continuo. Κατά την ακουστική αξιολόγηση, η απομόνωση δεν έχει γίνει ή μάλλον δεν έχει ανιχνευτεί από το μοντέλο η λειτουργία του εκκλησιαστικού οργάνου. Βέβαια αυτό δεν αναμένεται σαν μία από τις δυνατότητες του συγκεκριμένου μοντέλου στο παρόν στάδιο. Για αυτό το λόγο αναμένουμε να παραμείνει κενό το stem. Πράγματι έτσι συμβαίνει και επιβεβαιώνονται οι αντίστοιχες παρατηρήσεις οι οποίες έγιναν στο drums stem. (Δες προηγούμενη σελίδα)

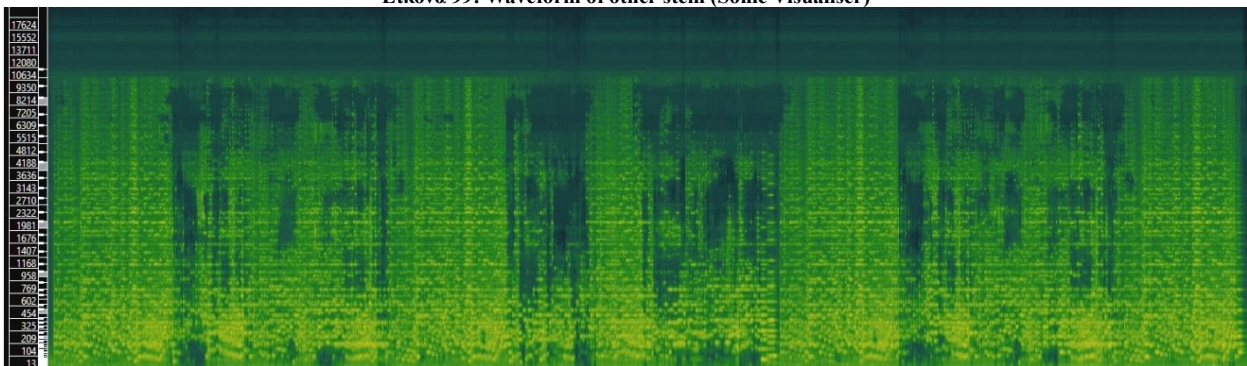
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 99: Waveform of other stem (Sonic Visualiser)



Εικόνα 100: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other θα έπρεπε να απομονωθούν τα έγχορδα, το εκκλησιαστικό και το φαγκότο. Ακουστικά, διακρίνεται η ορχήστρα, αλλά σε αρκετά «θολή» μορφή και σε χαμηλότερη ένταση όταν τραγουδά ο βαρύτονος στο αρχικό έργο. Επιπλέον, υπάρχει παρουσία του φωνητικού μέρους σε αρκετά σημεία, οδηγώντας σε αξιολόγηση της διαρροής ως 1^ο βαθμού και καθαρότητας 2^ο βαθμού καθώς η φωνή δεν ακούγεται σε όλη τη διάρκεια του διαχωρισμού για να καταστεί ως μηδενικού.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα		X		
Διαρροή		X		
Π.Δ.	0,5			

Συμπεράσματα Παραδείγματος

Το παράδειγμα 10 έδειξε ότι σε έργα μπαρόκ μουσικής, όπως το Christmas Oratorio του Bach, με περιορισμένο αριθμό οργάνων αλλά με πιο δεξιοτεχνική γραφή, το Spleeter κατάφερε και εδώ, όπως και στο παράδειγμα 4.8 να κάνει σχετική απομόνωση. Δυσκολεύτηκε σημαντικά να διαχωρίσει τη φωνή από την ορχήστρα λόγω της έντονης και συνεχούς αλληλεπίδρασης μεταξύ τους κάτι το οποίο εντοπίζουμε για πρώτη φορά όσον αφορά τη φωνή.

Σημειώνεται ότι το πιάνο και τα drums απουσιάζουν από τη σύνθεση με το δεύτερο μόνο να είναι εντελώς κενό. Το bass stem ήταν επίσης απόλυτα καθαρό (κενό), χωρίς όμως να εντοπιστεί η λειτουργία του basso continuo, καθώς το μοντέλο δεν αναγνώρισε το ρόλο του φαγκότου και του εκκλησιαστικού οργάνου.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα		X		
	Διαρροή		X		
	Π.Δ.	0,5			
Piano	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Drums	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Bass	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Other	Καθαρότητα		X		
	Διαρροή		X		
	Π.Δ.	0,5			

Συνολικά, εστοιάζοντας στα vocal και other stems, το μοντέλο δεν απέδωσε τον διαχωρισμό με την αναμενόμενη καθαρότητα, επιβεβαιώνοντας τους περιορισμούς του σε έργα της μπαρόκ εποχής, όπου η πολυφωνία και η πυκνή μουσική υφή καθιστούν την απομόνωση των stems ιδιαίτερα απαιτητική.

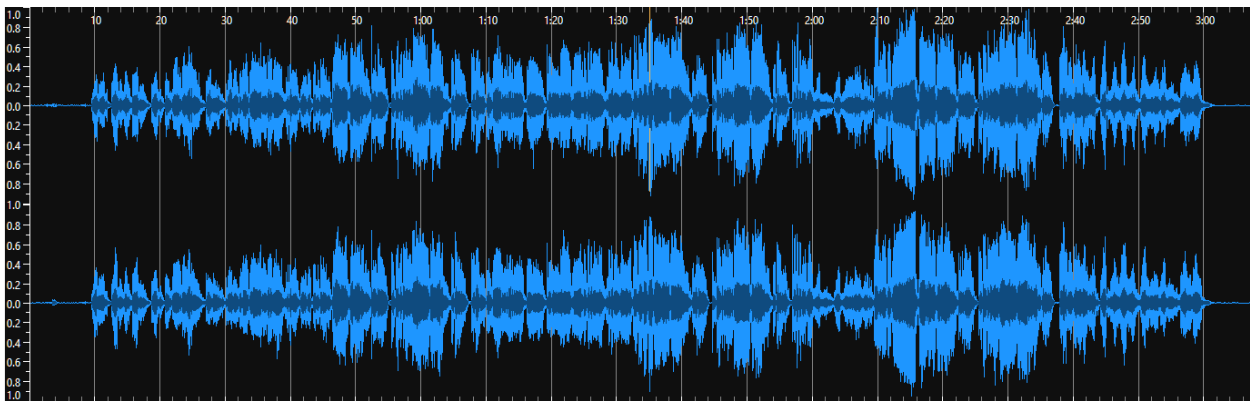
4.11 Μουσικό παράδειγμα 11: Kevin Penkin: Here, Tomorrow

Το ενδέκατο μουσικό παράδειγμα που αναλύθηκε είναι το Here, Tomorrow του Kevin Penkin, το οποίο αποτελεί μέρος από το soundtrack του βιντεοπαιχνιδιού League of Legends. Πρόκειται για ένα αρκετά ατμοσφαιρικό έργο βασισμένο σε αρμονία, η οποία πηγάζει από πεντατονικές κλίμακες. Το στοιχείο που κυριαρχεί είναι η γυναικεία φωνή, η οποία τραγουδά πάνω σε ένα «χαλί» συνοδευτικών εγχόρδων, κρουστών και διαφόρων ηλεκτρονικών στοιχείων, σε συνδυασμό με την επεξεργασία του βάθους (depth) και της αντήχησης (reverb).

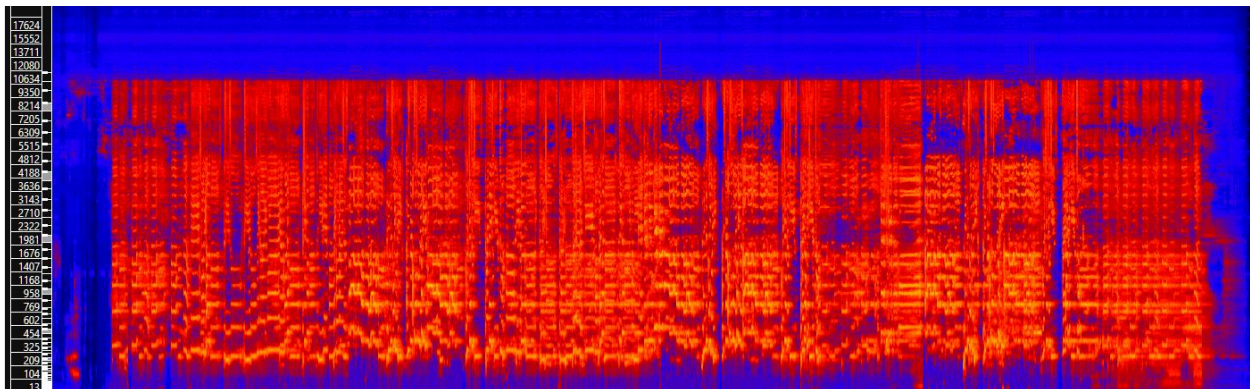
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη στο Spleeter είχε διάρκεια περίπου 3 λεπτών και 7 δευτερολέπτων. Ο διαχωρισμός έγινε χωρίς καμία ανάγκη για περαιτέρω επεξεργασία ή μετατροπή στο πρωτογενές αρχείο. Η εξαγωγή ολοκληρώθηκε χωρίς τεχνικά προβλήματα (επαναλαμβανόμενες προσπάθειες, «crush» του εργαλείου κτλ.), και για τα πέντε παραγόμενα stems καθώς έγινε η αποθήκευσή τους στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 11). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 101: Waveform of vocal stem (Sonic Visualiser)



Εικόνα 102: Spectrogram of vocal stem (Sonic Visualiser)

Λόγω της συνεχούς παρουσίας της φωνής καθ’ όλη σχεδόν τη διάρκεια του έργου, για το stem των φωνητικών δεν μπορεί να συμβάλει ιδιαίτερα το Sonic Visualiser. Παρόλα αυτά, μπορούμε να αντλήσουμε πληροφορίες σχετικά με το τι μπορεί να αναμένει κανείς κατά την ακρόαση. Στην περίπτωση αυτού του stem, η αξιολόγηση βασίζεται

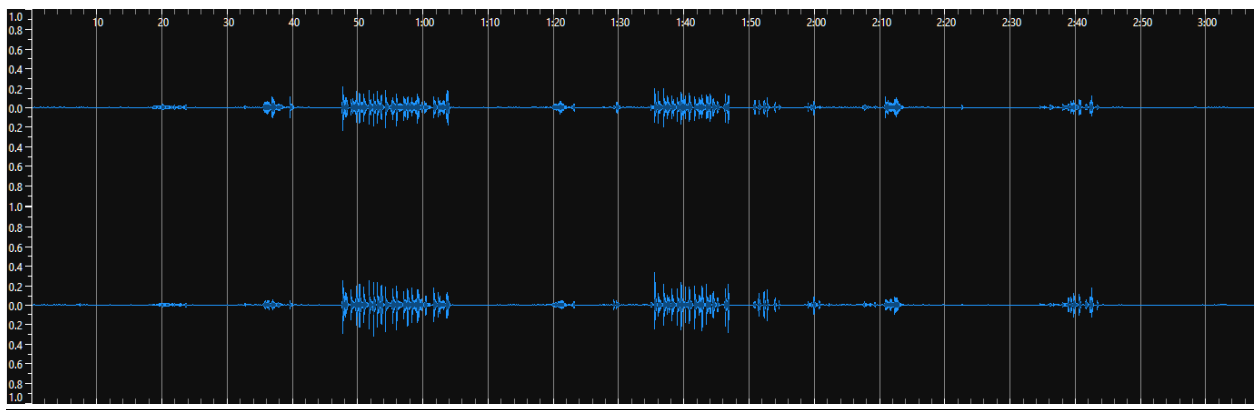
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

κυρίως στην ακρόαση. Κατόπιν αυτής, διαπιστώνεται ότι ο διαχωρισμός της φωνής επιτεύχθηκε σε μεγάλο βαθμό με επιτυχία, με σχετική απομόνωση και διαρροή από το other stem.

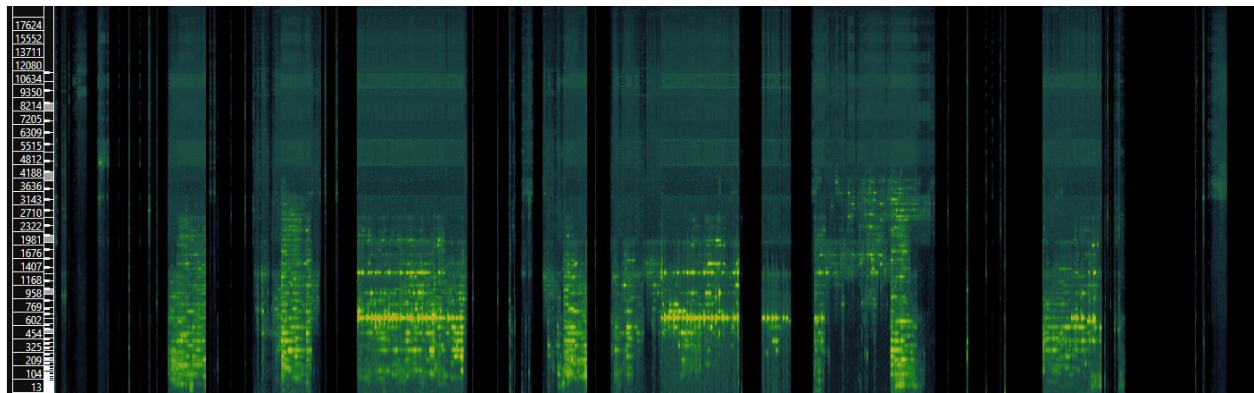
Αξίζει να σημειωθεί ότι η ύπαρξη αντήχησης και άλλων επιπρόσθετων εφέ, τα οποία είναι ενσωματωμένα στο αρχικό mixing, προσθέτουν μία επιπλέον δυσκολία στο διαχωρισμό. Λαμβάνοντας υπόψη τα παραπάνω και τα αποτελέσματα της ακρόασης, η διαρροή κρίνεται ως 1^ο βαθμού ενώ η καθαρότητα χαρακτηρίζεται ως 2ου βαθμού, λόγω της συχνότερης από την αναμενόμενη παρουσίας στοιχείων από το other stem.

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 103: Waveform of piano stem (Sonic Visualiser)



Εικόνα 104: Spectrogram of piano stem (Sonic Visualiser)

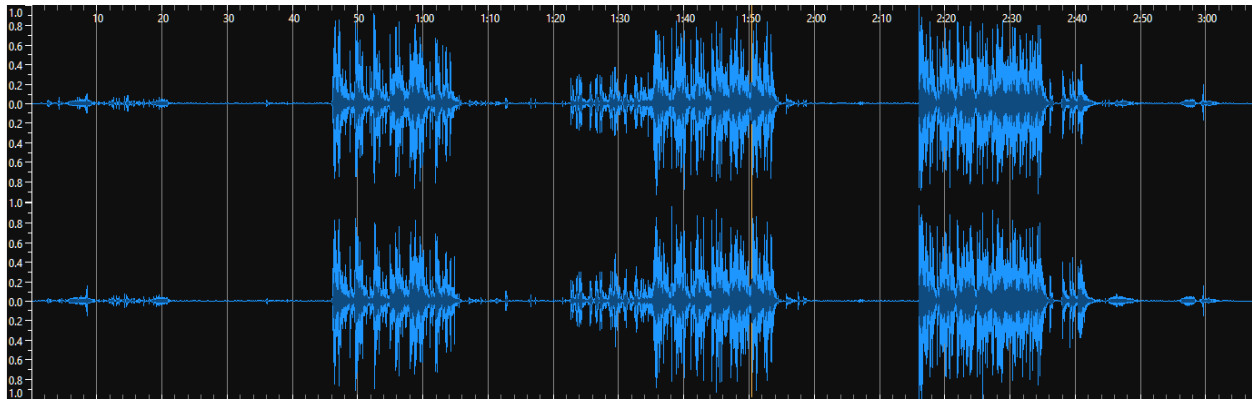
Αρχικά, πρέπει να διασαφηνίσουμε ότι για την περίπτωση του πιάνου, αυτό δεν υπάρχει σαν όργανο στο συγκεκριμένο παράδειγμα. Παρόλα αυτά, τόσο η κυματομορφή όσο και το φασματογράφημα υποδεικνύουν την παρουσία διαρροής στο συγκεκριμένο stem, το οποίο θεωρητικά θα έπρεπε να είναι κενό. Κατόπιν της ακρόασης, η διαρροή είναι αρκετά αισθητή, καθώς εντοπίζονται ευδιάκριτα στοιχεία από το other stem σε πολλαπλά σημεία.

Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα		X		
Διαρροή		X		
Π.Δ.				0,5

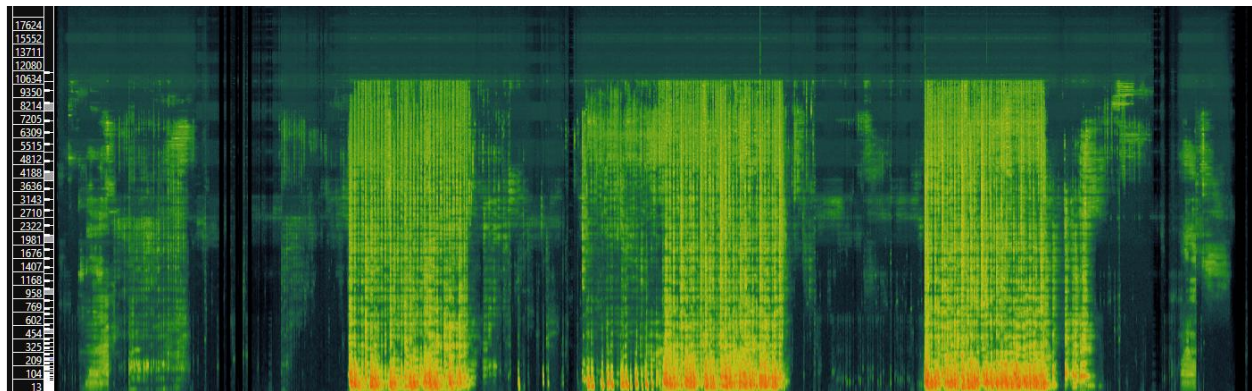
Συνεπώς, η διαρροή αξιολογείται ως 1^{ου} βαθμού, ενώ η καθαρότητά χαρακτηρίζεται και αυτή ως 1^{ου} βαθμού, λόγω της συχνής παρέμβασης από μέρη του other σε σχετικά αισθητή ένταση.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 105: Waveform of drums stem (Sonic Visualiser)



Εικόνα 106: Spectrogram of drums stem (Sonic Visualiser)

Τα κρουστά στο συγκεκριμένο παράδειγμα είναι ποικίλα και ηχούν σε αρκετά σημεία. Το φασματογράφημα μας δίνει ξεκάθαρα την απεικόνιση η οποία ανταποκρίνεται στον χαρακτήρα των κρουστών, χωρίς όμως να μπορούμε να αντλήσουμε κάποια επιπλέον πληροφορία όσο αναφορά τη διαρροή. Ακουστικά, το drum stem έχει διαχωριστεί σε γενικές γραμμές σωστά με ελάχιστα σημεία λίγο «θολά» και τέλος με πολύ μικρή διαρροή από μέρος των εγχόρδων. Λαμβάνοντας υπόψη τα παραπάνω, η διαρροή καθίσταται ως 1^{ου} βαθμού και η καθαρότητα ως 2^{ου}.

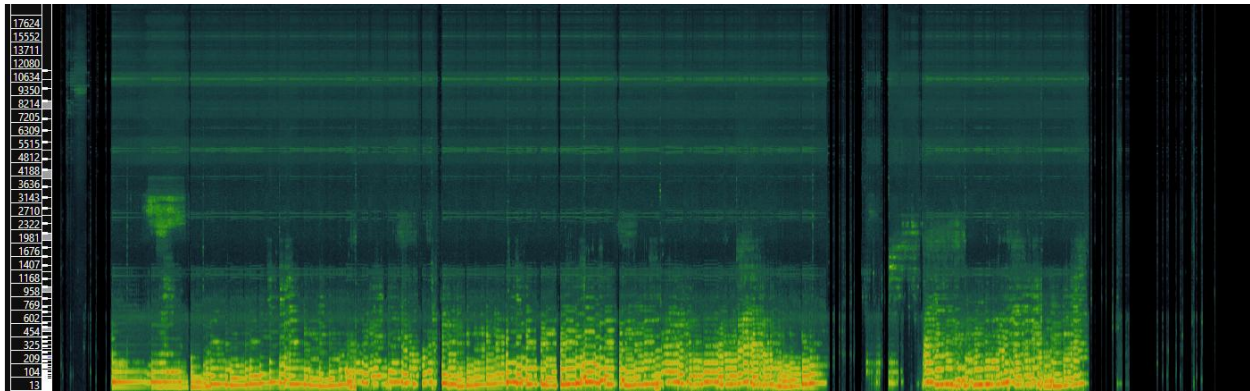
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 107: Waveform of bass stem (Sonic Visualiser)



Εικόνα 108: Spectrogram of bass stem (Sonic Visualiser)

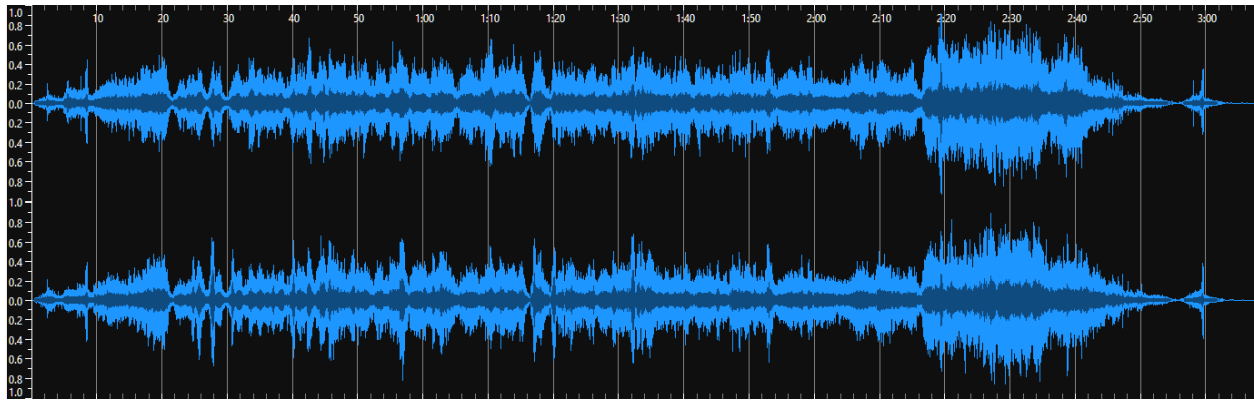
Ακουστικά, το stem του bass έχει διαχωριστεί σωστά σε σύγκριση με τα υπόλοιπα stems με σχεδόν μηδενική διαρροή εκτός του αναμενόμενου θορύβου. Και σε αυτό το stem, το φασματογράφημα μας δίνει ξεκάθαρα την εικόνα που ανταποκρίνεται στην απεικόνιση των συχνοτήτων του μπάσου, χωρίς όμως να μπορούμε να αντλήσουμε κάποια επιπλέον πληροφορία όσο αναφορά τη διαρροή.

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα			X	
Διαρροή	X			
Π.Δ.	0,83			

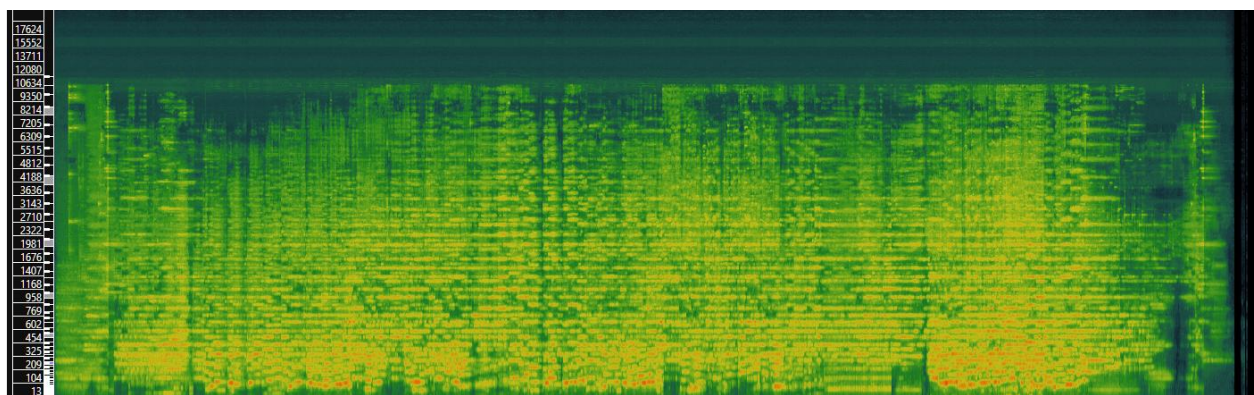
Κατά την ακουστική αξιολόγηση, θα μπορούσε να ειπωθεί ότι πράγματι η απομόνωση είναι αρχικά χωρίς παρεμβολές από τα υπόλοιπα stems, δεδομένου ότι αυτό το οποίο ακούγεται είναι ένας συνδυασμός ηλεκτρονικών μπάσων συχνοτήτων και των εγχόρδων. Το μόνο αρνητικό που εντοπίζεται είναι η χαμηλή ένταση και η θολότητα του stem σε ορισμένα σημεία. Για τον λόγο αυτό, στην ΚΚΔ ο διαχωρισμός κρίνεται με διαρροή μηδενικού βαθμού και η καθαρότητα ως 2^{ου} βαθμού.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 109: Waveform of other stem (Sonic Visualiser)



Εικόνα 110: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other αναμενόταν η συγκέντρωση των υπόλοιπων μελωδικών στοιχείων, κυρίως των εγχόρδων και ηλεκτρονικών ήχων. Από τη φύση του το συγκεκριμένο stem περιλαμβάνει ό,τι δεν μπόρεσε να διαχωριστεί προγενέστερα στα υπόλοιπα 4 stems. Ακουστικά, ως πρώτο θετικό στοιχείο το μπάσο και τα κρουστά απουσιάζουν πλήρως. Δυστυχώς όμως, παρατηρείται η παρουσία μικρών τμημάτων της φωνής, τα οποία, αν και βρίσκονται σε πολύ χαμηλή ένταση, γίνονται αντιληπτά κατά την διαδικασία της ακρόασης. Αυτό συνεπάγεται ακουστικά με καθαρότητα 2^{ον} βαθμού και με διαρροή 1^{ον}.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Συμπεράσματα Παραδείγματος

Ο διαχωρισμός στο συγκεκριμένο έργο απέδωσε ικανοποιητικά αποτελέσματα, ιδιαίτερα όσον αφορά τον διαχωρισμό της φωνής. Ακουστικά, στο stem των vocals παρατη-

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			

ρήθηκαν καθαρές περιοχές της φωνής και με σχετική ευκρίνεια όμως υπήρξε συχνότερη από την αναμενόμενη παρουσία στοιχείων από το other stem. Το εργαλείο κατόρθωσε να εντοπίσει σωστά το φωνητικό περιεχόμενο, αλλά δεν ήταν σε θέση να απομονώσει πλήρως τη φωνή από τα συνοδευόμενα εφέ, κάτι που είναι απόλυτα αναμενόμενο για το στάδιο εξέλιξης του μοντέλου.

Όσον αφορά τα drums και το other, το μοντέλο ενσωμάτωσε και θόρυβο με αποτέλεσμα «θολά» τμήματα στο διαχωρισμό. Το αντίθετο συνέβη στο stem του piano, το οποίο δεν υπήρχε σαν όργανο όμως είχε εμφανή διαρροή. Τέλος, το bass stem ήταν το μόνο που είχε μηδενική διαρροή όμως η καθαρότητά του δεν ήταν στο 3.

Piano	Καθαρότητα		X		
	Διαρροή		X		
	Π.Δ.	0,5			
Drums	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Bass	Καθαρότητα			X	
	Διαρροή	X			
	Π.Δ.	0,83			
Other	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			

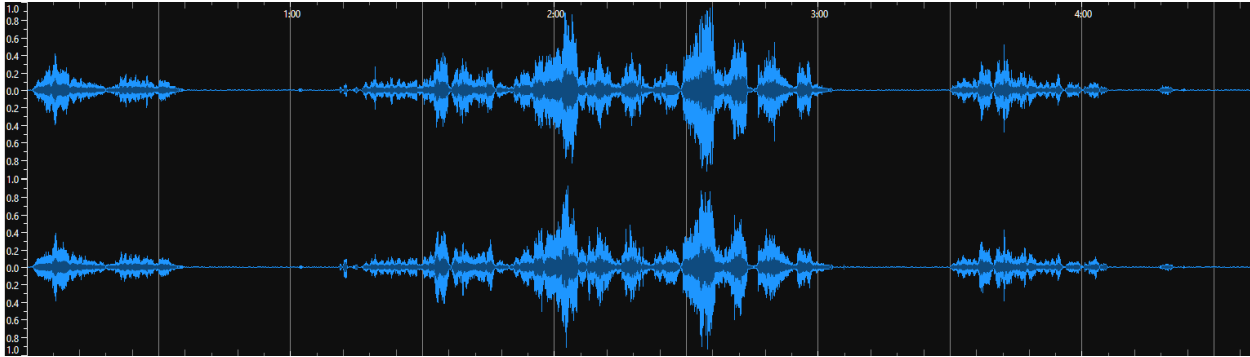
4.12 Μουσικό παράδειγμα 12: Ola Gjeilo: Across The Vast, Eternal Sky

Το δωδέκατο μουσικό παράδειγμα που αναλύθηκε είναι το Across The Vast, Eternal Sky του Ola Gjeilo. Πρόκειται για ένα κατεξοχήν φωνητικό έργο για χορωδία με μορφή SATB, συνοδευόμενο από ορχήστρα εγχόρδων και πιάνο. Το έργο το χαρακτηρίζει η πλούσια αρμονία του και οι εναλλαγές έντασης μεταξύ των επιμέρους μερών του. Επιλέχθηκε διότι περιλαμβάνει τρία βασικά stems (vocals, piano και other), επιτρέποντας την αξιολόγηση της αποτελεσματικότητας του διαχωρισμού σε πολυφωνικό χορωδιακό περιβάλλον.

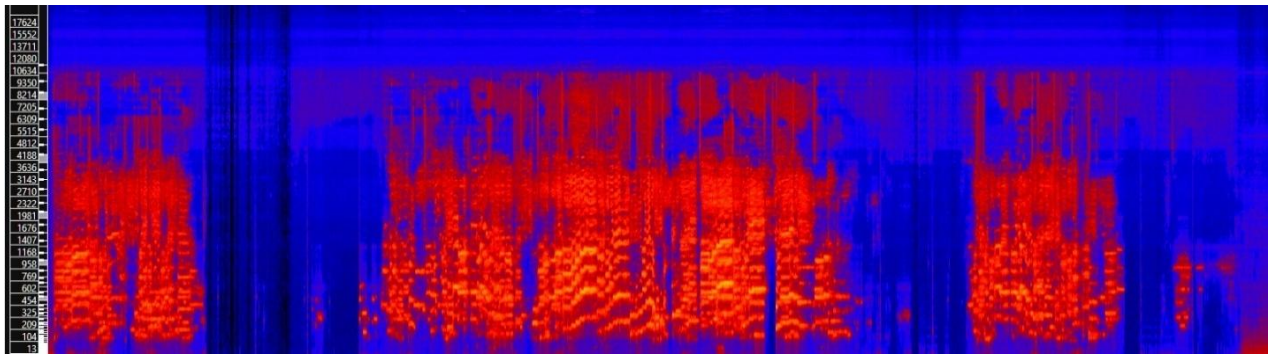
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη στο Spleeter είχε διάρκεια περίπου 3 λεπτών και 7 δευτερολέπτων. Ο διαχωρισμός έγινε χωρίς καμία ανάγκη για περαιτέρω επεξεργασία ή μετατροπή στο πρωτογενές αρχείο. Η εξαγωγή ολοκληρώθηκε χωρίς τεχνικά προβλήματα (επαναλαμβανόμενες προσπάθειες, «crush» του εργαλείου κτλ.), και για τα πέντε παραγόμενα stems καθώς έγινε η αποθήκευσή τους στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 12). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 111: Waveform of vocal stem (Sonic Visualiser)



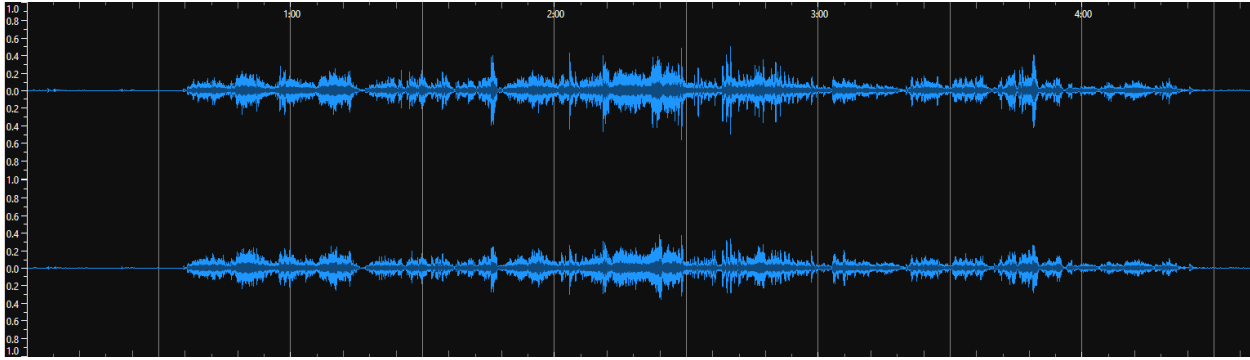
Εικόνα 112: Spectrogram of vocal stem (Sonic Visualiser)

Λόγω της συνεχούς παρουσίας της φωνής καθ' όλη σχεδόν τη διάρκεια του έργου, το stem των φωνητικών δεν μπορεί να αξιολογηθεί πλήρως μέσω του Sonic Visualiser. Παρόλα αυτά, μπορούμε να αντλήσουμε πληροφορίες σχετικά με το τι μπορεί να αναμένει κανείς κατά την ακρόαση. Στην περίπτωση αυτού του stem, η αξιολόγηση βασίζεται κυρίως στην ακρόαση. Κατόπιν αυτής, διαπιστώνεται ότι ο διαχωρισμός της φωνής δεν επιτεύχθηκε στο βαθμό που θα αναμενόταν. Η καθαρότητα βρίσκεται στο 1, καθώς δεν υπάρχει σωστή απομόνωση και των τεσσάρων φωνών ενώ πρόκειται για χορωδιακό έργο. Επιπλέον, εντοπίζεται η παρουσία θολότητας και θορύβου. Παράλληλα, δεν συμπεριλαμβάνεται πληροφορία προερχόμενη από τα stems του piano και του other, γεγονός το οποίο οδηγεί σε διαρροή μηδενικού βαθμού.

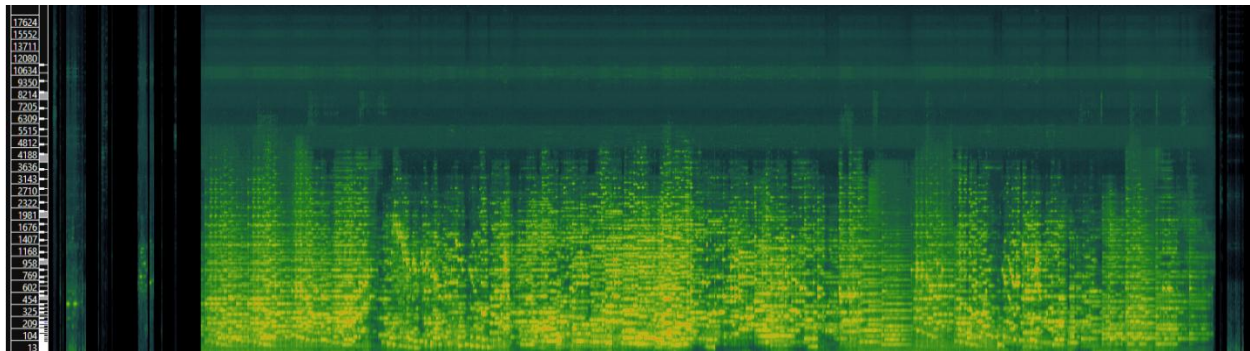
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα		X		
Διαρροή	X			
Π.Δ.	0,5			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 113: Waveform of piano stem (Sonic Visualiser)



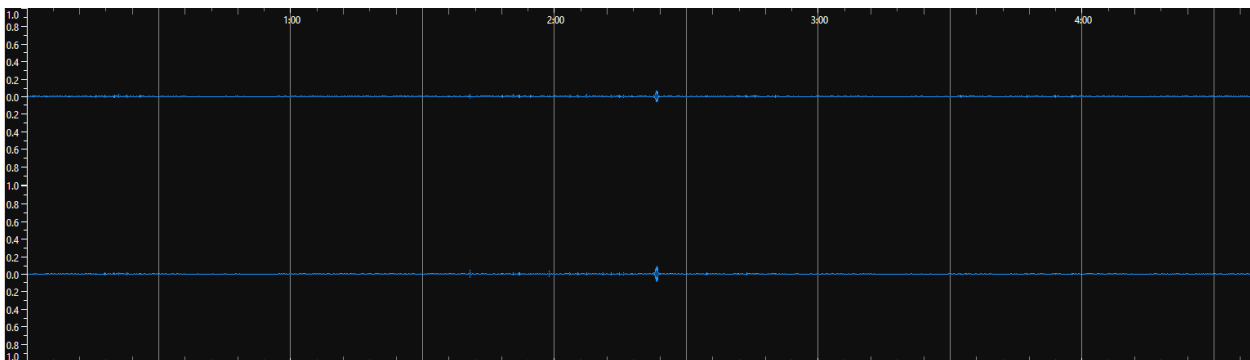
Εικόνα 114: Spectrogram of piano stem (Sonic Visualiser)

Το πιάνο είναι παρόν σχεδόν σε ολόκληρο το έργο (εκτός της αρχής) ως συνοδευτικό των εγχόρδων και της φωνής, ενώ κάποιες φορές ηχεί και μόνο του. Το φασματογράφημα πράγματι παραπέμπει στο διάστημα συχνοτήτων του πιάνου μεταξύ 27,5 Hz και 4.186 Hz (Fletcher & Rossing, 1991), ωστόσο δεν μπορούμε να ανιχνεύσουμε κάποια άλλη σημαντική πληροφορία που να αφορά την διαρροή. Ακουστικά, το piano stem αποδίδει σχεδόν αποκλειστικά το πιάνο, το οποίο όμως εμφανίζεται θολό σε ορισμένα σημεία, ενώ εντοπίζεται και μικρή διαρροή από τα έγχορδα σε χαμηλή ένταση. Συνολικά η καθαρότητα εκτιμάτε ως 2^ο βαθμού, ενώ η διαρροή ως 1^ο.

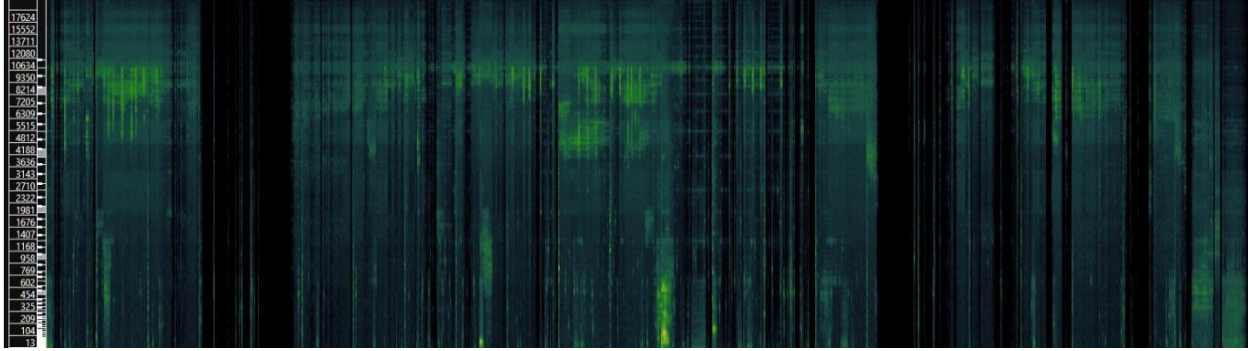
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 115: Waveform of drums stem (Sonic Visualiser)



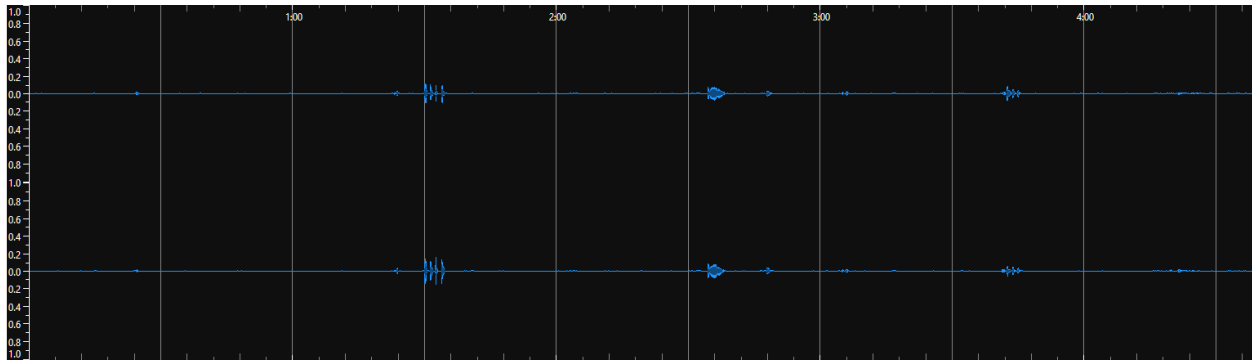
Εικόνα 116: Spectrogram of drums stem (Sonic Visualiser)

Στο συγκεκριμένο έργο δεν υπάρχουν κρουστά, συνεπώς το stem των drum θα έπρεπε να είναι εντελώς κενό. Ωστόσο, σύμφωνα με την κυματομορφή και το φασματογράφημα, εντοπίζονται συχνότητες σε πολύ χαμηλή ένταση γεγονός που υποδεικνύει ελάχιστη διαρροή. Ακουστικά, διακρίνονται σε πολύ σιγανή ένταση θολά τμήματα, κυρίως με θόρυβο αλλά όχι σημεία άλλων stems. Αυτό συνεπάγεται με μηδενική διαρροή και 3^ο βαθμού καθαρότητα, καθώς υπάρχει μόνο ελάχιστος θόρυβος τον οποίο κάποιος μπορεί να τον εντοπίσει κάτω από ιδανικές συνθήκες ακρόασης: με ποιοτικά ακουστικά, σε υψηλή ένταση και σε απόλυτα ήσυχο περιβάλλον.

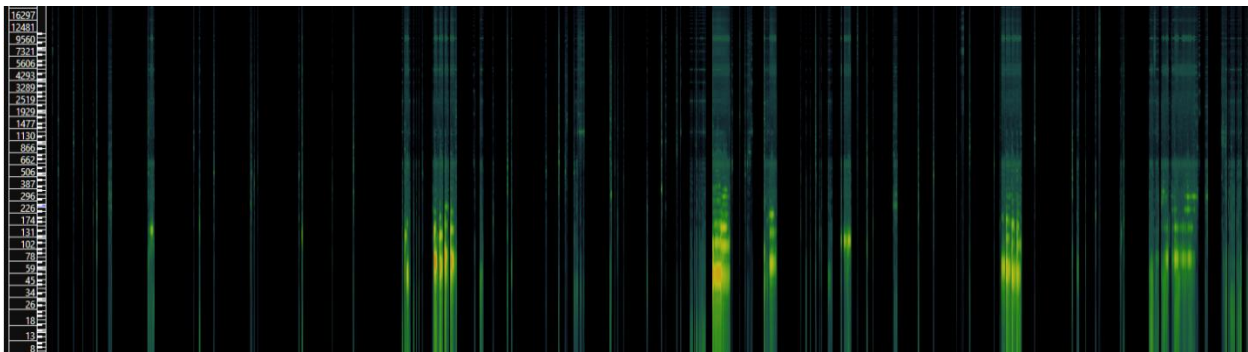
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 117: Waveform of bass stem (Sonic Visualiser)



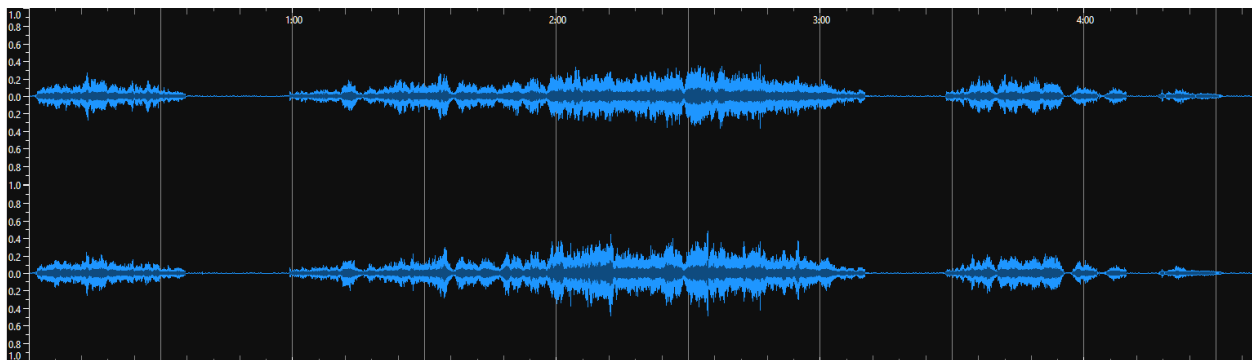
Εικόνα 118: Spectrogram of bass stem (Sonic Visualiser)

Το συγκεκριμένο παράδειγμα παρουσιάζει μία μικρή δυσκολία στην εύρεση του μπάσου καθώς το κοντραμπάσο δεν ακολουθεί σταθερά κάποιο ρυθμικό μοτίβο. Αντίθετα, εναλλάσσει ρόλους: αρχικά ως αρμονικό μέρος των εγχόρδων, αργότερα ως ισοκράτημα πάνω στην αρμονία και τέλος αναλαμβάνει τον ρόλο του αρμονικού μπάσου με την τεχνική «rizzicato». Το μοντέλο αναμένουμε να το εντοπίσει κυρίως με την τρίτη μορφή εκτέλεσης. Κατόπιν της ακρόασης, πράγματι διαπιστώνεται ότι τα σημεία που απομονώνει με μεγαλύτερη ευκρίνεια είναι αυτά με το rizzicato, ωστόσο ούτε σε αυτήν την περίπτωση η ανίχνευση είναι συνεπής σε όλη τη διάρκεια του έργου. Τέλος, διαρροή από άλλα stem δεν προκύπτει. Επομένως, το συγκεκριμένο stem αξιολογείται με 1^ο βαθμού καθαρότητα και μηδενική διαρροή.

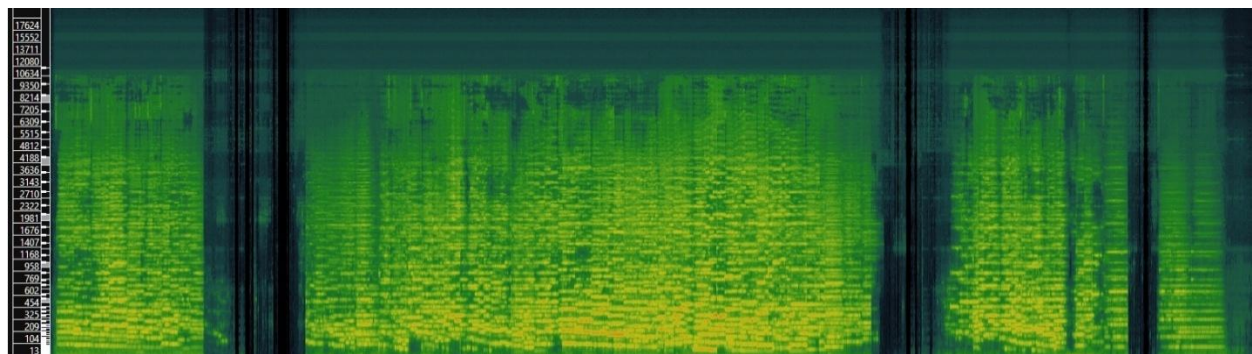
Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα		X		
Διαρροή	X			
Π.Δ.	0,66			

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 119: Waveform of other stem (Sonic Visualiser)



Εικόνα 120: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other, στο οποίο αναμενόταν να συγκεντρωθούν αποκλειστικά τα έγχορδα, περιλαμβάνονται ποικίλα θραύσματα από τα υπόλοιπα stems. Η ακρόαση καταδεικνύει ότι ο διαχωρισμός των ίδιων των εγχόρδων δεν είναι σε καμία περίπτωση πλήρης, καθώς σε αρκετά σημεία εμφανίζονται «θολά» και ανεπαρκώς απομονωμένα. Επιπλέον,

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα		X		
Διαρροή			X	
Π.Δ.	0,33			

ανιχνεύεται σημαντική διαρροή των φωνητικών, του πιάνου και του μπάσου, η οποία καθιστά το stem ιδιαίτερα επιβαρυσμένο σε αρκετά σημεία. Το παραπάνω οδηγεί στην αξιολόγηση της συνολικής διαρροής ως 2^ο, καθώς η απουσία των κρουστών αποτρέπει την αξιολόγηση της διαρροής ως 3ου βαθμού. Τέλος, σύμφωνα με τα παραπάνω, η καθαρότητα περιορίζεται στο 1.

Συμπεράσματα Παραδείγματος

Το πρώτο συμπέρασμα, το οποίο προέκυψε από το παράδειγμα 12, είναι ότι το μοντέλο διαχωρισμού δυσκολεύεται σημαντικά στην απομόνωση του vocal stem όταν πρόκειται για πολυφωνία, εξίσου και η καθαρότητα 1^ο βαθμού. Μέχρι τώρα, στα έργα που αναλύθηκαν η φωνή αποτελούσε συνήθως ένα από τα πιο «καθαρά» stems μετά τον διαχωρισμό μέσω του Spleeter. Σε αυτό το παράδειγμα αν και διαχωρίστηκε, δεν ήταν ξεκάθαρες και οι 4 χορωδιακές φωνές.

Παρόμοια προβλήματα παρατηρήθηκαν και στο stem του other, όπου ο διαχωρισμός ήταν ιδιαίτερα ανεπαρκής και μάλιστα με εκτεταμένη διαρροή από το πιάνο, το μπάσο και τη φωνή. Το συγκεκριμένο stem εμφάνισε τη μεγαλύτερη αποτυχία από τα υπόλοιπα κατά τον διαχωρισμό. Αντίθετα, το stem του ριάνο αναδείχθηκε ως το πιο καθαρό ακουστικά (καθαρότητα 2^ο βαθμού), με σχετικά ικανοποιητικό διαχωρισμό για τα δεδομένα του συγκεκριμένου παραδείγματος.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα		X		
	Διαρροή	X			
	Π.Δ.	0,5			
Piano	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Drums	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Bass	Καθαρότητα		X		
	Διαρροή	X			
	Π.Δ.	0,66			
Other	Καθαρότητα		X		
	Διαρροή			X	
	Π.Δ.	0,33			

4.13 Μουσικό παράδειγμα 13: Ola Gjeilo: Dark & Luminous Night

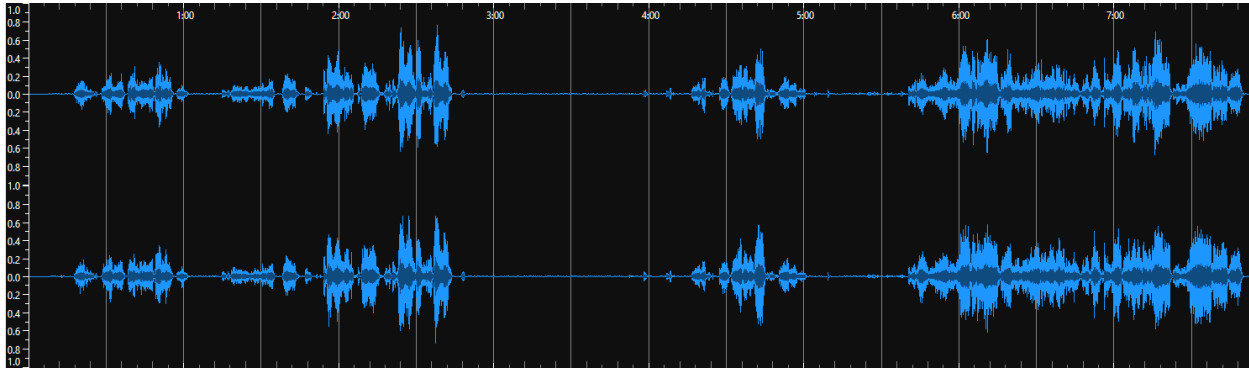
Το δέκατο τρίτο μουσικό παράδειγμα που αναλύθηκε είναι το Dark & Luminous Night του Ola Gjeilo. Πρόκειται και αυτό για ένα κατεξοχήν φωνητικό έργο για χορωδία με μορφή SATB, συνοδευόμενο από ορχήστρα εγχόρδων και πιάνο. Το έργο το χαρακτηρίζει η πλούσια αρμονία του και οι εναλλαγές έντασης μεταξύ των επιμέρους μερών του, αλλά επιπλέον και οι χαρακτηριστικές εναλλαγές τονικότητας και τέμπο. Επιλέχθηκε διότι περιλαμβάνει τρία βασικά stems (vocals, piano και other), επιτρέποντας την αξιολόγηση της αποτελεσματικότητας του διαχωρισμού σε πολυφωνικό χορωδιακό περιβάλλον όπως και το παράδειγμα 4.12.

Για την ανάλυση εφαρμόστηκε το μοντέλο των 5 stems. Το αρχείο είχε διάρκεια περίπου 7 λεπτών και 53 δευτερολέπτων. Ο διαχωρισμός πραγματοποιήθηκε χωρίς περαιτέρω μετατροπή του πρωτογενούς αρχείου και τα stems αποθηκεύτηκαν στον αντίστοιχο φάκελο της εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα

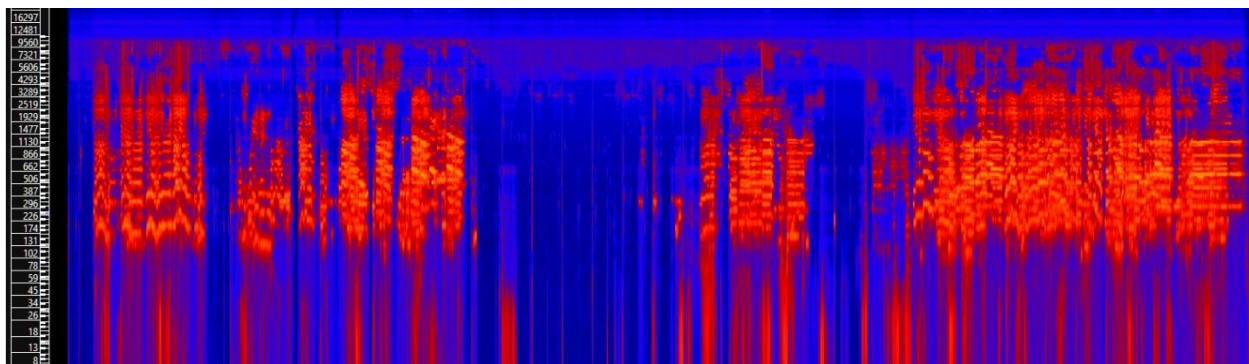
13). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 121: Waveform of vocal stem (Sonic Visualiser)



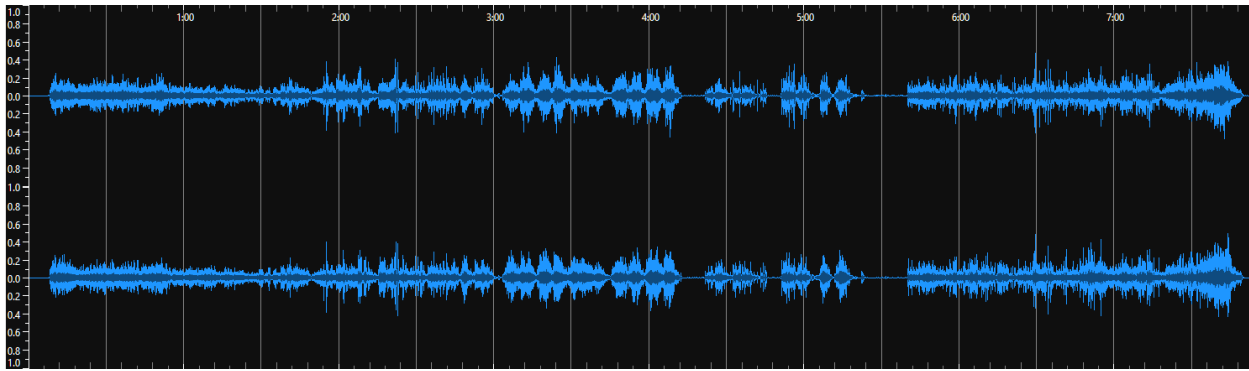
Εικόνα 122: Spectrogram of vocal stem (Sonic Visualiser)

Λόγω της συνεχούς παρουσίας της φωνής καθ' όλη σχεδόν τη διάρκεια του έργου, το stem των φωνητικών δεν μπορεί να αξιολογηθεί πλήρως μέσω του Sonic Visualiser. Παρόλα αυτά, μπορούμε να αντλήσουμε πληροφορίες σχετικά με το τι μπορεί να αναμένει κανείς κατά την ακρόαση. Στην περίπτωση αυτού του stem, η αξιολόγηση βασίζεται κυρίως στην ακρόαση. Κατόπιν αυτής, διαπιστώνεται ότι ο διαχωρισμός της φωνής δεν επιτεύχθηκε στο βαθμό που θα αναμενόταν. Η καθαρότητα βρίσκεται στο 1, καθώς δεν υπάρχει σωστή απομόνωση και των τεσσάρων φωνών ενώ πρόκειται για χορωδιακό έργο. Επιπλέον, εντοπίζεται έντονη η παρουσία θολότητας και θορύβου ενώ διακόπτεται συνεχώς το stem. Παράλληλα, συμπεριλαμβάνεται πληροφορία προερχόμενη από το stems του piano και του other, γεγονός το οποίο οδηγεί σε διαρροή 2^{ου} βαθμού.

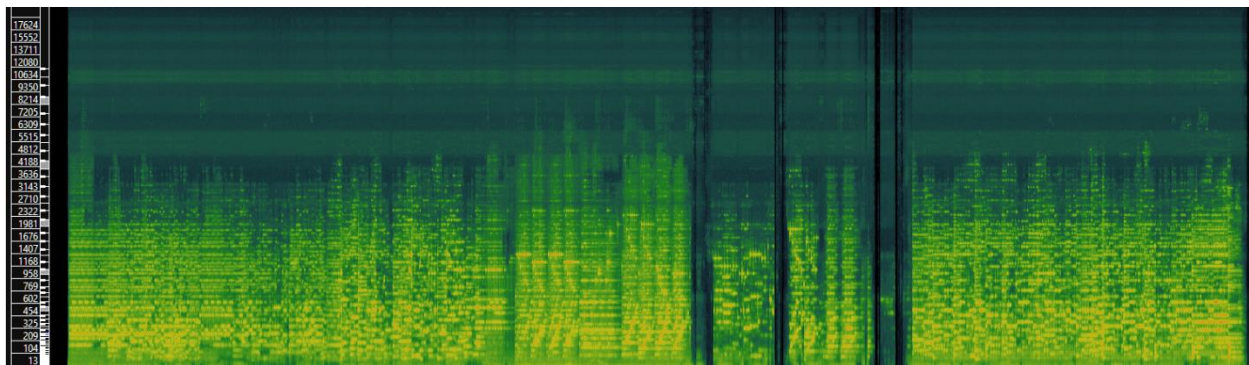
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα		X		
Διαρροή			X	
Π.Δ.	0,33			

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 123: Waveform of piano stem (Sonic Visualiser)



Εικόνα 124: Spectrogram of piano stem (Sonic Visualiser)

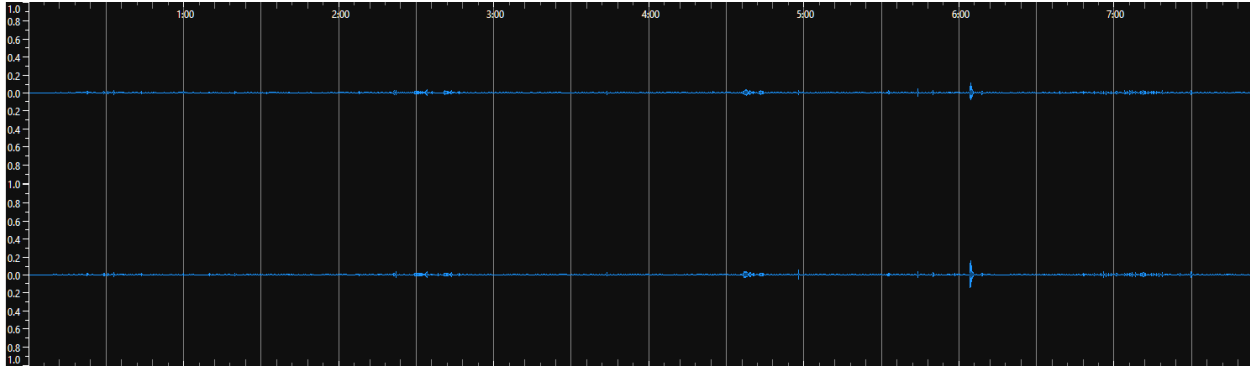
Το πιάνο είναι παρόν σε ολόκληρο το έργο, λειτουργώντας κυρίως ως συνοδευτικό των εγχόρδων και της φωνής, ενώ σε ορισμένα σημεία ηχεί και μόνο του. Το φασματογράφημα πράγματι παραπέμπει στο διάστημα συχνοτήτων του πιάνου μεταξύ 27,5 Hz και 4.186 Hz σύμφωνα και με τη βιβλιογραφία. (Fletcher & Rossing, 1991) Ωστόσο, παρατηρείται ότι σε τρία σημεία διακόπτεται η απεικόνιση, παρόλο που γνωρίζουμε ότι το όργανο ηχεί αδιάλειπτα σε όλη τη διάρκεια του έργου. Κατά συνέπεια, δεν μπορούμε να ανιχνεύσουμε κάποια άλλη σημαντική πληροφορία που να αφορά την διαρροή ή την καθαρότητα.

Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα		X		
Διαρροή	X			
Π.Δ.	0,66			

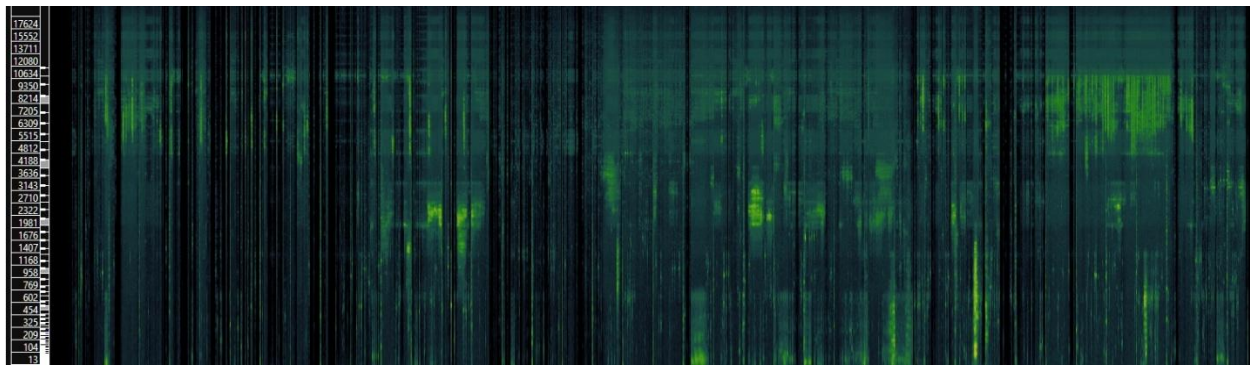
Ακουστικά, το piano stem αποδίδει σχεδόν αποκλειστικά το πιάνο, το οποίο όμως εμφανίζεται «θολό» σε πολλαπλά σημεία, ενώ κατά διαστήματα αυτό δεν εντοπίζεται σωστά, με ακουστικό αποτέλεσμα την εξαγωγή θορύβου αντί για καθαρό πιανιστικό περιεχόμενο. Τέλος, δεν εντοπίζεται διαρροή από τα κάποιο άλλο stem. Συνολικά η καθαρότητα εκτιμάτε ως 1^ο βαθμού, ενώ η διαρροή αξιολογείται ως μηδενικού.

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 125: Waveform of drums stem (Sonic Visualiser)



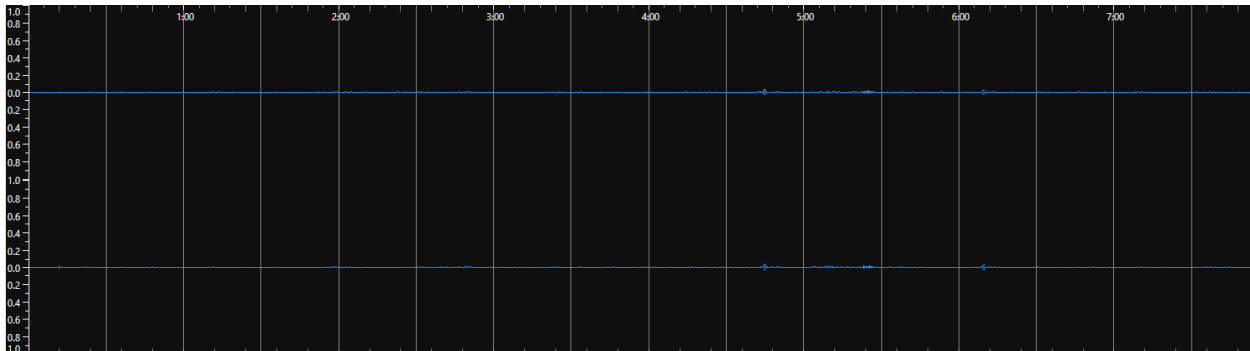
Εικόνα 126: Spectrogram of drums stem (Sonic Visualiser)

Στο συγκεκριμένο έργο δεν υπάρχουν κρουστά, συνεπώς το stem των drum θα έπρεπε να είναι εντελώς κενό. Ωστόσο, σύμφωνα με την κυματομορφή και το φασματογράφημα, εντοπίζονται συχνότητες σε πολύ χαμηλή ένταση γεγονός που υποδεικνύει ελάχιστη διαρροή. Ακουστικά, διακρίνονται σε πολύ σιγανή ένταση θολά τμήματα, κυρίως με θόρυβο αλλά όχι σημεία άλλων stems. Αυτό συνεπάγεται με μηδενική διαρροή και 3^ο βαθμού καθαρότητα, καθώς υπάρχει μόνο ελάχιστος θόρυβος τον οποίο κάποιος μπορεί να τον εντοπίσει κάτω από ιδανικές συνθήκες ακρόασης: με ποιοτικά ακουστικά, σε υψηλή ένταση και σε απόλυτα ήσυχο περιβάλλον.

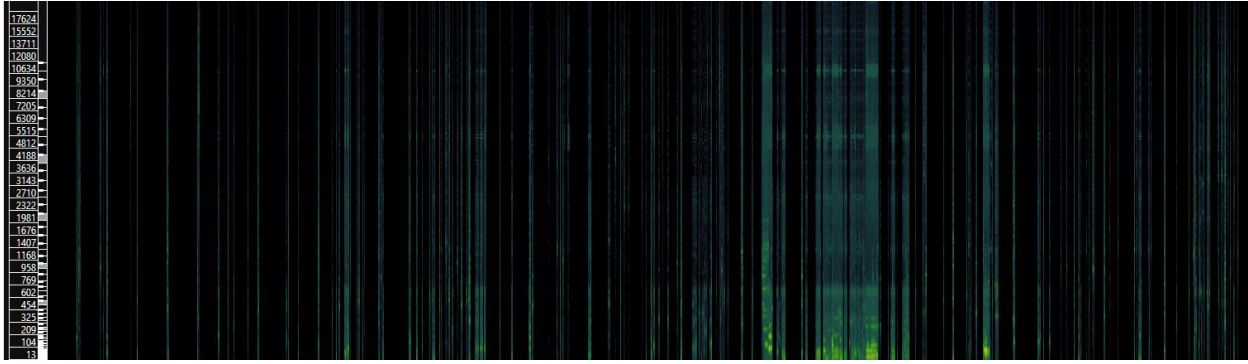
Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα				X
Διαρροή	X			
Π.Δ.	1			

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 127: Waveform of bass stem (Sonic Visualiser)



Εικόνα 128: Spectrogram of bass stem (Sonic Visualiser)

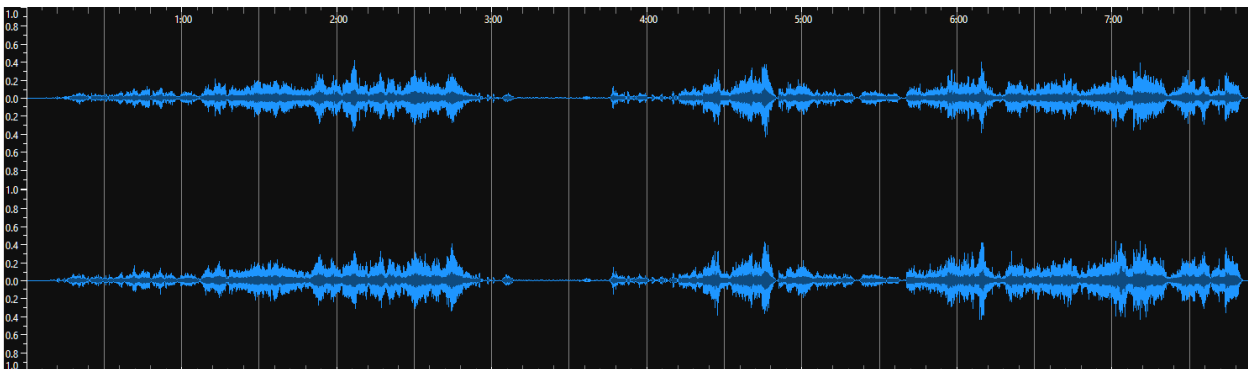
Παρομοίως με το παράδειγμα 4.12 ο διαχωρισμός του bass stem παρουσιάζει μία μικρή δυσκολία καθώς το κοντραμπάσο δεν ακολουθεί σταθερά κάποιο ρυθμικό μοτίβο. Αντίθετα, εναλλάσσει ρόλους: αρχικά ως αρμονικό μέρος των εγχόρδων, αργότερα ως ισοκράτημα πάνω στην αρμονία και τέλος αναλαμβάνει τον ρόλο του αρμονικού μπάσου με την τεχνική «pizzicato».

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα				
Διαρροή				
Π.Δ.				

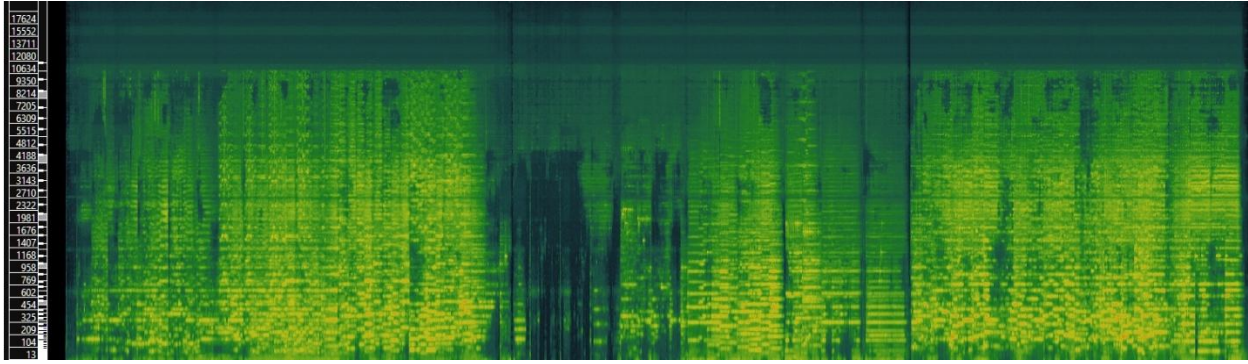
Ακουστικά, διαπιστώνεται ότι ο συνθέτης αξιοποιεί τους δύο πρώτους τρόπους στην γραφή του κοντραμπάσου. Συνεπώς, ούτε σε αυτό το παράδειγμα ανιχνεύτηκε το μπάσο σε όλη τη διάρκεια του έργου από το μοντέλο διαχωρισμού. Τέλος, διαρροή από άλλα stem δεν προκύπτει. Ωστόσο, λόγω της μη ανίχνευσης του κοντραμπάσου, η ΚΚΔ δεν έχει πρακτική εφαρμογή στην περίπτωση αυτού του stem. Επομένως, αν και η διαρροή αξιολογείται ως μηδενικού βαθμού, το stem καθίσταται ουσιαστικά μη αξιοποιήσιμο, καθώς το επιθυμητό όργανο δεν έχει απομονωθεί.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 129: Waveform of other stem (Sonic Visualiser)



Εικόνα 130: Spectrogram of other stem (Sonic Visualiser)

Στο stem του other, στο οποίο αναμενόταν να συγκεντρωθούν αποκλειστικά τα έγχορδα, περιλαμβάνονται ποικίλα θραύσματα από τα υπόλοιπα stems. Η ακρόαση καταδεικνύει ότι ο διαχωρισμός των ίδιων των εγχόρδων δεν είναι σε καμία περίπτωση πλήρης, καθώς σε αρκετά σημεία εμφανίζονται «θολά» και ανεπαρκώς απομονωμένα. Επιπλέον, ανιχνεύεται σημαντική διαρροή των φωνητικών και του πιάνου και του μπάσου, η οποία καθιστά το stem ιδιαίτερα επιβαρυνμένο σε αρκετά σημεία και μάλιστα σε υψηλή ένταση ενώ το πιάνο σε μικρότερο βαθμό.. Το παραπάνω οδηγεί στην αξιολόγηση της συνολικής διαρροής ως 2^ο, καθώς η απουσία των κρουστών αποτρέπει την αξιολόγηση της διαρροής ως 3ου βαθμού. Τέλος, σύμφωνα με τα παραπάνω, η καθαρότητα περιορίζεται στο 1.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα		X		
Διαρροή			X	
Π.Δ.	0,33			

Συμπεράσματα Παραδείγματος

Τα συμπεράσματα που προκύπτουν από το παράδειγμα 13 είναι σε μεγάλο βαθμό αντίστοιχα με εκείνα του παραδείγματος 4.12. Ωστόσο, αξίζει να επισημανθούν δύο διαφορές, οι οποίες εντοπίζονται στα vocal και piano stems. Στην περίπτωση των φωνητικών, έχουμε διαρροή 2^ο βαθμού σε αντίθεση με το προηγούμενο παράδειγμα όπου η διαρροή είχε εκτιμηθεί ως μηδενική. Στην περίπτωση του πιάνου, το stem εμφανίζεται ιδιαίτερα «θολό» σε πολλαπλά σημεία, ενώ κατά διαστήματα το όργανο δεν ανιχνεύεται σωστά, με ακουστικό αποτέλεσμα την εξαγωγή θορύβου αντί για καθαρό πιανιστικό περιεχόμενο. Συνεπώς, η καθαρότητα αξιολογείται στον 1ο βαθμό αντί για τον 2ο, όπως είχε καταγραφεί στο προηγούμενο παράδειγμα.

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα		X		
	Διαρροή			X	
	Π.Δ.	0,33			
Piano	Καθαρότητα		X		
	Διαρροή	X			
	Π.Δ.	0,66			
Drums	Καθαρότητα				X
	Διαρροή	X			
	Π.Δ.	1			
Bass	Καθαρότητα				
	Διαρροή				
	Π.Δ.				
Other	Καθαρότητα		X		
	Διαρροή			X	
	Π.Δ.	0,33			

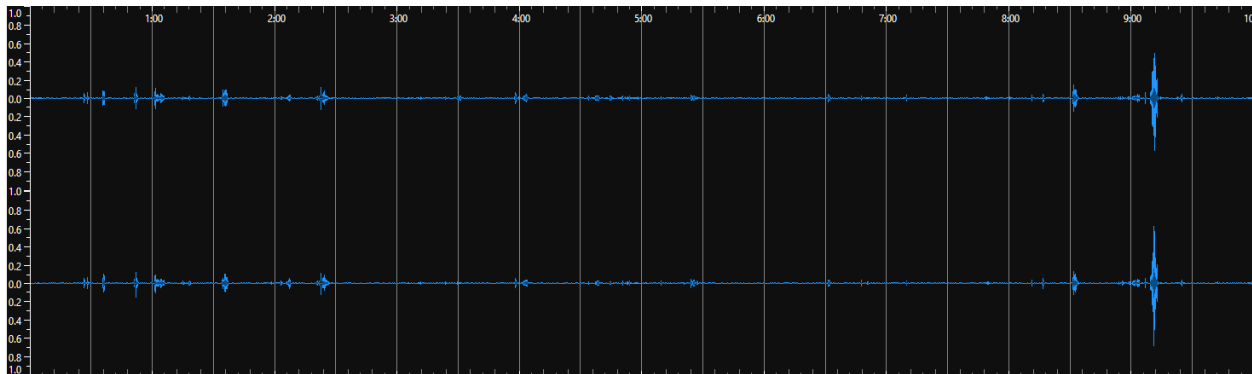
4.14 Μουσικό παράδειγμα 14: S. Rachmaninoff: Symphonic Dances (first movement)

Το δέκατο τέταρτο και τελευταίο μουσικό παράδειγμα που αναλύθηκε είναι πρώτο μέρος του Symphonic Dances του Sergei Rachmaninoff. Πρόκειται για ένα τριμερές συμφωνικό έργο το οποίο συνδυάζει την πλούσια και πολύπλοκη ενορχήστρωση, την δραματική ένταση και τις χαρακτηριστικές «ρομαντικές» μελωδικές γραμμές του συνθέτη. Η σύνθεση παρουσιάζει έντονη ρυθμικότητα, εναλλαγές θεμάτων, καθώς και άρτια κατασκευασμένες αντιθέσεις μεταξύ δυναμικών και λυρικών ενοτήτων. Το συγκεκριμένο έργο αποτελεί ένα ιδιαίτερα απαιτητικό παράδειγμα για ανάλυση, καθώς η ενορχήστρωση περιλαμβάνει πλήρη συμφωνική ορχήστρα, με πνευστά, έγχορδα, κρουστά, άρπα και πιάνο, τα οποία συχνά αλληλεπικαλύπτονται και δημιουργούν πολυσύνθετες μουσικές υφές.

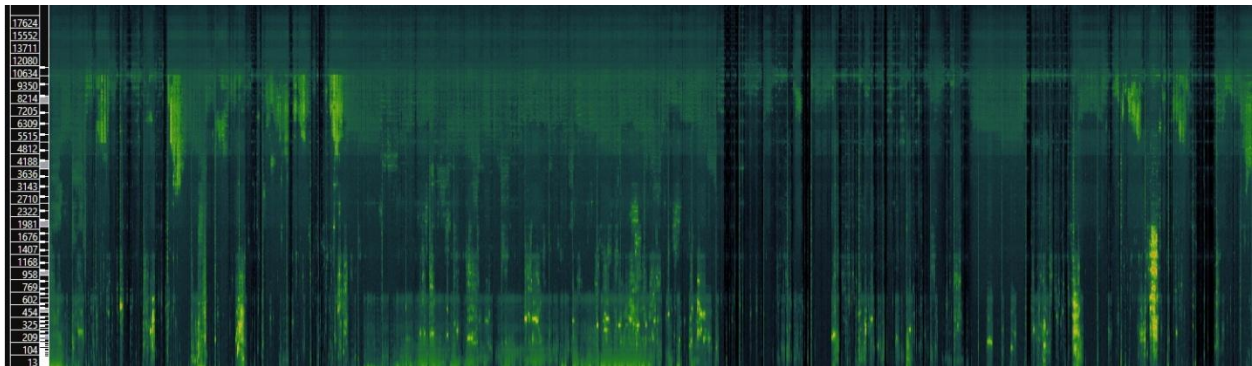
Για την ανάλυση χρησιμοποιήθηκε το μοντέλο των 5 stems. Το αρχείο που εισήχθη στο Spleeter είχε διάρκεια περίπου 12 λεπτών και 4 δευτερολέπτων. Ο διαχωρισμός έγινε χωρίς καμία ανάγκη για περαιτέρω επεξεργασία ή μετατροπή στο πρωτογενές αρχείο. Η εξαγωγή ολοκληρώθηκε με τεχνικά προβλήματα καθώς χρειάστηκε επαναλαμβανόμενες προσπάθειες εισαγωγής του ηχητικού αρχείου στο μοντέλο, λόγω των πέντε «crush» που επήλθαν σε αυτό. Τα παραγόμενα stems αποθηκεύτηκαν στο ανάλογο αρχείο με το όνομα του παραδείγματος που είναι και επισυναπτόμενο στον φάκελο της παρούσας εργασίας (Διπλωματική Εργασία → Sound Files → Παράδειγμα 14). Στη συνέχεια, εισήχθησαν ένα προς ένα στο Sonic Visualiser για περαιτέρω βοήθεια στην ανάλυση και αξιολόγηση της καθαρότητας και διαρροής του κάθε stem.

Vocals

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των vocals.



Εικόνα 131: Waveform of vocal stem (Sonic Visualiser)



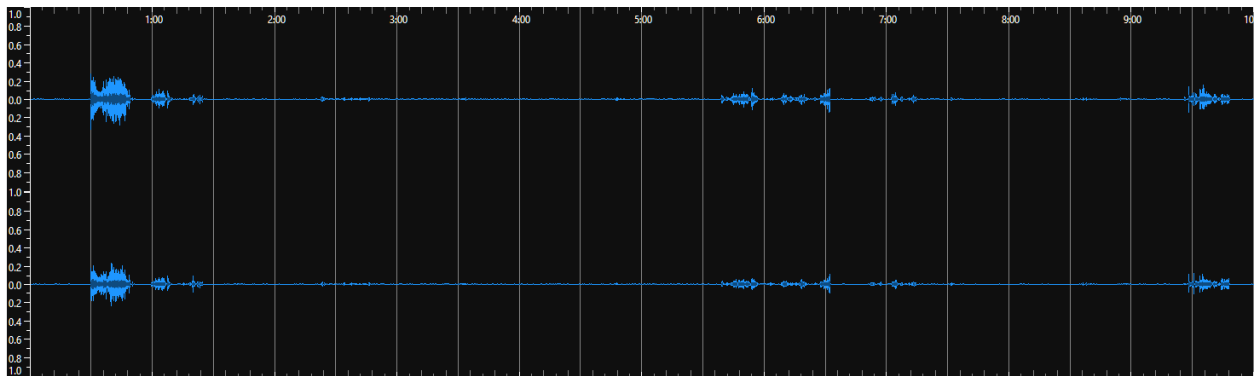
Εικόνα 132: Spectrogram of vocal stem (Sonic Visualiser)

Αρχικά, πρέπει να διασαφηνίσουμε ότι για την περίπτωση της φωνής, αυτή δεν υπάρχει σαν όργανο στο συγκεκριμένο παράδειγμα οπότε θεωρητικά θα έπρεπε να είναι κενό. Παρόλα αυτά, τόσο η κυματομορφή όσο και το φασματογράφημα υποδεικνύουν σημεία σε χαμηλή σχετικά ένταση στα οποία παρατηρείται διαρροή. Κατόπιν της ακρόασης, πράγματι η διαρροή είναι αισθητή, καθώς εντοπίζονται χαμηλά σε ένταση μικρά αποσπάσματα του έργου από τα υπόλοιπα 4 stem στο παρασκήνιο (background). Συνεπώς, διαρροή και καθαρότητα αξιολογούνται ως 2^{ου} βαθμού.

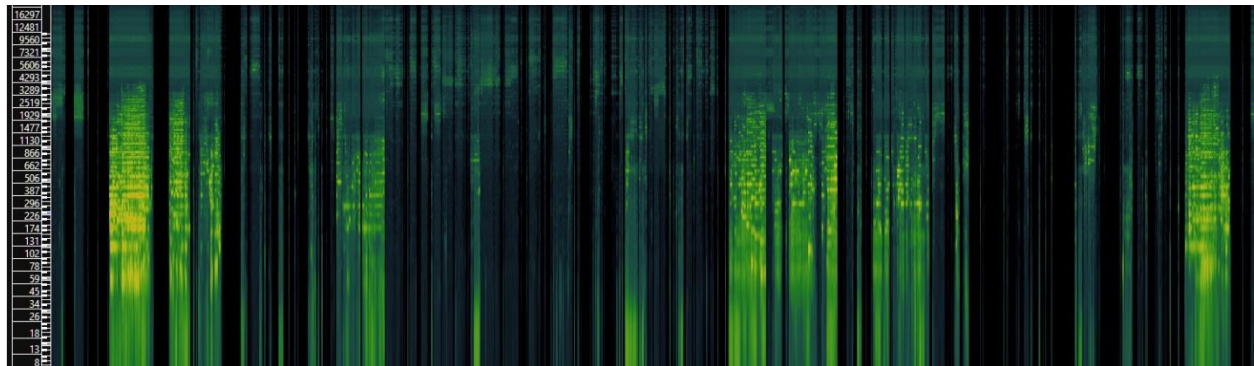
Κλίμακα ΚΚΔ				
Vocals	0	1	2	3
Καθαρότητα			X	
Διαρροή			X	
Π.Δ.				0,5

Piano

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του piano.



Εικόνα 133: Waveform of piano stem (Sonic Visualiser)



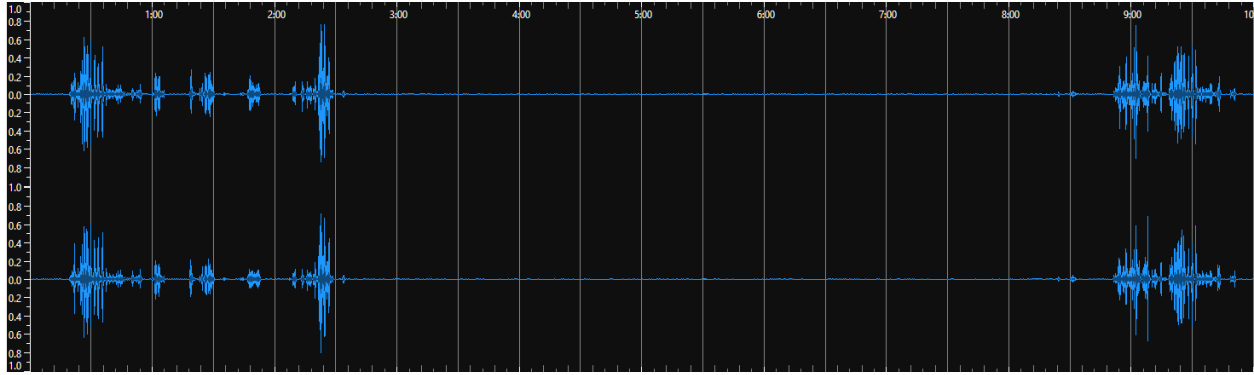
Εικόνα 134: Spectrogram of piano stem (Sonic Visualiser)

Το πιάνο ηχεί σε ελάχιστα σημεία μέσα στο έργο και κυρίως στα «tutti» μέρη. Έχει συνοδευτικό ρόλο και ποτέ δεν ηχεί μόνο του. Το φασματογράφημα πράγματι παραπέμπει στο διάστημα συχνοτήτων του πιάνου μεταξύ 27,5 Hz και 4.186 Hz (Fletcher & Rossing, 1991), ωστόσο δεν μπορούμε να ανιχνεύσουμε κάποια άλλη σημαντική πληροφορία που να αφορά την διαρροή. Ακουστικά, το piano stem αποδίδει σχεδόν αποκλειστικά το πιάνο, το οποίο όμως εμφανίζεται αρκετά θολό σε ορισμένα σημεία, ενώ εντοπίζεται και μικρή διαρροή από τα έγχορδα και τα πνευστά (other) σε χαμηλή ένταση. Συνολικά η καθαρότητα εκτιμάτε ως 1^{ου} βαθμού καθώς και η διαρροή.

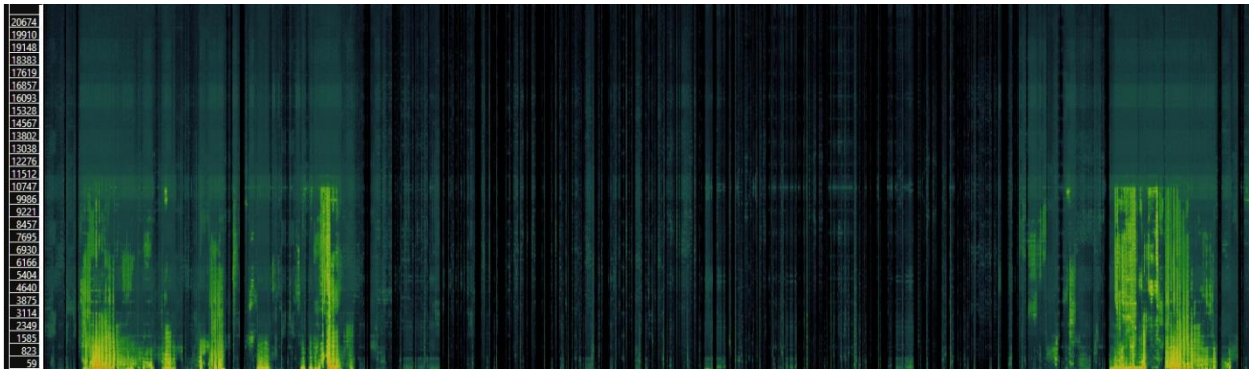
Κλίμακα ΚΚΔ				
Piano	0	1	2	3
Καθαρότητα		X		
Διαρροή		X		
Π.Δ.				0,5

Drums

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem των drums.



Εικόνα 135: Waveform of drums stem (Sonic Visualiser)



Εικόνα 136: Spectrogram of drums stem (Sonic Visualiser)

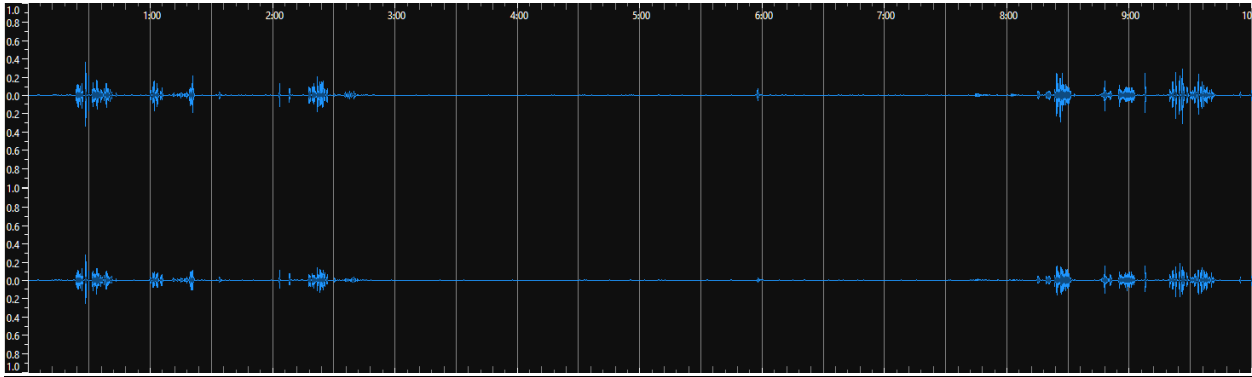
Για το συγκεκριμένο stem, τα ντραμς αναφέρονται σε κρουστά ορχήστρας όπως τα «τυμπάνια» και προφανώς δεν έχουν πρωταγωνιστικό ρόλο. Μέσω του φασματογραφήματος και της κυματομορφής δεν προκύπτουν σαφείς ενδείξεις για τη διαρροή. Συνεπώς, η αξιολόγηση βασίζεται κυρίως στην ακουστική παρατήρηση.

Κλίμακα ΚΚΔ				
Drums	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

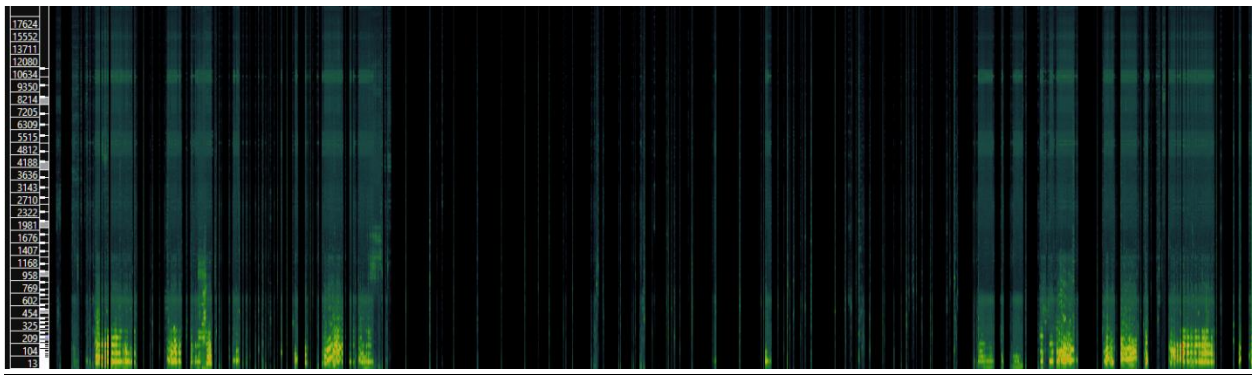
Το stem των drums παρουσιάζει ακουστικά μία όχι τόσο ευδιάκριτη παρουσία των κρουστών όσο αναμενόταν, ενώ παράλληλα ανιχνεύεται μια μικρή διαρροή από το other stem (έγχορδα και πνευστά). Το γεγονός αυτό μας οδηγεί στην αξιολόγηση της διαρροής ως 1^ο βαθμού. Θεωρητικά, το μοντέλο θα έπρεπε να τα εντοπίζει ευκολότερα, δεδομένου ότι τα κρουστά ηχούν σε δυναμική κυρίως «forte» και σε «tutti» μέρη του έργου. Όμως, παρόλο που ισχύει η παραπάνω συνθήκη, το αποτέλεσμα είναι θολό και όχι καθαρά απομονωμένο. Η διαρροή δεν είναι έντονα αισθητή και περιορίζεται σε ελάχιστα και συγκεκριμένα σημεία, επηρεάζοντας τη συνολική ποιότητα του διαχωρισμού. Συνεπώς η καθαρότητα παραμένει στο 2 λόγω της υπαρκτής θολότητας.

Bass

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του bass.



Εικόνα 137: Waveform of bass stem (Sonic Visualiser)



Εικόνα 138: Spectrogram of bass stem (Sonic Visualiser)

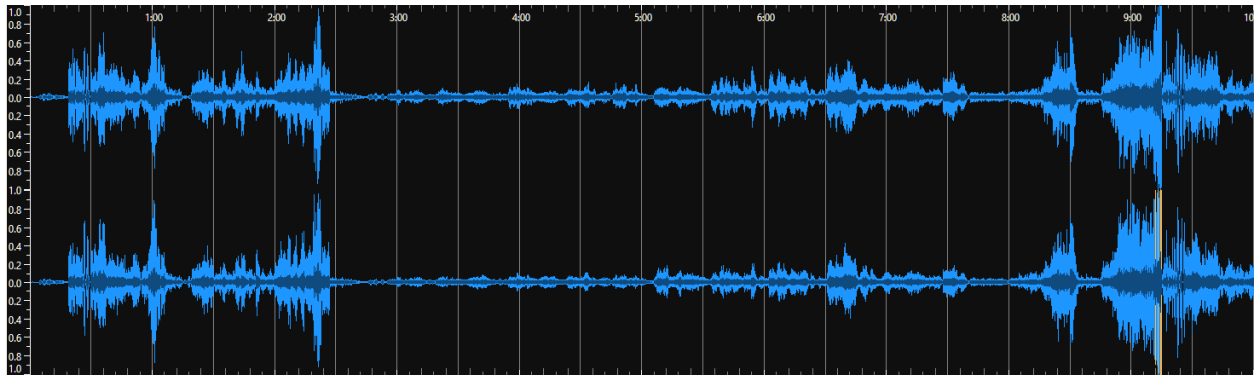
Στην περίπτωση του bass stem, η αξιολόγηση ξεκινά με την ακρόαση. Στο συγκεκριμένο έργο, το ρόλο του μπάσου αναλαμβάνουν κυρίως οι χαμηλές συχνότητες του πιάνου και των εγχόρδων, οι οποίες έχουν την λειτουργία της αρμονικής θεμελίου του μπάσου. Ωστόσο, το εργαλείο διαχωρισμού δεν είναι εκπαιδευμένο να εντοπίσει αυτή τη λειτουργία και να τα απομονώσει. Για τον λόγο αυτό, αναμένουμε το stem να εμφανίζεται κενό.

Κλίμακα ΚΚΔ				
Bass	0	1	2	3
Καθαρότητα			X	
Διαρροή		X		
Π.Δ.	0,66			

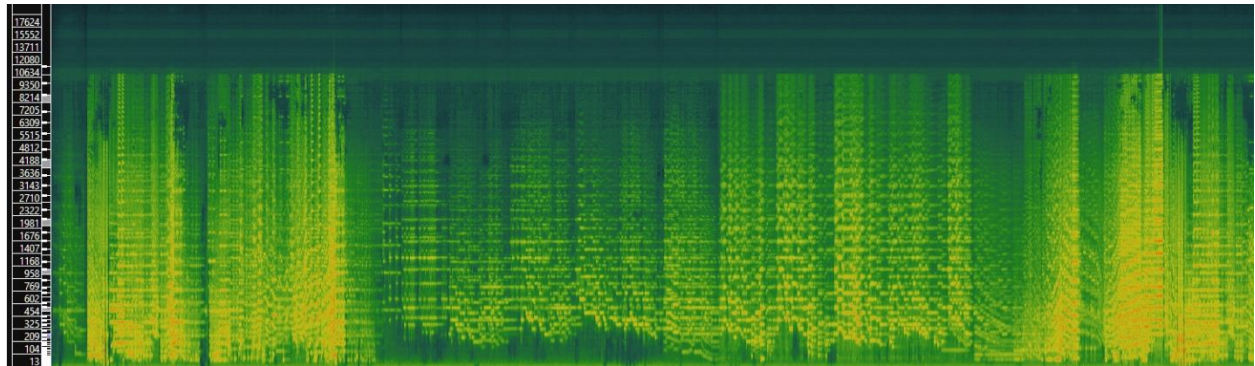
Ακουστικά, το αποτέλεσμα το οποίο ακούγεται περιορίζεται σε αποσπάσματα μόνο των εγχόρδων (βιολοντσέλα και κοντραμπάσα), στα οποία παρεμβάλλεται και η διαρροή από το other stem (κυρίως πνευστά όπως τρομπέτες και κόρνα). Η παρατήρηση αυτή οδηγεί σε διαρροή 1^{ου} βαθμού. Όσον αφορά την καθαρότητα, δεδομένου ότι αναμένονται μεγάλα κενά στο stem, αυτή αξιολογείται ως 2^{ου} βαθμού. Ένας επιπλέον λόγος είναι ότι, το μοντέλο δεν είναι ικανό να ανιχνεύσει την πλήρη λειτουργία του μπάσου σε όλη την έκτασή του, όπως εξηγήσαμε παραπάνω.

Other

Παρακάτω, παρατίθενται η κυματομορφή, το φασματογράφημα και η ΚΚΔ από το stem του other.



Εικόνα 139: Waveform of other stem (Sonic Visualiser)



Εικόνα 140: Spectrogram of other stem (Sonic Visualiser)

Θεωρητικά, το περιεχόμενο αυτού του stem θα έπρεπε να αποτελεί το «περίσσευμα» από την απομόνωση των υπολοίπων stems. Ακουστικά όμως, παρατηρείται πλήρης διαρροή. Αναλυτικότερα, το φασματογράφημα ενώ εμφανίζεται «πλούσιο» συχνοτικά, δεν μας δίνει κάποια επιπλέον πληροφορία για την ύπαρξη διαρροής. Αυτό πράγματι επαληθεύεται κατά την ακρόαση, όπου το αποτέλεσμα του διαχωρισμού θυμίζει το έργο πριν από αυτόν με επιπλέον κάποια προσθήκη θορύβου. Συνεπώς, δεν υπάρχει πρακτική εφαρμογή της ΚΚΔ λόγω του μη διαχωρισμού.

Κλίμακα ΚΚΔ				
Other	0	1	2	3
Καθαρότητα				
Διαρροή				
Π.Δ.				

Συμπεράσματα Παραδείγματος

Το τελευταίο παράδειγμα της συγκεκριμένης εργασίας, ανέδειξε με σαφήνεια τους περιορισμούς του Spleeter όταν εφαρμόζεται σε μεγάλης κλίμακας συμφωνικά έργα. Το συγκεκριμένο παράδειγμα είχε κυρίως ερευνητικό ενδιαφέρον για τον συγγραφέα, καθώς αποσκοπούσε στην κατανόηση της συμπεριφοράς του εργαλείου σε

Κλίμακα Διαρροής (ΚΚΔ)					
Stems	Αξιολόγηση	0	1	2	3
Vocals	Καθαρότητα			X	
	Διαρροή			X	
	Π.Δ.	0,5			

έργα με μεγάλη πολυπλοκότητα στην ενορχήστρωση και έντονη αλληλεπίδραση μεταξύ διαφορετικών οργάνων.

Το stem των vocals, που θεωρητικά θα έπρεπε να είναι εντελώς κενό, περιείχε εκτεταμένη διαρροή από όλα τα όργανα της ορχήστρας. Τα piano, drums και bass stems εμφάνισαν σημαντικά προβλήματα καθαρότητας και ακρίβειας, ενώ τα ντραμς είχαν πιο συνεπή αλλά όχι πλήρως επιτυχημένη απομόνωση.

Το stem του other απέτυχε να απομονώσει ξεχωριστά όργανα και λειτούργησε ουσιαστικά ως «συμπυκνωμένο» αντίγραφο του αρχικού έργου. Συνολικά, η ανάλυση κατέδειξε ότι το μοντέλο δεν μπορεί να ανταποκριθεί ικανοποιητικά. Αν για παράδειγμα υπήρχε Η παρουσία ακόμα και μίας επιπλέον χορωδίας ή άλλων πολυφωνικών στοιχείων θα περιέπλεκε περαιτέρω την απομόνωση, σε έργα με τόσο σύνθετη ενορχήστρωση και συνεχή αλληλεπίδραση μεταξύ πολλών ομάδων οργάνων.

Piano	Καθαρότητα		X		
	Διαρροή		X		
	Π.Δ.	0,5			
Drums	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Bass	Καθαρότητα			X	
	Διαρροή		X		
	Π.Δ.	0,66			
Other	Καθαρότητα				
	Διαρροή				
	Π.Δ.				

4.15 Μέση απόδοση του Spleeter

Παρακάτω παρατίθεται ένας πίνακας με την μέση απόδοση του μοντέλου συνδυάζοντας τον μέσο όρο της Π_Δ και στη συνέχεια η μετατροπή του σε ποσοστό για κάθε παράδειγμα. Στις περιπτώσεις όπου ο διαχωρισμός απέτυχε πλήρως όπως μη ανίχνευση ή πλήρη διαρροή (κυρίως στο other) και η ΚΚΔ δεν είχε πρακτικό αποτέλεσμα θεωρούμε ότι έχουμε 0 στην Π_Δ το οποίο συνεπάγεται με πλήρη αποτυχία.

Αριθμός παραδείγματος	Μέση απόδοση του μοντέλου	Αριθμός παραδείγματος	Μέση απόδοση του μοντέλου
1	66%	8	89%
2	33%	6	86%
3	53%	5	83%
4	46%	10	73%
5	83%	1	66%
6	86%	7	66%
7	66%	11	66%
8	89%	8	63%
9	56%	9	56%
10	73%	3	53%
11	66%	4	46%
12	63%	13	46%
13	46%	14	46%
14	46%	2	33%

5. Συμπεράσματα και Συζήτηση

5.1 Βασικά Συμπεράσματα

Η χρήση του Spleeter μας έδειξε σε γενικές γραμμές ότι είναι σε θέση να παράξει ένα σχετικά σωστό αποτέλεσμα στον διαχωρισμό των stems. Ιδιαίτερα, όταν η φωνή είναι εκείνη που κατέχει τον κεντρικό ρόλο σε μια σύνθεση, τα αποτελέσματα του διαχωρισμού είναι καλύτερα. Ωστόσο, ιδιαίτερο ενδιαφέρον παρουσιάζουν τα χρωδιακά έργα π.χ. του Ola Gjeilo, όπου η πολυφωνική φύση της φωνής και η συνύπαρξη με την ορχήστρα δυσχεραίνουν την απομόνωση. Σημαντικός παράγοντας επιβάρυνσης είναι επίσης η αντήχηση, ειδικά στα δύο χρωδιακά έργα, η οποία καθιστά το φάσμα «θολό» και αυξάνει την πιθανότητα διαρροής.

Σε γενικές γραμμές, τα stems των φωνητικών, του μπάσου και των ντραμς αποδίδονται με μεγαλύτερη καθαρότητα σε σύγκριση με εκείνα του πιάνου και του other. Εξαιρέση αποτελεί το αρμονικό μπάσο όπου δεν ανιχνεύεται αυτή η λειτουργία. Στην περίπτωση όπου αναμένεται ο διαχωρισμός μεμονωμένων οργάνων στο stem other, όπως για παράδειγμα το φλάουτο, συχνά παρατηρείται μεγάλη διαρροή. Επιπλέον, σε έργα με υψηλό βαθμό πολυφωνίας, όπως αυτά των Saint-Saëns και Rachmaninoff, η ταυτόχρονη παρουσία πολλαπλών streams οδηγεί σε επικαλύψεις συχνοτήτων, με αποτέλεσμα αυξημένα ποσοστά διαρροής. Το παραπάνω καθίσταται ιδιαίτερα εμφανές σε έργα κλασικής, τζαζ, μπαρόκ ή κινηματογραφικής μουσικής, όπου συχνά μέρη των stems εμφανίζονται ελάχιστα θολά, άλλοτε μερικώς και άλλοτε σε ολόκληρη την έκτασή τους.

Αντίθετα, σε μουσικά περιβάλλοντα χαμηλότερης πολυπλοκότητας, η μειωμένη πυκνότητα επιτρέπει πιο καθαρό και ακριβή διαχωρισμό συγκεκριμένων stem. Παράλληλα, καθοριστικό ρόλο στην ποιότητα του αποτελέσματος διαδραματίζει η φύση της αρχικής ηχογράφησης. Έργα που έχουν δημιουργηθεί με φυσικά όργανα κάτω από ποικίλες ακουστικές συνθήκες ενδέχεται να εμφανίσουν διαφορετικούς βαθμούς διαρροής σε σχέση με παραγωγές που έχουν δημιουργηθεί εξ ολοκλήρου σε ψηφιακά περιβάλλοντα DAW. Σύμφωνα με όλα τα παραπάνω, επαληθεύεται και η γενική υπόθεση πολυπλοκότητας, την οποία διατυπώσαμε στην υποενότητα [3.8](#).

Η αξιολόγηση με εργαλεία, όπως η οπτική παρατήρηση της κυματομορφής και του φασματογραφήματος στο Sonic Visualiser, ανέδειξε πιθανές διαρροές κυρίως θορύβου που δεν γίνονται άμεσα αντιληπτές ακουστικά. Σε αυτές τις περιπτώσεις απαιτείται μια δεύτερη πιο προσεκτική ακρόαση (πχ. Παράδειγμα 8, bass stem) για τον εντοπισμό διαρροής, πράγμα το οποίο υπογραμμίζει την ανάγκη συνδυασμού ακουστικής και οπτικής παρατήρησης για την ακριβέστερη αξιολόγηση της ποιότητας του διαχωρισμού. Συνοψίζοντας, τα ευρήματα καταδεικνύουν πως ενώ το Spleeter είναι χρήσιμο για ένα «πρώτο» και «βασικό» διαχωρισμό, η αποτελεσματικότητά του περιορίζεται σε πιο πολύπλοκα μουσικά περιβάλλοντα. Η περαιτέρω βελτίωση των αλγορίθμων διαχωρισμού σίγουρα είναι απαραίτητη για την καλύτερη διαχείριση και την ελαχιστοποίηση των διαρροών μεταξύ των stems.

5.2 Συζήτηση

Τα ευρήματα της παρούσας εργασίας προσπαθούν να συνδέσουν τη σχέση μεταξύ των υπολογιστικών μοντέλων της ASA και της ακουστικής αντίληψης του ανθρώπου. Στόχος της αρχικής σχεδίασης του Spleeter δεν ήταν για να αναπαράγει πιστά τις διαδικασίες που διέπουν την ASA. Αντίθετα, ήταν πιο πρακτικός και η εκπαίδευσή του, βασισμένη σε νευρωνικά δίκτυα μέσα από τη διαδικασία της βαθιάς εκμάθησης, το έχει τροφοδοτήσει με μεγάλα σύνολα δεδομένων μουσικής (κυρίως τραγούδια της ποπ-ροκ κουλτούρας) αντικατοπτρίζοντας ουσιαστικά ορισμένα από τα μουσικά είδη που οι άνθρωποι έρχονται σε επαφή και πραγματοποιούν την δική τους μη συνειδητή ανάλυση ακουστικής σκηνής. Αυτός είναι και ο λόγος για τον οποίο επιλέξαμε έργα μη συμβατικά για να δοκιμάσουμε τις δυνατότητες του μοντέλου.

Σε αντίθεση με το ανθρώπινο ακουστικό σύστημα, το οποίο μπορεί εύκολα να προσαρμοστεί, το Spleeter είναι τελείως εξαρτώμενο από το πλαίσιο και το σχεδιασμό της εκπαίδευσής του. Δεν είναι ικανό να προσαρμοστεί σε νέες ακουστικές συνθήκες, χωρίς επιπλέον εκπαίδευση και η απόδοση του παραμένει περιορισμένη λόγω των ενσωματωμένων εκπαιδευτικών δεδομένων και της ίδια του της κατασκευής. Αυτός ο περιορισμός οδηγεί σε ερωτήματα σχετικά με το πώς τα εργαλεία τεχνητής νοημοσύνης μπορούν τελικά να προσεγγίσουν περισσότερο τις ανθρώπινες δυνατότητες του διαχωρισμού. Υβριδικές αναλύσεις οι οποίες ίσως ενσωματώσουν στα υπολογιστικά μοντέλα καινούργιες μεθόδους βασισμένες σε πιο «ανθρώπινα» δεδομένα ίσως αποτελούν μία μελλοντική κατεύθυνση.

Όσον αφορά την εφαρμογή του Spleeter, η χρησιμότητα του επεκτείνεται πέρα από τους ερευνητικούς τομείς της μουσικολογίας και σε τομείς όπως η μουσική εκπαίδευση στα σχολεία, στα remixes και στη δημιουργία караόκε με τα παραγόμενα stems, αλλά ακόμη και για ανάλυση ήχου σε εγκληματολογικά εργαστήρια. (Hayward & Maas, 2021; Morris & Higgins, 2010) Ωστόσο, οι ατέλειές του υπογραμμίζουν την ανάγκη για κριτική και πολύ προσεκτική χρήση των εργαλείων τεχνητής νοημοσύνης είτε σε ακαδημαϊκά είτε σε επαγγελματικά πλαίσια. Η παρουσία σφαλμάτων είναι υπαρκτή και δείχνει ότι τα εξαγόμενα αποτελέσματα δεν πρέπει να αντιμετωπίζονται ως απόλυτη αλήθεια, αλλά ως προσεγγίσεις που απαιτούν ερμηνεία και, όπου χρειάζεται, ανθρώπινη παρέμβαση με σκοπό την διόρθωση.

5.3 Προτάσεις για Μελλοντική Έρευνα

Με βάση τα αποτελέσματα και τους περιορισμούς της παρούσας διπλωματικής εργασίας, διατυπώνονται οι ακόλουθες προτάσεις για μελλοντική έρευνα, με στόχο την μεγαλύτερη εμβάθυνση και διεύρυνση των δυνατοτήτων των μοντέλων αυτόματου διαχωρισμού σε stems. Ειδικότερα:

- 1) **Επέκταση του δείγματος και ενσωμάτωση διαφορετικών ειδών μουσικής:** Η διεύρυνση του αριθμού των μουσικών έργων, καθώς και η ένταξη περισσότερων και διαφορετικών μουσικών

ειδών (π.χ. παραδοσιακή μουσική, σύγχρονη πειραματική, ηλεκτροακουστική κτλ.), θα μπορούσε να ενισχύσει τη καθολικότητα των συμπερασμάτων και να δοκιμάσει την απόδοση του Spleeter σε εναλλακτικά μουσικά περιβάλλοντα.

- 2) **Σύγκριση με άλλα εργαλεία διαχωρισμού:** Η διεξαγωγή μίας σύγκρισης μεταξύ του Spleeter και άλλων εργαλείων διαχωρισμού π.χ. Demucs, Open-Unmix, VoiceFixer, RipX (Stöter et al., 2019), θα μπορούσε να αποκαλύψει τα μειονεκτήματα και πλεονεκτήματα κάθε μοντέλου, καθώς και να συμβάλει στην ανάπτυξη καινούργιων συστημάτων διαχωρισμού.
- 3) **Εκπαίδευση εξειδικευμένων μοντέλων:** Μία ακόμη ιδέα για έρευνα είναι η δημιουργία και εκπαίδευση νέων μοντέλων που να βασίζονται σε μουσικά δεδομένα συγκεκριμένων ειδών μουσικής π.χ. κλασικής, ώστε να εξασφαλιστεί μεγαλύτερη ακρίβεια στον διαχωρισμό.
- 4) **Αυτόματη αξιολόγηση της διαρροής:** Η ανάπτυξη εργαλείων ή αλγορίθμων για την αυτόματη υπολογιστική ανίχνευση της διαρροής, δηλαδή χωρίς την ανάγκη ακουστικής επαλήθευσης από το ανθρώπινο αυτί.
- 5) **Εμβάθυνση στις διαφορετικές τεχνικές ανάλυσης εκτός του STFT:** Η αξιολόγηση διαφορετικών τεχνικών φασματικής ανάλυσης, πέραν της STFT πχ. Wavelet Transform (Lu et al., 2024), (Sarwar & Shokri, 2025), Constant Q Transform (Holighaus et al., 2013), (Schörkhuber & Klaruri, 2010), θα μπορούσε να αναδείξει εναλλακτικές μεθοδολογίες για καλύτερη και ακριβέστερη ανάλυση.
- 6) **Ανάλυση παραδειγμάτων με διαφορετική μίξη και mastering:** Μελλοντικές έρευνες θα μπορούσαν να μελετήσουν πιο συστηματικά πώς το mixing και το mastering επηρεάζουν την ποιότητα του διαχωρισμού.

6. Επίλογος

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, διαπιστώσαμε ότι η τεχνολογία, όσο εντυπωσιακή κι αν φαίνεται αρχικά, εξακολουθεί να είναι άρρηκτα συνδεδεμένη με την ανθρώπινη αντίληψη και τους περιορισμούς της. Το Spleeter, ως ένα μουσικό παράδειγμα εφαρμογής τεχνητής νοημοσύνης, μας υπενθυμίζει ότι η προσπάθεια να μιμηθούμε τη λειτουργία του ακουστικού μας συστήματος δεν αποτελεί μόνο τεχνική πρόκληση, αλλά και τελικώς μια διαδικασία κατανόησης του ίδιου μας του εαυτού. Σε αυτό το σημείο, το επιστημονικό εργαλείο παύει να είναι απλώς ένα «απλό» πρόγραμμα που χρησιμοποιήθηκε στα πλαίσια της εργασίας αλλά λειτουργεί σαν αφορμή για μια πιο βαθιά σκέψη γύρω από τον τομέα της τεχνητής νοημοσύνης.

Τα αποτελέσματα στα οποία καταλήξαμε μας οδηγούν σε μια διπλή στάση. Από τη μία πλευρά, θαυμάζουμε τις δυνατότητες που ανοίγονται στον χώρο της μουσικής τεχνολογίας, από την άλλη αναγνωρίζουμε την αναγκαιότητα μιας διαρκούς κριτικής σκέψης. Η τεχνολογία μας «λύνει» τα χέρια όμως συνοδεύεται από προϋποθέσεις και περιορισμούς. Η αξιοποίηση των stems στην ανάλυση της ακουστικής σκηνής δεν είναι μόνο ένα τεχνολογικό επιτεύγματα, αλλά και ένα πρόσθετο κομμάτι μιας ευρύτερης βοήθειας στην αντίληψη του ίδιου του ήχου και κατά επέκταση της ίδιας της μουσικής. Μέσα από την ισορροπία ανάμεσα στον θαυμασμό και την κριτική σκέψη, διασφαλίζουμε ότι η τεχνολογική πρόοδος παραμένει συνδεδεμένη με τις πολιτισμικές μας αξίες και την ανθρώπινη μουσική εμπειρία, δηλώνοντας ότι τα εργαλεία είναι εδώ για να συνδράμουν στον τρόπο παραγωγής και κατανόησης της μουσικής, αλλά σε καμία περίπτωση την αντικατάσταση του ίδιου του ανθρώπινου παράγοντα.

Τελικά, η εργασία αυτή δεν καταλήγει σε ένα κλειστό συμπέρασμα, αλλά σε μια προτροπή στο να συνεχιστεί με κριτικό πνεύμα η εξερεύνηση των σχέσεων ανάμεσα στην τεχνολογία, την ακουστική αντίληψη και την καλλιτεχνική έκφραση των ίδιων των μουσικών δημιουργών. Ως ερευνητές, ακροατές και φυσικά καλλιτέχνες, καλούμαστε να δούμε την τεχνητή νοημοσύνη όχι ως αντικαταστάτη της ανθρώπινης υπόστασης, αλλά ως συνοδοιπόρο που μας ωθεί να σκεφτούμε ακόμη βαθύτερα τις προϋποθέσεις της μουσικής δημιουργίας. Σε αυτό το σημείο, θα θέλαμε επίσης να εκφράσουμε την ευγνωμοσύνη μας προς τον αναγνώστη για τον χρόνο που διέθεσε στην ανάγνωση της παρούσας διπλωματικής εργασίας. Μέσα από αυτόν τον επιστημονικό και καλλιτεχνικό διάλογο, ελπίζουμε να ανοίξουν νέοι δρόμοι, στους οποίους η επιστημονική έρευνα και η καλλιτεχνική έκφραση να συναντιούνται αμοιβαία.

Βιβλιογραφία

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems* (No. arXiv:1603.04467). arXiv. <https://doi.org/10.48550/arXiv.1603.04467>
- Al-Hawawreh, M., Bouadjeneq, M. R., Hacid, H., & Aryal, S. (2024). *A Comprehensive Study of Audio Profiling: Methods, Applications, Challenges, and Future Directions*. <https://doi.org/10.36227/techrxiv.171595948.84728317/v1>
- Allan, J. (2011). IMPORTANCE OF PITCH AND SPATIAL SEPARATION FOR THE COCKTAIL PARTY EFFECT. *Master's Thesis, Luleå University of Technology*. <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1022888&dswid=-4731>
- Balaban, M., Ebcioglu, K., & Laske, O. E. (Eds.). (1992). *Understanding music with AI: Perspectives on music cognition*. AAAI Press : MIT Press.
- Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C., & Bello, J. (2014). *MedleyDB: A MULTITRACK DATASET FOR ANNOTATION-INTENSIVE MIR RESEARCH*. <https://doi.org/10.5281/zenodo.1438309>
- Bregman, A. S. (2001). *Auditory scene analysis: The perceptual organization of sound* (2. paperback ed., repr). MIT Press.
- Cambouropoulos, E. (2008). Voice and Stream: Perceptual and Computational Modeling of Voice Separation. *Music Perception, 26*(1), 75–94. <https://doi.org/10.1525/mp.2008.26.1.75>
- Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. *Proceedings of the 18th ACM International Conference on Multimedia*, 1467–1468. <https://doi.org/10.1145/1873951.1874248>
- Cano, E., FitzGerald, D., Liutkus, A., Plumbley, M. D., & Stoter, F.-R. (2019). Musical Source Separation: An Introduction. *IEEE Signal Processing Magazine, 36*(1), 31–40. <https://doi.org/10.1109/MSP.2018.2874719>

- Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE*, 96(4), 668–696. <https://doi.org/10.1109/JPROC.2008.916370>
- Défossez, A., Usunier, N., Bottou, L., & Bach, F. (2021). *Music Source Separation in the Waveform Domain* (No. arXiv:1911.13254). arXiv. <https://doi.org/10.48550/arXiv.1911.13254>
- Deville, Y., Gannot, S., Mason, R., Plumbley, M. D., & Ward, D. (Eds.). (2018). *Latent Variable Analysis and Signal Separation* (Vol. 10891). Springer International Publishing. <https://doi.org/10.1007/978-3-319-93764-9>
- Ding, B., Zhang, T., Wang, C., Liu, G., Liang, J., Hu, R., Wu, Y., & Guo, D. (2024). Acoustic scene classification: A comprehensive survey. *Expert Systems with Applications*, 238, 121902. <https://doi.org/10.1016/j.eswa.2023.121902>
- Duggal, S. (2024). *Record, Mix and Master: A Beginner's Guide to Audio Production*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-40067-4>
- Engel, J., Hantrakul, L., Gu, C., & Roberts, A. (2020). *DDSP: Differentiable Digital Signal Processing* (No. arXiv:2001.04643). arXiv. <https://doi.org/10.48550/arXiv.2001.04643>
- Ewert, S., & Muller, M. (2012). Using score-informed constraints for NMF-based source separation. *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 129–132. <https://doi.org/10.1109/ICASSP.2012.6287834>
- Fletcher, N. H., & Rossing, T. D. (1991). *The Physics of Musical Instruments*. Springer New York. <https://doi.org/10.1007/978-1-4612-2980-3>
- Gutschalk, A., & Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hearing Research*, 307, 98–110. <https://doi.org/10.1016/j.heares.2013.08.003>
- Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture: An International Journal*, 17(2), 209–233. <https://doi.org/10.1177/1741659020917434>
- Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2019). Spleeter: A Fast and State-of-the-Art Music Source Separation Tool With Pre-Trained Models. *ISMIR 2019 Late-Breaking/Demo*, 36.

- Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, 5(50), 2154.
<https://doi.org/10.21105/joss.02154>
- Holighaus, N., Dorfler, M., Velasco, G. A., & Grill, T. (2013). A Framework for Invertible, Real-Time Constant-Q Transforms. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(4), 775–785.
<https://doi.org/10.1109/TASL.2012.2234114>
- Huang, J., Wang, J.-C., Smith, J. B. L., Song, X., & Wang, Y. (2021). Modeling the Compatibility of Stem Tracks to Generate Music Mashups. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1), 187–195. <https://doi.org/10.1609/aaai.v35i1.16092>
- Hummerson, C., Stokes, T., & Brookes, T. (2014). On the Ideal Ratio Mask as the Goal of Computational Auditory Scene Analysis. In G. R. Naik & W. Wang (Eds.), *Blind Source Separation* (pp. 349–368). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-55016-4_12
- Huron, D. (2001). Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. *Music Perception*, 19(1), 1–64. <https://doi.org/10.1525/mp.2001.19.1.1>
- Johnson, C., Rodríguez-Fernández, N., & Rebelo, S. M. (Eds.). (2023). *Artificial Intelligence in Music, Sound, Art and Design: 12th International Conference, EvoMUSART 2023, Held as Part of EvoStar 2023, Brno, Czech Republic, April 12–14, 2023, Proceedings* (Vol. 13988). Springer Nature Switzerland.
<https://doi.org/10.1007/978-3-031-29956-8>
- Joliveau, E., Smith, J., & Wolfe, J. (2004). Vocal tract resonances in singing: The soprano voice. *J. Acoust. Soc. Am.*, 116(4).
- Kaminskas, M., & Ricci, F. (2012). Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 6(2–3), 89–119. <https://doi.org/10.1016/j.cosrev.2012.04.002>
- Klapuri, A., & Davy, M. (Eds.). (2006). *Signal processing methods for music transcription*. Springer.
- Kong, Q., Wang, Y., Song, X., Cao, Y., Wang, W., & Plumbley, M. D. (2020). *Source separation with weakly labelled data: An approach to computational auditory scene analysis* (No. arXiv:2002.02065). arXiv.
<http://arxiv.org/abs/2002.02065>

- Kunchur, M. N. (2023). The human auditory system and audio. *Applied Acoustics*, *211*, 109507.
<https://doi.org/10.1016/j.apacoust.2023.109507>
- Li, H., Chen, K., Wang, L., Liu, J., Wan, B., & Zhou, B. (2022). Sound Source Separation Mechanisms of Different Deep Networks Explained from the Perspective of Auditory Perception. *Applied Sciences*, *12*(2), 832. <https://doi.org/10.3390/app12020832>
- Li, T. L., Chan, A. B., & Chun, A. H. (2010). Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 (IMECS 2010)*, *Vol I*, 546–550.
- Lu, W., Gong, D., Xue, X., & Gao, L. (2024). Improved multi-layer wavelet transform and blind source separation based ECG artifacts removal algorithm from the sEMG signal: In the case of upper limbs. *Frontiers in Bioengineering and Biotechnology*, *12*, 1367929. <https://doi.org/10.3389/fbioe.2024.1367929>
- Lupker, J. A. T., & Turkel, W. J. (2021). Music Theory, the Missing Link Between Music-Related Big Data and Artificial Intelligence. *Digital Humanities Quarterly (DHQ)*, *15*(1), 1-12 (article id 000520). <https://dhq-static.digitalhumanities.org/pdf/000520.pdf>.
- Magron, P., Vial, P.-H., Oberlin, T., & Fevotte, C. (2021). Phase Recovery with Bregman Divergences for Audio Source Separation. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 516–520. <https://doi.org/10.1109/ICASSP39728.2021.9413717>
- Mason, A. (2011). Terminology for loudness and level dBTP, LU, and all that. *BBC Research & Development White Paper*, WHP 202.
- McGill. (2025). *Albert S. Bregman (1936—2023)*. <https://www.mcgill.ca/psychology/people-0/faculty-0/memoriam-albert-s-bregman>
- Mitianoudis, N., & Davies, M. E. (2004). Audio source separation: Solutions and problems. *International Journal of Adaptive Control and Signal Processing*, *18*(3), 299–314. <https://doi.org/10.1002/acs.795>
- Morris, R. G., & Higgins, G. E. (2010). Criminological theory in the digital age: The case of social learning theory and digital piracy. *Journal of Criminal Justice*, *38*(4), 470–480.
<https://doi.org/10.1016/j.jcrimjus.2010.04.016>

- Pardo, B., Liutkus, A., Duan, Z., & Richard, G. (2018). Applying Source Separation to Music. In E. Vincent, T. Virtanen, & S. Gannot (Eds.), *Audio Source Separation and Speech Enhancement* (1st ed., pp. 345–376). Wiley. <https://doi.org/10.1002/9781119279860.ch16>
- Ren, I., Volk, A., Swierstra, W., & Veltkamp, R. C. (2020). *A Computational Evaluation of Musical Pattern Discovery Algorithms* (No. arXiv:2010.12325). arXiv. <https://doi.org/10.48550/arXiv.2010.12325>
- Riou, A., Gagneré, A., Hadjeres, G., Lattner, S., & Peeters, G. (2025). *Zero-shot Musical Stem Retrieval with Joint-Embedding Predictive Architectures* (No. arXiv:2411.19806). arXiv. <https://doi.org/10.48550/arXiv.2411.19806>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
- Saleem, R., Yuan, B., Kurugollu, F., Anjum, A., & Liu, L. (2022). Explaining deep neural networks: A survey on the global interpretation methods. *Neurocomputing*, *513*, 165–180. <https://doi.org/10.1016/j.neucom.2022.09.129>
- Sarwar, M. W., & Shokri, A. (2025). Application of wavelet transform for extracting and analysing evapotranspiration-induced diel fluctuations in streamflow records. *Stochastic Environmental Research and Risk Assessment*, *39*(8), 3263–3289. <https://doi.org/10.1007/s00477-025-03016-x>
- Schörkhuber, C., & Klapuri, A. (2010). Constant-Q Transform Toolbox for Music Processing. *Proceedings of the 7th Sound and Music Computing Conference*, 3–64.
- Schreiber, A., Sander, K., Kopiez, R., & Thöne, R. (2024). The creative performance of the AI agents ChatGPT and Google Magenta compared to human-based solutions in a standardized melody continuation task. *Jahrbuch Musikpsychologie*, *32*, e195. <https://doi.org/10.5964/jbdgm.195>
- Stöter, F.-R., Uhlich, S., Liutkus, A., & Mitsufuji, Y. (2019). Open-Unmix—A Reference Implementation for Music Source Separation. *Journal of Open Source Software*, *4*(41), 1667. <https://doi.org/10.21105/joss.01667>

- Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, 117(3), 1285–1298. <https://doi.org/10.1121/1.1854312>
- Tsougras, C., & Cambouropoulos, E. (2025). Textur, Stimme und Streaming (Musical Texture, Voice, and Auditory Streaming). *Handbuch Musikanalyse – Pluralität Und Methode*, 82–94.
- Volk, A., Wiering, F., & Kranenburg, van. (2011). Unfolding the potential of computational musicology. *Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations*, 137–144. Fryske Akademy.
- Zhang, Y., & Li, T. (2025). Music genre classification with parallel convolutional neural networks and capuchin search algorithm. *Scientific Reports*, 15(1), 9580. <https://doi.org/10.1038/s41598-025-90619-7>
- Γιάννου, Δ. (1995). *Ιστορία της μουσικής Σύντομη γενική επισκόπηση – Τόμος Α' (Μέχρι τον 16ο αιώνα): Vol. Τόμος Α'.* University Studio Press.