

Περίληψη

Η σύνθεση μουσικής με τη χρήση τεχνητής νοημοσύνης είναι ένας ραγδαία αναπτυσσόμενος κλάδος που επικεντρώνεται στην ανάπτυξη νευρωνικών δικτύων μεγάλης πολυπλοκότητας για αυτό το σκοπό. Σε αντίθεση με τα υπάρχοντα δίκτυα παραγωγής εικόνων, όπου τα αποτελέσματα είναι εκπληκτικά, καμία από τις υπάρχουσες υλοποιήσεις δικτύων συνθετών, δεν καταφέρνει να συνθέσει αληθοφανή μουσικά πρότυπα. Επίσης το γεγονός ότι δεν υπάρχει δυνατότητα ελέγχου των επιθυμητών χαρακτηριστικών της σύνθεσης δημιουργεί μία ακόμη δυσκολία στον σκοπό αυτό. Στη συγκεκριμένη εργασία παρουσιάζουμε δύο μεθόδους ελέγχου χαρακτηριστικών των παραγόμενων μουσικών προτύπων του δικτύου συνθέτη MuseGAN με την χρήση των οποίων, επιχειρούμε να βρούμε κατευθύνσεις συστηματικής αλλαγής των χαρακτηριστικών αυτών. Παράλληλα υλοποιούμε μετρικές οι οποίες απεικονίζουν και ποσοτικοποιούν κάποια από τα χαρακτηριστικά, που κατά την άποψή μας θα πρέπει να έχουν τα μουσικά πρότυπα για να είναι αληθοφανή.

Abstract

Artificial music generation is a vastly growing scientific field focused on creating Deep Neural Networks that can produce realistic sounding music. Unlike image generation, where the state of the art Deep Neural Networks are able to produce remarkable results, the existing music generating networks fail at this task. One of the greatest obstacles to realistic music generation is the inability to control and predict the characteristics of the generated output given the input. Also defining what is realistic music, in terms of human perspective is often a subjective task. In this paper, we present two existing methods for controlling the generation of Generative Adversarial Networks. We apply both methods in MuseGAN, the state of the art Neural Network for symbolic music generation, attempting to define directions of systematic change of the output. Moreover, we develop music related metrics, which are able to illustrate and quantify characteristics of the generated music, as a measure of realism, based on our musical experience and opinion.

Ευχαριστίες

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε στο Εθνικό Μετσόβιο Πολυτεχνείο, στο τμήμα των Εφαρμοσμένων Μαθηματικών και Φυσικών επιστημών στα πλαίσια του ΠΜΣ της Μαθηματικής Προτυποποίησης σε Σύγχρονες Τεχνολογίες και στην Οικονομία.

Ο επιβλέπων της εργασίας, Αν. Καθηγητής κύριος Γεώργιος Στάμου δεν δίστασε να μου αναθέσει ένα θέμα που προσπαθεί να προσεγγίσει και να αντιμετωπίσει ένα από τα πιο δύσκολα προβλήματα στην σύνθεση μουσικής μέσω τεχνητής νοημοσύνης. Τον ευχαριστώ βαθιά για την εμπιστοσύνη που μου έδειξε καθώς και για την πολύτιμη βοήθεια που μου παρείχε.

Επίσης θα ήθελα να ευχαριστήσω τους υποψήφιους Διδάκτορες Έντι Δερβάκο και Γιώργο Φιλανδριανό για την πολύτιμη καθοδήγηση τους καθ' όλη τη διάρκεια της ενασχόλησης μου με το συγκεκριμένο θέμα. Τα προβλήματα που αντιμετώπιζα γίνονταν και δικά τους και χωρίς την καθοδήγησή τους η ολοκλήρωση της εργασίας θα ήταν αδύνατη.

Για την πολύτιμη προσφορά τους όχι μόνο κατά τη διάρκεια της συγκεκριμένης εργασίας, αλλά και συνολικά όλου του ΠΜΣ, θα ήθελα να ευχαριστήσω Αλέξανδρο Γεωργίου και τον Χρήστο Μαυρίδη. Για τις επεξηγήσεις, τις κουβέντες, τις ιδέες και την προθυμότητα τους να με βοηθήσουν κάθε στιγμή που το είχα ανάγκη.

Τέλος ένα ευχαριστώ οφείλω στους κοντινούς μου ανθρώπους για την υπομονή και τη στήριξη τους. Στους γονείς μου, στο Κικάκι, στον Αλέξανδρο, στον Μπάμπη, στο Βαγγέλη.

Περιεχόμενα

Εισαγωγή	i
1 Χαρακτηριστικά Μουσικής Σύνθεσης και Μουσική Σημειογραφία	1
1.1 Προκλήσεις Σύνθεσης Μουσικής μέσω Τεχνητής Νοημοσύνης . . .	1
1.2 Μουσική Σημειογραφία και Έννοιες	3
1.3 Σύνθετες Μουσικές Έννοιες και Μουσικά Χαρακτηριστικά . . .	8
1.3.1 Μουσικές Κλίμακες	8
1.3.2 Πολυφωνία και Αρμονία - Η Έννοια της Συγχορδίας . .	10
2 Τεχνητή Νοημοσύνη στη Σύνθεση Μουσικής	12
2.1 Αναπαραστάσεις Μουσικής	12
2.1.1 Συμβολική Αναπαράσταση Μουσικής - Το Pianoroll . .	13
2.1.2 Αναπαράσταση Μουσικής Μέσω των Χαρακτηριστικών του Ήχου της	15
2.1.3 Ποιά Αναπαράσταση Είναι Καλύτερη;	16
2.2 Νευρωνικά Δίκτυα στη Σύνθεση Μουσικής	17
2.2.1 Recurrent Neural Networks (RNN's)	17
2.2.2 Variational AutoEncoders (VAEs)	23
2.2.3 Generative Adversarial Networks (GANs)	27
3 Μεθοδολογία και Ορισμοί	34
3.1 Έλεγχος Παραγωγής-Ένα Δύσκολο Πρόβλημα	34
3.2 Χαρακτηριστικά του Latent Space των GANs	35
3.2.1 Βασική Άλγεβρα Διανυσμάτων του Latent Space	36
3.3 Υπερεπίπεδα Διαχωρισμού Ταξινόμησης	38
3.3.1 Support Vector Machines Classifier	38
3.3.2 Η Τεχνική Latent Walking	41
3.4 Μείωση Διαστασιμότητας - PCA	42
3.4.1 Η Τεχνική PCA Walking	44
3.5 Σύγκριση των δύο μεθόδων	49
4 Μετρικές Αξιολόγησης Μουσικών Προτύπων	51
4.1 Η Μορφή των Παραγόμενων Προτύπων - Προεπεξεργασία . . .	51
4.2 Chord Contribution Metric	52
4.2.1 Note Contribution	52
4.2.2 Chord Metrics	52
4.2.3 Chord Contribution	54

4.2.4	Ερμηνεία και παρατηρήσεις	55
4.2.5	Chord Quality	56
4.3	Pypianoroll Metrics	56
4.3.1	Qualified Note Rate	57
4.3.2	Polyphonic Rate	57
4.3.3	Note Count	57
5	Πειραματική Διαδικασία	58
5.1	Δυσκολίες στη χρήση του MuseGAN	58
5.2	Υλοποίηση της Μεθόδου PCA Walking	58
5.2.1	Διαδικασία	59
5.2.2	Αποτελέσματα	61
5.2.3	Συμπεράσματα - Σχολιασμός	67
5.3	Υλοποίηση της Μεθόδου Latent Walking	68
5.3.1	Διαδικασία	68
5.3.2	Αποτελέσματα	70
5.3.3	Συμπεράσματα - Σχολιασμός	71
6	Παράρτημα	73
6.1	Πίνακες Αποτελεσμάτων της Chord Metric	73
6.2	Πίνακες Αποτελεσμάτων της Chord Quality	93
	Βιβλιογραφία	96

Εισαγωγή

Η μουσική σύνθεση αποτελεί μία από τις σημαντικότερες προκλήσεις στον κλάδο της τεχνητής νοημοσύνης. Πληθώρα νευρωνικών δικτύων έχουν αναπτυχθεί για το σκοπό αυτό, αποτυγχάνοντας παρόλαυτά να παράγουν αληθοφανή αποτελέσματα. Η σύνθεση σαφώς είναι μια ύψιστης μορφής ανθρώπινη τέχνη, αποτελώντας έτσι μία πολύ δύσκολη εφαρμογή για ένα νευρωνικό δίκτυο. Παρόλαυτά παρατηρούμε ότι αρκετές φορές και ανάλογα με το είδος της μουσικής, υπάρχουν διάφορα μοτίβα, στα οποία ο εκάστοτε συνθέτης βασίζεται, για τη δημιουργία μουσικής. Στο Κεφάλαιο 1 περιγράφονται αναλυτικά όλες οι δυσκολίες και οι προκλήσεις στη τεχνητή σύνθεση της μουσικής. Επιπλέον παρουσιάζουμε τον συστηματικό τρόπο γραφής της μουσικής και αρκετές σημαντικές μουσικές έννοιες, οι οποίες θα χρησιμοποιηθούν στην συνέχεια της εργασίας. Εξηγούμε επίσης τον τρόπο με τον οποίο αυτές οι έννοιες μπορούν να χρησιμοποιηθούν για τον εντοπισμό σημαντικών χαρακτηριστικών της μουσικής σύνθεσης.

Τα δίκτυα που έχουν υλοποιηθεί για παραγωγή μουσικής είναι αρκετά και ανήκουν σε διαφορετικούς τύπους. Οι πιο σημαντικοί τύποι δικτύων για το σκοπό αυτό είναι τα Recurrent Neural Networks / LSTMs, Variational AutoEncoders, Generative Adversarial Neural Networks. Ειδικότερα, τα δίκτυα συνθέτες τύπου GAN παράγουν μουσικά πρότυπα δειγματοληπτώντας τυχαία θόρυβο από έναν υψηλής διάστασης χώρο που ονομάζεται Latent Space. Το input περνώντας από τα layers του εκπαιδευμένου Generator μετατρέπεται σε μουσικό πρότυπο στην έξοδο. Το μεγαλύτερο πρόβλημα αυτού του είδους παραγωγής, είναι ότι δεν μπορούμε εκ των προτέρων να γνωρίζουμε τα χαρακτηριστικά του παραγόμενου τραγουδιού, γνωρίζοντας μόνο το αντίστοιχο input. Στο Κεφάλαιο 2 παρουσιάζονται η αρχή λειτουργίας των τριών βασικών παραπάνω τύπων δικτύων καθώς και υλοποιημένες εκδοχές για το σκοπό της μουσικής σύνθεσης. Επίσης παρουσιάζουμε τον τρόπο που η μουσική αναπαρίσταται στον υπολογιστή, καθώς με αυτόν τον τρόπο θα καθοριστεί και η φύση των δεδομένων εκπαίδευσης.

Σκοπός αυτής της εργασίας είναι η παρουσίαση μεθόδων για την εξερεύνηση του χώρου εισόδου των δικτύων GANs και η εφαρμογή των κύριων δύο υπάρχοντων μεθόδων στο δίκτυο MuseGAN, το οποίο είναι το state of the art δίκτυο στην παραγωγή συμβολικής μουσικής. Με την χρήση των μεθόδων καταφέρνουμε να ορίσουμε κατευθύνσεις στον 128-στατο χώρο εισόδου του MuseGAN, που κινούμενοι πάνω σε αυτές θα μπορούσαμε να παράγουμε πρότυπα που αλλάζουν συστηματικά κάποιο χαρακτηριστικό τους. Οι δύο αυτές μέθοδοι, καθώς και σημαντικές τεχνικές, ορισμοί και μαθηματικό υπόβαθρο, παρουσιάζονται στο Κεφάλαιο 3.

Ο καθορισμός των εξεταζόμενων χαρακτηριστικών γίνεται από προυπάρχουσα βιβλιογραφία καθώς και από την δική μας μουσική εμπειρία, υλοποιώντας μετρικές που απεικονίζουν σημαντικές πτυχές της αληθοφανούς μουσικής. Έννοιες όπως η πολυφωνικότητα ή η αρμονική συνοχή, μπορούν και εντοπίζονται επιτυχώς τις μετρικές αυτές, ποσοτικοποιώντας τα χαρακτηριστικά αυτά ως μέτρα συνεισφοράς. Στο Κεφάλαιο 4 παρουσιάζονται αναλυτικά όλες οι μετρικές αξιολόγησης των μουσικών προτύπων, ο τρόπος και το σκεπτικό που υλοποιήθηκαν. Επίσης εξηγείται ο λόγος που είναι σημαντικό να εξετάσουμε τα συγκεκριμένα χαρακτηριστικά ως προς την αξιολόγηση της αληθοφάνειας της παραγόμενης μουσικής.

Τέλος στο Κεφάλαιο 5 παρουσιάζεται η εφαρμογή των δύο μεθόδων στο pre-trained δίκτυο MuseGAN καθώς και η αξιολόγηση των αποτελεσμάτων από τις μετρικές που υλοποιήθηκαν παραπάνω. Πολλά από τα αποτελέσματα της πρώτης μεθόδου παρουσιάζονται στο Παράρτημα.

1 Χαρακτηριστικά Μουσικής Σύνθεσης και Μουσική Σημειογραφία

Πότε θεωρούμε έναν ήχο ως μουσική και πότε όχι; Υπάρχουν κοινά χαρακτηριστικά στην αλληλουχία των νοτών σε δύο διαφορετικά μουσικά κομμάτια που ανήκουν σε διαφορετικά είδη; Η μουσική σύνθεση είναι αδιαμφισβήτητα μια τέχνη που βασίζεται στην έμπνευση του συνθέτη. Είναι αδύνατον κανείς να εξηγήσει ποιοτικά και ποσοτικά τι συντελεί σε αυτή την έμπνευση. Παρολαυτά μπορούμε να εξετάσουμε την κοινή υπάρχουσα βάση πάνω στην οποία δουλεύουν όλοι όσοι συνθέτουν μουσική, να αναλύσουμε τις υπάρχουσες μουσικές δομές (μελωδία, ρυθμός, αρμονία κλπ) καθώς και τη μουσική σημειογραφία. Στο συγκεκριμένο κεφάλαιο επιχειρούμε να συνοψίσουμε όλα εκείνα τα σημαντικά χαρακτηριστικά που διαμορφώνουν τη μουσική σύνθεση παρουσιάζοντας ταυτόχρονα την πολυπλοκότητα της.

1.1 Προκλήσεις Σύνθεσης Μουσικής μέσω Τεχνητής Νοημοσύνης

Τα τελευταία χρόνια, με την ανάπτυξη γρηγορότερων και ικανότερων υπολογιστικών συστημάτων καθώς και τεχνικών μηχανικής μάθησης και τεχνητής νοημοσύνης γενικότερα, ο υπολογιστής έχει καταφέρει να υλοποιεί διαδικασίες που θα ήταν αδύνατον να πραγματοποιηθούν από τον ανθρώπινο εγκέφαλο. Η υπολογιστική δύναμη και οι δυνατότητες μνήμης έχουν αυξηθεί ραγδαία γεγονός που εξυπηρετεί την λειτουργία μεγάλης γκάμας αλγορίθμων προβλεψής, λήψης αποφάσεων, αλληλεπίδρασης με εξωτερικούς παράγοντες και παραγωγής προτύπων. Πιο συγκεκριμένα, όσον αφορά με την παραγωγή προτύπων, το δίκτυο StyleGAN2 αφού εκπαιδευτεί με τα κατάλληλα δεδομένα, καταφέρνει να 'σχεδιάσει' ανθρώπινα πρόσωπα, όπως φαίνεται στο σχήμα 1, τα οποία είναι σχεδόν αδύνατο κανείς να ξεχωρίσει από πραγματικές φωτογραφίες.

Δυστυχώς δεν συμβαίνει το ίδιο με την παραγωγή - σύνθεση μουσικής, μια διαδικασία που όλα τα δίκτυα που υπάρχουν μέχρι σήμερα αποτυγχάνουν να φέρουν εις πέρας σε τόσο ικανοποιητικό βαθμό. Οι τεχνικές δυσκολίες είναι μεγάλες και αυτές αποτελούν και το μεγαλύτερο κίνητρο της έρευνας σε αυτή τη κατεύθυνση. Για έμφαση παραθέτουμε μια φράση των ίδιων των δημιουργών του Jukebox, του δικτύου συνθέτη με τα πιο ρεαλιστικά μέχρι στιγμής αποτελέσματα :

“While Jukebox represents a step forward in musical quality, coherence, length of audio sample, and ability to condition on artist, genre, and lyrics,

there is a significant gap between these generations and human-created music. “[8]



Σχήμα 1: Ένα δείγμα από τα αποτελέσματα του StyleGAN2, παρουσιάζοντας την εκπληκτική ομοιότητα που πετυχαίνουν τέτοιου είδους δίκτυα με πραγματικές φωτογραφίες [20]

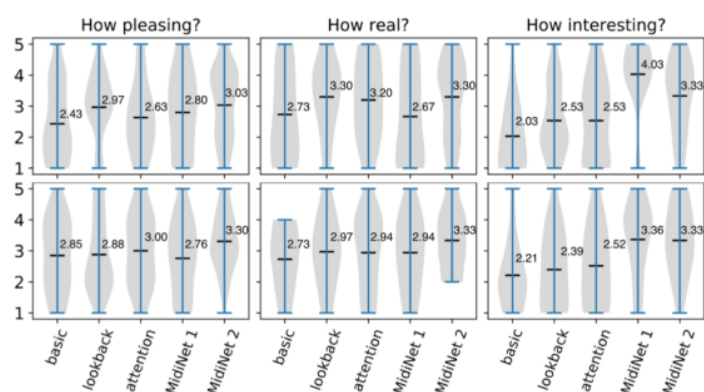
Καταλαβαίνουμε λοιπόν ότι η σύνθεση μουσικής αποτελεί μια από τις σημαντικότερες προκλήσεις στον χώρο της τεχνητής νοημοσύνης και της μηχανικής μάθησης καθώς, ενώ υπάρχει πλέον πληθώρα προηγμένων νευρωνικών δικτύων για αυτό το σκοπό, κανένα από αυτά δεν καταφέρνει να ‘ξεγελάσει’ το ανθρώπινο αυτί παρουσιάζοντας τόσο πειστικά αποτελέσματα. Στην εισαγωγή του άρθρου που πρωτοέφερε το MuseGAN, ένος από τα αποτελεσματικότερα δίκτυα παραγωγής μουσικής έως τώρα, οι συγγραφείς τονίζουν τρεις λόγους για τη δυσκολία του εγχειρήματος:

“First, music is an art of time, necessitating a temporal model. Second, music is usually composed of multiple instruments/tracks with their own temporal dynamics, but collectively they unfold over time interdependently. Lastly, musical notes are often grouped into chords, arpeggios or melodies in polyphonic music, and thereby introducing a chronological ordering of notes is not naturally suitable.”[11]

Η σημαντικότερη δυσκολία στην παραγωγή μουσικής με τεχνητή νοημοσύνη, που τελικά συνοψίζει όλες τις παραπάνω προκλήσεις, είναι η μη ύπαρξη κάποιας καλώς ορισμένης μετρικής που να αποτυπώνει την ποιότητα της μουσικής. Το τι είναι καλή μουσική και τι όχι, ακόμα και όταν έχει δημιουργηθεί από άνθρωπο συνθέτη, είναι κάτι εντελώς υποκειμενικό. Για παράδειγμα οι μουσικοί της κλασικής μουσικής σε αντίθεση με τους μουσικούς της Rock έχουν εντελώς διαφορετική άποψη για το τί κάνει ένα μουσικό κομμάτι ‘σωστό’ ή ‘ποιοτικό’. Συνεπώς μπορούμε να καταλάβουμε ότι χωρίς έναν τρόπο τυποποίησης και ποσοτικοποίησης της ποιότητας των παραγόμενων μουσικών

προτύπων από τα διαθέσιμα δίκτυα συνθέτες, πολλά από τα αποτελέσματα δεν θα μπορούν να είναι αληθοφανή, διότι η ‘αληθοφάνεια’ είναι μια παράμετρος που κανένα δίκτυο ποτέ δεν θα μάθει.

Οι ερευνητές συχνά δίνουν ερωτηματολόγια σε ομάδες ανθρώπων σχετικά με την ποιότητα των παραγόμενων προτύπων των δικτύων τους, ως μία προσπάθεια αποτύπωσης της με στατιστικό τρόπο. Ένα παράδειγμα αξιολόγησης της ποιότητας των μουσικών προτύπων από τους δημιουργούς του MidiNet φαίνεται στο Σχήμα 2.



Σχήμα 2: Οι μετρικές αξιολόγησης του MidiNet, σε μουσικούς (πάνω) και σε μη μουσικούς (κάτω). [33]

1.2 Μουσική Σημειογραφία και Έννοιες

Η μουσική εκτός από τέχνη, είναι και ένα γνωστικό αντικείμενο για το οποίο έχει δημιουργηθεί εκτεταμένη βιβλιογραφία. Φυσικά οτιδήποτε μελετάται συστηματικά θα πρέπει να μπορεί να αποτυπώνεται γραπτώς. Έτσι και η μουσική έχει το δικό της ‘αλφάβητο’ με τη χρήση του οποίου δύο διαφορετικοί άνθρωποι μπορούν να αναπαράγουν ένα μουσικό κομμάτι πιστά και με τον ίδιο τρόπο.

Δύο από τις βασικότερες έννοιες που αποτυπώνονται στο χαρτί είναι η μελωδία και ο ρυθμός. Σε αυτό το κεφάλαιο θα εξηγήσουμε συνοπτικά τις δύο έννοιες αυτές και θα δώσουμε μια εικόνα για το πως αυτές αποτυπώνονται στην παρτιτούρα.

Μελωδία

Η μουσική αποτελεί αναπαραγωγή ήχου. Και όπως γνωρίζουμε από τη Φυσική, η διαφορά που αντιλαμβάνεται το ανθρώπινο αυτί σε δύο ήχους σταθερής έντασης, έγκειται στην πραγματικότητα στη διαφορά των συχνοτήτων των

δύο ηχητικών κυμάτων που φτάνουν σε αυτό. Μελωδία λοιπόν στη μουσική, εννοούμε τη συλλογή μιας σειράς από ήχους με πολύ συγκεκριμένες συχνότητες (τουλάχιστον στη δυτική μουσική) οι οποίες αναπαριστώνται από τις νότες. Οι νότες εκτός από τη διάρκεια του ήχου (που θα τη συζητήσουμε αργότερα) κυρίως αναπαριστούν στο χαρτί ποιο θα είναι το τονικό ύψος του ήχου που θα παιχτεί καθορίζοντας και έτσι τη συχνότητά του. Στο Σχήμα 3 παρουσιάζεται ένας πίνακας με τα ονόματα των νοτών και τη συχνότητα τους:

	C	C#	D	E ^b	E	F	F#	G	G#	A	B ^b	B
0	16.35	17.32	18.35	19.45	20.60	21.83	23.12	24.50	25.96	27.50	29.14	30.87
1	32.70	34.65	36.71	38.89	41.20	43.65	46.25	49.00	51.91	55.00	58.27	61.74
2	65.41	69.30	73.42	77.78	82.41	87.31	92.50	98.00	103.8	110.0	116.5	123.5
3	130.8	138.6	146.8	155.6	164.8	174.6	185.0	196.0	207.7	220.0	233.1	246.9
4	261.6	277.2	293.7	311.1	329.6	349.2	370.0	392.0	415.3	440.0	466.2	493.9
5	523.3	554.4	587.3	622.3	659.3	698.5	740.0	784.0	830.6	880.0	932.3	987.8
6	1047	1109	1175	1245	1319	1397	1480	1568	1661	1760	1865	1976
7	2093	2217	2349	2489	2637	2794	2960	3136	3322	3520	3729	3951
8	4186	4435	4699	4978	5274	5588	5920	6272	6645	7040	7459	7902

Σχήμα 3: Νότες και οι συχνότητες τους

Οι νότες ονομάζονται ως Ντο, Ντο#, Ρε, Ρε#, Μι, Φα, Φα#, Σολ, Σολ#, Λα ή αντίστοιχα ως C, C#, D, D#, E, F, F#, G, G#, A, A#, B, όπως μπορούμε να δούμε σε μία από τις γραμμές του σχήματος 3. Η ελάχιστη συχνοτική διαφορά μεταξύ δύο νοτών δεδομένης της γραμμής ονομάζεται ημιτόνιο και όπως παρατηρούμε αυτή δεν είναι σταθερή, αλλά αυξάνεται όσο αυξάνεται το τονικό ύψος, δηλαδή ο αριθμός της γραμμής. Οι νότες οι οποίες βρίσκονται στην ίδια γραμμή του σχήματος 3 λέμε ότι ανήκουν στην ίδια οκτάβα. Μετά την τελευταία νότα της οκτάβας, ακολουθεί η πρώτη νότα της καινούριας οκτάβας και τα ονόματα των νοτών επαναλαμβάνονται στην καινούρια γραμμή.

Συνεπώς μπορούμε να πούμε ότι οι νότες C0 και C1 απέχουν απόσταση μίας οκτάβας. Οι νότες που απέχουν οκτάβα, έχουν πολύ όμοιο άκουσμα, που παρόλαυτα διαφέρει στο τονικό τους ύψος. Μπορούμε να πούμε ότι είναι ίδιοι ήχοι με τη διαφορά ότι ο ένας είναι πιο μπάσος ενώ ο άλλος πιο πρίμος. Γιαυτό το λόγο επιλέγεται και το ίδιο όνομα με διαφορετικό παρόλαυτα αριθμό, ο οποίος δηλώνει την οκτάβα.

Είναι σημαντικό λοιπόν να καταλάβει κανείς ότι η μουσική γράφεται και ερμηνεύεται σε ένα σχετικό σύστημα αναφοράς και μετρικής που γνώμονα δεν έχει τη συχνοτική απόσταση των νοτών, μιας και αυτή όπως βλέπουμε δεν είναι

σταθερή. Οι αποστάσεις στη μουσική μετρώνται με την χρήση του διαστήματος το οποίο στην πραγματικότητα είναι ένας ακέραιος αριθμός που δείχνει την ονομαστική απόσταση μεταξύ δύο νοτών. Για παράδειγμα οι νότες E0 και F0 λέμε ότι απέχουν διάστημα 1ης, ενώ οι νότες C0 και E0 απέχουν διάστημα 3ης.

Τα διαστήματα μπορεί επίσης να είναι μικρά ή μεγάλα ανάλογα με τον αριθμό των ημιτονίων που απέχουν οι δύο νότες. Για παράδειγμα οι νότες C0 και E0 απέχουν διάστημα τρίτης μεγάλο (4 ημιτόνια), ενώ οι νότες A0 και C1 απέχουν διάστημα τρίτης μικρό (3 ημιτόνια). Η απόσταση δύο ημιτονίων ονομάζεται αλλιώς και τόνος, και παρατηρούμε ότι όλες οι νότες χωρίς αλλοιώσεις (δηλαδή χωρίς διέσεις ή υφέσεις) που βρίσκονται στην ίδια οκτάβα απέχουν απόσταση τόνου εκτός από τα ζευγάρια Σι - Ντο και Μι - Φα που απέχουν ημιτόνιο. Το σύμβολο της διέσης (#) συνοδεύει μια νότα, ανεβάζοντάς την ένα ημιτόνιο ενώ το σύμβολο της υφέσης (b) κατεβάζει ένα ημιτόνιο τη νότα που συνοδεύει.

Είναι προφανές ότι ο πίνακας του Σχήματος 3 μπορεί εκταθεί προς τα πάνω και προς τα κάτω προσθέτοντας οκτάβες προς τις υψηλές ή προς τις χαμηλές συχνότητες θεωρητικά. Πρακτικά το ανθρώπινο αυτί μπορεί να αντιληφθεί ήχους από 20 Hz έως περίπου 20 KHz συνεπώς κάτι τέτοιο δεν θα είχε κανένα πρακτικό νόημα.

Συνοψίζοντας, όταν μιλάμε για μελωδικά χαρακτηριστικά του μουσικού κομματιού, αναφερόμαστε στην αλληλουχία των νοτών που το απαρτίζουν, τις σχέσεις μεταξύ τους και τα διαστήματα που τις χωρίζουν.

Ρυθμός

Η συχνότητα της εκάστοτε νότας, όπως είδαμε παραπάνω, απεικονίζεται με το όνομα που δίνεται στη νότα καθώς και την οκτάβα που ανήκει. Παρόλαυτά οι νότες σε ένα μουσικό κομμάτι ανεξαρτήτως του ποιές είναι, έχουν και κάποια συγκεκριμένη διάρκεια. Ονομάζουμε ρυθμό τον τρόπο που οι νότες ενός μουσικού κομματιού είναι χρονικά διατεταγμένες. Στην μουσική, η διάρκεια μιας νότας, είναι σχετική. Η απόλυτη διάρκεια καθορίζεται από το τέμπο (beats per minute(bpm)) που ο συνθέτης του μουσικού κομματιού ορίζει για το συγκεκριμένο κομμάτι. Στην πραγματικότητα γρήγορο τέμπο (δηλαδή πολλά bpm) αποτελούν μικρότερη διάρκεια σε όλες τις νότες συνολικά ανεξαρτήτως από τη χρονική σχέση που έχουν μεταξύ τους.

Δεδομένου ότι το τέμπο είναι σταθερό, η χρονική σχέση μεταξύ των νοτών σε ένα κομμάτι καθορίζεται από τη χρονική αξία της κάθε νότας και το μέτρο του κομματιού. Στη μουσική, το μέτρο αναφέρεται σε μια διάταξη ιεραρχημένων χτύπων στο χρόνο και επαναλαμβάνονται σε σταθερό, συνήθως, μοτίβο. Το μέτρο του κομματιού αρχικά δηλώνεται ως ένα κλάσμα που υποδεικνύει πόσοι χτύποι στο χρόνο θα πρέπει να περάσουν έως ώστε να επαναληφθεί το μοτίβο μέτρησης. Στη συνέχεια οι νότες γράφονται στα μέτρα τα οποία χω-

ρίζονται από τις διαστολές. Το μέτρο σε πολλές περιπτώσεις καθορίζει και το ύψος της μουσικής, για παράδειγμα τα τραγούδια τύπου βάλς έχουν μέτρο 3/4, τα τραγούδια τύπου Blues έχουν συνήθως μέτρο 6/8 ενώ στην ελληνική μουσική ο χορός του ζειμπέχικου έχει μέτρο 9/8. Μέσα σε ένα μέτρο, η χρονική διάρκεια των νοτών καθορίζεται από τις χρονικές αξίες τους. Οι χρονικές αξίες ενδεικτικά στη μουσική είναι το ολόκληρο, (4 χτύποι), το μισό (2 χτύποι) και οι υπόλοιπες υποδιαίρέσεις τους (τέταρτο - 1 χτύπος, όγδοο μισός - μισός χτύπος κλπ). Ένα μέτρο 4/4 θα πρέπει να συμπληρωθεί από νότες που αθροίζουν στους 4 χτύπους δηλαδή 1 ολόκληρο ή 2 μισά ή 4 τέταρτα κλπ. Οι παύσεις δηλώνουν την απουσία ήχου και χαρακτηρίζονται και αυτές από συγκεκριμένη χρονική αξία (παύση ολόκληρου, παύση τετάρτου κλπ).

Άλλη μια σημαντική έννοια που πρέπει να αναφερθεί είναι η έννοια του ισχυρού χτύπου στο μέτρο. Συνήθως οι ισχυροί χτύποι διαφέρουν ανάλογα με το μέτρο. Φέρνουμε ως παράδειγμα το πιο διαδεδομένο μέτρο, αυτό των 4/4ων. Συγκεκριμένα ο πρώτος και ο τρίτος χτύπος ορίζουν το ισχυρό του μέτρου (θέσεις), ενώ ο δεύτερος και ο τέταρτος ορίζουν τις άρσεις. Ο τρόπος που είναι διατεταγμένες χρονικά οι νότες πάνω στα ισχυρά ή στις άρσεις ενός μέτρου, σε πολλές περιπτώσεις καθορίζει το ρυθμικό σχήμα του κομματιού και κατά συνέπεια το μουσικό του ύψος. Ως παράδειγμα φέρνουμε το ύψος της Reagge μουσικής, της οποίας το ρυθμικό σχήμα βασίζεται στις άρσεις του μέτρου των 4/4ων σε αντίθεση με την Ρόκ μουσική που το ρυθμικό της σχήμα, αν το μέτρο είναι 4/4ων, συνήθως τονίζει τις θέσεις.

Τέλος είναι σημαντικό να τονίσουμε, ότι η μουσική και το ύψος της, καθορίζεται και από τις νότες της σε συνδιασμό με τον τρόπο που είναι διατεταγμένες. Για παράδειγμα κοινότητες μουσικές φράσεις, πχ στην ελληνική παραδοσιακή μουσική, απαιτούν την ύπαρξη συγκεκριμένων νοτών διατεταγμένων χρονικά με συγκεκριμένο τρόπο για να ορίσουν ένα μουσικό μοτίβο που χρησιμοποιείται στα φινάλε των μουσικών κομματιών. Ένα τέτοιο παράδειγμα απεικονίζεται στο Σχήμα 4.



Σχήμα 4: Μουσικό μοτίβο που χρησιμοποιείται στο τέλος παραδοσιακών ελληνικών τραγουδιών. Εδώ θεωρήθηκε ότι το τραγούδι παίζεται στην κλίμακα Λα.

Μουσική Σημειογραφία

Σε αυτό το σημείο κρίνουμε σκόπιμο να παρουσιάσουμε συνοπτικά τον πιο συνηθισμένο τρόπο αναπαράστασης της μουσικής, το πεντάγραμμο. Φυσικά το πεντάγραμμο δεν είναι ο μοναδικός τρόπος αναπαράστασης παρόλαυτά είναι ευρέως διαδεδομένο στους μουσικούς, και περιέχει όλη την πληροφορία που χρειάζεται ένας μουσικός για να εκτελέσει το μουσικό έργο.

Το πεντάγραμμο αποτελούν 5 παράλληλες γραμμές που πάνω τους ή ενδιάμεσα τους γράφονται οι νότες ως μικροί κύκλοι. Κάθε πεντάγραμμο ξεκινάει με ένα σύμβολο, όπως παρατηρείται στο Σχήμα 4 και στο Σχήμα 5. Το σύμβολο αυτό ονομάζεται κλειδί και το συγκεκριμένο, κλειδί του Σολ. Στην ουσία το κλειδί είναι μία σύμβαση για το μουσικό, που καθορίζει το πού θα γράφονται πάνω στο πεντάγραμμο οι διαφορετικές νότες. Πολλοί μουσικοί διαβάζουν σε διαφορετικά κλειδιά ανάλογα με το όργανο που παίζουν. Στη συνέχεια ενδεχομένως να υπάρχει οπλισμός, ο οποίος καθορίζει την κλίμακα στην οποία θα παιχτεί το κομμάτι προσθέτοντας διέσεις και υφέσεις σε όλες τις νότες που δηλώνονται στο σημείο του οπλισμού, όπως φαίνεται στο Σχήμα 6. Στη συνέχεια βλέπουμε το κλάσμα που καθορίζει τη χρονική αξία του κάθε μέτρου. Η συχνότητα των νοτών καθορίζεται από τη θέση τους πάνω στο πεντάγραμμο, και παρατηρούμε ότι πολλές νότες γράφονται και εκτός πενταγράμμου (κάτω ή πάνω) με τη χρήση βοηθητικών γραμμών. Ο τρόπος που ζωγραφίζεται η κάθε νότα θεωρώντας δεδομένη τη θέση της στο πεντάγραμμο καθορίζει την χρονική αξία της νότας. Στο σχήμα 5 παρατηρούμε μια σειρά από ζωγραφισμένα όγδοα τα οποία έχουν χρονική διάρκεια μισό χτύπο και έτσι 8 από αυτά συμπληρώνουν ένα μέτρο 4/4ων.

Όπως παρατηρούμε στο σχήμα 4 οι παύσεις ζωγραφίζονται με διαφορετικά σύμβολα που υποδηλώνουν τη χρονική αξία της παύσης. Επίσης από το σχήμα 4 παρατηρούμε ότι στο δεύτερο μέτρο 6 νότες έχουν γραφτεί κάθετα. Αυτές οι νότες θα συνηχήσουν την ίδια χρονική στιγμή δημιουργώντας στο μουσικό κομμάτι τη διάσταση της πολυφωνικότητας και πιο συγκεκριμένα της αρμονίας, έννοιες που θα μας απασχολήσουν αργότερα στην παρούσα εργασία.



Σχήμα 5: Νότες πάνω στο πεντάγραμμο

Σχήμα 6: Μέρος από το προλούδιο της πρώτης σπουδής του Heitor Villa Lobos για κλασική κιθάρα

Παράλληλα μία παρτιτούρα μπορεί να αναφέρει το επιθυμητό τέμπο του μουσικού κομματιού, το ύφος εκτέλεσης μιας συγκεκριμένης φράσης, και διάφορα άλλα τεχνικά χαρακτηριστικά που αφορούν ειδικότερα το όργανο που προορίζεται να την εκτελέσει.

Καταλαβαίνουμε λοιπόν από την ενδεικτική παρουσίαση της σημειογραφίας, ότι η μουσική αποτυπώνεται με πολύ συγκεκριμένο τρόπο στο χαρτί και ότι έχει αναπτυχθεί μια αρκετά εκτεταμένη γλώσσα η οποία μπορεί να απεικονίσει τις σημαντικότερες πτυχές μίας σύνθεσης.

1.3 Σύνθετες Μουσικές Έννοιες και Μουσικά Χαρακτηριστικά

Αφού είδαμε τον τρόπο με τον οποίο οι νότες αναπαριστούν το μουσικό άκουσμα, πλέον μπορούμε να αναλύσουμε κάποιες πιο σύνθετες μουσικές έννοιες οι οποίες αποτελούν και σημαντικά χαρακτηριστικά της σύνθεσης.

1.3.1 Μουσικές Κλίμακες

Με τον όρο κλίμακα ή σκάλα, εννοούμε μία σειρά από μουσικούς φθόγγους, την οποία χρησιμοποιεί ο μουσικός/συνθέτης μουσικής. Ο αριθμός των μουσικών φθόγγων ποικίλει από μια κλίμακα σε άλλη: μια πεντατονική κλίμακα περιέχει πέντε νότες, ενώ αντίστοιχο αριθμό περιέχουν οι εξατονικές (έξι νότες) και επτατονικές (επτά νότες). Κάθε μουσική κλίμακα χαρακτηρίζεται από μια

συγκεκριμένη διάταξη ως προς τα διαστήματα μεταξύ των φθόγγων της.

Οι κλίμακες εν γένει ονομάζονται σύμφωνα με τη νότα που αποτελεί τη βάση τους. Για παράδειγμα αν η βάση μιας κλίμακας είναι το Ντο, τότε αυτή η κλίμακα θα ονομάζεται Ντο. Οι σχέσεις μεταξύ των νοτών της κλίμακας είναι αυτές που δίνουν τον πρόσθετο χαρακτηρισμό του ονόματος. Ενδεικτικά μια κλίμακα ονομάζεται Ντο Μείζονα (C Major or C+) όταν ξεκινώντας από την βάση της (Ντο) οι επόμενες νότες προκύπτουν από τα συγκεκριμένα διαδοχικά διαστήματα : Τ,Τ,Η,Τ,Τ,Τ,Η όπου Τ εννοούμε απόσταση τόνου και όπου Η εννοούμε απόσταση ημιτονίου αντίστοιχα.

Οποιαδήποτε άλλη Μείζονα κλίμακα είναι η αυτούσια μεταφορά της σχέσης των διαστημάτων που είδαμε παραπάνω, απλώς από άλλη νότα βάσης. Και δεδομένου ότι οι νότες ονομαστικά είναι 12, έχουμε 12 αποτυπώσεις της Μείζονας κλίμακας ανάλογα με τη εκάστοτε βάση. Το ίδιο ισχύει και με όλα τα υπόλοιπα ήδη κλιμάκων. Όταν λέμε λοιπόν ότι η συγκεκριμένη μουσική φράση ανήκει σε κάποια κλίμακα, εννοούμε ότι οι νότες της φράσης ανήκουν όλες ονομαστικά, στις νότες που περιέχει η κλίμακα. Παράλληλα εδώ μπορούμε να αναφέρουμε την έννοια των σχετικών κλιμάκων. Σχετικές Κλίμακες είναι 2 κλίμακες, η 1η μείζονα και η 2η ελασσονα, οι οποίες έχουν τον ίδιο οπλισμό δηλαδή έχουν τις ίδιες νότες αλλά διαφορετική βάση. Για να βρούμε μια σχετική μιας κλίμακας σκεφτόμαστε ότι οι τονικές βαθμίδες των 2 κλιμάκων, έχουν απόσταση διαστήματος 3ης μικρής, παράδειγμα Ντο Μείζονα - Λα Ελάσσονα.

Προφανώς οι διαφορετικές φράσεις της μουσικής σύνθεσης ενδέχεται να ανήκουν και σε διαφορετικές κλίμακες, κάνοντας την περισσότερο πολυδιάστατη και ενδιαφέρουσα. Οι μουσικές κλίμακες είναι ένα σημαντικό κεφάλαιο για τον σπουδαστή μουσικής και είναι και ένα σημαντικό εργαλείο σύνθεσης και αυτοσχεδιασμού. Στην πραγματικότητα η ομαδοποίηση των νοτών σε κλίμακες είναι μια ανθρώπινη διαδικασία η οποία προκύπτει από τη μουσική εμπειρία. Για παράδειγμα η μουσική της ανατολής χρησιμοποιεί πολύ διαφορετικά τις σχέσεις μεταξύ των νοτών από την δυτική μουσική. Συνεπώς η παραδοσιακή ελληνική, τούρκικη,αραβική μουσική κλπ, έχει τις δικές της κλίμακες που στην Ελλάδα ονομάζουμε δρόμους.

Ένα άλλο παράδειγμα είναι η μουσική Τζαζ στην οποία χρησιμοποιούνται πολλές διαφορετικές και 'έξωτικές' κλίμακες, σε διάφορες μουσικές φράσεις, οι οποίες χρησιμοποιούνται κατά κύριο λόγο σε αυτό το είδος μουσικής. Αυτός είναι και ο λόγος που η τεχνική Τζαζ πολλές φορές ξενίζει τον ακροατή που δεν έχει εκτεθεί σε αυτό το είδος αρχικά.

Οι κλίμακες που χρησιμοποιούνται σε ένα μουσικό κομμάτι καθώς και η αλληλουχία τους στις μουσικές φράσεις αποτελούν ένα σημαντικό χαρακτηριστικό της μουσικής σύνθεσης, για το χαρακτηρισμό του είδους της και της πολυπλοκότητάς της, και είναι ένα αναπόσπαστο κομμάτι για τη μελέτη μιας σύνθεσης.

1.3.2 Πολυφωνία και Αρμονία - Η Έννοια της Συγχορδίας

Η μουσική μέχρι και τις αρχές του 12ου αιώνα είναι μονοφωνική. Δηλαδή το μουσικό έργο αποτελεί μία μελωδική γραμμή που κομμάτια του ή εξ ολοκλήρου ανήκει σε συγκεκριμένες μουσικές κλίμακες (όπως τις χαρακτηρίζουμε εκ των υστέρων). Ως πολυφωνία ή πολυφωνική μουσική, χαρακτηρίζεται κάθε μουσικό είδος το οποίο βασίζεται σε πολλές μελωδικές γραμμές, οι οποίες συμπλέκονται κατάλληλα σχηματίζοντας μία πολυγραμμική υφή. Οι φωνές αυτές είναι ανεξάρτητες μεταξύ τους έχοντας την δικιά τους ξεχωριστή μελωδία και παρόλαυτά όταν συνηχούν όλες μαζί δημιουργούν ένα ενιαίο μουσικό νόημα.

Η μουσική είναι πολυφωνική από τον 12ο αιώνα μέχρι την εποχή του Μπαρόκ όπου τότε αρχίζει και εισάγεται η έννοια της αρμονίας. Στη νεότερη μουσική ορολογία η λέξη αρμονία χαρακτηρίζει γενικά το σύνολο των κανόνων που διέπουν το ταυτόχρονο άκουσμα δύο ή περισσότερων φθόγγων, τη συνήχηση. Έτσι, ενώ η αντίστιξη ενδιαφέρεται κυρίως για τις μεμονωμένες φωνές και τη μελωδικά και ρυθμικά αυτόνομη ανάπτυξη της καθεμιάς σ' ένα πολυφωνικό μουσικό έργο, η αρμονία μελετά τις συνηχήσεις δύο ή περισσότερων φθόγγων που προκύπτουν κάθε στιγμή. Στη διάρκεια του 16ου αι. οι συνθέτες άρχισαν να προσέχουν ολοένα και περισσότερο αυτή τη συνήχηση και να εκμεταλλεύονται τις εκφραστικές δυνατότητες που παρείχε.

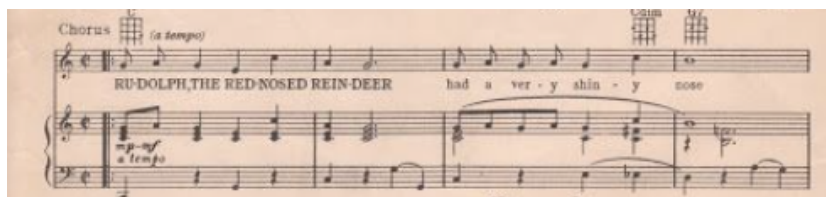
Ως αποτέλεσμα, η μουσική κατά κύριο λόγο από την εποχή του Μπαρόκ μέχρι και σήμερα, δομείται με την λογική αυτή, τείνεται μια ισχυρή αρμονική βάση πάνω στην οποία χτίζεται ξεχωριστή φωνή που περιέχει τη μελωδία του κομματιού. Στη σύγχρονη εποχή και στην πλειοψηφία των πιο μοντέρνων και διαδεδομένων ειδών η σύνθεση βασίζεται σε αρμονικά μοτίβα, ενώ όργανα ή φωνές, αναπαράγουν την κύρια μελωδική γραμμή.

Η Έννοια της Συγχορδίας

Συγχορδία ονομάζεται στη θεωρία της μουσικής το σύνολο τριών τουλάχιστον φθόγγων που ηχούν ταυτόχρονα. Στο τονικό μουσικό σύστημα η συγχορδία σχηματίζεται με επάλληλους φθόγγους που απέχουν μεταξύ τους διάστημα τρίτης, δηλ. με την προσθήκη της τρίτης, της πέμπτης, της έβδομης κ.ο.κ πάνω από ένα βασικό φθόγγο. Ο φθόγγος αυτός ονομάζεται θεμέλιος και αποτελεί τη βάση και το χαρακτηριστικότερο φθόγγο της συγχορδίας. Οι υπόλοιποι φθόγγοι ονομάζονται τρίτη, πέμπτη, έβδομη, ένατη, κ.ο.κ. Αρμονικά πλήρης είναι η συγχορδία που περιέχει τουλάχιστον τρεις φθόγγους, η τρίφωνη συγχορδία.

Η συγχορδία αποτελεί το βασικότερο χαρακτηριστικό στη σύνθεση καθώς δημιουργεί τη βάση πάνω στην οποία θα κινηθεί η μελωδική γραμμή του κομματιού. Η μουσική σύνθεση χωρίζεται νοητά σε δύο μέρη, στο αρμονικό υπόβαθρο που ορίζει η αλληλουχία των συγχορδιών του κομματιού και στη μελωδία που βασίζεται πάνω σε αυτό. Αυτή η εικόνα μπορεί να γίνει αντιληπτή

ακόμα και στον τρόπο που γράφεται η μουσική στην παρτιτούρα του Σχήματος 7. Η έννοια αυτή γίνεται ακόμα περισσότερο αισθητή στα σύγχρονα τραγούδια που εκτελούνται από μπάντες ή ορχήστρες. Εκεί ο ενορχηστρωτής επιλέγει ποιά όργανα θα συνεισφέρουν στο αρμονικό υπόβαθρο της σύνθεσης και ποια θα αναλάβουν ή θα συμπληρώνουν την κύρια μελωδία.

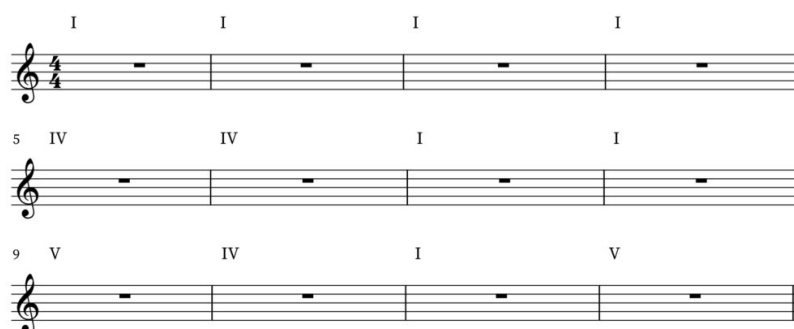


Σχήμα 7: Απόσπασμα από παρτιτούρα πιάνο. Στο πάνω πεντάγραμμο διακρίνουμε τη μελωδική γραμμή, στα δύο κάτω ορίζεται η αρμονική αλληλουχία του κομματιού.

Στην παρούσα εργασία χρησιμοποιούμε τον όρο της τρίφωνης συγχορδίας με νότες τη θεμέλιο, την τρίτη και την πέμπτη. Και πιο συγκεκριμένα τις μείζονες και τις ελάσσονες συγχορδίες. Στις μείζονες συγχορδίες, η τρίτη νότα απέχει από τη θεμέλιο διάστημα τρίτης μεγάλο. Για παράδειγμα στη συγχορδία Ντο - Μι - Σολ ή αλλιώς Ντο μείζονα ή (Ντο+ , C+), η τρίτη νότα (Μι) απέχει από τη θεμέλιο (Ντο) απόσταση 4 ημιτονίων δηλαδή διάστημα τρίτης μεγάλο. Αντίθετα στις ελάσσονες συγχορδίες η τρίτη νότα απέχει από τη θεμέλιο διάστημα τρίτης μικρό. Για παράδειγμα η Λα ελάσσονα συγχορδία θα αποτελείται από τις νότες Λα - Ντο - Μι και η τρίτη νότα (Ντο) θα απέχει από τη θεμέλιο (Λα) τρία ημιτόνια. Αξίζει να αναφέρουμε ότι οι μείζονες συγχορδίες, στο άκουσμα είναι χαρούμενες, ενώ οι ελάσσονες συγχορδίες μοιάζουν μελαγχολικές.

Εξηγήσαμε λοιπόν ότι η αρμονική δομή ενός μουσικού κομματιού χτίζεται από τις συγχορδίες του. Και η αρμονική δομή αυτή καθέ αυτή είναι κατά την άποψή μας το σημαντικότερο χαρακτηριστικό της μουσικής (τουλάχιστον της σημερινής που δημιουργείται με αυτόν τον τρόπο). Ήχος με αρμονική δομή τρίφωνων συγχορδιών, θεωρούμε ότι πληρεί την ικανή συνθήκη να αναγνωρίζεται από το ανθρώπινο αυτί ως μουσική. Παράλληλα ολόκληρα είδη μουσικής χαρακτηρίζονται από την αρμονική δομή τους. Στη Blues μουσική για παράδειγμα, τα λεγόμενα 12 Bar Blues έχουνε πολύ συγκεκριμένη αρμονική δομή με τις βαθμίδες των συγχορδιών (I, IV, V) να διατάσσονται χρονικά σε μια κοινή και πολύ συγκεκριμένη φόρμα. Η φωνή ή το όργανο που αναλαμβάνει την μελωδική γραμμή, βασίζεται πάνω σε αυτή τη φόρμα για να την τοποθετήσει χρησιμοποιώντας κατάλληλες κλίμακες. Στο Σχήμα 8 παρουσιάζεται ένα παράδειγμα της φόρμας των 12 Bar Blues.

12 Bar Blues



Σχήμα 8: Η πιο συνηθισμένη αρμονική φόρμα των 12 Bar Blues.

2 Τεχνητή Νοημοσύνη στη Σύνθεση Μουσικής

Πληθώρα αλγορίθμων και τεχνολογιών τεχνητής νοημοσύνης έχουν αναπτυχθεί για να φέρουν εις πέρας συγκεκριμένες διαδικασίες σχετικά με τον ήχο και γενικότερα με τη μουσική σαν πρότυπο. Με την εξαγωγή σημαντικών χαρακτηριστικών του ήχου έχουμε καταφέρει να υλοποιήσουμε τεχνολογίες που αναγνωρίζουν την πηγή από την οποία προέρχεται, κατατάσσουν μουσικά τραγούδια σε διαφορετικά είδη, συνθέτουν τη δικιά τους μουσική και ενορχηστρώνουν υπάρχουσες μελωδικές γραμμές.

Στο συγκεκριμένο κεφάλαιο θα μελετήσουμε ποιες είναι εκείνες οι αναπαραστάσεις μουσικής που χρησιμοποιούνται ως είσοδος στις διάφορες τεχνολογίες που έχουν αναπτυχθεί. Στη συνέχεια θα εξηγήσουμε την αρχή λειτουργίας των σημαντικότερων υπάρχοντων δικτύων σύνθεσης εικόνων και μουσικής.

2.1 Αναπαραστάσεις Μουσικής

Η μουσική μπορεί να αναπαρασταθεί με διάφορους τρόπους. Όσον αφορά την εκπαίδευση νευρωνικών δικτύων και εφαρμογές μουσικής σύνθεσης, δύο αναπαραστάσεις είναι οι χρησιμότερες. Η συμβολική αναπαράσταση και η αναπαράσταση των χαρακτηριστικών του ήχου της. Και οι δύο μέθοδοι αναπαράστασης έχουν πλεονεκτήματα και μειονεκτήματα, με την συμβολική αναπαράσταση να είναι απλούστερη και υπολογιστικά φθηνή, ενώ η αναπαράσταση

των χαρακτηριστικών του ήχου να είναι περισσότερο ακριβής αλλά υπολογιστικά απαιτητική.

2.1.1 Συμβολική Αναπαράσταση Μουσικής - Το Pianoroll

Η συμβολική αναπαράσταση της μουσικής χρησιμοποιείται ευρέως εδώ και αρκετούς αιώνες. Σε αυτόν τον τρόπο αναπαράστασης, η μουσική, κάθε νότα της, αναπαριστάται με διάφορα σύμβολα. Στις συμβολικές αναπαραστάσεις συχνά υπάρχει η πληροφορία της έντασης της κάθε νότας, του τονικού της ύψους και της διάρκειάς της. Δίνει δηλαδή μεγάλο όγκο πληροφορίας για την ίδια τη σύνθεση και το πώς ο συνθέτης θα επιθυμούσε να εκτελεστεί. Παρόλαυτά τέτοιες αναπαραστάσεις αφήνουν χώρο σε μεγάλο εύρος διαφορετικής ερμηνείας. Για παράδειγμα δύο εκτελέσεις του ίδιου έργου από δύο διαφορετικούς μουσικούς, είναι σχεδόν σίγουρο ότι θα διαφέρουν ελαφρώς σε σημεία. Αυτή η διαφορά οφείλεται στο γεγονός ότι οι συμβολικές μουσικές αναπαραστάσεις απεικονίζουν τη σύνθεση αυτή καθέ αυτή αλλά δεν εμπεριέχουν την πληροφορία της εκτέλεσης του έργου, του ηχοχρώματος του κάθε εκτελεστή και την μουσική του προσέγγιση στο έργο. Επίσης καθώς η συμβολική αναπαράσταση βασίζεται στην ερμηνεία των συμβόλων της, ενδεχομένως πολλά διαφορετικά μουσικά όργανα θα μπορούσαν να την εκτελέσουν, ακόμα και ανθρώπινες φωνές. Συνεπώς η χρήση συμβολικής μουσικής αναπαράστασης παρέχει περιορισμένο μέγεθος ακρίβειας στην ηχητική πληροφορία που κωδικοποιεί.

Από την άλλη πλευρά, οι συμβολικές αναπαραστάσεις αποδίδουν πολύ καλά τη μελωδία, την αρμονία και τον ρυθμό μιας σύνθεσης και αυτός είναι και ένας από τους λόγους που χρησιμοποιούνται τόσο συχνά σε εφαρμογές σύνθεσης μέσω τεχνητής νοημοσύνης. Η κωδικοποίηση του ήχου σε σύμβολα, αποτελεί και έναν εύκολο και αποδοτικό τρόπο για την παροχή πληροφορίας σε έναν δίκτυο και γενικότερα στον υπολογιστή.

Μια από τις ευρέως διαδεδομένες συμβολικές μουσικές αναπαραστάσεις είναι το Pianoroll. Στο Σχήμα 9 μπορούμε να δούμε μία μουσική σύνθεση να αναπαριστάται σε Pianoroll. Όπως παρατηρούμε το Pianoroll είναι ένας πίνακας, P , που τα στοιχεία του έχουν τιμές 0 ή 1. Οι γραμμές του πίνακα αντιστοιχούν σε διαφορετικές συχνότητες και πιο συγκεκριμένα σε διαφορετικές νότες. Οι διαφορετικές στήλες του πίνακα αναπαριστούν διαφορετικές χρονικές στιγμές μέσα στο έργο. Συνεπώς αν

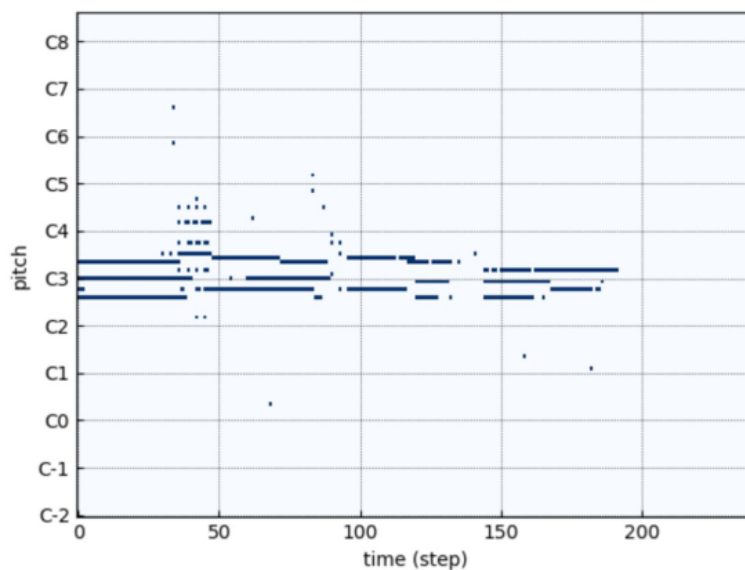
$$p_{ij} = 1, \quad p \in P \quad (1)$$

τότε η i -th νότα παίχτηκε την j -th χρονική στιγμή. Αν για δεδομένη στήλη θ όλα τα

$$p_{ij} = 0, \quad \forall i, \quad p_i \in P \quad (2)$$

τότε την j -th χρονική στιγμή υπάρχει πάυση. Όπως εύκολα γίνεται κατανοητό, για να έχει μία νότα δεδομένη διάρκεια, θα πρέπει το στοιχείο του πίνακα που την αναπαριστά να έχει τιμή ίση με 1 στις διαφορετικές στήλες για όλη τη διάρκεια αυτή.

Καταλαβαίνουμε ότι με τη χρήση αυτής της αναπαράστασης, το μουσικό πρότυπο μεταφράζεται σε έναν πίνακα παρέχοντας πολλά οφέλη και ευκολία όσον αφορά τη χρήση ήδη ανεπτυγμένων τεχνικών και αλγορίθμων που αφορούν τα πρότυπα σε μορφή πινάκων, όπως για παράδειγμα οι εικόνες. Το δίκτυο δεν χρειάζεται να κωδικοποιήσει την πληροφορία σχετικά με τον τρόπο που η μουσική στην πραγματικότητα παίζεται, αλλά κυρίως χρειάζεται απλά να παράγει μια ευχάριστη ακολουθία από σύμβολα. Εν συνεχεία, απαιτείται η χρησιμοποίηση κάποιου άλλου εργαλείου που να μετατρέψει αυτήν την αναπαράσταση σε ήχο, στην δική μας περίπτωση σε MIDI αρχείο.

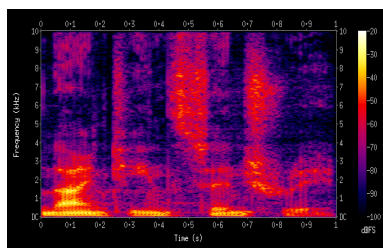


Σχήμα 9: *Pianoroll* - Μία συμβολική μουσική αναπαράσταση [12]

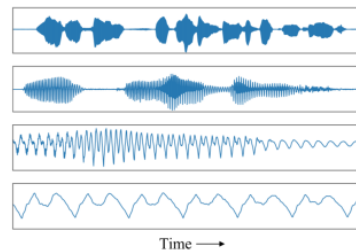
2.1.2 Αναπαράσταση Μουσικής Μέσω των Χαρακτηριστικών του Ήχου της

Ενώ πολλά δίκτυα έχουν χρησιμοποιήσει συμβολικές αναπαραστάσεις μουσικής, με την εξέλιξη της τεχνολογίας και την αύξηση της υπολογιστικής δύναμης των υπολογιστών, έχει δημιουργηθεί η δυνατότητα στην χρησιμοποίηση των χαρακτηριστικών του ήχου μιας μουσικής σύνθεσης. Η ηχητική αναπαράσταση της μουσικής έγκειται στην απευθείας μετάδοση ηχητικών σημάτων, ή στην κωδικοποίηση και στην αποκωδικοποίηση του ήχου με ελάχιστο κόστος ανακατασκευής. Ένα σύνθητες παράδειγμα είναι τα CDs ή τα ηλεκτρονικά αρχεία όπως το mp3. Σε αυτές τις μεθόδους ο ήχος καταγράφεται απευθείας και κωδικοποιείται σε ένα συγκεκριμένο format που επιτρέπει στους καταγραφόμενους ήχους να μπορούν να ανακατασκευαστούν σχεδόν αψεγάδιαστα.

Οι ακουστικές αναπαραστάσεις της μουσικής αποτελούν οποιαδήποτε χαρακτηριστικά εξάγονται ή κατασκευάζονται από τα raw audio files και συνήθως είναι οι κυματομορφές του ήχου, σπεκτρογράμματα και χρωμογράμματα, ανάλογα με την εφαρμογή που απαιτείται να υλοποιηθεί.



(α') Ακουστική Αναπαράσταση
Σπεκτρογράμματος



(β') Ακουστική Αναπαράσταση
Κυματομορφής

Σχήμα 10: Οι ακουστικές αναπαραστάσεις του Σπεκτρογράμματος και της Κυματομορφής

Οι κυματομορφές είναι σχετικά απλές στη χρήση, όμως απαιτούν μεγάλο όγκο δεδομένων για να δημιουργηθεί ακόμα και ένα δευτερόλεπτο ήχου. Συνήθως αποθηκεύονται σε αρχεία τύπου .WAV ως μια σειρά του πλάτους του ηχητικού κύματος συναρτήσει με το χρόνο. Ο χρόνος εδώ δεν είναι συνεχής, διακριτοποιείται με μια συχνότητα δειγματοληψίας. Συνεπώς ενώνοντας όλες αυτές τις τιμές στο χρόνο, παράγεται μια αναπαράσταση του αναλογικού ακουστικού κύματος που είναι και η κυματομορφή. Συνήθως τα αρχεία τύπου .WAV εμπεριέχουν 44.100 δείγματα για την παραγωγή ήχου ενός δευτερολέπτου. Η ποσότητα των δεδομένων είναι τεράστια, και μπορεί να οδηγήσει σε αδυναμία του δικτύου να αναγνωρίσει τις εξαρτήσεις σημείων που απέχουν μερικά δευτε-

ρόλεπτα κατά την εκπαίδευση του.

Τα σπεκτρογράμματα είναι αναπαραστάσεις ήχου στο πεδίο των συχνοτήτων. Είναι δισδιάστατες αναπαραστάσεις χρόνου - συχνότητας που περιέχουν πληροφορία σχετικά με τον τρόπο που το συχνοτικό περιεχόμενο ενός ήχου αλλάζει στο χρόνο [30]. Τα σπεκτρογράμματα δεν είναι τόσο δημοφιλή σε εφαρμογές σύνθεσης μουσικής μέχρι τώρα, παρόλαυτά μέσω αυτών μπορούν να εξαχθούν αρκετά ποιοτικά χαρακτηριστικά για εφαρμογές κατηγοριοποίησης ήχου (αναγνώριση φωνής , music genre classification κ.α.)

2.1.3 Ποιά Αναπαράσταση Είναι Καλύτερη;

Το κύριο μειονέκτημα στην χρησιμοποίηση των ακουστικών αναπαραστάσεων της μουσικής είναι η τεράστια υπολογιστική πολυπλοκότητα που προκύπτει. Το μέγεθος της πληροφορίας που κωδικοποιείται στις κυματομορφές ή στα σπεκτρογράμματα είναι τάξεις μεγέθους μεγαλύτερο από αυτό των Pianorolls. Ως αποτέλεσμα είναι δυσκολότερο για ένα δίκτυο να μάθει τις σημαντικές δομές που κρύβονται στα δεδομένα εκπαίδευσης, πράγμα το οποίο μπορεί να οδηγήσει σε μη ρεαλιστικά αποτελέσματα, αν το δίκτυο δεν είναι αρκετά εύρωστο σε ακουστικά δεδομένα.

Από την άλλη πλευρά, οι ακουστικές αναπαραστάσεις της μουσικής είναι ικανές να απεικονίσουν με μεγαλύτερη ακρίβεια τα πραγματικά ηχητικά δεδομένα που θα χρησιμοποιηθούν για την εκπαίδευση ενός δικτύου, μιας και προκύπτουν κατευθείαν από το ηχητικό σήμα [11]. Μέσω αυτού του τρόπου αναπαράστασης, μπορούν να εμπεριέχονται πληροφορίες σχετικά με το όργανο που εκτελεί τη σύνθεση, τις φράσεις που μπορεί να τραγουδάει ο τραγουδιστής, τα ηχοχρώματα της απόδοσης, χαρακτηριστικά τα οποία δεν μπορούν με κάποιο τρόπο να συμπεριληφθούν σε συμβολικές αναπαραστάσεις. Γιαυτό το λόγο, ενώ δίκτυα όπως το MuseGap έχουν περίπλοκες αρχιτεκτονικές για να μπορούν να μάθουν τις αλληλοεξαρτήσεις των διαφορετικών οργάνων των συμβολικών δεδομένων εκπαίδευσης, τα ακουστικά δεδομένα όπως η κυματομορφή , εξαρχής εμπεριέχουν τέτοιες πληροφορίες και χαρακτηριστικά.

Όσον αφορά τους δικούς μας σκοπούς σε αυτή την εργασία, κρίθηκε σκόπιμο η χρησιμοποίηση του δικτύου MuseGap το οποίο όπως αναφέραμε χρησιμοποιεί συμβολικές αναπαραστάσεις μουσικής. Ο λόγος είναι αρχικά τεχνικός, τα δίκτυα συνθέτες που χρησιμοποιούν ακουστικές αναπαραστάσεις έχουν μεγάλες υπολογιστικές απαιτήσεις, που για τα πλαίσια της συγκεκριμένης διπλωματικής, δεν ήταν εύκολα διαθέσιμες. Παράλληλα στη συγκεκριμένη δουλειά ερευνάται περισσότερο η ποιότητα της σύνθεσης σε όρους θεωρίας της μουσικής, μελωδίας και αρμονικής συνοχής, χαρακτηριστικά τα οποία εμπεριέχονται στις συμβολικές αναπαραστάσεις. Τα κομμάτια της ερμηνείας και του ηχοχρώματος είναι σημαντικά στην εκτέλεση της μουσικής, παρόλαυτά στην

παρούσα εργασία δεν λαμβάνονται υπόψιν ως παράμετροι της σύνθεσης λόγω της χρήσης της συγκεκριμένης αναπαράστασης.

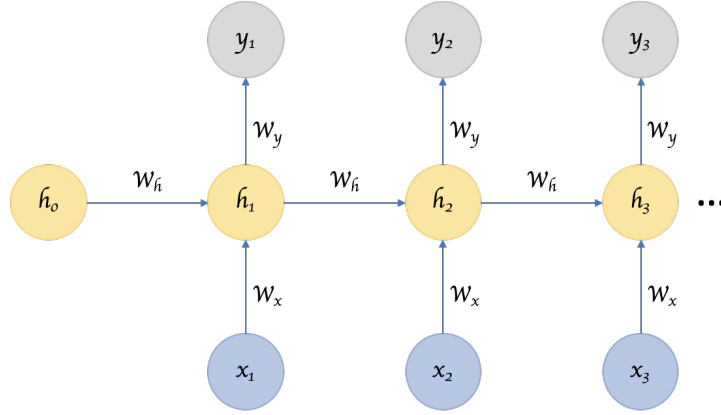
2.2 Νευρωνικά Δίκτυα στη Σύνθεση Μουσικής

Εδώ και εκατοντάδες χρόνια ο η ανθρωπότητα προσπαθεί και ερευνά τρόπους σύνθεσης και παραγωγής μουσικής. Μία ιστορική επισκόπηση όλων αυτών των τεχνικών παρέχεται με μεγάλη λεπτομέρεια στο βιβλίο *Algorithmic Composition: Paradigms of Automated Music Generation by Gerhard Nierhaus* [25]. Μέσα στις 8 τεχνικές που αναλύονται στο συγκεκριμένο βιβλίο, μία από αυτές είναι και η χρήση της Τεχνητής Νοημοσύνης και των Νευρωνικών Δικτύων για αυτό το σκοπό. Στο συγκεκριμένο κεφάλαιο θα εξηγηθούν οι τρεις σημαντικότεροι τύποι Νευρωνικών Δικτύων που έχουν αναπτυχθεί για τη σύνθεση μουσικής, τα Recurrent Neural Networks (RNNs), Variational Autoencoders (VAE) και Generative Adversarial Neural Networks (GANs), η αρχή λειτουργίας τους και οι αρχιτεκτονικές που έχουν υλοποιηθεί ως εργαλεία σύνθεσης.

2.2.1 Recurrent Neural Networks (RNN's)

Τα συνηθισμένα νευρωνικά δίκτυα αποτελούνται από επίπεδα νευρώνων. Τα πρότυπα εισόδου εισάγονται στο δίκτυο στους νευρώνες εισόδου, και όλοι οι νευρώνες αυτοί συνέονται με τους νευρώνες του επόμενου επιπέδου. Η έξοδος του κάθε επιπέδου εξαρτάται μόνο από την δεδομένης χρονικής στιγμής είσοδο, δεν έχουν δηλαδή μνήμη και αυτά τα δίκτυα χαρακτηρίζονται ως 'feed forward' δίκτυα. Τα RNN's διαφέρουν από τα feed - forward δίκτυα, συμπεριλαμβάνοντας κάποια μορφή 'κρυφής' πληροφορίας η οποία χρησιμοποιείται για να αποθηκεύσει γνώση σχετικά με το παρελθόν. Η πιο κοινή πρακτική είναι η χρησιμοποίηση των Long Short-Term Memory cells [15] που επιτρέπουν στο δίκτυο να έχει μνήμη αντίθετα με τα συμβατικά νευρωνικά δίκτυα ή τα συνελκτικά δίκτυα. Η μνήμη αυτή μπορεί να χρησιμοποιηθεί σε εφαρμογές όπου υπάρχει εξάρτηση με το χρόνο, όπως για παράδειγμα σε δεδομένα χρονοσειρών και ειδικότερα, στην περίπτωση μας σε παραγωγή μελωδίας και μουσικής. Οι συγγραφείς ενός εκ των πρώτων άρθρων που χρησιμοποίησαν RNN's στη σύνθεση μουσικής γράφουν στο άρθρο τους :

'Lacking the ability to store any information about the past, [a feedforward] network would be unable to keep track of where it is in a song. In principle an RNN does not suffer from this limitation. With recurrent connections it can use hidden layer activations as memory and thus is capable of exhibiting (seemingly arbitrary) temporal dynamics. In practice, however, RNNs do not perform very well at this task.' [14]



Σχήμα 11: Η αρχιτεκτονική των *Recurrent Neural Networks*

Τα RNN's εκπαιδεύονται προσπαθώντας να προβλέψουν την επόμενη τιμή σε μία δωσμένη ακολουθία $X = \{x_1, x_2, \dots, x_T\}$ σύμφωνα με τη σχέση (3)

$$p(X) = \prod_{i=0}^{T-1} p(x_{i+1}|x_1, \dots, x_i) \quad (3)$$

Στο σχήμα 11 μπορούμε να δούμε την αρχιτεκτονική των RNN's. Η ακολουθία εισόδου, η οποία στην δική μας εφαρμογή είναι μια ακολουθία από νότες, περνάει στους νευρώνες εισόδου. Για κάθε νότα εισόδου x_i προβλέπεται η επόμενη νότα y_i , ενώ ενδιάμεσα, το επίπεδο h_0, h_1, \dots, h_T περιέχει όλη την πληροφορία των προηγούμενων νοτών εισόδου. Το επίπεδο H δηλαδή, αναπαριστά το κελί μνήμης όπως το *LSTM* [23].

$$h_t = H(h_{t-1}, x_{i=t}) \quad (4)$$

$$h_t = f(W^{(hh)}h_{t-1} + W^{(hx)}x_t) \quad (5)$$

Στο επίπεδο εξόδου εφαρμόζεται η Softmax συνάρτηση έτσι ώστε να επιλεγεί η προβλεπόμενη νότα με τη μεγαλύτερη πιθανότητα και να προστεθεί στην ακολουθία:

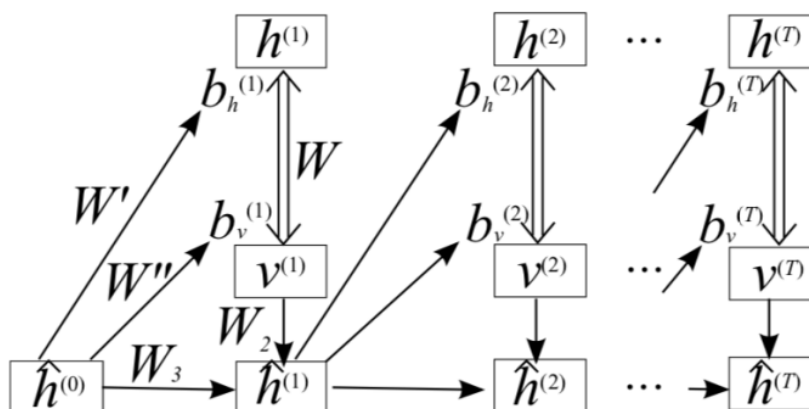
$$p(x_{i+1}|x_1, \dots, x_i) = \text{softmax}(MLP(h_t)) \quad (6)$$

Συνήθως για την εκπαίδευση επιλέγεται η ελαχιστοποίηση μιας συνάρτησης κόστους (συνήθως cross entropy) μεταξύ της προβλεπόμενης νότας και της νότας που βρίσκεται πράγματι στα δεδομένα εκπαίδευσης. Για τη μουσική σύνθεση δε, ένα εκπαιδευμένο *RNN* ξεκινώντας από μία δωσμένη, ή τυχαία νότα, μπορεί και αναπαράγει τις επόμενες δεσμεύοντας κάθε μία με την προηγούμενη.

Στη συνέχεια θα παρουσιάσουμε τα σημαντικότερα *RNN* δίκτυα για την παραγωγή μουσικής που έχουν υλοποιηθεί τα τελευταία χρόνια παραθέτοντας και πληροφορίες για την αρχή λειτουργίας τους.

RNN - RBM

Μία από τις σημαντικότερες προκλήσεις στην παραγωγή μουσικής με *RNN*'s χρησιμοποιώντας συμβολικές αναπαραστάσεις μουσικής, είναι η ικανότητα του δικτύου να μπορεί να συνθέτει πολυφωνική μουσική. Το 2012 ο Boulanger, et. al προσπάθησε να αντιμετωπίσει αυτό το πρόβλημα εισάγωντας στο *RNN* μία δομή που χρησιμοποιεί ένα ενεργειακό μοντέλο που ονομάζεται Restricted Boltzman Machine (RBM).



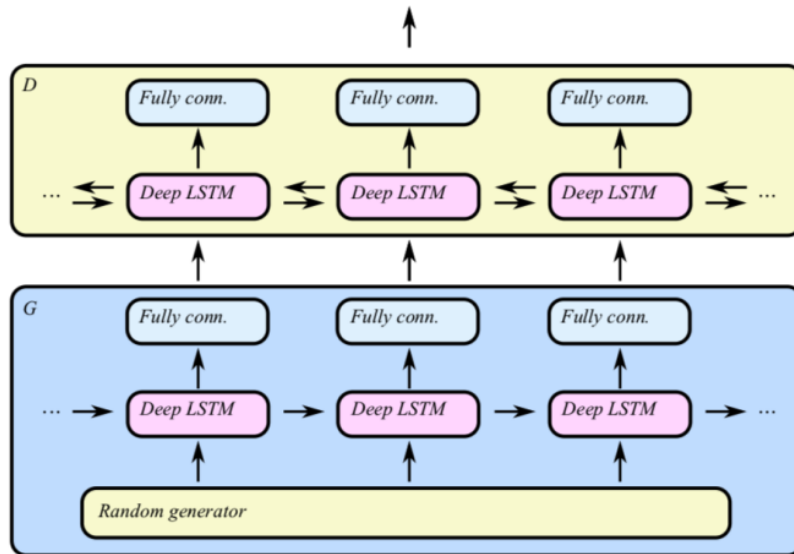
Σχήμα 12: Η αρχιτεκτονική των *RNN - RBM*. Στο κάτω μέρος της αρχιτεκτονικής βρίσκονται οι νευρώνες του *RNN*, ενώ οι νευρώνες του ανώτερου μέρους αποτελούν το *RBM*. [2]

Το *RNN - RBM* δίκτυο χρησιμοποίησε τα restricted Boltzman machines για να μπορέσει να μοντελάρει πολυφωνικές δομές όπως οι συγχορδίες. Παρόλαυτά η κύρια αδυναμία του είναι η έλλειψη της μακρόχρονης μνήμης. Επίσης οι αλληλεπιδράσεις των δύο *RNN* και *RBM* κάνουν την αρχιτεκτονική εξαιρετικά σύνθετη όπως μπορεί κανείς να καταλάβει από το Σχήμα 12.

C-RNN-GAN

Το πρώτο δίκτυο *RNN* το οποίο χρησιμοποίησε ακουστική αναπαράσταση για σύνθεση μουσικής, ήταν το C-RNN-GAN το 2016 [24] η αρχιτεκτονική του οποίου παρουσιάζεται στο σχήμα 13. Οι συνδέσεις του ακολουθούν το μοτίβο ενός κλασικού *RNN* αλλά αντί να προβλέπει μία νότα κάθε φορά, χρησιμοποιείται μια ανταγωνιστική μέθοδος (όμοια με αυτή στα GANs). Στην

πραγματικότητα χρησιμοποιείται ένας Generator και ένας Discriminator ακριβώς όπως και στα GANs. Η εκπαίδευση εγκείται σε ένα min - max πρόβλημα βελτιστοποίησης. Και τα δύο κομμάτια, απαρτίζονται από LSTM's τα οποία στη συνέχεια ενώνονται με πλήρως συνδεδεμένα επίπεδα.



Σχήμα 13: Η αρχιτεκτονική του δικτύου C-RNN-GAN [24]

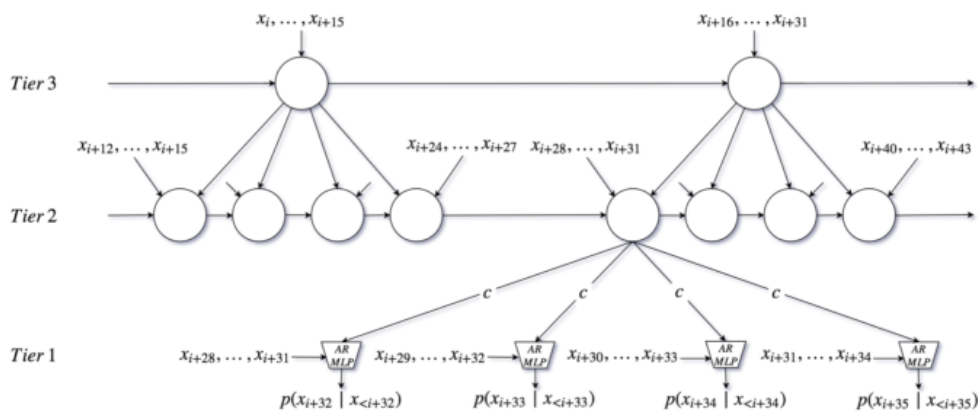
Τα αποτελέσματα του συγκεκριμένου δικτύου μπορούν να χαρακτηριστούν επιτυχημένα, αν και κατά ομολογία των δημιουργών του, δεν είναι σε ικανοποιητικό βαθμό ρεαλιστικά. Στο Σχήμα 14 παρουσιάζεται μια μουσική σύνθεση του C-RNN-GAN.



Σχήμα 14: Μουσική σύνθεση του δικτύου C-RNN-GAN [24]

SampleRNN

Το SampleRNN εισήγαγε την δυνατότητα της μακροπρόθεσμης μνήμης στην παραγωγή μουσικής, χρησιμοποιώντας μια ιεραρχική δομή [23]. Όπως φαίνεται στο Σχήμα 15, στο δίκτυο αυτό χρησιμοποιήθηκαν τρία tiers τα οποία έχουν την RNN δομή ένα για κάθε διαφορετική κλίμακα του χρόνου. Κατά την εκπαίδευση χρησιμοποιήθηκε η τεχνική Truncated Back Propagation Through Time (Truncated BPTT)[26]. Τα αποτελέσματα ήταν αρκετά καλά, παρόλαυτά δεν μπορούσαν να συγκριθούν με πραγματικές συνθέσεις.



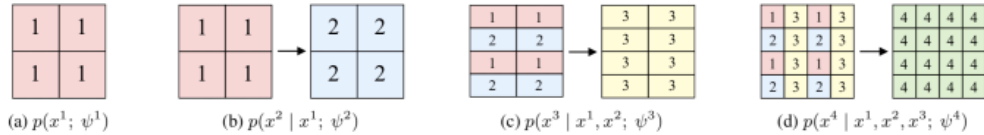
Σχήμα 15: Η αρχιτεκτονική του SampleRNN [23]

MelNet

Η ίδια ιεραρχική δομή στην αρχιτεκτονική του SampleRNN εμφανίζεται και στο MelNet. Σε αυτή την περίπτωση χρησιμοποιείται η αναπαράσταση του σπεκτρογράμματος του ήχου αντί για την κυματομορφή, με σκοπό την καλύτερη απεικόνιση της μακροπρόθεσμης δομής στα δεδομένα. Η ιεραρχική προσέγγιση που υλοποιήθηκε στο MelNet είναι αρκετά πολύπλοκη. Όπως φαίνεται και στο Σχήμα 16, το δίκτυο χρησιμοποιεί διαβαθμίσεις. Στην πρώτη διαβάθμιση (a) το output παράγεται χωρίς να δεσμεύεται από κάποια προηγούμενη πληροφορία. Όσο προχωράμε σε μεγαλύτερο αριθμό tier, οι επόμενες διαβαθμίσεις δεσμεύονται από τις προηγούμενες, δημιουργώντας έτσι το χαρακτηριστικό της μακράς μνήμης.

Το MelNet παράγει σχετικά καλά αποτελέσματα από άποψη σύνθεσης, παρόλαυτά χαμηλής ηχητικής ποιότητας. Οι ίδιοι οι δημιουργοί του προτείνουν το συνδιασμό του MelNet μαζί με κάποιο άλλο δίκτυο όπως πχ το WaveNet για καλύτερα αποτελέσματα. Όπως και να έχει στο MelNet χρησιμοποιήθηκε

για πρώτη φορά το σπεκτρόγραμμα ως μουσική αναπαράσταση σε κάποιο δίκτυο και τα αποτελέσματά του κάνουν εμφανή τη χρησιμότητα χαρακτηριστικών από το πεδίο των συχνοτήτων στην μουσική σύνθεση.

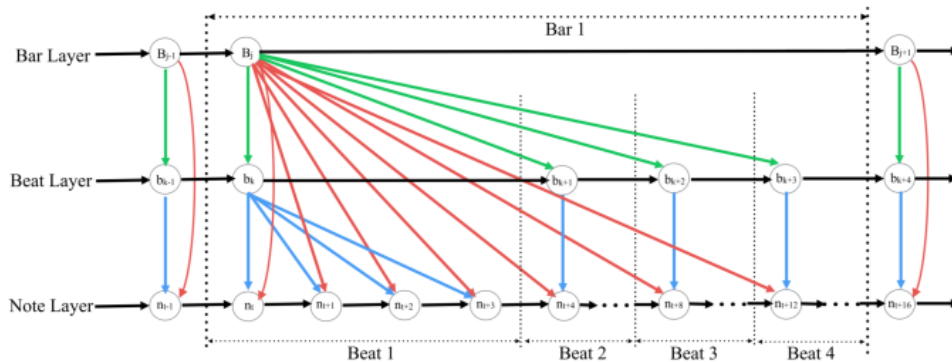


Σχήμα 16: Η αρχιτεκτονική του MelNet [30]

HRNN

Το νεότερο RNN δίκτυο για σύνθεση μουσικής είναι το HRNN το οποίο χρησιμοποιεί συμβολική αναπαράσταση μουσικής στην είσοδό του [32]. Η αρχιτεκτονική του δικτύου αυτού είναι όμοια με αυτή του SampleRNN. Έχει και αυτό τρεις διαβαθμίσεις, χρησιμοποιώντας RNN's υπεύθυνα για την παραγωγή νοτών, ρυθμού και μέτρων.

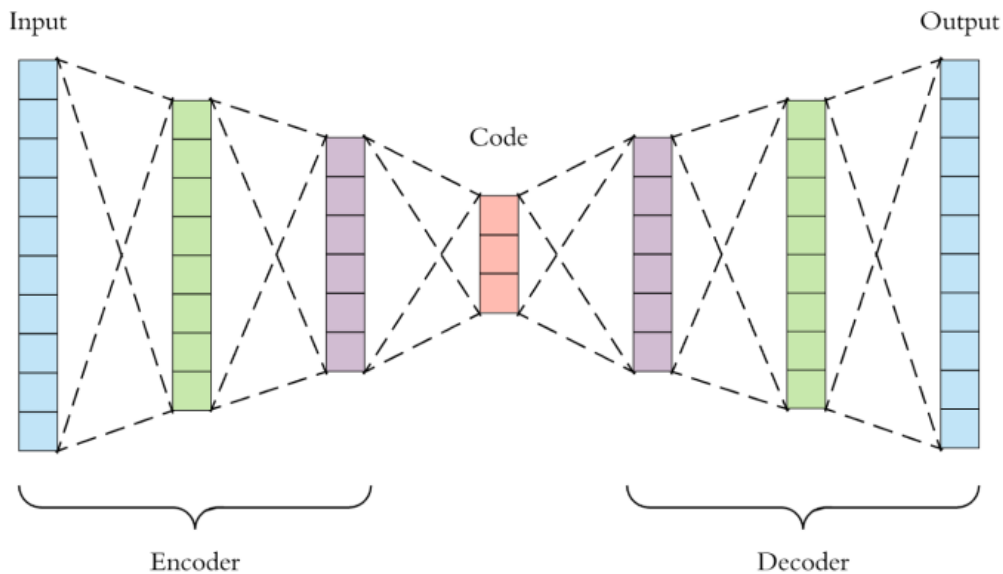
Τα αποτελέσματα του HRNN παρουσιάζουν εξαιρετικό ενδιαφέρον, καθώς σε περίπου 2000 μουσικά δείγματα που δημιουργήθηκαν από το δίκτυο, το 33.69 % αυτών χαρακτηρίστηκε από 659 ανθρώπους ως πραγματικά δείγματα.



Σχήμα 17: Η αρχιτεκτονική του HRNN [32]

2.2.2 Variational AutoEncoders (VAEs)

Το πρώτο χαρακτηριστικό που συνήθως μας έρχεται στο μυαλό όταν αναφερόμαστε στους AutoEncoders είναι το ιδιαίτερο σχήμα της δομής τους τους, όπως φαίνεται στο Σχήμα 18.



Σχήμα 18: Η συνηθισμένη δομή ενός AutoEncoder [6]

Τα δίκτυα αυτού του είδους αποτελούνται από δύο μέρη, τον Encoder και τον Decoder. Συγκεκριμένα ο Encoder επιχειρεί να συμπίεσει τα πρότυπα εισόδου (όπως εικόνα ή ήχο) σε μία ασαφή δομή που στο σχήμα 18 συμβολίζεται ως Code. Το συμπιεσμένο δείγμα ανήκει σε έναν χώρο αρκετά μικρότερων διαστάσεων από τον χώρο που βρίσκονται τα δείγματα εισόδου που ονομάζεται λανθάνων χώρος ή Latent Space.

Ο Decoder χρησιμοποιεί στην είσοδό του τα συμπιεσμένα πρότυπα του Latent Space και σκοπό έχει να τα ανακατασκευάσει, με τη λιγότερη δυνατή απόκλιση. Το 2013 οι Kingma, et. al, εισήγαγαν τον Variational AutoEncoder, ο οποίος σε αντίθεση με τους συνηθισμένους AutoEncoders, υποθέτει μια άγνωστη κατανομή πάνω στα δεδομένα και ως στόχο έχει τον προσδιορισμό παραμέτρων στην κατανομή αυτή. Θεωρούμε λοιπόν ως x τη μεταβλητή που αναπαριστά τα δεδομένα τα οποία παράγει ο Decoder και υποθέτουμε ότι κάθε ένα από τα x αντιστοιχεί σε ένα διάνυσμα z του Latent Space.

Αρχικά ένα διάνυσμα z επιλέγεται τυχαία από την prior κατανομή $p(z)$. Στη συνέχεια ένα πρότυπο x παράγεται από την δεσμευμένη συνάρτηση πυκνότητας πιθανότητας $p(x|z)$. Προφανώς τώρα ο Decoder αντιπροσωπεύει το μηχανισμό παραγωγής προτύπων δεδομένου του λανθάνοντος διανύσματος z , $p(x|z)$ ενώ ο Encoder αντιπροσωπεύει την ακριβώς ανάποδη διαδικασία, $p(z|x)$. Συμβολίζοντας με θ τις άγνωστες παραμέτρους της κατανομής αυτής έχουμε από το θεώρημα Bayes:

$$p_\theta(z|x) = \frac{p_\theta(x|z)p(z)}{p_\theta(x)} \quad (7)$$

όπου

$$p_\theta(x) = \int p_\theta(x|z)p_\theta(z)dz \quad (8)$$

Εν γένει δεν υπάρχει κλειστής μορφής λύση για τον υπολογισμό της posterior και γιαυτό το λόγο χρησιμοποιείται η τεχνική Variational Inference. Σύμφωνα με αυτή, η ζητούμενη κατανομή, προσεγγίζεται μέσω μιας οικογένειας κατανομών, $q_\lambda(z|x)$ όπου λ το σύνολο των παραμέτρων της οικογένειας αυτής.

Ένα μέτρο σύγκρισης της ομοιότητας δύο κατανομών είναι η απόσταση Kullback - Leibler divergence. Συνεπώς η $q_\lambda(z|x)$ που ελαχιστοποιεί την KL divergence θα προσεγγίζει καλύτερα την $p(z|x)$:

$$q_\lambda(z|x) = \operatorname{argmin}_\lambda KL\{q_\lambda(z|x)||p(x|z)\} \quad (9)$$

Κατασκευάζουμε την ποσότητα:

$$J(\lambda) = E_{q_\lambda}\{\log(p(z, x))\} - E_{q_\lambda}\{\log(q_\lambda(z|x))\} \quad (10)$$

Και καταλήγουμε :

$$\log p(x) = J(\lambda) + KL\{q_\lambda(z|x)||p(x|z)\} \quad (11)$$

Από την ανισότητα Jensen ξέρουμε ότι ο δεύτερος όρος της παραπάνω εξίσωσης είναι μεγαλύτερος ή ίσος με το μηδέν, συνεπώς η ελαχιστοποίηση του είναι ισοδύναμη με τη μεγιστοποίηση της $J(\lambda)$. Αντιστοιχίζοντας τώρα τους επιμέρους όρους της $J(\lambda)$ με τους Encoder και Decoder έχουμε και τη συνάρτηση κόστους των VAE:

$$L_{VAE} = E_{q_\theta(z|x)}[\log p_\phi(x|z)] - \beta D_{KL}[(q_\theta(z|x)||p(z))] \quad (12)$$

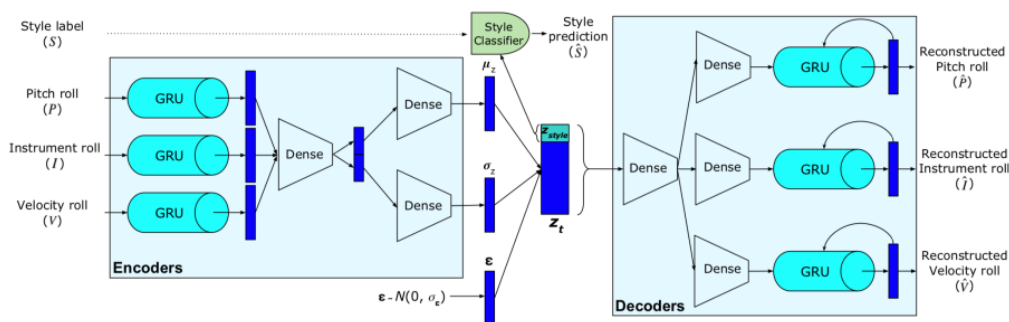
Η παραπάνω συνάρτηση κόστους ελαχιστοποιείται κατά την εκπαίδευση του VAE και το σύνολο των παραμέτρων ϕ και θ μαθαίνονται. Οι δύο όροι είναι το κόστος της ανακατασκευής του προτύπου και η KL divergence και αυτοί εξισορροπούνται με τη ρυθμιστική παράμετρο β . Μέσω της ελαχιστοποίησης

του κόστους ανακατασκευής, πετυχαίνουμε ελάχιστη αλλαγή στα δύο πρότυπα, στο αρχικό και στο τελικά παραγόμενο από τον Decoder, ενώ η KL divergence φροντίζει έτσι ώστε ο Latent Space να είναι συμπαγής και να έχει δομή, δηλαδή εξασφαλίζει ότι τα δεδομένα που προέρχονται από την ίδια κλάση, βρίσκονται κοντά και παράλληλα.

Στη συνέχεια θα παρουσιάσουμε τα σημαντικότερα VAE δίκτυα για την παραγωγή μουσικής που έχουν υλοποιηθεί τα τελευταία χρόνια παραθέτοντας και πληροφορίες για την αρχή λειτουργίας τους.

MIDI-VAE

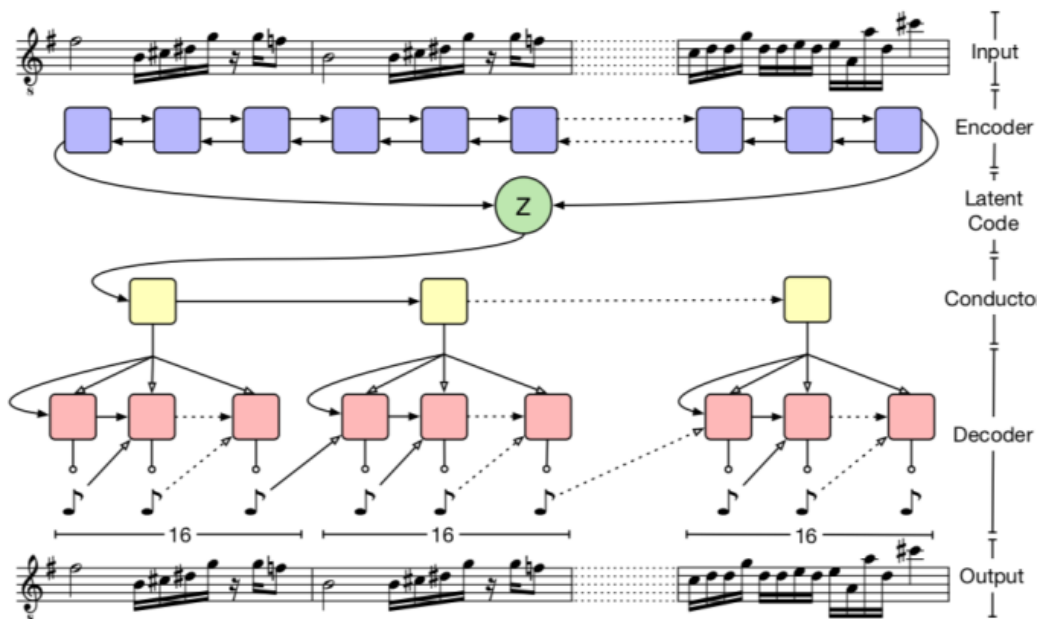
Το MIDI-VAE είναι ένα VAE δίκτυο που εκπαιδεύεται σε δεδομένα συμβολικής μουσικής, και ειδικότερα σε MIDI αρχεία [3]. Χρησιμοποιεί GRUs τα οποία είναι μια βελτιωμένη εκδοχή των RNN layers όπως φαίνεται στο σχήμα 19. Η αρχιτεκτονική του, αν και απλή, του επιτρέπει να χρησιμοποιήσει τον Latent Space έτσι ώστε να μπορεί να τη μουσική δωσμένων MIDI αρχείων σε μουσική διαφορετικού ύφους.



Σχήμα 19: Η δομή του MIDI-VAE [3]

MusicVAE

Το MusicVAE είναι ένα ιεραρχικό VAE μοντέλο που χρησιμοποιεί τη συμβολική αναπαράσταση μουσικής. Οι δημιουργοί του χρησιμοποιούν τον latent space του δικτύου και καταφέρνουν μεταξύ άλλων τη μίξη δύο διαφορετικών μουσικών κομματιών σε ένα. Ιεραρχική δομή παρατηρείται μόνο στον Decoder του δικτύου, ενώ ο Encoder είναι ένα επίπεδο δομής RNN, όπως φαίνεται στο σχήμα 20.



Σχήμα 20: Η δομή του MusicVAE [27]

Στο συγκεκριμένο δίκτυο επίσης υπάρχει η δυνατότητα ελέγχου των στοιχείων της μουσικής σύνθεσης, με την προσπάθεια εύρεσης σημαντικών κατευθύνσεων στον latent space που είναι υπεύθυνες για τα διάφορα χαρακτηριστικά της παραγόμενης μουσικής. Η συγκεκριμένη εφαρμογή είναι και το κύριο αντικείμενο της παρούσας εργασίας.

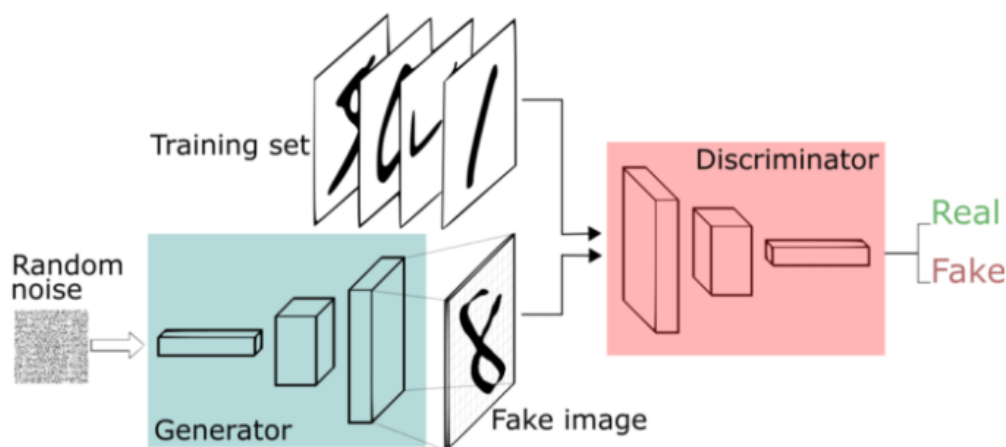
Jukebox

Το Jukebox χρησιμοποιεί την ακουστική αναπαράσταση της μουσικής και είναι ένα VQ-VAE μοντέλο που μπορεί και παράγει όχι μόνο μουσική αλλά και τραγούδι. Το μοντέλο αυτό χρησιμοποιεί μια συγκεκριμένη εφαρμογή των VAE, τα Vector - Quantized VAE[29]. Η δομή του δικτύου είναι σχετικά πολύπλοκη και εξηγείται αναλυτικά από τους δημιουργούς του στο άρθρο [8].

Το δίκτυο αυτό, μπορεί να παράγει μουσική επιθυμητού είδους καθώς και συγκεκριμένου καλλιτέχνη. Η κύρια αδυναμία του είναι η έλλειψη δομής για μακροπρόθεσμη μνήμη καθώς και η ταχύτητα παραγωγής μουσικής. Ενδεικτικά το Jukebox χρειάζεται περίπου 9 ώρες για την παραγωγή 1 λεπτού μουσικής.

2.2.3 Generative Adversarial Networks (GANs)

Τα GANs είναι ένας καινούριος τύπος δικτύων που εισήγαγε ο Ian Goodfellow το 2014 [16]. Τα GANs αποτελούνται από δύο επί μέρους δίκτυα, τον Generator και τον Discriminator. Σκοπός του Generator είναι η παραγωγή προτύπων που φαίνονται όσο το δυνατόν περισσότερο αληθινά. Κοιτάζοντας το Σχήμα 1, μπορούμε να δούμε τα εκπληκτικά αποτελέσματα του Generator του StyleGan-2 όπου τα παραγόμενα πρόσωπα φαίνονται σαν πραγματικές φωτογραφίες πραγματικών προσώπων και είναι δύσκολο να καταλάβει κανείς ότι έχουν δημιουργηθεί από δίκτυο με τεχνητό τρόπο. Ο Generator έχει ως είσοδο θόρυβο, δηλαδή κάποιο τυχαίο διάνυσμα χαμηλότερης διάστασης από το πρότυπο που θα δημιουργηθεί τελικά. Η είσοδος αυτή, όπως και στα VAE ζεί σε έναν χώρο χαμηλής διάστασης που ονομάζεται latent space. Το διάνυσμα αυτό θα εισαχθεί στο Generator και στη συνέχεια, παίρνοντας μέσα από τα layers του, τα οποία θα είναι τέτοια ώστε να εξυπηρετούν το είδος των επιθυμητών παραγόμενων προτύπων, θα δημιουργήσει ένα πρότυπο στην έξοδο. Μια σχηματική αναπαράσταση της δομής του δικτύου παρουσιάζεται στο Σχήμα 21.



Σχήμα 21: Η δομή του GAN

Ο Discriminator, δέχεται πραγματικά πρότυπα στην εισοδό του καθώς και πρότυπα τα οποία έχουν δημιουργηθεί από τον Generator. Σκοπός του είναι να κατατάξει τα πρότυπα σε πραγματικά ή μη. Συνεπώς το δίκτυο αυτό είναι ένας Classifier και αποτελείται από layers, το είδος των οποίων καθορίζεται ανάλογα με τα πρότυπα που καλείται να διαχωρίσει.

Συχνά η εκπαίδευση αυτού του δικτύου χαρακτηρίζεται ως ένα παιχνίδι "κλέφτη και αστυνόμου", όπου ο Generator επιχειρεί να παράγει τόσο πειστικά πρότυπα έτσι ώστε να εξαπατήσει τον Discriminator, ενώ αντίθετα ο δεύτερος εκπαιδεύεται για να μάθει πότε ένα πρότυπο δεν είναι πραγματικό και προέρχεται από τον πρώτο.

Για να παρουσιάσουμε τον τρόπο που εκπαιδεύεται ένα GAN θα ακολουθήσουμε το notation που χρησιμοποιείται στο άρθρο του Goodfellow και θα θεωρήσουμε ότι τα δύο δίκτυα των Generator και Discriminator είναι δύο Multi Layer Perceptrons που αντιστοιχίζονται σε δύο παραγωγίσιμες συναρτήσεις $G(z, \theta_g)$ και $D(x, \theta_d)$. Για τη μάθηση της κατανομής του Generator p_g πάνω στα πρότυπα x , θεωρείται η ύπαρξη μιας prior $p_z(z)$ κατανομής που αφορά τις κρυφές εκείνες μεταβλητές z του latent space.

Η $D(x)$ αναπαριστά την πιθανότητα το πρότυπο x να ήρθε από τα πραγματικά δεδομένα. Ο D εκπαιδεύεται με σκοπό τη μεγιστοποίηση της πιθανότητας ο D να κατηγοριοποιήσει ορθά, τα δείγματα που δέχεται, πραγματικά η μή. Η πιθανότητα ο D να μαντέψει ορθά ότι ένα δείγμα που προέρχεται από τον G είναι μή πραγματικό είναι $D(G(z))$. Συνεπώς η πιθανότητα ο D να ξεγελιέται κάθε φορά από τον G θα είναι $1 - D(G(z))$ και αυτή θέλουμε να ελαχιστοποιήσουμε.

Με άλλα λόγια οι D και G παίζουν ένα minimax παίγνιο δύο παικτών με συνάρτηση $V(G, D)$:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (13)$$

Για τη μεγιστοποίηση της $V(D, G)$ προτείνεται η χρήση της τεχνικής ανάβασης κλίσης, ενώ για την ελαχιστοποίηση του δεύτερου όρου της παραπάνω εξίσωσης προτείνεται η τεχνική της κατάβασης κλίσης από το άρθρο του Goodfellow.

Συχνά παρατηρείται ότι η βελτιστοποίηση της συνάρτησης του Generator δεν γίνεται αποτελεσματικά, αφού κατά πάσα πιθανότητα το παραγόμενο δείγμα, τουλάχιστον κατά την αρχή της εκπαίδευσης, θα κατηγοριοποιηθεί από τον Discriminator ως μή αληθινό. Συνεπώς η πιθανότητα $1 - D(G(z))$ θα είναι πολύ μικρή και τότε η συνάρτηση πρακτικά δεν θα έχει gradients. Και εφόσον η μεγιστοποίηση του όρου γίνεται με ανάβαση κλίσης αυτή δεν θα είναι δυνατή. Για να προσπεράσουμε το πρόβλημα αυτό συχνά μεγιστοποιούμε την συμπληρωματική πιθανότητα $D(G(z))$.

$$\max_G E_{z \sim p_z(z)}[\log(D(G(z)))] \quad (14)$$

Ο αλγόριθμος εκπαίδευσης του δικτύου GAN όπως τουλάχιστον πρωτοπαρουσιάστηκε στο άρθρο του Goodfellow φαίνεται στο Σχήμα 22. Συνοπτικά παρατηρούμε ότι δειγματολειτουργούν πακέτα - batches z διανυσμάτων από τον

latent space καθώς και ίδιος αριθμός πραγματικών προτύπων x από τα διαθέσιμα δεδομένα εκπαίδευσης. Στη συνέχεια χρησιμοποιείται η μέθοδος Stochastic Gradient Ascend για να καθοριστούν οι παράμετροι του Discriminator στην αντίστοιχη συνάρτηση κόστους του. Αυτό προτείνεται να γίνει k φορές, παρόλαυτά για λόγους οικονομίας υπολογιστικού κόστους, στο συγκεκριμένο άρθρο τέθηκε το $k = 1$.

Στη συνέχεια, δειγματοληπτείται τυχαία πάλι κάποιος m αριθμός διανυσμάτων z από τον latent space. Οι παράμετροι του Generator ανανεώνονται με τη μέθοδο της Stochastic Gradient Descend στην αντίστοιχη συνάρτηση κόστους. Όλη η διαδικασία επαναλαμβάνεται N φορές έως ότου υπάρξει σύγκλιση.

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{\text{data}}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[\log D(x^{(i)}) + \log(1 - D(G(z^{(i)}))) \right].$$

end for

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)}))).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

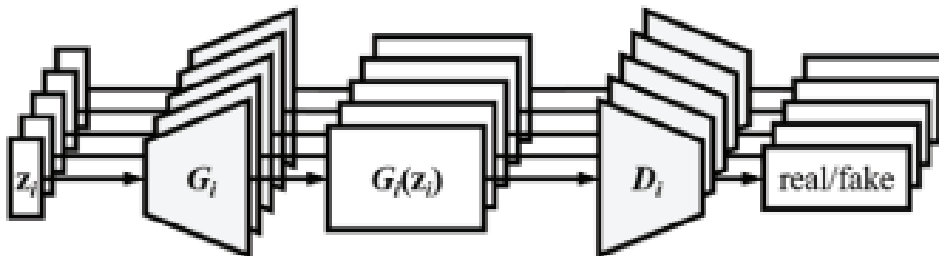
Σχήμα 22: Ο αλγόριθμος εκπαίδευσης του GAN

Το κριτήριο πάυσης της εκπαίδευσης δεν είναι σαφώς ορισμένο. Ιδανικά ένα τέτοιο δίκτυο θα φτάσει στην ισορροπία Νας, παρόλαυτά στην πράξη πολλά απλά GANs δεν θα καταφέρουν ποτέ να συγκλίνουν. Συνήθως δημιουργοί διαφόρων τέτοιων δικτύων χρησιμοποιούν μετρικές που ελέγχουν το κατά πόσο τα παραγόμενα πρότυπα απέχουν από τα πραγματικά που χρησιμοποιήθηκαν κατά την εκπαίδευση. Για παράδειγμα το WGAN, το οποίο εκπαιδεύεται για την παραγωγή εικόνων, ελαχιστοποιεί την απόσταση Wasserstein μεταξύ των κατανομών των παραγόμενων εικόνων και των πραγματικών. Η συγκεκριμένη συνάρτηση κόστους είναι ερμηνεύσιμη συνεπώς η ελαχιστοποίησή της μπορεί να χρησιμοποιηθεί ως κριτήριο πάυσης της εκπαίδευσης.

MuseGAN

Το συγκεκριμένο δίκτυο είναι εκείνο το οποίο χρησιμοποιήθηκε για την υλοποίηση των πειραμάτων αυτής της εργασίας. Αποτελεί αυτή τη στιγμή το state of the art δίκτυο τύπου GAN για δημιουργία μουσικής χρησιμοποιώντας συμβολική αναπαράσταση. Υλοποιήθηκε από τους Hao-Wen Dong, et al. [11,12,13] και πάνω στη συγκεκριμένη υλοποίηση δημοσιεύτηκαν τρία άρθρα το 2017 και 2018.

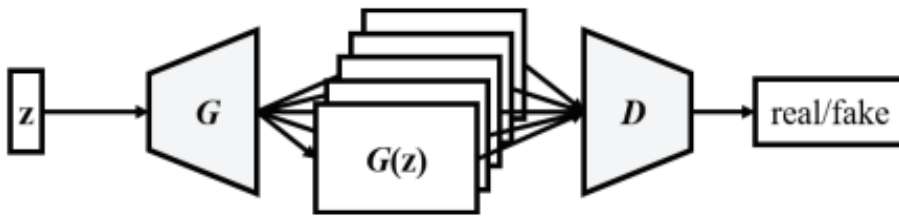
Οι συγγραφείς, προσπαθώντας να προσεγγίσουν τους διαφορετικούς τρόπους που μια σύνθεση δημιουργείται, κατέληξαν σε δύο βασικά μοντέλα σύνθεσης. Το Jamming Model προσπαθεί να μιμηθεί το κόνσεπτ σύνθεσης όπου κάθε μουσικός συνθέτει τη μουσική για το όργανο το οποίο παίζει ανεξάρτητα από τους άλλους. Η ένωση των επί μέρους συνθέσεων στο χρόνο, αποτελεί και την συνολική σύνθεση. Το Jamming Model καταφέρνει να υλοποιήσει αυτήν την ιδέα χρησιμοποιώντας ξεχωριστές εισόδους z_i από τον latent space για κάθε διαφορετικό όργανο, κάθε μία από τις οποίες περνάει στη συνέχεια στον αντίστοιχο G_i Generator. Η σύνθεση για κάθε όργανο στο Jamming Model θα ελέγχεται από τον αντίστοιχο Discriminator D_i ξεχωριστά όπως φαίνεται στο σχήμα 23.



Σχήμα 23: Η δομή του Jamming Model

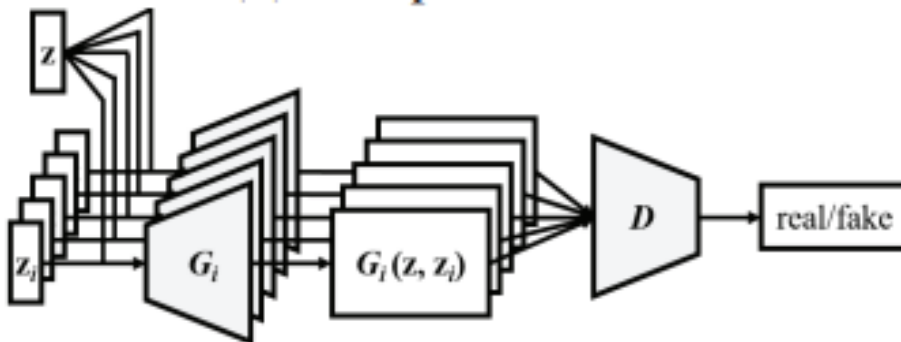
Ένας άλλος τρόπος σύνθεσης, κατά τους δημιουργούς του Musegan, είναι ο εξής. Όλες οι μελωδικές γραμμές, όλων των οργάνων γράφονται από έναν μοναδικό συνθέτη, ο οποίος έχει συνολική γνώση της αρμονικής δομής της παραγόμενης μουσικής καθώς και της ενορχήστρωσης. Για αυτό το σκοπό προτείνουν το αποκαλούμενο Composer Model. Στο μοντέλο αυτό υπάρχει μία μοναδική είσοδος z από τον latent space, η οποία όπως αναφέρουν οι συγγραφείς αντικατοπτρίζει την αρχική πρόθεση του συνθέτη. Η είσοδος αυτή θα περάσει σε έναν μοναδικό Generator ο οποίος θα την χρησιμοποιήσει πολλές φορές αλλά δρώντας κάθε φορά διαφορετικά, έτσι ώστε να συνθέσει τη μουσική

για κάθε γραμμή οργάνου. Στη συνέχεια η παραγόμενη μουσική σαν σύνολο από όλα τα όργανα αξιολογείται από έναν μοναδικό Discriminator ο οποίος αποφασίζει αν το δείγμα που δέχεται είναι πραγματικό ή τεχνητό. Με τη χρήση του Composer Model επιτυγχάνεται η περισσότερη συνοχή και καλύτερη αρμονική δομή ανάμεσα στις μελωδικές γραμμές του κάθε οργάνου. Η δομή του μοντέλου φαίνεται στο Σχήμα 24.



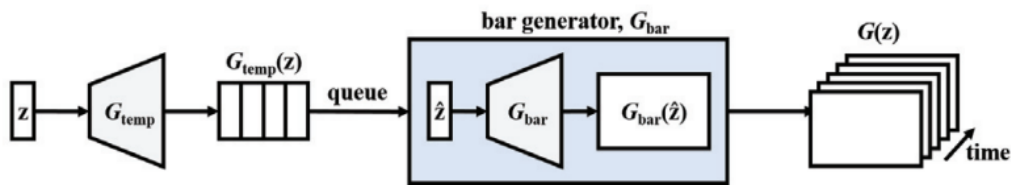
Σχήμα 24: Εκδοχή Composer Model του MuseGAN

Τέλος οι συγγραφείς προτείνουν το συνδιασμό των δύο αυτών βασικών ιδεών στην παραγωγή μουσικής, δημιουργώντας το Hybrid Model η δομή του οποίου παρουσιάζεται στο Σχήμα 25. Η βασική διαφορά μεταξύ του Hybrid Model και του Composer Model έγκειται στην ευελιξία όσον αφορά την αρχιτεκτονική. Στο Hybrid Model μπορούμε να έχουμε πολλές διαφορετικές αρχιτεκτονικές (πχ αριθμό νευρώνων, διαφορετικά φίλτρα) καθώς και διαφορετικές εισόδους z_i στον Generator καταφέρνοντας έτσι για παράδειγμα να παράγουμε το ίδιο μουσικό τραγούδι αλλάζοντας μέρος μόνο ενός οργάνου.



Σχήμα 25: Το Hybrid Model, ένας συνδιασμός των προηγούμενων δύο εκδοχών

Τα παραπάνω προτεινόμενα μοντέλα είναι ικανά να παράγουν μέτρα μουσικής τα οποία όμως είναι ανεξάρτητα μεταξύ τους. Για να καταφέρουν η παραγόμενη μουσική να έχει συνοχή από μέτρο σε μέτρο, εισάγουν ένα προσωρινό μοντέλο G_{temp} και το μοντέλο που αντιστοιχεί στην παραγωγή μουσικής των μέτρων G_{bar} , θεωρώντας την χρονική ακολουθία των μέτρων σαν ξεχωριστή διάσταση όπως φαίνεται και στο Σχήμα 26. Παράλληλα προτάθηκε η δημιουργία του Track Conditional Model το οποίο δέχεται από το χρήστη μία πραγματική μελωδική γραμμή και ενορχηστρώνει όλα τα υπόλοιπα όργανα γύρω από αυτή. Η συγκεκριμένη εκδοχή δεν χρησιμοποιήθηκε, συνεπώς δεν θα αναλυθεί περαιτέρω.



Σχήμα 26: Δομή Generator για Generation from scratch

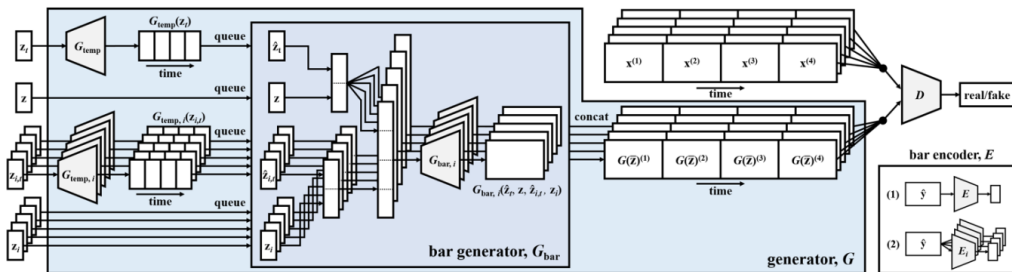
Συγκεντρώνοντας λοιπόν όλες αυτές τις ιδέες που εξηγούνται σε αυτό το κομμάτι, οι δημιουργοί προτείνουν το MuseGAN. Το δίκτυο αυτό όπως βλέπουμε στο Σχήμα 27 έχει εξαιρετικά περίπλοκη δομή, με μια πιο προσεκτική ματιά παρόλαυτά, παρατηρούμε όλα τα σημαντικά επί μέρους κομμάτια τα οποία αναλύθηκαν παραπάνω. Συνεπώς συγκεκριμένα κομμάτια του δικτύου χρησιμοποιούνται, ανάλογα με την εφαρμογή που επιθυμούμε να υλοποιήσουμε.

Σε αυτή τη δουλειά χρησιμοποιήσαμε την εκδοχή του Composer Model, δίνοντας στο δίκτυο ένα διάνυσμα του latent space ως είσοδο, για το task του Generation from scratch με τον Generator να είναι ήδη εκπαιδευμένος. Γιαυτό το λόγο τα τεχνικά χαρακτηριστικά της εκπαίδευσης δεν θα συμπεριληφθούν. Παρόλαυτά θεωρούμε σημαντικό να αναφέρουμε τη φύση των δεδομένων εκπαίδευσης, μιας και όπως έχει γίνει ήδη αντιληπτό, η ποιότητα των παραγόμενων μουσικών προτύπων έχει άμεση εξάρτηση από τα δεδομένα αυτά.

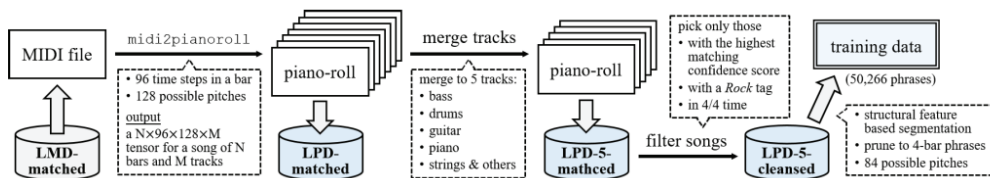
Το αρχικό Dataset που χρησιμοποιήθηκε ήταν το Lakh MIDI dataset (LMD) (Raffel 2016), μία μεγάλη συλλογή από 176.581 MIDI αρχεία. Οι συγγραφείς κατάφεραν να μετατρέψουν με τη χρήση της βιβλιοθήκης pretty_midi όλα αυτά τα αρχεία σε Multitrack Pianoroll πίνακες. Κάθε pianoroll πίνακας έχει 128 γραμμές που αντιστοιχούν στις 128 διαφορετικές MIDI νότες και η χρονική ανάλυση τέθηκε να είναι 96 έτσι ώστε να υπάρχουν τα χρονικά χαρακτηριστικά των τριήχων ή των 16ών. Κάποια τραγούδια εμφανίζουν πολύ λίγες

νότες σε συγκεκριμένα όργανα, κάνοντας τους πίνακες τους εξαιρετικά αραιούς. Το πρόβλημα αυτό αντιμετωπίστηκε ενώνοντας τους πίνακες των κοινών οργάνων με απλή άθροιση, και έτσι δεν θα υπάρχουν τραγούδια, που θα έχουν κενά μέτρα.

Το είδος των τραγουδιών εκπαίδευσης, άνηκε στη Ροκ μουσική ενώ το μέτρο των τραγουδιών ήταν για όλα 4/4. Το τελικό παραγόμενο μουσικό πρότυπο είναι ένας γενικευμένος πίνακας της μορφής 4(bars) X 96(time resolution) X 84(possible notes) X 5(tracks). Μιά σχηματική απεικόνιση των σταδίων προετοιμασίας των δεδομένων που χρησιμοποιήθηκαν παρουσιάζεται στο Σχήμα 28.



Σχήμα 27: Η περίπλοκη δομή του MuseGAN



Σχήμα 28: Η προετοιμασία των δεδομένων του MuseGAN

3 Μεθοδολογία και Ορισμοί

Στα προηγούμενο κεφάλαιο είδαμε ότι τα δίκτυα GAN έχουν τη δυνατότητα να παράγουν μουσικά πρότυπα, χρησιμοποιώντας ως είσοδο τυχαία διανύσματα του latent space. Η σχέση του διανύσματος εισόδου z και του μουσικού προτύπου x στην έξοδο είναι ένα προς ένα, με άλλα λόγια δεδομένου του εκπαιδευμένου Generator το μουσικό πρότυπο που θα παραχτεί εξαρτάται μόνο από το διάνυσμα εισόδου z . Στο παρόν κεφάλαιο θα ορίσουμε τον χώρο εισόδου των GANs και θα εξηγήσουμε τις δυσκολίες που υπάρχουν στην παραγωγή μουσικής συγκεκριμένων χαρακτηριστικών. Επίσης θα εξηγήσουμε αναλυτικά τεχνικές που θα χρησιμοποιηθούν αργότερα στην πειραματική διαδικασία αυτής της δουλειάς.

3.1 Έλεγχος Παραγωγής-Ένα Δύσκολο Πρόβλημα

Όπως λοιπόν καταλαβαίνουμε η δομή του χώρου εισόδου και το διάνυσμα που δειγματοληπτείται από αυτόν, είναι το μοναδικό στοιχείο που καθορίζει τη μορφή του παραγόμενου προτύπου, δεδομένου του δικτύου. Παρόλαυτά όλα τα διανύσματα εισόδου, ζουν σε έναν χώρο 128 διαστάσεων και δεν υπάρχει σαφής και άμεσος τρόπος να μπορούμε να γνωρίζουμε τα χαρακτηριστικά του παραγόμενου προτύπου, μόνο από το αντίστοιχο διάνυσμα εισόδου του, πριν δούμε το αποτέλεσμα.

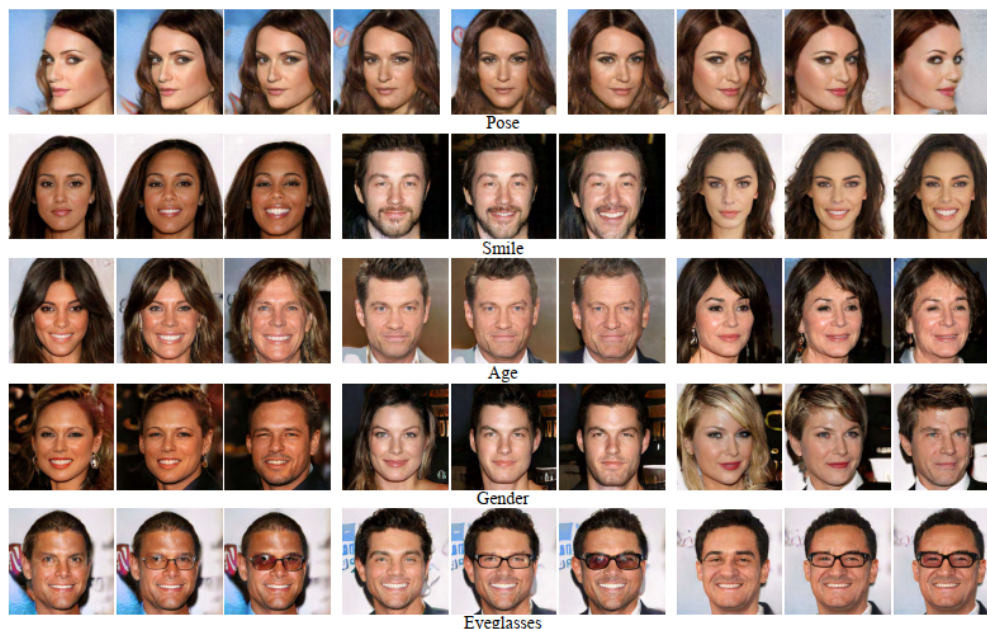
Στην ουσία, το MuseGAN, και όλα τα Generative δίκτυα τύπου GAN είναι τεχνολογίες που από τη μία, μπορούν να παράγουν πειστικά πρότυπα, αλλά έρχονται χωρίς οδηγίες χρήσης! Για παράδειγμα, δεν υπάρχει κανένας εύκολος και άμεσος τρόπος, να μπορεί κανείς με τη χρήση του MuseGAN να παράγει δύο τραγούδια ακριβώς ίδια, που όμως να διαφέρουν κατά επιθυμητό τρόπο σε επιθυμητό αριθμό νοτών. Δεν υπάρχει επίσης κανένας εύκολος και άμεσος τρόπος να παράγει κανείς ίδια μουσική σε άλλη τονικότητα. Τέλος δεν υπάρχει κανένας άμεσος και εύκολος τρόπος, να παράγει κανείς μουσική συγκεκριμένης αρμονίας και μελωδίας.

Το συγκεκριμένο πρόβλημα, επιχειρούμε να προσεγγίσουμε και εμείς στη συγκεκριμένη εργασία. Σκοπός μας είναι η εξερεύνηση του 128 διαστάσεων χώρου εισόδου του MuseGAN με σκοπό την εύρεση κατευθύνσεων η περιοχών του που τα παραγόμενα μουσικά τραγούδια θα αλλάζουν, με διαισθητικά κατανοητό τρόπο.

Όπως τονίζεται στο άρθρο του InterfaceGan[28], έχει παρατηρηθεί ότι οποιαδήποτε γραμμική παρεμβολή μεταξύ δύο διανυσμάτων του latent space z_1 και z_2 , οδηγεί σε παραγόμενα πρότυπα που τα βασικά ποιοτικά χαρακτηριστικά

τους αλλάζουν σταδιακά και ομαλά. Τα βασικά αυτά χαρακτηριστικά αναφέρονται στο συγκεκριμένο άρθρο με τον όρο semantics και στο Σχήμα 29 παρατηρείται αυτή η ομαλή και σταδιακή αλλαγή. Κοινά semantic χαρακτηριστικά, στα δίκτυα παραγωγής εικόνων είναι ο προσανατολισμός της εικόνας, το contrast και η μεγέθυνση. Παρατηρούμε στο Σχήμα 29 ότι τέτοια χαρακτηριστικά μπορεί να είναι και αρκετά συγκεκριμένα όπως γνωρίσματα του περιεχομένου των εικόνων. Πχ βλέπουμε ότι μπορεί να είναι η πόζα, το χαμόγελο του προσώπου ή η ηλικία και το φύλο.

Καταλαβαίνουμε λοιπόν από τα παραπάνω ότι είναι απαραίτητο να ορίσουμε αυστηρά τον latent space των δικτύων αυτών, να εξετάσουμε τις βασικές ιδιότητες καθώς και να παρουσιάσουμε τρόπους ελέγχου του παραγόμενου μουσικού προτύπου.



Σχήμα 29: Στην εικόνα παρουσιάζεται μια γραμμική παρεμβολή ανάμεσα σε δύο διανύσματα εισόδου που οδηγούν σε παραγόμενα πρότυπα που διαφέρουν σε ένα semantic χαρακτηριστικό σε κάθε γραμμή. Παρατηρούμε ότι οι εικόνες αλλάζουν ομαλά και με συνεχή τρόπο InterfaceGAN [28]

3.2 Χαρακτηριστικά του Latent Space των GANs

Για τον ορισμό του χώρου εισόδου των GANs θα χρησιμοποιήσουμε την βάση που θέτει το άρθρο του InterfaceGAN[28]. Σύμφωνα με το άρθρο, ένας

εκπαιδευμένος Generator ενός GAN, αναπαριστά μία ντετερμινιστική συνάρτηση G :

$$G : Z \rightarrow X, \quad Z \subseteq \mathbb{R}^d \quad (15)$$

Το σύνολο $Z \subseteq \mathbb{R}^d$ αποτελεί τον latent space και συνήθως χρησιμοποιείται η πολυδιάστατη κανονική κατανομή $N(0, I_d)$. Όπως αναφέραμε, στο MuseGAN το οποίο χρησιμοποιήσαμε στα πειράματά μας, ο latent space είναι 128-διάστατος, δηλαδή $d = 128$. Επίσης το MuseGAN δειγματοληπτεί από την truncated 128 διάστατη κανονική κατανομή $N_{128}(-2, 2)$. Δηλαδή, με άλλα λόγια το κάθε διάνυσμα z θα είναι μία συλλογή από 128 αριθμούς, οι οποίοι, ο καθένας τους θα βρίσκονται στο διάστημα $[-2, 2]$. Το σύνολο X αναπαριστά το σύνολο των παραγόμενων προτύπων, που στην δική μας περίπτωση είναι οι μουσικές συνθέσεις που παράγει ο Generator σε μορφή 5-track pianorolls.

3.2.1 Βασική Άλγεβρα Διανυσμάτων του Latent Space

Υποθέτοντας συνέχεια και ομαλότητα στον χώρο εισόδου, η οποία συνήθως προκύπτει από έναν πολύ καλά εκπαιδευμένο Generator, μπορούμε να εκτελέσουμε βασική άλγεβρα μεταξύ δύο η περισσοτέρων διανυσμάτων εισόδου. Τα καινούρια διανύσματα που θα προκύψουν θα θέλαμε να φέρουν και κάποια αναμενόμενη αλλαγή στα semantics των προτύπων που θα προκύψουν [5,20,28]. Αντίθετα αν ο χώρος εισόδου δεν είναι ομαλός, τότε τα αποτελέσματα θα είναι μη προβλέψιμα και συνεπώς είναι πολύ δύσκολο κανείς να ελέγξει το αποτέλεσμα της παραγωγής.

Μία κοινή πρακτική είναι ο μέσος όρος. Άς υποθέσουμε ότι επιλέγουμε k διανύσματα από τον latent space. Τα διανύσματα αυτά θα χρησιμοποιηθούν το κάθε ένα από τον Generator του δικτύου για να παράχθούν k πρότυπα. Έστω ότι εξετάζουμε τα πρότυπα, όλα για ένα semantic, πχ το πόσο πολύ αποκλίνει το εικονιζόμενο πρόσωπο από την ευθεία (πόσο στρίβει το κεφάλι - πόζα) της κάμερας. Έστω ότι ποσοτικοποιούμε αυτό χαρακτηριστικό, πχ σε μοίρες. Τα k λοιπόν πρότυπα - εν προκειμένω εικονιζόμενα πρόσωπα, θα έχουν και k τιμές που θα αντικατοπτρίζουν το επιλεγμένο semantic s_k

Το μέσο διάνυσμα εισόδου θα δίνεται από τη σχέση:

$$z_{avg} = \frac{1}{k} \sum_{i=1}^k z_i \quad (16)$$

Ενώ αντίστοιχα η μέση τιμή του semantic:

$$s_{avg} = \frac{1}{k} \sum_{i=1}^k s_i \quad (17)$$

Παίρνοντας τον μέσο όρο ενός μεγάλου αριθμού διανυσμάτων εισόδου με κοντινές τιμές σε κάποιο semantic, μπορεί να οδηγηθούμε στην εύρεση κεντροειδών στον latent space, που θεωρητικά τουλάχιστον, μπορούν να παράγουν τα αντίστοιχα κεντροειδή στον semantic space. Αυτό μπορεί να επιτευχθεί επιλέγοντας ένα υποσύνολο του Z , Z_x όπου:

$$f(z_x) = \beta \pm \delta, \forall z_x \in Z_x \quad (18)$$

Το β είναι η επιθυμητή τιμή του semantic s και το δ είναι μία αποδεκτή τιμή διασποράς. Η τεχνική του μέσου όρου χρησιμοποιείται στον χώρο Z_x για την εύρεση κεντροειδούς. Τα διανύσματα που βρίσκονται κοντά σε αυτό το κεντροειδές, είναι αρκετά πιθανό να παράγουν πρότυπο που θα έχει την επιθυμητή τιμή β με μικρή διασπορά για το χαρακτηριστικό s . Η τεχνική αυτή χρησιμοποιήθηκε στο MusicVAE, για την εύρεση κεντροειδών μεγάλης και μικρής τιμής σε 5 διαφορετικά χαρακτηριστικά [27].

Μία άλλη αλγεβρική τεχνική που χρησιμοποιείται για τον έλεγχο της παραγωγής, είναι η γραμμική παρεμβολή, η οποία παρουσιάστηκε και παραπάνω και η μέθοδος της βασίζεται σε απλή αφαίρεση διανυσμάτων.

Η γραμμική παρεμβολή χρησιμοποιείται με τον εξής τρόπο. Έστω ότι έχουμε δυο διανύσματα εισόδου με το πρώτο να έχει την επιθυμητή τιμή σε ένα εξεταζόμενο χαρακτηριστικό και το δεύτερο να είναι τυχαίο. Αφαιρώντας τα δύο διανύσματα, πετυχαίνουμε καλύτερη ποιότητα προτύπου εξόδου για το τυχαίο διάνυσμα. Το μόνο σίγουρο είναι ότι για να λειτουργήσει αυτή η μέθοδος, απαιτείται συνέχεια και ομαλότητα του εξεταζόμενου χαρακτηριστικού στον χώρο εισόδου.

Ένας άλλος τρόπος να γίνει αντιληπτή η έννοια του interpolation είναι ο εξής. Έστω ότι έχουμε δύο διανύσματα εισόδου, z_1 και z_2 που τα πρότυπα που παράγουν διαφέρουν σημαντικά σε ένα χαρακτηριστικό. Το διάνυσμα που προκύπτει από την αφαίρεση των z_1 και z_2 θεωρούμε ότι εν γένει μπορεί να κωδικοποιεί το χαρακτηριστικό αυτό, $V_{feature} = z_2 - z_1$. Η πιο σημαντική ταχτική είναι η εύρεση κεντροειδών υψηλής τιμής σε ένα χαρακτηριστικό, και η εύρεση κεντροειδών αντίθετης, ή χαμηλής τιμής στο ίδιο χαρακτηριστικό, με τη χρήση της τεχνικής του μέσου όρου.

Οι παραπάνω τεχνικές έχουν σημαντικούς περιορισμούς. Προβλήματα κυρίως προκύπτουν λόγω της μη γραμμικής δομής του latent space των νευρωνικών δικτύων. Οι κατευθύνσεις αλλαγής ενός χαρακτηριστικού σε αυτήν την περίπτωση, δεν θα είναι οι ίδιες για όλα τα διανύσματα εισόδου. Επίσης το εύρος αυτών των κατευθύνσεων δεν θα είναι απεριόριστο. Για παράδειγμα ας σκεφτούμε ότι με κάποιο τρόπο έχει προσδιοριστεί μία κατεύθυνση στον χώρο εισόδου που κινούμενοι πάνω σε αυτή και ξεκινώντας από ένα συγκεκριμένο διάνυσμα z οδηγούμαστε σε αύξηση της τονικότητας των παραγόμενων προ-

τύπων. Όταν ο χώρος εισόδου είναι μη γραμμικός και μη ομαλός, τότε είναι πολύ πιθανό, ξεκινώντας από κάποιο άλλο διάνυσμα και κινούμενοι όμως προς την ίδια κατεύθυνση να μην μπορούμε να επιτύχουμε την ίδια αλλαγή στα παραγόμενα πρότυπα. Επίσης είναι πολύ πιθανό, η συγκεκριμένη κατεύθυνση να επιφέρει την επιθυμητή αλλαγή για δεδομένο αριθμό βημάτων στην καθορισμένη κατεύθυνση ή ακόμα και σε ασυνεχή τμήματα της τροχιάς αυτής.

3.3 Υπερεπίπεδα Διαχωρισμού Ταξινόμησης

Μία από τις πιο δημοφιλείς και αναγκαίες εφαρμογές της Μηχανικής Μάθησης και των Νευρωνικών Δικτύων είναι η ταξινόμηση προτύπων σε κλάσεις. Πληθώρα αλγορίθμων έχουν αναπτυχθεί για το σκοπό αυτό. Ο Nearest Neighbors αλγόριθμος για παράδειγμα, ταξινομεί ένα πρότυπο σε μία κλάση, σύμφωνα με την κλάση που ανήκουν οι κοντινότεροι (ελάχιστη Ευκλείδεια απόσταση) σε αυτό γείτονες του, χρησιμοποιώντας τον κανόνα της πλειοψηφίας. Άλλοι αλγόριθμοι όπως ο ταξινομητής του Bayes θεωρεί την ύπαρξη κατανομής $p(\theta)$ για κάθε κλάση και ταξινομεί το πρότυπο x στην κλάση που η πιθανότητα $p(x|\theta)p(\theta)$ γίνεται μέγιστη. Οι συγκεκριμένοι αλγόριθμοι διαχωρισμού (καθώς και αρκετοί άλλοι) δεν παρέχουν μία άμεση γεωμετρική διαίσθηση στο πώς πραγματικά δύο κλάσεις διαχωρίζονται στον χώρο που ανήκουν τα δεδομένα τους. Τελικά θα οριστούν οι περιοχές απόφασης, και για τους γραμμικούς ταξινομητές (Bayes Classifier, Logistic Regression κλπ.) θα είναι πράγματι γραμμικές, παρόλαυτά ο τρόπος που η απόφαση λαμβάνεται δεν είναι άμεσα γεωμετρικός.

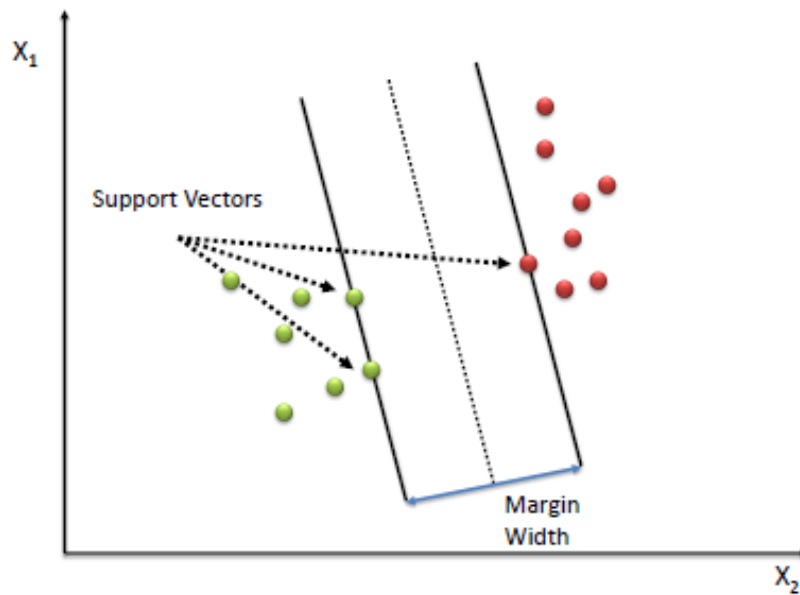
Είδαμε παραπάνω ότι η χρήση της άλγεβρας και της αναλυτικής γεωμετρίας στα διανύσματα του Latent Space των GANS, ανοίγει το δρόμο στην ικανότητα ελέγχου του παραγόμενου προτύπου. Συνεπώς αν για αυτό το σκοπό θα υπήρχε η ανάγκη της ταξινόμησης των παραγόμενων προτύπων σε κλάσεις (πχ ανάλογα με τα semantics των προτύπων), τότε ο αλγόριθμος ταξινόμησης θα θέλαμε να βασίζεται σε έναν γεωμετρικό τρόπο λειτουργίας.

3.3.1 Support Vector Machines Classifier

Ο ταξινομητής Support Vector Machines ή εν συντομία SVM διαχωρίζει δύο κλάσεις βρίσκοντας το υπερεπίπεδο εκείνο που θα μεγιστοποιήσει το margin μεταξύ των κλάσεων, όπως φαίνεται στο Σχήμα 30. Στο συγκεκριμένο σημείο θα παρουσιάσουμε τον τρόπο λειτουργίας του γραμμικού SVM ταξινομητή.

Το υπερεπίπεδο διαχωρισμού των δύο κλάσεων μπορεί να εκφραστεί ως :

$$WX + b = 0 \tag{19}$$



Σχήμα 30: Ο γραμμικός SVM ταξινομητής βρίσκει με τη χρησιμοποίηση των Support Vectors το υπερεπίπεδο διαχωρισμού που μεγιστοποιεί το margin μεταξύ των δύο κλάσεων[28]

Όπου $W = \{w_1, w_2, \dots, w_n\}$ είναι το διάνυσμα των βαρών, X τα διανύσματα εκπαίδευσης και b ο σταθερός όρος bias. Το n είναι η διάσταση του χώρου, όπου στα σχήματα 30 και 31 υπονοείται ότι $n = 2$, πράγμα που θα υιοθετήσουμε και εμείς για απλότητα. Θεωρώντας τον όρο b ως ένα πρόσθετο βάρος w_0 η εξίσωση γράφεται:

$$w_0 + w_1x_1 + w_2x_2 = 0 \quad (20)$$

Συνεπώς, για όλα τα σημεία στα δεδομένα μας που βρίσκονται δεξιά του υπερεπιπέδου διαχωρισμού θα πρέπει :

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (21)$$

ενώ για όλα τα σημεία στα δεδομένα μας που βρίσκονται αριστερά του υπερεπιπέδου διαχωρισμού θα πρέπει:

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (22)$$

Το επίπεδο διαχωρισμού λοιπόν θα ορίσει δύο πλευρές. Θα λέμε ότι ένα data-point θα ανήκει στην πλευρά H_1 και στην πλευρά H_2 αντίστοιχα, όταν

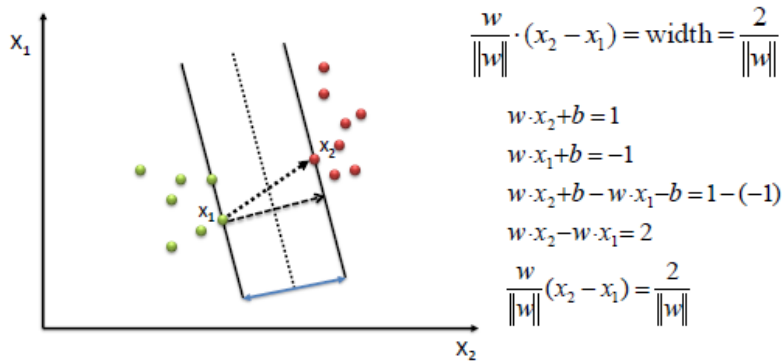
$$H_1 : w_0 + w_1x_1 + w_2x_2 \geq 1, y_i = +1 \quad (23)$$

$$H_2 : w_0 + w_1x_1 + w_2x_2 \leq 1, y_i = -1 \quad (24)$$

Στις παραπάνω εξισώσεις η τιμή του y_i κωδικοποιεί την κλάση που ανήκει το κάθε πρότυπο. Συνδιάζοντας τις δύο αυτές ανισότητες έχουμε

$$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1 \quad (25)$$

Τα σημεία που ικανοποιούν την ισότητα της παραπάνω σχέσης ονομάζονται Support Vectors και αντιπροσωπεύουν τα κοντινότερα σημεία από κάθε κλάση στην διαχωριστική επιφάνεια. Αυτά τα σημεία χρησιμοποιεί ο αλγόριθμος για την εύρεση του βέλτιστου υπερεπιπέδου διαχωρισμού. Η απόσταση των σημείων αυτών από το υπερεπίπεδο θα είναι $\frac{1}{|W|}$. Συνεπώς το συνολικό εύρος του margin θα είναι $\frac{2}{|W|}$, όπως φαίνεται και στο Σχήμα 31.



Σχήμα 31: Υπολογισμός μέγιστου margin για τον καθορισμό του βέλτιστου υπερεπιπέδου στα SVMs

Η εύρεση των κατάλληλων παραμέτρων έγκειται στην μεγιστοποίηση του όρου $\frac{1}{|W|}$ και είναι ένα πρόβλημα τετραγωνικής βελτιστοποίησης, η λύση του οποίου δεν θα παρουσιαστεί στο παρόν κείμενο.

Όπως βλέπουμε η ερμηνεία της λειτουργίας των SVMs είναι γεωμετρική, αφού όταν ο αλγόριθμος εκπαίδευσης συγκλίνει, το βέλτιστο υπερεπίπεδο θα διαχωρίζει τις κλάσεις στον χώρο εισόδου. Σε γραμμικά διαχωρίσιμα προβλήματα, επιτυγχάνεται σύγκλιση σε ολικό ελάχιστο. Οι συντεταγμένες του υπερεπιπέδου θα προκύπτουν από τις τιμές του διανύσματος W και του αριθμού b .

3.3.2 Η Τεχνική Latent Walking

Η δυνατότητα εύρεσης του υπερεπιπέδου διαχωρισμού δύο κλάσεων με τη χρήση των Support Vector Machines, μπορεί να βρεί εφαρμογή στην εύρεση επιθυμητών κατευθύνσεων στον latent space των GANs. Ας σκεφτούμε ότι έχουμε στη διαθεσή μας δύο σύνολα διανυσμάτων $Z_a = \{z_{a1}, z_{a2}, \dots, z_{am}\}$ και $Z_b = \{z_{b1}, z_{b2}, \dots, z_{bm}\}$ όπου $Z_a \subseteq Z$ και $Z_b \subseteq Z$. Έστω ότι όλα τα διανύσματα του Z_a οδηγούν παραγόμενα πρότυπα με όμοια τιμή σε ένα συγκεκριμένο semantic και ομοίως τα διανύσματα του Z_b οδηγούν σε παραγόμενα πρότυπα με διαφορετική τιμή στο ίδιο semantic. Με τη χρήση των SVMs, μας παρέχεται η δυνατότητα διαχωρισμού του χώρου εισόδου Z σε δύο περιοχές, ως προς το εξεταζόμενο semantic. Η απόσταση ενός διανύσματος z από την διαχωριστική υπερεπιφάνεια θα δίνεται από τη σχέση :

$$d = zW^T + b \quad (26)$$

Η απόσταση αυτή θα είναι θετική στην περίπτωση που το z βρίσκεται στην πλευρά που αναπαριστά την κλάση $y = 1$ ενώ αρνητική αντίθετα.

Κάνοντας μικρά βήματα από κάποιο z_a διάνυσμα στην κατεύθυνση κάθετα στο υπερεπίπεδο, και δεδομένου ότι τα SVMs έχουν πετύχει καλή σύγκλιση και ο χώρος είναι γραμμικά διαχωρίσιμος ως προς την τιμή του semantic, παίρνουμε διανύσματα z_{new} τα οποία δημιουργούν πρότυπα για τα οποία η τιμή του semantic αλλάζει ολοένα και περισσότερο προσεγγίζοντας την κλάση b . Με άλλα λόγια κινούμαστε κάθετα στο διαχωριστικό υπερεπίπεδο, μέχρι τη στιγμή που το τελικό διάνυσμα θα έχει περάσει από την άλλη πλευρά. Με μια τέτοια κίνηση, στην ουσία καταφέρνουμε να βρούμε μια κατεύθυνση του λανθάνοντος χώρου, στην οποία για τα παραγόμενα πρότυπα, αλλάζει η τιμή του εξεταζόμενου χαρακτηριστικού με τρόπο συνεχόμενο και ομαλό.

Η τεχνική αυτή ονομάζεται latent walking και εν γένει το καινούριο διάνυσμα που θα προκύπτει μετά από κάθε βήμα θα είναι δίνεται από τη σχέση :

$$z_{new} = z + \beta \frac{\lambda}{|\lambda|} \quad (27)$$

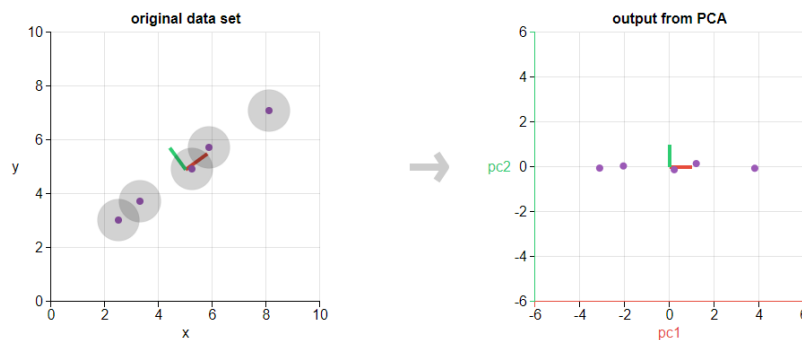
Όπου z το διάνυσμα εκίνησης, β το μέγεθος του βήματος και λ η το διάνυσμα κάθετο στο υπερεπίπεδο. Τη συγκεκριμένη τεχνική χρησιμοποιούμε και αργότερα στην πειραματική διαδικασία της άσκησης αυτής για την εύρεση σημαντικών κατευθύνσεων στο χώρο εισόδου του MuseGAN.

3.4 Μείωση Διαστασιμότητας - PCA

Η μεγάλη διαστασιμότητα του πίνακα των δεδομένων είναι ένα από τα μεγαλύτερα προβλήματα στο χώρο της μηχανικής μάθησης. Σχετικά απλοί υπολογιστικά αλγόριθμοι δυσκολεύονται να είναι χρονικά αποδοτικοί σε datasets που χρησιμοποιούν μεγάλο αριθμό επεξηγηματικών μεταβλητών. Τα νευρωνικά δίκτυα, λόγω της περίπλοκης δομής τους καλούνται να μάθουν εκατομμύρια παραμέτρων όταν εκπαιδεύονται σε δεδομένα μεγάλης διαστασιμότητας απαιτώντας μεγάλο υπολογιστικό χρόνο και κόστος.

Από την άλλη πλευρά η χρήση πολλών επεξηγηματικών μεταβλητών, παρέχει στο μοντέλο σημαντική πληροφορία που χρειάζεται να λάβει υπόψιν του κατά την εκπαίδευση για να κάνει μία σωστή πρόβλεψη.

Η Principal Component Analysis (PCA) είναι μια τεχνική του χώρου της στατιστικής, που χρησιμοποιείται κυρίως για τη μείωση της διαστασιμότητας, σε προβλήματα που εμπεριέχουν μεγάλο αριθμό επεξηγηματικών μεταβλητών. Σκοπός της είναι η δημιουργία ενός dataset που περιέχει λιγότερες μεταβλητές, διαφορετικές από τις πρώτες, οι οποίες όμως περιέχουν το μεγαλύτερο ποσό πληροφορίας του αρχικού dataset. Για να το πετύχει αυτό, μετασχηματίζει τα δεδομένα ορίζοντας σαν μεταβλητές, τις κατευθύνσεις εκείνες στις οποίες προκύπτει η μέγιστη διασπορά. Μία σχηματική απεικόνιση αυτής της ιδέας παρουσιάζεται στο σχήμα 32.



Σχήμα 32: Μία σχηματική απεικόνιση του μετασχηματισμού της PCA σε δεδομένα δύο διαστάσεων. Στην εικόνα δεν παρουσιάζεται μείωση της διαστασιμότητας, παράλαυτα μπορούμε να δούμε ότι οι καινούριες μεταβλητές (άξονες) βρίσκονται στις κατευθύνσεις μέγιστης διασποράς των δεδομένων

Συνοψίζοντας, η ιδέα της PCA είναι απλή, μειώνει τον αριθμό των επεξηγηματικών μεταβλητών ενός dataset ταυτόχρονα διατηρώντας όσο το δυνατόν

περισσότερη πληροφορία. Η PCA αποτελείται από μία σειρά απλών βημάτων τα οποία παρουσιάζονται στη συνέχεια.

Standardization

Το πρώτο βήμα είναι η τυποποίηση του εύρους των τιμών σε κάθε μεταβλητή. Συχνά, διαφορετικές μεταβλητές έχουν και διαφορετικό εύρος τιμών. Για παράδειγμα μία μεταβλητή μπορεί να περιέχει παρατηρήσεις εύρους 0-100, και μία άλλη παρατηρήσεις εύρους 0-1. Η τεχνική, όπως αναφέραμε, θα βρεί τις σημαντικές κατευθύνσεις στο dataset που μεγιστοποιούν τη διασπορά. Ο υπολογισμός της διασποράς επηρεάζεται αρκετά από την κλίμακα των δεδομένων, συνεπώς όλες οι μεταβλητές θα είναι καλό να έχουν τα ίδια εύρη τιμών, για παράδειγμα [0,1] ή [-1,1].

Για την τυποποίηση πολύ απλά χρησιμοποιούμε την εξής σχέση:

$$z = \frac{value - mean}{std} \quad (28)$$

Υπολογισμός Πίνακα Συνδιασποράς

Έχοντας τις μεταβλητές μας στην ίδια κλίμακα, μπορούμε να υπολογίσουμε τον πίνακα συνδιασποράς με σκοπό να καταλάβουμε τον τρόπο που οι μεταβλητές, διαφέρουν από το μέσο, δεδομένων των υπολοίπων. Με άλλα λόγια θέλουμε να καθορίσουμε τη συσχέτιση των μεταβλητών. Ένα παράδειγμα ενός πίνακα συνδιασποράς τριών μεταβλητών x, y, z παρατίθεται στο Σχήμα 33.

$$\begin{bmatrix} Cov(x, x) & Cov(x, y) & Cov(x, z) \\ Cov(y, x) & Cov(y, y) & Cov(y, z) \\ Cov(z, x) & Cov(z, y) & Cov(z, z) \end{bmatrix}$$

Σχήμα 33: Ο πίνακας συνδιασποράς τριών μεταβλητών x, y, z .

Αν το $Cov(y, x)$ είναι θετικό, αυτό σημαίνει ότι οι μεταβλητές y, x είναι θετικά συσχετισμένες (όταν αυξάνεται η μία, τότε αυξάνεται και η άλλη), ενώ αλλιώς αντίθετα. Ο πίνακας επίσης είναι συμμετρικός αφού $Cov(a, b) = Cov(b, a)$. Στη διαγώνιο έχουμε τις διασπορές των μεταβλητών αφού $Cov(a, a) = Var(a)$.

Υπολογισμός Ιδιοδιανυσμάτων και Ιδιοτιμών.

Στη συνέχεια υπολογίζουμε τα ιδιοδιανύσματα και τις ιδιοτιμές του πίνακα συνδιασποράς έτσι ώστε να καθορίσουμε τα Principal Components. Δεν

θα εξηγήσουμε τον τρόπο αναλυτικού υπολογισμού των ιδιοδιανυσμάτων και των ιδιοτιμών ενός πίνακα στη συγκεκριμένη παράγραφο. Το σημαντικό παράλλατο, είναι να καταλάβουμε ότι τα ιδιοδιανύσματα του πίνακα συνδιασποράς, είναι εκείνες οι κατευθύνσεις που δείχνουν τη μέγιστη διασπορά του αρχικού dataset ενώ οι αντίστοιχες ιδιοτιμές περιέχουν την πληροφορία του μεγέθους της διασποράς που υπάρχει σε κάθε κατεύθυνση.

Συνεπώς κατατάσσουμε τα ιδιοδιανύσματα από την μεγαλύτερη στη μικρότερη τιμή ιδιοτιμής. Κάθε ιδιοτιμή στη συνέχεια διαιρείται με το άθροισμα όλων των ιδιοτιμών, δίνοντάς μας ένα ποσοστό του συνολικού Variance που φέρει η αντίστοιχη PCA κατεύθυνση.

Μετασχηματισμός Αρχικού Dataset

Το τελευταίο βήμα είναι να μετασχηματίσουμε το αρχικό dataset πολλαπλασιάζοντας το εσωτερικά με πίνακα των PCA κατευθύνσεων όπως φαίνεται στη σχέση (29).

$$X_{transformed} = W_{pca}^T X_{data} \quad (29)$$

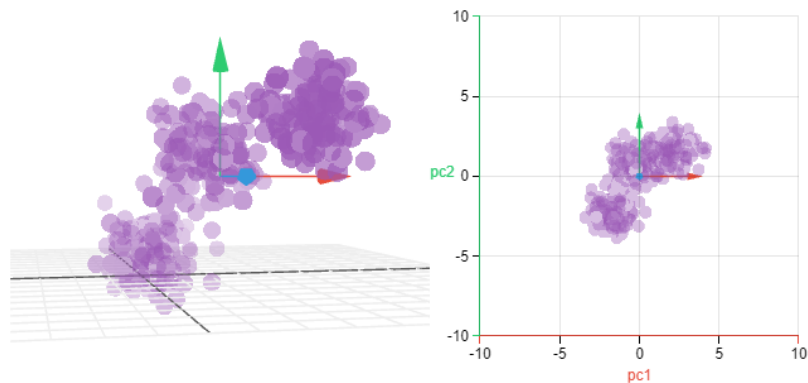
Ο πίνακας των PCA κατευθύνσεων θα μετασχηματίσει τα δεδομένα μας. Τώρα κάθε μεταβλητή στο καινούριο dataset $X_{transformed}$ θα είναι περιέχει ανάλογα στη θέση που βρίσκεται στον πίνακα και ανάλογο ποσοστό της συνολικής πληροφορίας. Η πρώτη επεξηγηματική μεταβλητή θα περιέχει το μεγαλύτερο ποσοστό η δεύτερη λιγότερο κλπ.

Εδώ είναι σημαντικό να τονίσουμε ότι οι καινούριες μεταβλητές που προέκυψαν, δεν είναι απαραίτητο ότι θα έχουν κάποιο φυσικό νόημα. Τα χαρακτηριστικά αυτά παρόλα αυτά που κωδικοποιούν, είναι με σειρά εκείνα που περιέχουν τη σημαντικότερη πληροφορία του dataset.

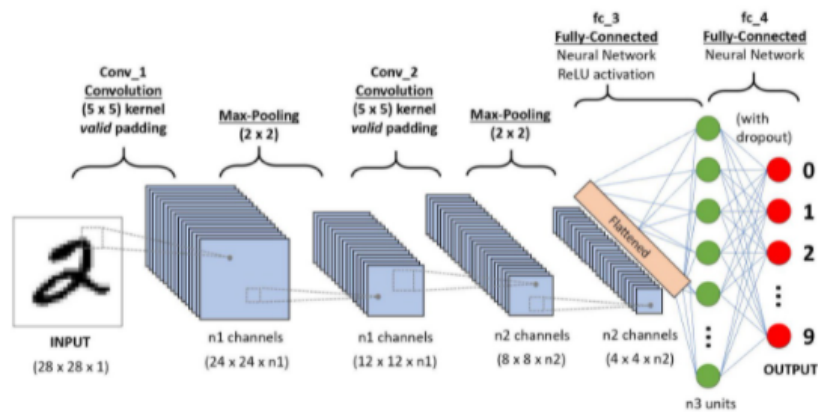
3.4.1 Η Τεχνική PCA Walking

Είδαμε ότι με την χρήση της τεχνικής PCA εξάγουμε τα ιδιοδιανύσματα του πίνακα συνδιασποράς ενός συνόλου δεδομένων. Τα ιδιοδιανύσματα αυτά υποδηλώνουν τις κατευθύνσεις που η διασπορά των δεδομένων γίνεται μέγιστη. Πως όμως αυτή η τεχνική μπορεί να χρησιμοποιηθεί για τον έλεγχο παραγωγής ενός GAN;

Η ουσία της μεθόδου κρύβεται στον τρόπο που τα Convolutional Neural Networks (CNNs) εξάγουν χαρακτηριστικά από τα πρότυπα εκπαίδευσης. Για να γίνει περισσότερο κατανοητό, ας υποθέσουμε ένα συνελικτικό νευρωνικό δίκτυο, το οποίο εκπαιδεύεται με εικόνες, με σκοπό την κατηγοριοποίηση τους σε κλάσεις. Ένα τυπικό παράδειγμα της αρχιτεκτονικής ενός τέτοιου δικτύου παρουσιάζεται στο Σχήμα 35.



Σχήμα 34: Ένα παράδειγμα της PCA σε δεδομένα 3ων διαστάσεων. Παρατηρούμε ότι μετά την εφαρμογή της τεχνικής οι σημαντικές κατευθύνσεις μέγιστης διασποράς είναι δύο, μειώνοντας τη διάσταση των δεδομένων κατά ένα.



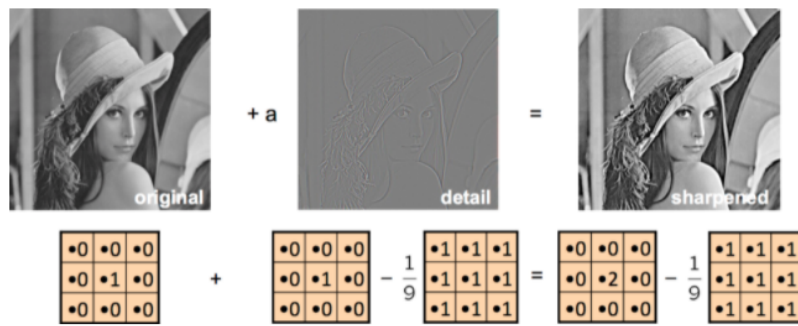
Σχήμα 35: Η δομή ενός CNN Classifier στο MNIST dataset.

Παρατηρούμε ότι ένα τέτοιο δίκτυο αποτελείται στην ουσία από δύο υπο-δίκτυα. Το πρώτο υποδίκτυο δομεί μία αλληλουχία από συνελικτικά layers. Ενδιάμεσα από κάθε τέτοιο layer εφαρμόζεται κάποιο pooling ανάλογα με την αρχιτεκτονική του δικτύου. Τα συνελικτικά στρώματα είναι μικροί (συνήθως 3×3 ή 5×5) πίνακες οι οποίοι ονομάζονται φίλτρα. Κάθε φίλτρο σαρώνει την εικόνα εισόδου με ένα μικρό βήμα (stride) και εφαρμόζει το μετασχηματισμό. Τα outputs δηλαδή του κάθε συνελικτικού νευρώνα είναι στην ουσία το αποτέλεσμα της εφαρμογής ενός συγκεκριμένου φίλτρου πάνω στην εικόνα.

Η εκπαίδευση δηλαδή αυτών των δικτύων έγκειται στον καθορισμό των φίλτρων κάθε νευρώνα, τα οποία στη συνέχεια θα εφαρμοστούν στην test εικόνα

για πρόβλεψη. Την κατηγοριοποίηση των εικόνων σε κλάσεις θα την εκτελέσει το δεύτερο υποδίκτυο, το οποίο δεν είναι τίποτε άλλο από ένα συμβατικό MLP. Συνεπώς τα outputs των συνελικτικών δικτύων είναι και τα χαρακτηριστικά αυτά των εικόνων που θα χρειαστεί ο MLP ταξινομητής για να κάνει την ταξινόμηση.

Τα φίλτρα λοιπόν εφαρμόζονται στα πρότυπα εισόδου για την εξαγωγή χαρακτηριστικών από τις εικόνες. Ένα παράδειγμα χρήσης δύο διαφορετικών φίλτρων σε μία εικόνα παρουσιάζεται στο Σχήμα 36. Διαφορετικά φίλτρα μπορούν να κάνουν smoothing σε μια εικόνα, να εξάγουν τις κάθετες ή τις οριζόντιες γραμμές καθώς επίσης να εξάγουν τις λεπτομέριες μιας εικόνας.

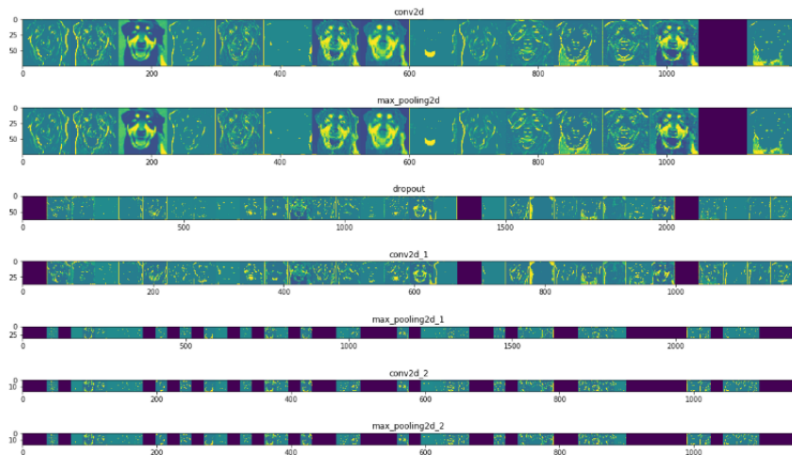


Σχήμα 36: Εφαρμογή δύο διαφορετικών φίλτρων σε εικόνες και τα χαρακτηριστικά που εξάγονται

Καταλαβαίνουμε δηλαδή ότι η χρήση των συνελικτικών layers γίνεται για την εξαγωγή αυτών των χαρακτηριστικών που το δίκτυο θα θεωρεί σημαντικά για την κατηγοριοποίηση των εικόνων και η εκπαίδευση τους γίνεται για τον καθορισμό των φίλτρων αυτών. Στη βιβλιογραφία, η δράση των συνελικτικών στρωμάτων αναφέρεται ως feature mapping.

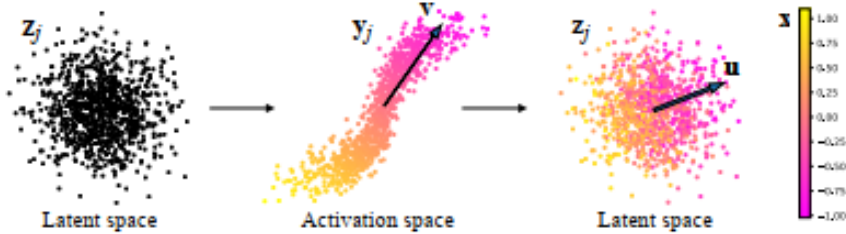
Συγκεκριμένα, έχει παρατηρηθεί ότι τα πρώτα συνελικτικά λαφερς εξάγουν πολύ βασικά χαρακτηριστικά μετασχηματίζοντας τις εικόνες εισόδου σε μικρό βαθμό (πόζα, ακμές κλπ) ενώ τα ανώτερα layers μαθαίνουν πολύ συγκεκριμένα χαρακτηριστικά που είναι δύσκολο να αναγνωριστούν. Ένα παράδειγμα ενός feature map παρουσιάζεται στο σχήμα 37, από ένα δίκτυο που εκπαιδεύτηκε για την κατηγοριοποίηση εικόνων από σκύλους και γάτες.

Ας σκεφτούμε τώρα ένα GAN το οποίο είναι εκπαιδευμένο για να παράγει εικόνες. Το GAN θα χρησιμοποιεί την πολυδιάστατη κανονική κατανομή $p(z)$



Σχήμα 37: Εξαγωγή χαρακτηριστικών στα συνεκτικά layers.

η οποία είναι ιστροπική. Δηλαδή αν θεωρήσουμε ότι δειγματοληπτούμε τυχαία από αυτή έναν μεγάλο αριθμό διανυσμάτων εισόδου, η εικόνα που θα πάρουμε θα μοιάζει με την αριστερά εικόνα του σχήματος 38. Τα διανύσματα είναι πυκνά στο κέντρο της κατανομής, ενώ απλώνονται ιστροπικά, με την ίδια πιθανότητα να βρίσκονται σε οποιαδήποτε κατεύθυνση όσο απομακρυνόμαστε από το κέντρο. Καταλαβαίνουμε ότι εδώ δεν υπάρχει κάποια κατεύθυνση μέγιστης διασποράς, αφού η διασπορά της κανονικής κατανομής $p(z)$ είναι ο μοναδιαίος πίνακας.



Σχήμα 38: Έλεγχος παραγωγής με τη χρήση PCA. Αριστερά δειγματοληπτούμε τυχαία μεγάλο αριθμό διανυσμάτων του latent space. Στη συνέχεια το output του πρώτου layer του δικτύου θα μετασχηματίσει τα διανύσματα δίνοντας κάποια κατεύθυνση μέγιστης διασποράς. Με τη χρήση της PCA προβάλλουμε τις κατευθύνσεις πίσω στο latent space.

Όταν τα διανύσματα αυτά περάσουν από κάποιο από τα πρώτα layers ενός GAN, θα μετασχηματιστούν με τέτοιο τρόπο, που θα αποκτήσουν κατευθύνσεις μέγιστης διασποράς. Αυτή την ιδέα απεικονίζει η δεύτερη εικόνα του

Σχήματος 38. Συνδιάζοντας το γεγονός ότι στα πρώτα layers γίνεται η εξαγωγή των βασικότερων χαρακτηριστικών από τα πρότυπα εκπαίδευσης, μπορούμε να υποθέσουμε ότι οι κατευθύνσεις μεγάλης διασποράς θα αντιπροσωπεύουν και τις σημαντικότερες κατευθύνσεις πάνω στις οποίες το δίκτυο θα διαχωρίζει τα αρχικά semantics που μπορεί να παράγει. Με άλλα λόγια προσπαθούμε να εντοπίσουμε τις κατευθύνσεις αυτές που θεωρούμε ότι είναι υπεύθυνες για τον έλεγχο των βασικών χαρακτηριστικών, που το δίκτυο μαθαίνει στα πρώτα layers του.

Ο εντοπισμός των σημαντικότερων κατευθύνσεων θα γίνει με τη χρήση της μεθόδου PCA η οποία εξηγήθηκε παραπάνω. Αν θεωρήσουμε ότι δειγματοληπτούμε N αριθμό διανυσμάτων εισόδου $z_{1:N}$, η έξοδος από το i layer για το κάθε ένα θα είναι $y_j = \hat{G}_i(z_j)$. Η μέθοδος PCA θα δώσει έναν low rank πίνακα βάσης V . Οι PCA κατευθύνσεις x_j μπορούν να υπολογιστούν από τη σχέση:

$$x_j = V^T(y_j - \mu) \quad (30)$$

Στη συνέχεια μπορούμε να προβάλουμε τη βάση V πίσω στον latent space. Η διαδικασία αυτή θα γίνει με τη χρήση linear regression, μεταξύ των διανυσμάτων εισόδου z και των PCA διανυσμάτων x , ελαχιστοποιώντας τη συνάρτηση κόστους:

$$\operatorname{argmin}_j \sum_k \|u_k x_j^k - z_k\|^2 \quad (31)$$

Στην ουσία έχουμε καταφέρει με αυτόν τον τρόπο να μάθουμε τον πίνακα U ο οποίος έχει διάσταση $L \times P$ όπου L η διάσταση του latent space και P ο αριθμός των PCA components που κρατήσαμε. Ο πίνακας U καταλαβαίνουμε ότι μετασχηματίζει οποιαδήποτε κίνηση πάνω σε κάποια κατεύθυνση x_j του PCA χώρου, στον χώρο εισόδου. Συνεπώς ξεκινώντας από ένα διάνυσμα αναφοράς z μπορούμε να ‘περπατήσουμε’ πάνω σε μία PCA κατεύθυνση κάνοντας μικρά βήματα β σύμφωνα με τη σχέση:

$$z_{new} = z + \beta U x \quad (32)$$

Όπου το x θα είναι ένα διάνυσμα μήκους L με τιμή 1 στην κατεύθυνση που επιλέγουμε να κινηθούμε και παντού αλλού μηδέν. Η τεχνική αυτή υλοποιήθηκε από τους Erik Härkönen, et al. [34] στα δίκτυα StyleGAN και BigGAN, τα οποία παράγουν πρότυπα εικόνων και παρουσίασε εκπληκτικά αποτελέσματα. Οι συγγραφείς κατάφεραν κινούμενοι στις PCA κατευθύνσεις ξεχωριστά να αλλάζουν semantics όπως το orientation της εικόνας, το χρώμα, την έκφραση του προσώπου και πολλά άλλα. Μερικά από τα αποτελέσματα της τεχνικής παρουσιάζονται στο σχήμα 39.



Σχήμα 39: Εύρεση κατευθύνσεων για τον έλεγχο παραγωγής του StyleGAN2 με τη χρήση της παρουσιαζόμενης τεχνικής. Κινούμενοι σε διαφορετικές κατευθύνσεις παρατηρείται η αλλαγή διαφορετικών semantics.

Στη συνέχεια της εργασίας θα εφαρμοστεί η συγκεκριμένη τεχνική για την εύρεση κατευθύνσεων στο latent space του MuseGAN.

3.5 Σύγκριση των δύο μεθόδων

Οι δύο μέθοδοι που παρουσιάστηκαν σε αυτό το κεφάλαιο προσπαθούν να λύσουν το πρόβλημα του ελέγχου στην παραγωγή προτύπων των δικτύων GAN. Η κύρια διαφορά τους έγκειται στον τρόπο με τον οποίο προσπαθεί η κάθε μία να εξάγει τις σημαντικές κατευθύνσεις από το latent space.

Η τεχνική latent walking επιχειρεί να εντοπίσει κατευθύνσεις με supervised τρόπο. Στην ουσία αφού ορίσουμε το semantic πάνω στο οποίο θέλουμε να προκύψουν οι αλλαγές με την χρησιμοποίηση κάποιας μετρικής που θα ποσοτικοποιεί την αλλαγή αυτή, χρησιμοποιούμε έναν αλγόριθμο μηχανικής μάθησης (SVMs) για να προσδιορίσουμε την κατεύθυνση αυτή. Η μέθοδος αυτή θα εφαρμοστεί ξεχωριστά για κάθε semantic που μας ενδιαφέρει εξαναγκάζοντας στην ουσία το δίκτυο να παράγει πρότυπα στις προσδιοριζόμενες κατευθύνσεις. Παρόλαυτά, αν ο latent space δεν είναι γραμμικός ως προς το εξεταζόμενο semantic, τα SVMs δεν θα καταφέρουν να κάνουν καλό διαχωρισμό και τα α-

ποτελέσματα δεν θα είναι τόσο ικανοποιητικά. Η τεχνική αυτή παρόλαυτά μας επιτρέπει να ορίσουμε μετρικές που μπορούν να ποσοτικοποιούν σημαντικά χαρακτηριστικά της μουσικής και να επιχειρήσουμε να καθορίσουμε τις κατευθύνσεις του latent space πάνω στις οποίες η τιμή των μετρικών αλλάζει, αν αυτό είναι δυνατό. Μας παρέχει δηλαδή μεγαλύτερη ελευθερία στον πειραματισμό, δίνοντας μας τη δυνατότητα να προσπαθήσουμε να ελέγξουμε εκείνο το χαρακτηριστικό που εμείς θεωρούμε σημαντικό.

Αντίθετα, η μέθοδος PCA walking είναι μια unsupervised μέθοδος. Επιχειρούμε να βρούμε τις κατευθύνσεις αλλαγής τις οποίες το ίδιο το δίκτυο θεωρεί σημαντικές. Η τεχνική αυτή εφαρμόστηκε το 2020 στα δίκτυα Stylegan2 και BigGAN παρουσιάζοντας εκπληκτικά αποτελέσματα. Με τη χρήση της τεχνικής δεν χρειάζεται μεγάλος χρόνος πειραματισμού για τον καθορισμό σημαντικών κατευθύνσεων, με τη χρησιμοποίηση πολλών διαφορετικών μετρικών. Οι κατευθύνσεις θα είναι τόσες, όσες και τα Principal Components που θα αποφασίσουμε να κρατήσουμε. Η δυσκολία παρόλαυτά έγκειται στην αξιολόγηση αυτών των κατευθύνσεων, διότι δεν είναι απαραίτητο οι αλλαγές που προκύπτουν σε αυτές, να έχουν και διαισθητικό νόημα, πόσο μάλλον μουσικό νόημα. Η καθαρότητα των κατευθύνσεων έχει άμεση εξάρτηση από την ποιότητα του Generator καθώς και των δεδομένων εκπαίδευσης που χρησιμοποιήθηκαν. Ο καθορισμός του τί αλλάζει σε κάθε κατεύθυνση θα πρέπει να γίνει αρχικά με το αυτί (στα μουσικά πρότυπα) για να μπορούμε στη συνέχεια να σκεφτούμε την υλοποίηση μιας μετρικής που θα ποσοτικοποιήσει αυτήν την αλλαγή.

4 Μετρικές Αξιολόγησης Μουσικών Προτύπων

Πρωτού προχωρήσουμε στην παρουσίαση της εφαρμογής των δύο μεθόδων που αναλύσαμε στο προηγούμενο κεφάλαιο, θεωρούμε σκόπιμο να παρουσιάσουμε τον τρόπο αξιολόγησης της παραγόμενης μουσικής. Σε μία προσπάθεια να καθορίσουμε τις αλλαγές στην παραγόμενη μουσική, είναι απαραίτητο να οριστούν μετρικές οι οποίες θα τις απεικονίζουν και αν είναι δυνατόν θα τις ποσοτικοποιούν.

Όπως εξηγήσαμε, η μέθοδος Latent Walking επιχειρεί να βρει κάποια κατεύθυνση του Latent Space, που κινούμενοι πάνω σε αυτή θα επιτύχουμε την αύξηση ενός συγκεκριμένου χαρακτηριστικού. Παρομοίως η μέθοδος PCA Walking θα μας δώσει σημαντικές κατευθύνσεις, πάνω στις οποίες εμείς θα πρέπει να εντοπίσουμε την αλλαγή. Σε κάθε περίπτωση προσπαθώντας να σκεφτούμε βασικά χαρακτηριστικά της ανθρώπινης μουσικής σύνθεσης στο συγκεκριμένο κεφάλαιο θα παρουσιαστούν όλες εκείνες οι μετρικές που υλοποιήθηκαν και χρησιμοποιήθηκαν για την αξιολόγηση των αλλαγών της παραγόμενης μουσικής.

4.1 Η Μορφή των Παραγόμενων Προτύπων - Προεπεξεργασία

Πρωτού προχωρήσουμε στην παρουσίαση των μετρικών που χρησιμοποιήθηκαν, καθώς και στον τρόπο υλοποίησής τους, είναι σημαντικό να εξηγήσουμε την μορφή της παραγόμενης μουσικής.

Το output του δικτύου είναι ένα multitrack, που περιέχει 5 pianoroll πίνακες, έναν για κάθε όργανο που συμμετέχει στη σύνθεση. Τα όργανα είναι drums, piano, guitar, bass, strings. Κάθε pianoroll είναι ένας 128x240 πίνακας. Οι 128 γραμμές αναφέρονται στις 128 διαφορετικές νότες που μπορεί να χρησιμοποιηθούν, ενώ οι 240 στήλες είναι τα διαφορετικά timestamps του τραγουδιού.

Η πρώτη παρατήρηση μετά από πειραματισμό, ήταν η εξής. Όλα τα παραγόμενα πρότυπα, σε όλα τα tracks, 'παίζανε μουσική' μέχρι το 189ο timestamp. Έπειτα από αυτό, για όλα τα όργανα και για μεγάλο αριθμό μουσικών προτύπων, ο pianoroll πίνακας δεν είχε πουθενά την τιμή True.

Για την εξαγωγή του τελευταίου timestamp, κατασκευάστηκε και αντίστοιχη συνάρτηση και δοκιμάστηκε σε παραπάνω από 400 παραγόμενα πρότυπα. Συνεπώς το επόμενο βήμα ήταν η κατασκευή ενός trimmer που θα κόβει τον πίνακα μετά από το συγκεκριμένο timestamp.

Παράλληλα, υλοποιήσαμε με απλό τρόπο μία συνάρτηση που διαχωρίζει ένα multitrack σε μια λίστα η οποία περιέχει τα pianorolls για το κάθε όργανο.

Τέλος, ρυθμίσαμε το MuseGAN να παράγει τραγούδια τεσσάρων μέτρων των τεσσάρων τετάρτων. Μία απλή συνάρτηση υλοποιήθηκε, η οποία διαχωρίζει τον συνολικό πίνακα του κάθε οργάνου, σε 4 μικρότερους πίνακες, έναν για κάθε μέτρο.

4.2 Chord Contribution Metric

Όπως εξηγήσαμε στο πρώτο κεφάλαιο, η σύγχρονη δυτική μουσική έχει αρμονική δομή. Η αρμονία θα καθοριστεί στα μέτρα και στη συνέχεια οι μελωδικές γραμμές θα κινούνται σε κλίμακες που ‘ταιριάζουν’ στην αρμονική δομή. Αυτό βέβαια δεν αποτελεί κανόνα, παρόλαυτά, εν γένει, μουσική που παρουσιάζει αυτή τη δομή, είναι προσφιλής στο ανθρώπινο αυτί και αποτελεί μια κοινή πρακτική σύνθεσης. Για παράδειγμα, αν η αρμονία ενός μέτρου σε ένα μουσικό κομμάτι βασίζεται στην Ντο μείζονα συγχορδία, οι νότες που είναι πιθανότερο να βρούμε στο μέτρο αυτό, είναι και οι νότες της συγχορδίας αυτής, δηλαδή Ντο, Μι και Σολ.

Γιαυτό το λόγο θεωρήσαμε σημαντικό να κατασκευάσουμε μία μετρική που να απεικονίζει το ποσοστό συνεισφοράς μιας συγχορδίας, στις νότες του κάθε μέτρου του μουσικού προτύπου. Ονομάζουμε την συγκεκριμένη μετρική Chord Contribution. Για την υλοποίηση της Chord Contribution εργαστήκαμε ως εξής:

4.2.1 Note Contribution

Υλοποιήσαμε τη συνάρτηση Note Contribution η οποία δέχεται ως input ένα πίνακα pianoroll. Η συνάρτηση αυτή κατασκευάζει μία λίστα 12 θέσεων. Κάθε θέση περιέχει το άθροισμα όλων των νοτών, που αντιστοιχούν στη θέση. Για παράδειγμα η πρώτη θέση θα περιέχει το άθροισμα όλων των Ντο, όλων των οκτάβων που παίχτηκαν στο συγκεκριμένο pianoroll, στη δεύτερη θέση θα βρίσκονται όλα τα Ντο # κ.ο.κ.

4.2.2 Chord Metrics

- Στη συνέχεια κατασκευάστηκε η συνάρτηση Chord Metrics. Η συγκεκριμένη συνάρτηση δέχεται ως input ένα multitrack καθώς και άλλες δύο παραμέτρους, root_node και mode. Αρχικά χωρίζουμε το input σε 5 tracks (pianorolls) από τα οποία κρατάμε μόνο τα 4. Τα drums δεν συνεισφέρουν στην μελωδία ή στην αρμονία, συνεπώς δεν θα χρησιμοποιηθούν στη συγκεκριμένη μετρική.

- Ύστερα, για κάθε όργανο, διαχωρίζουμε τον αρχικό πίνακα σε 4, έναν για κάθε μέτρο του τραγουδιού και αθροίζουμε τους πίνακες όλων των οργάνων στο χρόνο. Δηλαδή καταλήγουμε σε 4 πίνακες, έναν για κάθε μέτρο, οι οποίοι περιέχουν όλες τις νότες που παίχτηκαν από τα 4 μελωδικά όργανα μαζί, σαν να εκτελέστηκαν μόνο από ένα όργανο.
- Για κάθε τέτοιο πίνακα - χρονικό segment, υπολογίζουμε την τιμή της Note Contribution συνάρτησης που ορίστηκε παραπάνω. Συνεπώς παίρνουμε 4 λίστες μήκους 12 η κάθε μία. Σε κάθε θέση της εκάστοτε λίστας, περιέχεται ο αριθμός των νοτών που αντιστοιχούν στη συγκεκριμένη θέση.
- Θέλουμε αυτοί οι αριθμοί, από αθροίσματα να μετατραπούν σε δειγματικές πιθανότητες. Στην ουσία θα πρέπει να διαιρέσουμε κάθε στοιχείο της λίστας με το άθροισμα των συνολικών νοτών. Τώρα, το κάθε στοιχείο της λίστας θα απεικονίζει το ποσοστό συνεισφοράς της συγκεκριμένης νότας στο εκάστοτε μέτρο. Δηλαδή αν το πρώτο στοιχείο της λίστας του πρώτου μέτρου για παράδειγμα είναι 0.2, αυτό σημαίνει ότι το 20% των νοτών που παίχτηκαν από όλα τα όργανα σε όλες τις οκτάβες είναι Ντο. Ομοίως για όλα τα 12 στοιχεία της λίστας.
- Κατασκευάζουμε την κλάση Circular List μήκους 12 για την οποία θα ισχύει $list[end + 1] = list[start]$, $list[end + 2] = list[start + 1]$ κλπ. Μετατρέπουμε τις 4 λίστες σε κυκλικές, $[note_contribution_bar1, note_contribution_bar2, note_contribution_bar3, note_contribution_bar4]$.
- Οι πιθανές root notes που θα εισαχθούν από το χρήστη θα μπορεί να είναι μία από τις $possible_roots = [C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B]$ δηλαδή όλες τις πιθανές (ονομαστικά) νότες που υπάρχουν. Δεδομένης της root note, βρίσκουμε τη θέση της - $root_note_index$ στην λίστα $possible_roots$.
- Αν το $mode = Major$ τότε ορίζουμε 4 ποσοστά με τον ίδιο τρόπο για κάθε μέτρο ως εξής. Η μείζονα συγχορδία αποτελείται από τρεις νότες. Την βάση της συγχορδίας (root note), την τρίτη νότα, η οποία απέχει διάστημα 3ης μεγάλο (4 ημιτόνια) από τη βάση, και η 5ή νότα η οποία απέχει διάστημα 5ης καθαρό (7 ημιτόνια) από τη βάση. Ο δείκτης i στην παρακάτω σχέση αναφέρεται στον αριθμό του μέτρου, $i = \{1, 2, 3, 4\}$.

$$percentage^i = note_contribution^i[root_note_index] \\ + note_contribution^i[root_note_index + 4] \\ + note_contribution^i[root_note_index + 7]$$

- Αν το mode = Minor τότε επαναλαμβάνουμε τη διαδικασία αναζητώντας διάστημα τρίτης μικρό για την τρίτη νότα της συγχορδίας:

$$\begin{aligned}
 percentage^i &= note_contribution^i[root_note_index] \\
 &+ note_contribution^i[root_note_index + 3] \\
 &+ note_contribution^i[root_note_index + 7]
 \end{aligned}$$

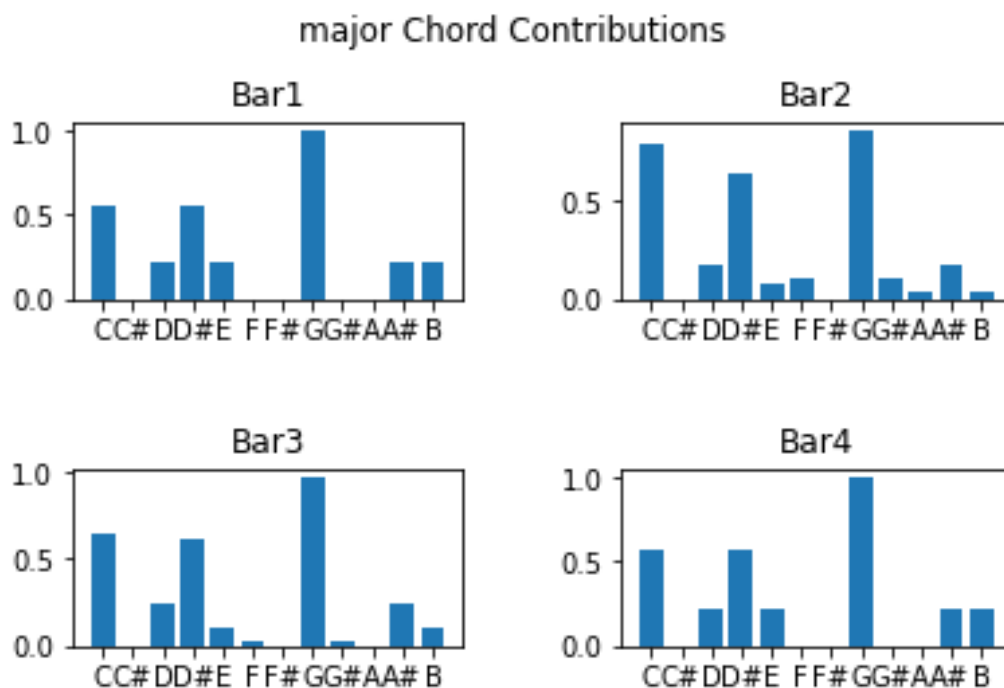
- Η συνάρτηση θα επιστρέφει μία λίστα από τις τέσσερις τιμές των ποσοστών αυτών, στρογγυλοποιημένες στο τρίτο δεκαδικό ψηφείο.

4.2.3 Chord Contribution

Τέλος, χρησιμοποιώντας την τελευταία συνάρτηση, κατασκευάζουμε την μετρική μας. Στην Chord Contribution εφαρμόζεται η προηγούμενη συνάρτηση Chord Metrics σε όλες τις πιθανές root notes και στα δύο πιθανά modes. Τα ποσοστά για τα 4 μέτρα κατασκευάζονται ξεχωριστά από την Chord Metrics για κάθε πιθανό συνδυασμό root note και mode. Παρέχεται επίσης η δυνατότητα καθορισμού ενός threshold για τη συνεισφορά. Το threshold ορίζεται από το χρήστη, και η μετρική θα επιστρέφει την ένδειξη "No Chord" όταν η συνεισφορά κάποιας συγχορδίας στο μέτρο είναι μικρότερη από το threshold. Η μετρική σε αρχικό στάδιο υπολογίζει τις συνεισφορές όλων των δυνατών triads σε κάθε μέτρο και παρέχει χρήσιμα barplots για την κάθε συγχορδία. Ένα παράδειγμα εμφανίζεται στο σχήμα 40.

Στο συγκεκριμένο σημείο θα πρέπει να τονίσουμε το εξής. Διαφορετικές συγχορδίες, ενδεχομένως να έχουν κοινές νότες. Για παράδειγμα η συγχορδία Ντο μείζονα έχει τις νότες C,E,G ενώ η συγχορδία Σολ μείζονα έχει τις νότες G,B,D. Γιαυτό το λόγο παρατηρούμε ότι στο δεύτερο μέτρο του σχήματος 40, η συνεισφορά των C major και G major είναι και οι δύο αρκετά ψηλές, ενδεχομένως στο συγκεκριμένο μέτρο να παιχτήκαν αρκετές νότες G. Παρόλαυτα παρατηρούμε ότι η συγχορδία G έρχεται με μεγαλύτερη πιθανότητα - συνεισφορά, υποδηλώνοντας μας ότι οι νότες B,D παιχτήκαν περισσότερο.

Το τελευταίο βήμα της μετρικής είναι η επιλογή της συγχορδίας με τη μεγαλύτερη συνεισφορά για το κάθε μέτρο. Για παράδειγμα στο συγκεκριμένο πρότυπο του σχήματος 40 η μετρική θα τυπώσει ('G', 1.0) , ('G', 0.856) , ('G', 0.972) , ('G', 1.0) , όπου το γράμμα αναφέρεται στην συγχορδία με τη μεγαλύτερη συνεισφορά στο αντίστοιχο μέτρο, ενώ ο αριθμός είναι η τιμή της συνεισφοράς αυτής.



Σχήμα 40: Εφαρμογή της Chord Contribution μετρικής.

4.2.4 Ερμηνεία και παρατηρήσεις

Τιμή συνεισφοράς 1 σημαίνει ότι όλες οι νότες που παιχτήκαν στο αντίστοιχο μέτρο, ήταν οι νότες που άνηκαν στην επικρατέστερη συγχορδία του μέτρου. Συγκεκριμένα στο παράδειγμα του σχήματος 40, όλες οι νότες που παίχτηκαν στο 1ο και στο 4ο μέτρο, θα είναι και οι νότες της Σολ μείζονας συγχορδίας, δηλαδή G,B,D. Όσο η τιμή της συνεισφοράς μειώνεται, τόσο θεωρούμε ότι μειώνεται και η καθαρότητα της αρμονίας του μέτρου. Δηλαδή, διαφορετικές νότες προστίθενται στο μέτρο ξεφεύγοντας από τα συγκεκριμένα όρια που θέτει η triad συγχορδία. Φυσικά κάποιες νότες μπορούν να σχηματίσουν πιο σύνθετες συγχορδίες όπως έβδομες, ελαττωμένες, ένατες κ.ο.κ. Παρόλαυτά τέτοιες δομές δεν εξετάζονται στην παρούσα μετρική, αφήνοντας σημαντικό έδαφος εξέλιξης.

Για τη χρήση της συγκεκριμένης μετρικής, θεωρείται ότι το μουσικό πρότυπο δεν αλλάζει συγχορδία εντός μέτρου. Παρατηρήσαμε ότι η πλειοψηφία των παραγόμενων προτύπων ικανοποιεί αυτή τη συνθήκη, κρατώντας κατά κανόνα μία αρμονική δομή μέσα σε κάθε μέτρο. Με αφορμή αυτήν την παρατήρηση προχωρήσαμε και σε αυτόν τον τρόπο υλοποίησης.

Οι νότες αθροίστηκαν ως κοινές ανεξαρτήτως του τονικού τους ύψους

ή των οργάνων που τις παίζανε. Με άλλα λόγια η μετρική αντιμετωπίζει τις νότες με κοινό όνομα (πχ Nτο) με τον ίδιο τρόπο ανεξαρτήτως αν ανήκουν σε διαφορετικές οκτάβες. Στην πραγματική σύνθεση και ενορχήστρωση, οι συγχορδίες τοποθετούν την βάση τους στην πιο μπάσα φωνή. Αν δεν συμβαίνει αυτό, τότε λέμε ότι η συγχορδία είναι σε αναστροφή. Συνήθως το όργανο του μπάσου αναλαμβάνει την εκτέλεση της πιο μπάσας φωνής. Μία συνθήκη που επιχειρήθηκε είναι η απόδοση μεγαλύτερου βάρους στον καθορισμό της επικρατέστερης συγχορδίας, στην περίπτωση που το μπάσο παίζει την root note. Παρόλαυτά συγκρίνοντας την λειτουργία της μετρικής με το αυτί, κρίναμε ότι τα αποτελέσματα δεν είναι ικανοποιητικότερα. Αν τα δεδομένα εκπαίδευσης είχαμε φροντίσει να αποτελούνται από καθαρή αρμονική δομή βασισμένη στις triads συγχορδίες χωρίς αναστροφή, τότε ίσως το δίκτυο θα είχε μάθει αυτήν την παράμετρο και η συγκεκριμένη ενέργεια να έφερνε καλύτερα αποτελέσματα.

4.2.5 Chord Quality

Υλοποιήθηκε παράλληλα και μια παραλλαγή της Chord Contribution την οποία ονομάσαμε Chord Quality. Η συγκεκριμένη παραλλαγή επιστρέφει τον μέσο όρο των συνεισφορών όλων των μέτρων και υλοποιήθηκε για να ποσοτικοποιήσει την ποιότητα της αρμονικής δομής όλου του μουσικού προτύπου.

Με τη μετρική αυτή, δεν αναζητάμε την επικρατέστερη συνεισφορά συγχορδίας σε κάθε μέτρο. Απαντάμε στο ερώτημα του κατά πόσο (κατά μέσο όρο) συνολικά το παραγόμενο πρότυπο σέβεται την επικρατέστερη συγχορδία ανα μέτρο, όποια και αν είναι αυτή. Για παράδειγμα αν σε ένα μουσικό πρότυπο έχουμε τιμή της Chord Quality =1 αυτό σημαίνει ότι ανεξάρτητα του ποιές συγχορδίες εντοπίστηκαν σε κάθε μέτρο, όλες οι νότες θα ανήκουν στις αντίστοιχες συγχορδίες, παράγοντας μία αρμονική δομή καλής ποιότητας.

Η Chord Quality αντικατοπτρίζει κατά την άποψή μας ένα πολύ σημαντικό στοιχείο της μουσικής σύνθεσης. Το γεγονός ότι είναι διατάξιμη είναι επίσης πολύ σημαντικό διότι με τη χρήση της μπορεί να γίνει σύγκριση διαφορετικών μουσικών προτύπων που προήλθαν από διαφορετικά διανύσματα εισόδου.

4.3 Pypianoroll Metrics

Για τους σκοπούς της εργασίας χρησιμοποιήθηκαν επίσης μερικές από τις μετρικές αξιολόγησης των αποτελεσμάτων του MuseGAN. Οι συγγραφείς δημιούργησαν την pypianoroll βιβλιοθήκη η οποία χρησιμοποιεί την pretty-midi για τη διαχείριση και την επεξεργασία των pianorolls. Επίσης στη Pypianoroll περιέχονται συναρτήσεις με τις οποίες μπορούμε να εξάγουμε διάφορα χαρακτηριστικά από τα μουσικά πρότυπα. Από αυτές τις ήδη υλοποιημένες μετρικές,

κάποιες δουλεύουν και χρησιμοποιήθηκαν ως έχουν, ενώ κάποιες άλλες παρουσιάζουν bugs και δεν ήταν δυνατόν να χρησιμοποιηθούν.

4.3.1 Qualified Note Rate

Η συγκεκριμένη μετρική υπήρχε μέσα στην PyrianoRoll παρόλαυτά δεν δούλεψε. Γιαυτό το λόγο υλοποιήθηκε από την αρχή. Με τον όρο qualified note αναφερόμαστε σε μία νότα η οποία έχει χρονική διάρκεια παραπάνω από κάποιον αριθμό n timestamps. Εμείς θέσαμε το $n = 2$ θεωρώντας ότι οι νότες που είχαν χρονική διάρκεια μόνο ένα timestamp δεν είναι qualified. Η μετρική επιστρέφει τον αριθμό των qualified νοτών διά τον αριθμό όλων των νοτών σε ένα κομμάτι. Ο λόγος που αυτή η μετρική είναι σημαντική έγκειται στο γεγονός ότι οι νότες οι οποίες ακούγονται μόνο για ένα timestamp μπορούν να θεωρηθούν ως artefacts του αλγορίθμου. Καμία νότα σε μια πραγματική σύνθεση δεν είναι στιγμιαία, συνεπώς πρότυπα με μεγάλο qualified note rate μπορούμε να πούμε πως ως προς αυτή την παράμετρο προσεγγίζουν πραγματικές μουσικές συνθέσεις. Για την υλοποίηση σημειώνουμε ότι παραλήφθηκαν τα drums, τα οποία δικαιολογημένα μπορεί να έχουν στιγμιαίους χτύπους χωρίς διάρκεια.

4.3.2 Polyphonic Rate

Η Polyphonic Rate μετρική, αναζητά τον αριθμό των timestamps ανά όργανο, στα οποία δύο η παραπάνω νότες συνηγούν ταυτόχρονα. Στη συνέχεια διαιρεί τον αριθμό των πολυφωνικών timestamps με τη συνολική διάρκεια, επιστρέφοντας ένα ποσοστό πολυφωνικότητας του οργάνου. Λαμβάνουμε τον μέσο όρο στα 4 μελωδικά όργανα δίνοντας έτσι ένα συνολικό μέτρο της πολυφωνικότητας του μουσικού προτύπου. Η μετρική αυτή αντικατοπτρίζει μία σημαντική ιδιότητα της ανθρώπινης μουσικής σύνθεσης η οποία τείνει να είναι εν γένει πολυφωνική.

4.3.3 Note Count

Η πιο προφανής και απλή μετρική είναι αυτή που επιστρέφει το συνολικό αριθμό των νοτών ενός παραγόμενου μουσικού προτύπου. Η μετρική αυτή ενδεχομένως να μην απεικονίζει κάποιο ποιοτικό χαρακτηριστικό της μουσικής, παρόλαυτά παρατηρούμε ότι οι αλλαγές που παράγει το δίκτυο, οδηγούν σχεδόν πάντα σε αλλαγές στον αριθμό των νοτών των κομματιών. Συνεπώς θεωρήσαμε ότι ο αριθμός των νοτών που παίχτηκαν στο παραγόμενο πρότυπο, είναι ένα χαρακτηριστικό που οπωσδήποτε πρέπει να εξετάσουμε στην έρευνά μας.

5 Πειραματική Διαδικασία

Όπως είναι εύκολα κατανοητό μία εξαντλητική εξερεύνηση του χώρου είναι απαγορευτική. Δεδομένου ότι ξεκινάμε από κάποιο διάνυσμα αναφοράς z οι πιθανές κινήσεις μόνο για ένα βήμα στον 128 διάστατο χώρο είναι σε αριθμό 2^{128} . Ο αριθμός αυτός είναι τεράστιος και απαιτεί πρακτικά άπειρη υπολογιστική δύναμη. Είναι αναγκαία λοιπόν η χρησιμοποίηση κάποιας πρακτικότερης μεθόδου για την εξερεύνηση του χώρου. Στο κεφάλαιο αυτό θα παρουσιαστούν όλα τα βήματα της υλοποίησης των δύο μεθόδων που παρουσιάστηκαν στα προηγούμενα κεφάλαια για την εύρεση κατευθύνσεων στον 128 χώρο εισόδου του δικτύου MuseGAN.

5.1 Δυσκολίες στη χρήση του MuseGAN

Το MuseGAN μας δημιούργησε δυσκολίες στη χρησιμοποίησή του. Το documentation στον κώδικα είναι σχεδόν ανύπαρκτο. Το MuseGAN επίσης δεν υπάρχει ως κλάση σε κάποια από τις γνωστές βιβλιοθήκες νευρωνικών δικτύων της Python (Pytorch/Tensorflow). Ο κώδικας παρέχεται στο Github από τους δημιουργούς τους, σε τέτοια μορφή έτσι ώστε να εκτελεί το δίκτυο πολύ συγκεκριμένες εφαρμογές (training, inference, interpolation) με τη χρήση bash αρχείων. Τα python αρχεία δομούνται κατά τέτοιο τρόπο (χρησιμοποιώντας το κάθε ένα τα υπόλοιπα) που ήταν αρκετά δύσκολο, να βρούμε τί εκτελεί το κάθε τί.

Μια άλλη τεχνική δυσκολία που παρουσιάζει το MuseGAN βασίζεται στο γεγονός ότι ο κώδικας του όσον αφορά το δίκτυο, είναι γραμμένος σε αρκετά παλαιότερη version της Tensorflow (1.15). Στη συγκεκριμένη εκδοχή, ο κώδικας γράφεται με εντελώς διαφορετικό τρόπο σε σχέση με τις πιο σύγχρονες εκδοχές (≥ 2.0). Αυτό σημαίνει ότι η χρήση του MuseGAN θα γίνει σε περιβάλλον Linux στο οποίο η Python θα γίνει downgrade για να είναι συμβατή με την παλαιότερη έκδοση του Tensorflow.

5.2 Υλοποίηση της Μεθόδου PCA Walking

Η υλοποίηση της παραπάνω μεθόδου έγινε με τη χρήση της Python ακολουθώντας ακριβώς τα βήματα που παρουσιάστηκαν στο άρθρο [34] καθώς και στο κεφάλαιο 3.4.1. Τα βήματα της μεθόδου παρουσιάζονται και εξηγούνται στη συνέχεια.

5.2.1 Διαδικασία

Δειγματοληψία του Συνόλου Z

Το πρώτο βήμα της υλοποίησης είναι η κατασκευή του συνόλου Z , το οποίο θα αποτελείται από 1 εκατομμύριο 128-διάστατα διανύσματα εισόδου z_i . Δειγματολειπούμε κάθε z_i από την 128-διάστατη κανονική κατανομή με κέντρο το μηδενικό διάνυσμα και πίνακα συνδιασποράς τον μοναδιαίο.

Περιγραφή και Κατασκευή του Generator

Το επόμενο βήμα για την υλοποίηση της μεθόδου είναι να πάρουμε output από κάποιο ενδιάμεσο Layer του Generator του δικτύου. Στο άρθρο [35] οι συγγραφείς, χρησιμοποίησαν το BigGAN για αυτό το σκοπό. Το πρώτο layer του δικτύου αυτού είναι linear, τα διανύσματα του συνόλου Z πέρασαν από αυτό, και για κάθε ένα κατασκευάστηκε το output σύνολο Y . Στο άρθρο γενικά προτείνεται η χρήση των αρχικών layers του δικτύου, διότι σε εκείνα τα αρχικά στάδια, το δίκτυο δεν μαθαίνει συγκεκριμένα χαρακτηριστικά που αφορούν το κάθε όργανο της σύνθεσης ξεχωριστά, αλλά διαχωρίζει βασικά χαρακτηριστικά της σύνθεσης ως σύνολο.

Στη δική μας περίπτωση, ο Generator του MuseGAN δεν διαθέτει linear layer. Αντιθέτως δομείται από blocks. Κάθε τέτοιο block αποτελείται από ένα convolutional-3D-transpose layer μεταβλητών διαστάσεων, και ένα batch normalization layer. Το κάθε διάνυσμα εισόδου, περνάει με τη σειρά από τρία τέτοια blocks τα οποία αποτελούν το Shared Network του Generator. Σε αυτό το στάδιο είναι σημαντικό να τονίσουμε ότι δεν γίνεται κανένας διαχωρισμός του input σε κομμάτια υπεύθυνα για τη σύνθεση διαφορετικών οργάνων ή οποιοσδήποτε άλλος διαχωρισμός που να αφορά το χρονικό στοιχείο (πχ των μέτρων). Θεωρήσαμε ότι τα blocks του Shared Network είναι αυτά που πρέπει να χρησιμοποιηθούν, για τη μέθοδο, σαν ένα ανάλογο του linear layer του BigGAN.

Καταλαβαίνουμε από τα προαναφερθέντα, ότι στην πραγματικότητα είχαμε αρκετά μεγάλη δυσκολία να προσαρμόσουμε τον κώδικα του MuseGAN στις δικές μας ανάγκες. Γιαυτό το λόγο κατασκευάσαμε ξεχωριστά μια Generator κλάση στην Pytorch, η οποία αποτελείται ακριβώς από τα τρία blocks που δομούσαν το Shared Network. Τα βάρη του δικτύου με τη χρήση κατάλληλης βιβλιοθήκης διαβάστηκαν, και ενσωματώθηκαν στο δικό μας δίκτυο.

Στη συνέχεια τα διανύσματα του συνόλου Z τροφοδοτήθηκαν στο δικό μας Shared Network και πήραμε ξεχωριστά το output από κάθε block. Για τη συνέχεια της εργασίας, θα παρουσιαστεί μόνο η διαδικασία για το πρώτο block. Τα αποτελέσματα αρχικά αξιολογήθηκαν με το αυτί και για την περίπτωση του δεύτερου και του τρίτου block ήταν αρκετά χαμηλής ποιότητας, οπότε δεν θα παρουσιαστούν.

Κατασκευάστηκε έτσι το σύνολο Y (activation space) που αποτελείται από 1 εκατομμύριο outputs του πρώτου block του Generator του MuseGAN.

Εφαρμογή της PCA

Το επόμενο βήμα έγκειται στην εφαρμογή της PCA τεχνικής στα σύνολο των διανυσμάτων του activation space Y . Για το σκοπό αυτό χρησιμοποιήθηκε το αντίστοιχο πακέτο της sklearn. Κρατήσαμε τα σημαντικότερα 10 Principal Components για ευκολία στον πειραματισμό. Επίσης δοκιμάστηκε και η μέθοδος της kernel PCA με τη χρήση RBF kernel, παρόλαυτά ήταν υπολογιστικά ακριβή για τις δικές μας δυνατότητες, συνεπώς εγκαταλείφθηκε. Ο πίνακας βάσης V που προέκυψε από την μέθοδο αυτή αποθηκεύτηκε σε μορφή pickle.

Προβολή στο Χώρο Εισόδου

Οι PC κατευθύνσεις x_j υπολογίστηκαν σύμφωνα όπως εξηγήθηκε. Για να προβάσουμε τις κατευθύνσεις πίσω στον χώρο εισόδου, χρησιμοποιήσαμε τη μέθοδο γραμμικής παλινδρόμησης ανάμεσα στο σύνολο Z και στο σύνολο των κατευθύνσεων X . Ο συντελεστής του γραμμικού μοντέλου U , αποτελεί έναν πίνακα 128×10 διαστάσεων ο οποίος προβάλλει τα διανύσματα του χώρου 10 διαστάσεων, στα διανύσματα του χώρου εισόδου. Για τον αλγόριθμο της γραμμικής παλινδρόμησης χρησιμοποιήθηκε το αντίστοιχο πακέτο της βιβλιοθήκης sklearn.

PCA Walking

Για το τελευταίο στάδιο της μεθόδου, υλοποιήσαμε μία συνάρτηση η οποία δέχεται ως input ένα διάνυσμα του latent space z_{ref} , τον αριθμό x της PC κατεύθυνσης που θέλουμε να κινηθούμε (ο αριθμός αυτός παίρνει τιμές από το 0 έως το 9), και το βήμα της κίνησης β .

Κάθε φορά, ο αλγόριθμος θα παράγει καινούρια διανύσματα στην κατεύθυνση που επιθυμούμε με τον τρόπο που εξηγούμε στη σχέση (32). Τα πειράματα έγιναν σε μία από τις 10 κατευθύνσεις ξεχωριστά. Χρησιμοποιήθηκε μετά από πειραματισμό, η τιμή του $\beta=1$. Μικρότερες τιμές του β κρίναμε ότι οδηγούν σε πάρα πολύ αργές αλλαγές στα παραγόμενα πρότυπα.

Αντίθετα μεγαλύτερες τάξεις μεγέθους του β οδηγούσαν σε απότομες αλλαγές στα μουσικά τραγούδια, δυσκολεύοντας τον εντοπισμό της αλλαγής. Παράλληλα θα πρέπει να τονίσουμε ότι κάνοντας βήματα προς κάποια από τις PC κατευθύνσεις, συνεχώς αυξάνεται η Ευκλείδεια απόσταση των διανυσμάτων που παίρνουμε από το κέντρο της κατανομής p_G . Έτσι σύντομα θα οδηγηθούμε σε διανύσματα εισόδου, που ποτέ δεν είδε το δίκτυο για να εκπαιδευτεί. Στην πράξη τέτοια διανύσματα οδήγησαν σε πολύ κακής ποιότητας πρότυπα, με πολύ μεγάλο αριθμό νοτών, που σε καμία περίπτωση δεν μπορούν να χαρακτηριστούν ως μουσική. Η κίνηση έγινε σε κάθε κατεύθυνση μέχρι 30 βήματα μεγέθους 1.

Παράλληλα, όλη η διαδικασία υλοποιήθηκε δύο φορές, για δύο διαφορετικά αρχικά διανύσματα z_{ref} . Το ένα ήταν το κέντρο της κατανομής p_G , δηλαδή το

128-στατο μηδενικό διάνυσμα, ενώ το δεύτερο επιλέχθηκε τυχαία.

5.2.2 Αποτελέσματα

Τα αποτελέσματα εξετάστηκαν με τη βοήθεια των μετρικών που εξηγήθηκαν στο 4ο κεφάλαιο.

Chord Contribution / Chord Quality Results

Όσον αφορά την Chord Contribution, τα αποτελέσματα δεν παρουσιάζουν κάποια σαφήνεια ως προς την αλλαγή. Παρατηρούμε από το παράρτημα 6.1, ότι σε κάθε κατεύθυνση προκύπτουν συγκεκριμένες αλλαγές ως προς την κυρίαρχη συγχορδία σε κάθε μέτρο, παρόλαυτά δεν παρατηρείται κάποιο συγκεκριμένο μοτίβο αλλαγής.

Τα αποτελέσματα που εξάγουμε με τη συγκεκριμένη μετρική είναι περιγραφικά. Παρατηρούμε ξεκινώντας από το διάνυσμα z_{ref0} το οποίο είναι το διάνυσμα του κέντρου της κατανομής, η κυρίαρχη συγχορδία και στα 4 μέτρα είναι το D#, με μέτρο συνεισφοράς 1. Σε κάθε βήμα, ανά κατεύθυνση οι κυρίαρχες συγχορδίες σε κάθε μέτρο, διαφοροποιούνται από την D# σταδιακά. Επίσης υπάρχουν περιοχές, που σε ορισμένα μέτρα η μετρική μας δεν καταφέρνει να εντοπίσει κυρίαρχη συγχορδία. Αυτό σημαίνει ότι το συγκεκριμένο μέτρο δεν έχει καμία αρμονική δομή και εν γένει το πρότυπο δεν βγάζει κάποιο μουσικό νόημα. Παρατηρούμε ότι αυτή η συμπεριφορά εμφανίζεται στα μέτρα σε μεσαίο ή μεγάλο αριθμό βημάτων, συνήθως πάνω από τα 20 βήματα.

Η ποιότητα της αρμονικής συνοχής, υπολογίζεται ως ο μέσος όρος των συνεισφορών των κυρίαρχων συγχορδιών στα 4 μέτρα του προτύπου. Τα αποτελέσματα παρουσιάζονται στους πίνακες του παραρτήματος 6.2. Εν γένει παρατηρούμε ότι η συνολική ποιότητα της αρμονικής δομής πέφτει όσο απομακρυνόμαστε πολύ από το z_{ref0} . Ξεκινώντας δε από το δεύτερο σημείο αναφοράς z_{ref0} , η Chord Quality μετρική έχει μικρότερη τιμή. Στο σχήμα 41 παρουσιάζεται η τιμή της Chord Quality για 40 βήματα στις 10 PCA κατευθύνσεις, ξεκινώντας από τα δύο σημεία αναφοράς.

Τα αποτελέσματα, για τη συγκεκριμένη μετρική είναι μη ερμηνεύσιμα. Δηλαδή τελικά, δεν υπάρχει κάποια κατεύθυνση η οποία να μπορούμε να πούμε ότι διατηρεί σταθερή ή αυξάνει την αρμονική συνοχή με τον τρόπο που εμείς την ορίσαμε. Κινούμενοι και από τα δύο διαφορετικά σημεία εκίνησης, η τιμή της Chord Quality πέφτει σταδιακά, για κάθε κατεύθυνση με διαφορετικό τρόπο. Παράλληλα παρατηρούμε μικρές περιοχές αύξησης χωρίς όμως να μπορούμε να εντοπίσουμε κάποια συστηματικότητα σε αυτή.

Παράλληλα μπορούμε να εξετάσουμε αν η κάθε κατεύθυνση δρά με κάποιο συστηματικό τρόπο στα δύο διαφορετικά σημεία εκκίνησης εξετάζοντας την τιμή της μετρικής για τα δύο references κινούμενοι στην ίδια κατεύθυνση κάθε φορά όπως παρουσιάζεται στο σχήμα 44.

Παρατηρούμε κάποια συστηματικότητα ως προς τη μεταβολή της μετρικής, μόνο στην κατεύθυνση PCA3. Κατά τα άλλα, οι υπόλοιπες κατευθύνσεις δεν φαίνεται να δρουν κοινά στην μεταβολή της αρμονικής συνοχής στα παραγόμενα πρότυπα των δύο διαφορετικών σημείων εκκίνησης.

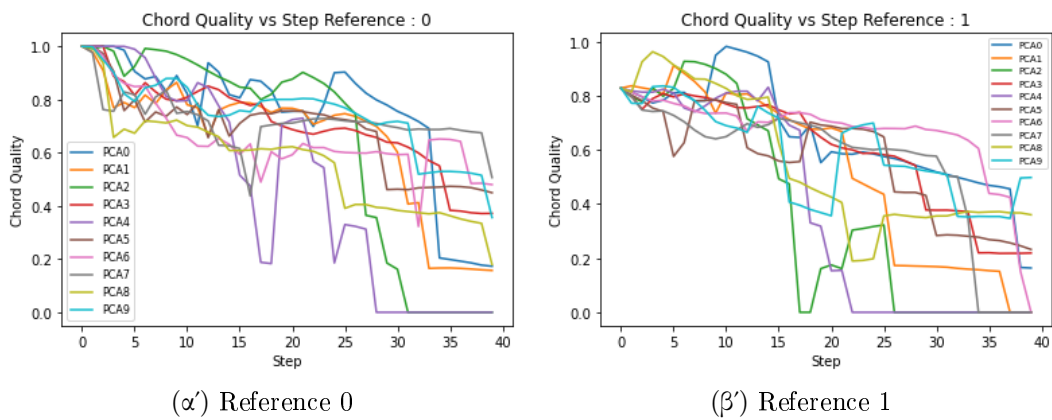
Qualified Note Rate Results

Εξετάζουμε στη συνέχεια, όλες τις κατευθύνσεις με τον ίδιο τρόπο για τη μετρική Qualified Note Rate. Τα αποτελέσματα παρουσιάζονται στο σχήμα 42 για κάθε reference ξεχωριστά. Παρατηρούμε ότι η μετρική παρουσιάζει μεγάλες διακυμάνσεις σε κάθε κατεύθυνση και στα δύο σημεία εκκίνησης γεγονός που δεν μας επιτρέπει να εξάγουμε κάποιο συμπέρασμα για την συστηματικότητα της αλλαγής.

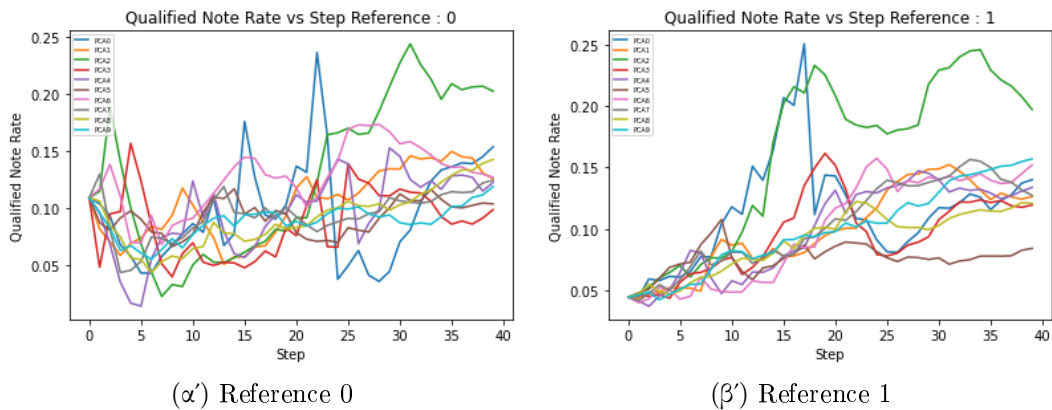
Στο σχήμα 45 παρουσιάζεται, ομοίως με προηγουμένως η αλλαγή της Qualified Note Rate σε κάθε κατεύθυνση ξεχωριστά, σε κοινό γράφημα για τα δύο references. Μπορούμε να δούμε ότι σε αρκετές περιπτώσεις η τιμή της QNR εμφανίζει κοινή συμπεριφορά για τα δύο σημεία αναφοράς. Σε αυτές τις περιπτώσεις, η κάθε κατεύθυνση αλλάζει την τιμή της συγκεκριμένης μετρικής με τον δικό της τρόπο, όπως φαίνεται στα γραφήματα. Κοινές συμπεριφορές μπορούμε να εντοπίσουμε στις PCA0, PCA8, PCA4 κατευθύνσεις. Παρόλαυτά είναι αδύνατον να εντοπίσουμε καί σε αυτή την περίπτωση κάποια κατεύθυνση που να αλλάζει την τιμή της QNR με τρόπο συστηματικό. Ενδιαφέρον παρουσιάζεται στην κατεύθυνση PCA2 στην οποία η τιμή της QNR αυξάνεται όσο αυξάνεται το βήμα. Εν γένει η κίνηση σε όλες τις κατευθύνσεις φαίνεται να παρουσιάζει αυξητική τάση στην QNR σε αντίθεση με την Chord Quality.

Polyphonic Rate Results

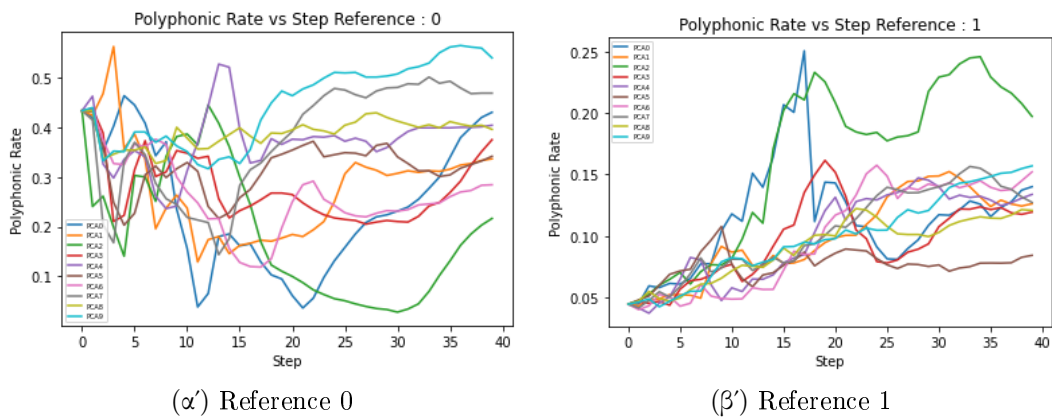
Τέλος στα σχήματα 43 και 46 εξετάζουμε τις παραγόμενες κατευθύνσεις με την μετρική Polyphonic Rate. Παρατηρώντας το γράφημα της PCA8 στο σχήμα 46, μπορούμε να δούμε ότι η τιμή της μετρικής και για τα δύο σημεία εκκίνησης είναι σχεδόν σταθερή με μία πολύ μικρή αυξητική τάση. Μπορούμε δηλαδή να πούμε ότι η PCA8 κατεύθυνση, παραβλέποντας τις μικρές διακυμάνσεις που εμφανίζονται κυρίως στο πρώτο σημείο εκκίνησης (μπλέ γραμμή) είναι υπεύθυνη για την διατήρηση της τιμής της πολυφωνικότητας των προτύπων. Για την επικύρωση αυτού του συμπεράσματος, θα χρειαστεί να γίνουν περαιτέρω πειράματα στη συγκεκριμένη κατεύθυνση, από μεγάλο αριθμό διαφορετικών σημείων εκκίνησης έτσι ώστε να μπορεί να εντοπιστεί η συγκεκριμένη συμπεριφορά αρκετές φορές. Κατά τα άλλα, όσον αφορά τη δράση των υπόλοιπων κατευθύνσεων, τα γραφήματα παρουσιάζουν μεγάλες διακυμάνσεις και είναι αρκετά δύσκολο να εντοπισθεί μια συστηματική αλλαγή.



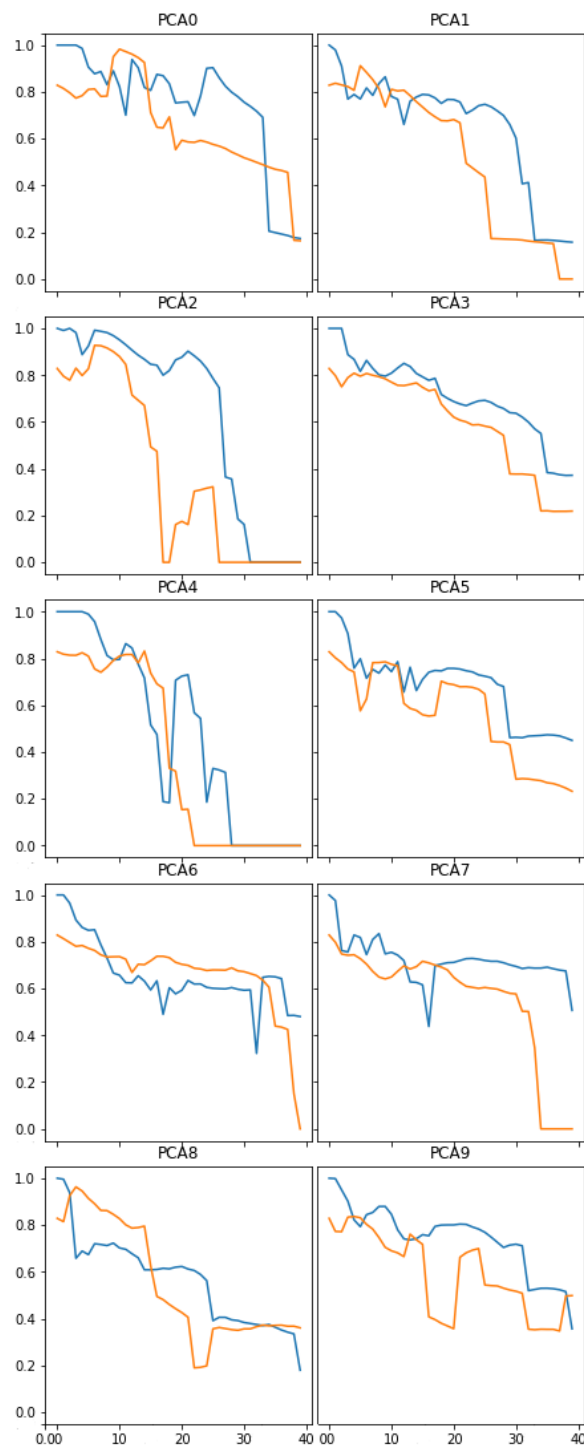
Σχήμα 41: Τα αποτελέσματα της μεθόδου PCA walking για 40 βήματα σε κάθε μία από τις 10 PCA κατευθύνσεις για τη μετρική Chord Quality



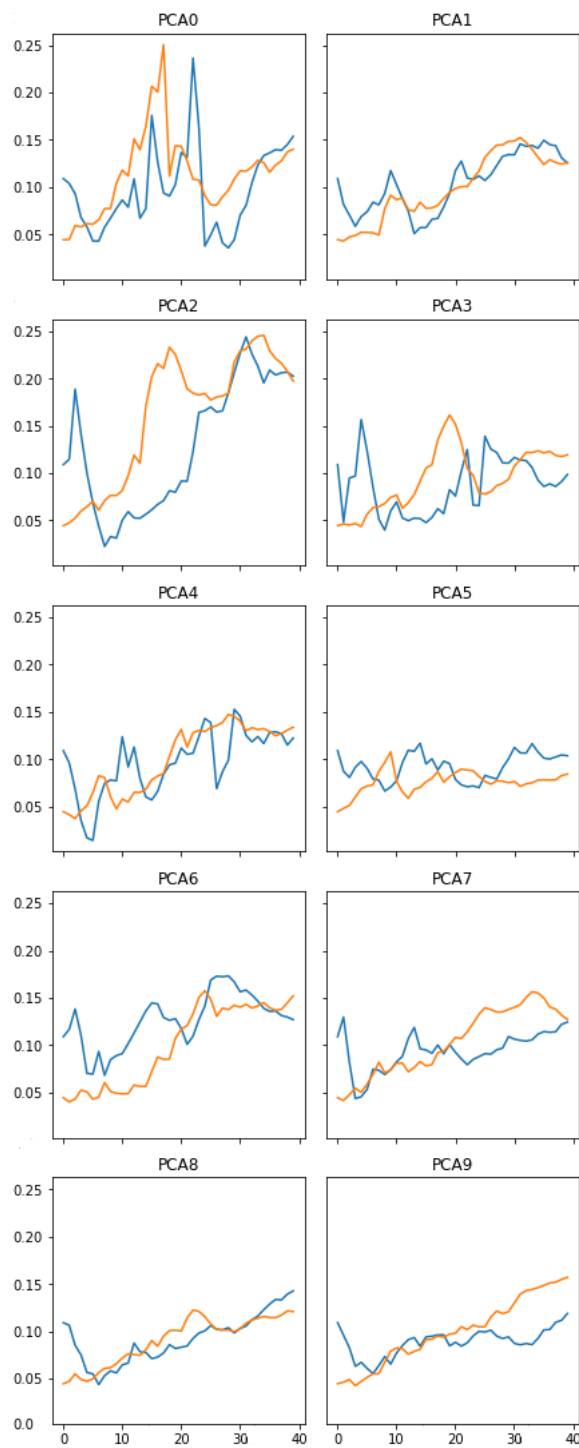
Σχήμα 42: Τα αποτελέσματα της μεθόδου PCA walking για 40 βήματα σε κάθε μία από τις 10 PCA κατευθύνσεις για τη μετρική Qualified Note Rate



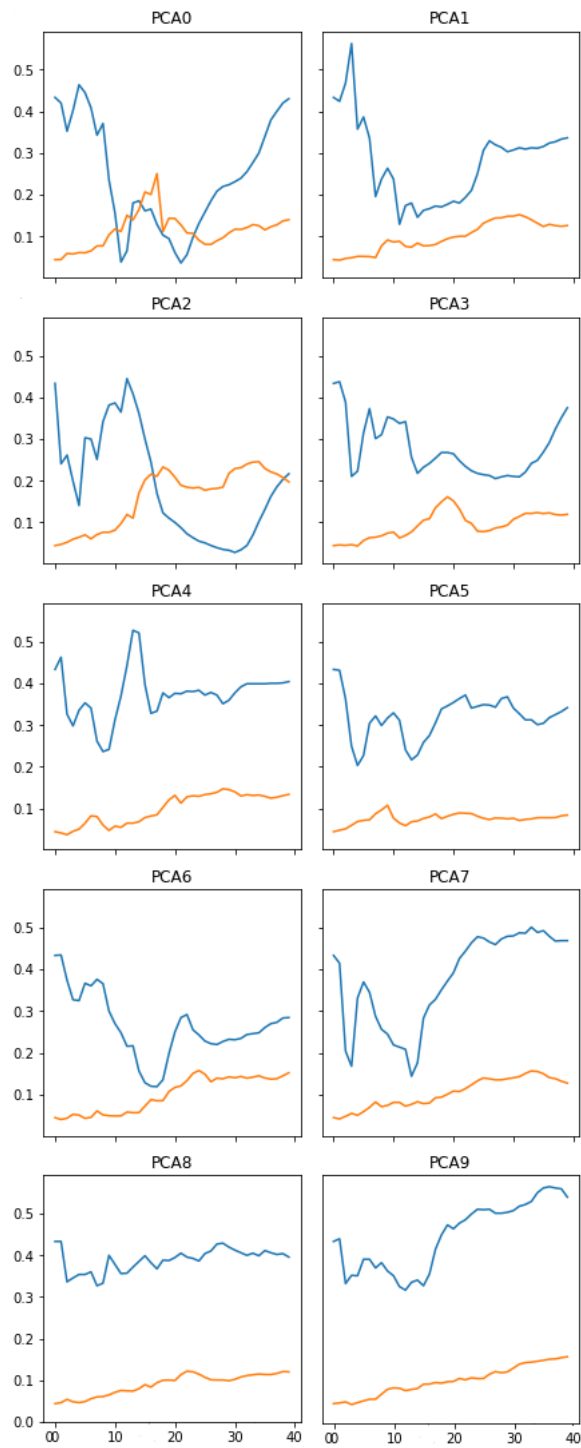
Σχήμα 43: Τα αποτελέσματα της μεθόδου PCA walking για 40 βήματα σε κάθε μία από τις 10 PCA κατευθύνσεις για τη μετρική Polyphonic Rate



Σχήμα 44: Η μεταβολή της Chord Quality σε κοινό γράφημα στα δύο σημεία εκκίνησης



Σχήμα 45: Η μεταβολή της *Qualified Note Rate* σε κοινό γράφημα στα δύο σημεία εκκίνησης



Σχήμα 46: Η μεταβολή της Polyphonic Rate σε κοινό γράφημα στα δύο σημεία εκκίνησης

5.2.3 Συμπεράσματα - Σχολιασμός

Με τη μέθοδο PCA Walking, καταφέραμε να βρούμε τις κατευθύνσεις μέγιστης διασποράς που προκύπτουν από την δράση του πρώτου block του MuseGAN, σε 1 εκατομμύριο κανονικά δειγματοληπτημένων διανυσμάτων του latent space. Η κίνηση στις κατευθύνσεις αυτές ξεχωριστά, αναμέναμε να αποτυπώνουν σημαντικές αλλαγές στα παραγόμενα πρότυπα, τις οποίες όμως δεν καταφέραμε να τις ονοματίσουμε. Οι τρεις μετρικές που δοκιμάστηκαν, απέτυχαν να μας δώσουν κάποια σαφή εικόνα για το ποιά επι μέρους αλλαγή επιτυγχάνει η κάθε κατεύθυνση.

Το γεγονός αυτό, ενδεχομένως να οφείλεται στην αδυναμία του Generator του MuseGAN να παράγει πειστικά πρότυπα. Οι Erik Härkönen, et. al εφαρμόσαν τη συγκεκριμένη μέθοδο σε ένα δίκτυο που καταφέρνει να παράγει εικόνες εκπληκτικής ποιότητας, που δύσκολα τις ξεχωρίζει κανείς από πραγματικές φωτογραφίες. Είναι λογικό λοιπόν, τα πρώτα layers του Generator του BigGAN, να μαθαίνουν χαρακτηριστικά, τα οποία όχι μόνο θα είναι σημαντικά στην παραγωγή των προτύπων, αλλά ενδεχομένως να έχουν και κάποιο διαισθητικό νόημα. Συνεπώς και η αξιολόγηση της αλλαγής σε κάθε κατεύθυνση θα γίνεται με τρόπο φυσικό και όπως παρατηρούμε από το συγκεκριμένο άρθρο φαίνεται με το μάτι!

Εν αντιθέσει, στην πλειοψηφία των παραγόμενων προτύπων του MuseGAN είναι εμφανής η αδυναμία του δικτύου να παράγει αληθοφανείς συνθέσεις. Οι παραγόμενες κατευθύνσεις από τη μία θα εμπεριέχουν και την αλλαγή των σημαντικότερων χαρακτηριστικών που το δίκτυο έμαθε στα αρχικά layers. Παρόλαυτά αυτές οι κατευθύνσεις δεν είναι απαραίτητο απεικονίζουν αλλαγές με διαισθητικό νόημα. Οι μετρικές που υλοποιήθηκαν και χρησιμοποιήθηκαν, που κατά την αποψη μας συνοψίζουν σε μεγάλο βαθμό την ποιότητα της σύνθεσης (από την ανθρώπινη σκοπιά) απέτυχαν να εντοπίσουν αλλαγές που να μπορούν να χαρακτηρίσουν τη δράση της κάθε κατεύθυνσης.

Οι επί μέρους αλλαγές που επιφέρει η κίνηση σε κάθε κατεύθυνση, ενδεχομένως να απεικονίζονται από μετρικές, οι οποίες όμως δεν θα έχουν κάποιο μουσικό νόημα. Όπως εξηγήσαμε και στα προηγούμενα κεφάλαια, η μέθοδος είναι unsupervised και ενώ θα δώσει κατευθύνσεις αλλαγών, η καθαρότητα των αλλαγών αυτών εξαρτάται από την ποιότητα των δεδομένων εκπαίδευσης και την ποιότητα του Generator.

Η PCA8 κατεύθυνση φαίνεται να διατηρεί την πολυφωνικότητα των προτύπων σταθερή, εξεταζόμενη από δύο σημεία εκκίνησης. Εξετάζοντας επίσης τη συμπεριφορά των άλλων δύο μετρικών σε αυτή τη κατεύθυνση παρατηρούμε τα εξής. Η QNR μετρική αυξάνεται με όμοιο κατά μέσο όρο τρόπο για τα δύο σημεία εκκίνησης. Συνεπώς μπορούμε να σκεφτούμε ότι αυτή η κατεύθυνση παράγει πρότυπα με σταθερή κατά μέσο όρο πολυφωνικότητα και με διάρκεια

νοτών μεγαλύτερη από 1 timestamp, δηλαδή με μικρότερο αριθμό artefacts στις παραγόμενες συνθέσεις. Η τιμή παρόλαυτά της Chord Quality πέφτει απότομα σε μικρό αριθμό βημάτων και για τα δύο refernces. Αυτό πρακτικά σημαίνει ότι ενώ τα κομμάτια που παράγονται θα έχουν αντιληπτής διάρκειας νότες, οι οποίες μάλιστα θα συνηχούν σε σταθερό ποσοστό, θα χάνουν γρήγορα και απότομα την αρμονική τους συνοχή οδηγώντας σε κακής ποιότητας μουσικές συνθέσεις.

5.3 Υλοποίηση της Μεθόδου Latent Walking

Σε αυτό το σημείο εξηγείται αναλυτικά ο τρόπος υλοποίησης της μεθόδου που παρουσιάστηκε στο κεφάλαιο 3.3.2 καθώς και στο άρθρο [35].

5.3.1 Διαδικασία

Δειγματοληψία του συνόλου Z

Το πρώτο βήμα της διαδικασίας είναι η κατασκευή του συνόλου Z των διανυσμάτων του latent space. Γιαυτό το λόγο κατασκευάστηκε μία συνάρτηση η οποία δειγματοληπτεί 50 τυχαία διανύσματα διαστάσεων (1,128) από την truncated κανονική κατανομή 128 διαστάσεων $N(-2,2)$. Η δειγματοληψία των 50 διανυσμάτων έγινε σε 1000 διαφορετικά seeds. Συνεπώς το σύνολο Z θα αποτελείται από 50.000 διανύσματα εισόδου.

Παραγωγή Προτύπων και Αξιολόγηση

Τα διανύσματα εισόδου του Z τροφοδοτήθηκαν στον Generator του MuseGAN έτσι ώστε να παραχθούν 50.000 διαφορετικά μουσικά πρότυπα. Στη συνέχεια κάθε μουσικό πρότυπο αξιολογήθηκε από τις μετρικές note count, polyphonic rate, qualified note rate και chord quality.

Κατασκευάζονται λοιπόν 4 διαφορετικά datasets που το κάθε ένα περιέχει 50.000 ζευγάρια των διανυσμάτων εισόδου z_i με την τιμή της αντίστοιχης μετρικής m_i , (z_i, m_i) .

Και οι 4 μετρικές που επιλέχθηκαν, όπως εξηγήσαμε στο κεφάλαιο 4, επιστρέφουν αριθμό, ποσοτικοποιώντας το χαρακτηριστικό που εξετάζουν. Από κάθε dataset που κατασκευάστηκε, κρατάμε το 3% των προτύπων τα οποία έχουν τις μικρότερες τιμές της αντίστοιχης μετρικής και το 3% που αντιστοιχεί στις μεγαλύτερες τιμές μετρικής. Με άλλα λόγια διατάσσουμε τα διανύσματα εισόδου από την μικρότερη τιμή μετρικής στην μεγαλύτερη. Κρατάμε τα 1500 διανύσματα εισόδου που χαρακτηρίζονται από τις μικρότερες τιμές της μετρικής καθώς και τα 1500 διανύσματα εισόδου που οδηγούν σε μεγάλες τιμές της εξεταζόμενης μετρικής.

Στη συνέχεια δημιουργούμε τα classification labels y . Τα 1500 διανύσματα εισόδου μικρής τιμής μετρικής θα έχουν $y_{low} = 0$ ενώ αντίστοιχα τα άλλα $y_{high} = 1$. Η διαδικασία αυτή θα γίνει 4 φορές ξεχωριστά, για κάθε εξεταζόμενη μετρική κατασκευάζοντας 4 τελικά classification datasets με τιμές (z_i, y_i) . Στην πραγματικότητα κάθε μία από τις 128 διαστάσεις των z_i θα χρησιμοποιείται ως feature από τον αλγόριθμο κατηγοριοποίησης. Μια άλλη παρατήρηση είναι ότι το δείγμα εκπαίδευσης είναι balanced περιέχοντας τον ίδιο αριθμό δεδομένων για τις δύο κλάσεις (1500).

Classification - SVMs

Χρησιμοποιούμε την αντίστοιχη βιβλιοθήκη της sklearn για να εκπαιδεύσουμε τον αλγόριθμο Linear-Support Vector Machines. Για να πετύχουμε σύγκλιση αυξάνουμε τον αριθμό των iterations στις 7000.

Ο αλγόριθμος εκπαιδεύεται στα 4 datasets των 4 μετρικών ξεχωριστά, συγκλίνει σε όλες τις περιπτώσεις και τοποθετεί το βέλτιστο υπερεπίπεδο διαχωρισμού κάθε φορά. Τα training accuracy scores παρουσιάζονται στον πίνακα 1:

SVM Training Accuracy of 4 Metrics	
Metric	Accuracy
Note Count	0.8103
Polyphonic Rate	0.7516
Qualified Note Rate	0.7996
Chord Quality	0.6946

Πίνακας 1: To Classification Accuracy για τα Datasets των 4 μετρικών

Η απόδοση των ταξινομητών θεωρείται σχετικά καλή και έτσι για κάθε ένα από τα 4 μοντέλα μπορούμε να εξάγουμε το διάνυσμα του υπερεπίπεδου που έχει τοποθετήσει, καθώς και τον σταθερό όρο με τη χρησιμοποίηση αντίστοιχων μεθόδων της sklearn. Το διάνυσμα που παίρνουμε από την sklearn είναι κάθετο στο υπερεπίπεδο και έχει φορά στην κατεύθυνση της κλάσης που $y=1$.

Latent Walking

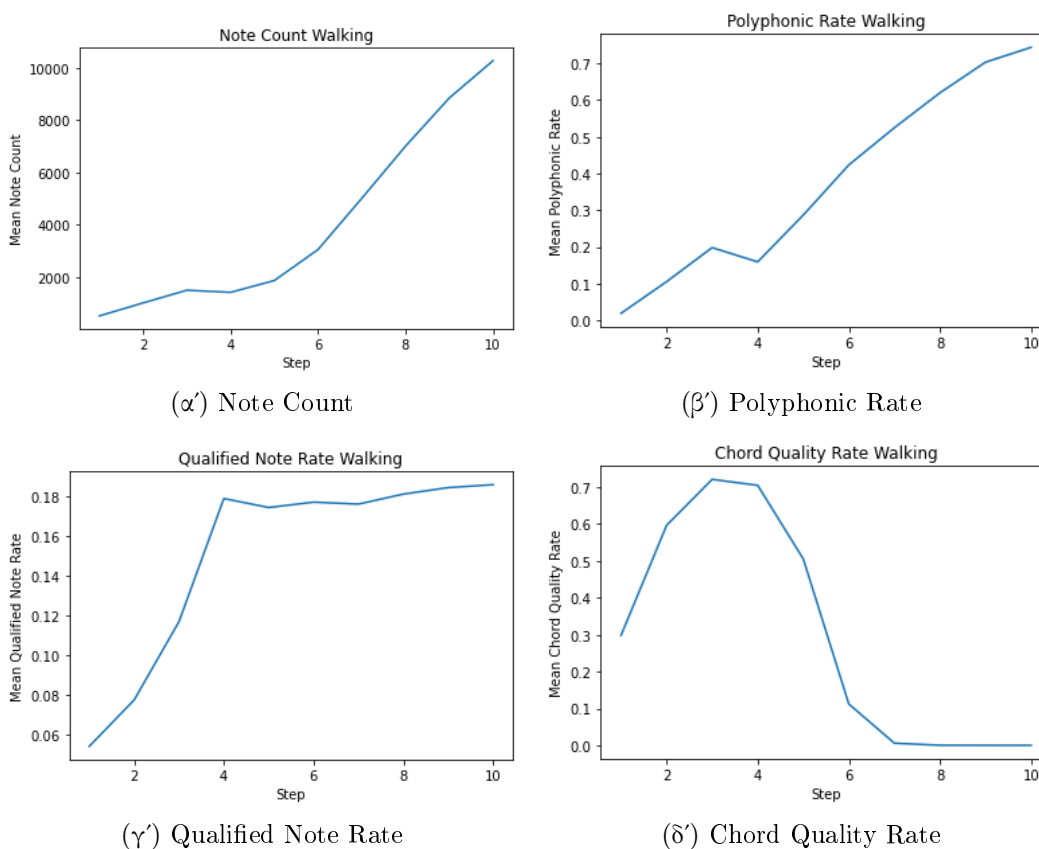
Το τελευταίο βήμα της μεθόδου είναι η επιλογή των διανυσμάτων αναφοράς στα οποία θα ξεκινήσει το Walking. Επιλέξαμε το σύνολο των διανυσμάτων που αντιστοιχούν στις μικρές τιμές της εξεταζόμενης μετρικής, και από αυτά κρατήσαμε μόνο όσα είχαν κατηγοριοποιηθεί σωστά από τον αλγόριθμο SVM.

Ξεκινώντας λοιπόν από κάθε ένα από αυτά τα διανύσματα, χρησιμοποιούμε τη σχέση (27) κάνοντας 10 βήματα μεγέθους $\beta=1$.

5.3.2 Αποτελέσματα

Χρησιμοποιώντας την παραπάνω μέθοδο, δεδομένης της μετρικής που εξετάζουμε, ξεκινήσαμε το walking από το σύνολο των διανυσμάτων εισόδου, τα οποία οδήγησαν σε πρότυπα μικρής τιμής της μετρικής. Κάθε βήμα στην ευρισκόμενη κατεύθυνση θα μας οδηγεί ολοένα και πιο κοντά στο διαχωριστικό υπερπίπεδο και σε κάποιο αριθμό βημάτων το παραγόμενο z θα βρίσκεται στην περιοχή των μεγάλων τιμών μετρικής. Προφανώς το βήμα που θα συμβεί αυτό δεν είναι σταθερό για όλα τα πρότυπα, αλλά εν γένει εξαρτάται από την απόσταση του κάθε αρχικού σημείου εκκίνησης από το υπερπίπεδο.

Σε κάθε περίπτωση, κατά μέσο όρο για όλα τα πρότυπα που χρησιμοποιούνται ως σημείο εκκίνησης, θα θέλαμε σε κάθε βήμα να αυξάνεται η τιμή της εξεταζόμενης μετρικής. Τα αποτελέσματα είναι ικανοποιητικά και παρουσιάζονται στα παρακάτω διαγράμματα.



Σχήμα 47: Τα αποτελέσματα της μεθόδου Latent Walking στις τέσσερις δοκιμαζόμενες μετρικές

5.3.3 Συμπεράσματα - Σχολιασμός

Με τη χρήση της μεθόδου Latent Walking καταφέραμε να βρούμε κατευθύνσεις που αλλάζουν με συστηματικό τρόπο συγκεκριμένα χαρακτηριστικά των παραγόμενων μουσικών προτύπων.

Όσον αφορά τον αριθμό των νοτών των προτύπων, φαίνεται ότι ο αλγόριθμος των SVMs διαχώρισε σε ικανοποιητικό βαθμό τις δύο κλάσεις. Ο latent space φαίνεται να εμφανίζει εν γένει γραμμικότητα ως προς αυτή τη μετρική και έτσι η καμπύλη (α') του σήματος 47 είναι σχεδόν σε όλα τα βήματα αύξουσα.

Αυτό σημαίνει ότι καθορίστηκε με επιτυχία μία κατευθυνση στον 128 διάστατο χώρο εισόδου, όπου ξεκινώντας από έναν μεγάλο αριθμό διανυσμάτων εκκίνησης και κινούμενοι σε αυτή, καταφέρνουμε συνεχώς να αυξανουμε τις συνολικές νότες των παραγόμενων προτύπων.

Ομοίως στις περιπτώσεις των Polyphonic Rate και Qualified Note Rate οι αντίστοιχες καμπύλες (β') και (γ') εμφανίζουν αυξητική τάση και τα classification accuracies είναι ικανοποιητικά. Στην καμπύλη της πολυφωνικότητας βλέπουμε μία μικρή περιοχή μείωσης ή οποία μπορούμε να θεωρήσουμε ότι οφείλεται στο μικρότερο σκορ κατηγοριοποίησης σε σχέση με τις υπόλοιπες δύο που προαναφέρθηκαν.

Εν γένει ένα μικρό σκορ διαχωρισμού σε κάποια μετρική σημαίνει ότι ο χώρος δεν είναι γραμμικά διαχωρίσιμος για αυτή τη μετρική. Ο ταξινομητής που χρησιμοποιήσαμε είναι γραμμικός, συνεπώς θα εμφανίσει αδυναμία στην κατηγοριοποίηση. Παράλληλα όσο μεγαλώνει το δείγμα του συνόλου Z τόσο περισσότερο μπορούμε να είμαστε περισσότερο σίγουροι για τη ισχύ της μεθόδου σε όλον τον latent space.

Παρατηρούμε ότι η αδυναμία καλού διαχωρισμού απεικονίζεται στο σχήμα (δ') της μετρικής της αρμονικής συνοχής. Ενώ το Chord Quality Rate αρχικά αυξάνεται, από το βήμα 4 και μετά παρατηρείται πτώση, μέχρι το βήμα 7 στο οποίο μηδενίζεται. Παρόλαυτά η περιοχή αύξησης υπάρχει και μπορεί να χρησιμοποιηθεί για να παράγει πρότυπα ολοένα αυξανόμενης αρμονικής συνοχής, μικραίνοντας απλώς το βήμα του walking, κάτι που ήδη έχουμε αρχίσει να εξετάζουμε παρόλαυτά δεν θα συμπεριληφθεί στην παρούσα εργασία.

Παράλληλα, η μέθοδος αυτή μας δίνει επιθυμητές κατευθύνσεις αύξησης της κάθε μετρικής ξεχωριστά χωρίς να εξετάζει τι συμβαίνει στις υπόλοιπες. Συνεπώς ως μελλοντική εργασία, θα είχε ενδιαφέρον η εύρεση μίας μεθόδου, βασισμένη σε αυτή, η οποία θα μπορούσε να επιτυγχάνει την αλλαγή του συνδιασμού των εξεταζόμενων μετρικών κατά επιθυμητό τρόπο.

Τέλος το γεγονός ότι η μέθοδος είναι supervised ως προς την επιλεγόμενη μετρική, παρέχει τη δυνατότητα υλοποίησης οποιασδήποτε μετρικής θεωρεί ο ερευνητής ότι είναι σημαντική για την αξιολόγηση της μουσικής. Η μόνη

προυπόθεση είναι η μετρική να επιστρέφει διατάξιμο μέγεθος, έτσι ώστε ο αλγόριθμος να μπορεί να διαχωρίσει τον latent space σε περιοχές μικρών και μεγάλων τιμών της μετρικής.

6 Παράρτημα

6.1 Πίνακες Αποτελεσμάτων της Chord Metric

Reference : 0 , PCA direction : 0			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',0.98)	('D#',1.0)	('D#',1.0)	('D#',0.963)
('D#',0.827)	('D#',0.996)	('D#',1.0)	('D#',0.801)
('D#',0.771)	('D#',0.985)	('D#',1.0)	('D#',0.754)
('D#',0.827)	('D#',0.908)	('D#',0.993)	('D#',0.82)
('D#m',0.831)	('D#m',0.84)	('D#',0.864)	('F#',0.788)
('D#m',0.929)	('B',0.922)	('D#m',0.822)	('D#m',0.888)
('D#m',0.953)	('B',0.996)	('D#m',0.488)	('D#m',0.852)
('D#m',0.73)	('B',0.995)	('D#m',0.324)	('D',0.753)
('D',0.895)	('Bm',0.95)	('Bm',0.978)	('D',0.931)
('D',0.979)	('Bm',0.832)	('Bm',0.836)	('D',0.966)
('D',0.982)	('Em',0.696)	('Bm',0.619)	('D',0.978)
('D',1.0)	('Em',0.764)	('Bm',0.462)	('D',1.0)
('D',1.0)	('Em',0.747)	('G',0.753)	('D',1.0)
('D',1.0)	('Em',0.735)	('G',0.804)	('D',0.937)
('D',0.935)	('G',0.711)	('G',0.8)	('D',0.893)
('D',0.913)	('G',0.654)	('D',0.488)	('D',0.956)
('D',0.955)	('G',0.6)	('D',0.468)	('D',1.0)
('D',1.0)	('A',0.605)	('G',0.432)	('D',0.994)
('A#',0.889)	('A',0.818)	('D',0.205)	('Dm',0.883)
('A#',0.97)	('A#m',0.717)	('A#',0.5)	('Dm',0.941)
('A#',1.0)	('A#',0.918)	('A#',0.705)	('A#',0.98)
('A#',0.963)	('D#',0.946)	('A#',0.74)	('A#',0.965)
('A#',0.861)	('D#',0.946)	('A#',0.734)	('A#',0.902)
('A#',0.782)	('D#',0.924)	('A#',0.755)	('A#',0.837)
('A#',0.746)	('D#m',0.9)	('A#',0.767)	('A#',0.782)

Reference : 0 , PCA direction : 1			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',0.946)	('D#',0.971)	('D#',1.0)
('D#',0.974)	('D#',0.833)	('D#',0.866)	('D#',0.968)
('D#',0.814)	('Gm',0.713)	('D#',0.747)	('D#',0.803)
('D#',0.82)	('A#',0.952)	('D#',0.597)	('D#',0.787)
('D#',0.81)	('A#',0.981)	('D#',0.505)	('D#',0.783)
('D#',0.786)	('A#',0.986)	('Gm',0.734)	('D#',0.762)
('D#',0.757)	('A#',0.983)	('Gm',0.664)	('D#',0.737)
('D#',0.84)	('A#',0.985)	('Gm',0.682)	('D#',0.837)
('D#',0.869)	('A#',0.986)	('Gm',0.755)	('D#',0.848)
('Gm',0.679)	('A#',0.994)	('Gm',0.785)	('Gm',0.664)
('C',0.643)	('A#',0.993)	('Gm',0.778)	('C',0.661)
('C',0.752)	('A#',0.994)	('C',0.172)	('C',0.726)
('C',0.977)	('A#',0.983)	('C',0.154)	('C',0.927)
('C',1.0)	('A#',0.981)	('C',0.172)	('C',0.959)
('C',1.0)	('A#',0.969)	('C',0.221)	('C',0.967)
('C',1.0)	('A#',0.947)	('C',0.233)	('C',0.968)
('C',1.0)	('A#',0.871)	('C',0.259)	('C',0.967)
('C',1.0)	('Dm',0.762)	('C',0.284)	('C',0.958)
('C',1.0)	('Dm',0.762)	('Em',0.344)	('C',0.965)
('C',1.0)	('Dm',0.793)	('C',0.319)	('C',0.956)
('C',0.96)	('Dm',0.832)	('C',0.315)	('C',0.92)
('C',0.912)	('Dm',0.729)	('C',0.305)	('C',0.883)
('C',0.898)	('F',0.706)	('C',0.422)	('C',0.862)
('C',0.903)	('F',0.715)	('C',0.472)	('C',0.874)
('C',0.906)	('F',0.729)	('C',0.474)	('C',0.879)
('C',0.903)	('F',0.708)	('C',0.461)	('C',0.874)
('C',0.874)	('F',0.689)	('C',0.469)	('C',0.843)
('C',0.84)	('F',0.681)	('C',0.471)	('C',0.804)
('C',0.771)	('F',0.667)	('Cm',0.465)	('C',0.739)

Reference : 0 , PCA direction : 2			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',0.997)	('D#',0.984)	('D#',0.984)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',0.983)	('D#',0.966)	('D#',0.979)
('D#',0.87)	('Gm',0.907)	('Gm',0.922)	('D#',0.852)
('Gm',0.891)	('Gm',1.0)	('Gm',0.976)	('Gm',0.832)
('Gm',0.988)	('Gm',1.0)	('Gm',0.993)	('Gm',0.988)
('Gm',1.0)	('Gm',0.959)	('G',1.0)	('Gm',0.991)
('G',1.0)	('G',0.936)	('G',1.0)	('G',0.992)
('G',0.978)	('G',0.902)	('G',0.996)	('G',1.0)
('G',0.967)	('G',0.884)	('G',0.953)	('G',1.0)
('G',0.9)	('G',0.877)	('G',0.943)	('G',1.0)
('G',0.856)	('G',0.876)	('G',0.947)	('G',0.949)
('Em',0.824)	('G',0.863)	('G',0.958)	('G',0.897)
('Em',0.806)	('G',0.857)	('G',0.968)	('G',0.84)
('Em',0.787)	('G',0.852)	('G',0.974)	('G',0.774)
('Em',0.806)	('G',0.833)	('G',0.983)	('Em',0.748)
('Em',0.865)	('G',0.826)	('Em',0.725)	('Em',0.784)
('Em',0.9)	('G',0.787)	('Em',0.762)	('Em',0.831)
('Em',0.918)	('G',0.817)	('Em',0.862)	('Em',0.864)
('Em',0.906)	('G',0.82)	('Em',0.876)	('Em',0.905)
('Em',0.893)	('G',0.838)	('G',0.968)	('Em',0.911)
('Em',0.867)	('G',0.884)	('G',0.899)	('Em',0.881)
('Em',0.815)	('G',0.92)	('G',0.858)	('Em',0.848)
('Em',0.774)	('G',0.924)	('G',0.813)	('Em',0.804)
('G',0.707)	('G',0.924)	('G',0.776)	('Em',0.738)
('G',0.667)	('G',0.867)	('G',0.752)	('Em',0.698)
('NoChord',0)	('G',0.792)	('NoChord',0)	('G',0.667)
('NoChord',0)	('Gm',0.763)	('NoChord',0)	('G',0.662)
('NoChord',0)	('Gm',0.741)	('NoChord',0)	('NoChord',0)

Reference : 0 , PCA direction : 3			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',0.949)	('D#',0.855)	('D#',0.837)	('D#',0.911)
('Gm',0.993)	('Gm',0.759)	('Gm',0.729)	('Gm',0.989)
('G',0.931)	('Gm',0.689)	('Gm',0.677)	('G',0.968)
('G',1.0)	('Gm',0.765)	('Gm',0.721)	('G',0.967)
('G',1.0)	('Gm',0.692)	('Gm',0.645)	('G',0.979)
('G',1.0)	('G',0.679)	('G',0.548)	('G',0.979)
('G',1.0)	('G',0.664)	('Em',0.538)	('G',0.981)
('G',1.0)	('C',0.706)	('Em',0.549)	('G',0.98)
('G',1.0)	('C',0.794)	('Em',0.55)	('G',0.979)
('G',0.981)	('C',0.889)	('Em',0.583)	('G',0.949)
('G',0.922)	('C',0.949)	('Em',0.58)	('G',0.897)
('G',0.882)	('C',0.926)	('Em',0.572)	('G',0.85)
('G',0.86)	('C',0.892)	('Em',0.576)	('G',0.843)
('G',0.852)	('C',0.86)	('Em',0.562)	('G',0.838)
('G',0.841)	('C',0.87)	('Em',0.612)	('G',0.824)
('G',0.826)	('C',0.856)	('G',0.363)	('G',0.826)
('G',0.818)	('C',0.8)	('G',0.362)	('G',0.824)
('G',0.811)	('C',0.783)	('G',0.337)	('G',0.818)
('G',0.814)	('C',0.768)	('G',0.319)	('G',0.807)
('G',0.816)	('C',0.764)	('G',0.275)	('G',0.824)
('G',0.841)	('C',0.776)	('G',0.277)	('G',0.829)
('G',0.852)	('C',0.795)	('G',0.265)	('G',0.848)
('G',0.86)	('C',0.81)	('G',0.264)	('G',0.836)
('G',0.841)	('C',0.811)	('G',0.257)	('G',0.825)
('G',0.825)	('C',0.806)	('G',0.237)	('G',0.802)
('G',0.808)	('C',0.809)	('G',0.225)	('G',0.788)
('G',0.782)	('C',0.8)	('G',0.223)	('G',0.754)

Reference : 0 , PCA direction : 4			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',0.989)	('D#',1.0)	('D#',1.0)	('D#',0.969)
('D#',0.948)	('D#',1.0)	('D#',0.986)	('D#',0.896)
('D#',0.878)	('D#',0.973)	('D#',0.92)	('D#',0.754)
('D#',0.783)	('D#',0.943)	('D#',0.831)	('D#',0.693)
('D#',0.757)	('D#',0.868)	('D#',0.796)	('Gm',0.757)
('Gm',0.771)	('D#',0.825)	('D#',0.769)	('Gm',0.819)
('Gm',0.878)	('Gm',0.843)	('Gm',0.856)	('Gm',0.877)
('Gm',0.879)	('Gm',0.784)	('Gm',0.856)	('Gm',0.86)
('Gm',0.826)	('Gm',0.726)	('Gm',0.8)	('Gm',0.752)
('Gm',0.741)	('Gm',0.73)	('Gm',0.715)	('Gm',0.68)
('Gm',0.688)	('Gm',0.697)	('Gm',0.678)	('NoChord',0)
('Gm',0.601)	('C',0.713)	('Gm',0.587)	('NoChord',0)
('NoChord',0)	('C',0.752)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('C',0.733)	('NoChord',0)	('NoChord',0)
('C',0.603)	('C',0.795)	('C',0.783)	('C',0.646)
('C',0.636)	('C',0.816)	('C',0.807)	('C',0.639)
('C',0.666)	('C',0.836)	('C',0.813)	('C',0.606)
('C',0.656)	('C',0.82)	('C',0.795)	('NoChord',0)
('C',0.617)	('Em',0.791)	('Em',0.768)	('NoChord',0)
('NoChord',0)	('Em',0.743)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Em',0.719)	('NoChord',0)	('C',0.601)
('NoChord',0)	('Em',0.691)	('NoChord',0)	('Em',0.602)
('NoChord',0)	('Em',0.64)	('NoChord',0)	('Em',0.611)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)

Reference : 0 , PCA direction : 5			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',0.901)	('D#',0.992)	('D#',1.0)	('D#',1.0)
('D#',0.708)	('D#',0.924)	('D#',1.0)	('D#',0.994)
('A#',0.858)	('Gm',0.703)	('Gm',0.5)	('D#',0.973)
('A#',0.946)	('A#',0.723)	('Gm',0.599)	('D#',0.927)
('A#',1.0)	('Dm',0.647)	('Gm',0.583)	('D#',0.632)
('A#',0.988)	('Dm',0.662)	('Gm',0.586)	('G#',0.777)
('A#',0.97)	('Dm',0.707)	('Gm',0.491)	('G#',0.781)
('A#',0.982)	('D',0.834)	('Gm',0.451)	('Fm',0.821)
('A#',0.95)	('D',0.98)	('A#',0.161)	('Fm',0.884)
('A#',0.853)	('D',0.992)	('Dm',0.387)	('Fm',0.913)
('A#',0.681)	('D',0.996)	('A#',0.185)	('Fm',0.765)
('F',0.8)	('D',0.996)	('Dm',0.37)	('F',0.884)
('F',0.758)	('D',0.991)	('F',0.199)	('F',0.703)
('Am',0.859)	('D',0.956)	('F',0.232)	('A',0.795)
('A',0.968)	('D',0.902)	('F#m',0.246)	('A',0.848)
('A',0.991)	('D',0.885)	('Am',0.239)	('A',0.88)
('A',1.0)	('D',0.875)	('Am',0.237)	('A',0.875)
('A',1.0)	('D',0.859)	('Am',0.258)	('A',0.913)
('A',1.0)	('D',0.837)	('Am',0.263)	('A',0.93)
('A',0.992)	('D',0.816)	('Am',0.282)	('A',0.929)
('A',0.971)	('D',0.79)	('Am',0.296)	('A',0.933)
('A',0.937)	('D',0.771)	('Am',0.328)	('A',0.934)
('A',0.898)	('D',0.732)	('Am',0.352)	('A',0.937)
('A',0.882)	('D',0.695)	('Am',0.374)	('A',0.946)
('A',0.866)	('D',0.671)	('Am',0.384)	('A',0.95)
('A',0.857)	('D',0.634)	('F#m',0.325)	('A',0.943)
('A',0.859)	('D',0.605)	('F#m',0.333)	('A',0.923)
('A',0.858)	('NoChord',0)	('A',0.08)	('A',0.908)

Reference : 0 , PCA direction : 6			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',0.966)	('D#',0.989)	('D#',0.905)
('D#',1.0)	('D#',0.877)	('D#',0.917)	('D#',0.778)
('D#',1.0)	('D#',0.768)	('D#',0.84)	('A#',0.837)
('D#',1.0)	('D#',0.724)	('D#',0.738)	('A#',0.933)
('D#',1.0)	('A#',0.801)	('D#',0.646)	('A#',0.958)
('D#',0.966)	('A#',0.763)	('D#',0.53)	('A#',0.888)
('D#',0.947)	('A#',0.706)	('D#',0.466)	('A#',0.801)
('D#',0.888)	('Fm',0.667)	('D#',0.371)	('A#',0.739)
('D#',0.866)	('C#',0.762)	('D#',0.309)	('A#',0.691)
('D#',0.803)	('C#',0.805)	('D#',0.258)	('A#',0.634)
('D#',0.748)	('C#',0.81)	('D#m',0.188)	('F',0.75)
('D#',0.803)	('C#',0.778)	('D#m',0.201)	('F',0.836)
('D#',0.751)	('C#',0.737)	('D#m',0.222)	('F',0.816)
('D#',0.713)	('C#',0.714)	('D#m',0.275)	('F',0.674)
('D#m',0.751)	('A#m',0.665)	('D#m',0.307)	('A',0.809)
('D#m',0.79)	('NoChord',0)	('D#m',0.316)	('A',0.853)
('D#m',0.652)	('C#m',0.612)	('D#m',0.328)	('A',0.823)
('B',0.718)	('A',0.648)	('B',0.164)	('A',0.779)
('B',0.665)	('A',0.698)	('B',0.251)	('A',0.756)
('E',0.644)	('A',0.704)	('E',0.429)	('A',0.761)
('E',0.645)	('A',0.688)	('E',0.42)	('A',0.724)
('E',0.654)	('A',0.68)	('Em',0.455)	('A',0.689)
('Em',0.659)	('A',0.644)	('Em',0.47)	('A',0.647)
('E',0.664)	('A',0.632)	('Em',0.481)	('A',0.627)
('Em',0.668)	('A',0.645)	('Em',0.461)	('A',0.624)
('Em',0.65)	('A',0.674)	('E',0.446)	('A',0.624)
('Em',0.663)	('A',0.686)	('E',0.447)	('A',0.62)
('Em',0.652)	('A',0.688)	('E',0.433)	('A',0.614)

Reference : 0 , PCA direction : 7			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',0.979)	('D#',0.926)	('D#',1.0)
('D#',1.0)	('A#',0.664)	('D#',0.539)	('D#',0.847)
('D#',1.0)	('A#',0.884)	('D#',0.473)	('A#',0.667)
('D#',1.0)	('A#',0.98)	('D#',0.444)	('A#',0.891)
('D#',1.0)	('A#',0.982)	('D#',0.435)	('A#',0.856)
('D#',1.0)	('A#',0.98)	('D#',0.346)	('A#',0.654)
('D#',1.0)	('A#',0.963)	('D#',0.352)	('Fm',0.924)
('D#',1.0)	('A#',0.863)	('Gm',0.631)	('Fm',0.845)
('D#',0.986)	('A#',0.762)	('D#',0.349)	('F',0.896)
('D#',0.965)	('A#',0.759)	('D#',0.361)	('F',0.935)
('D#',0.957)	('A#',0.697)	('D#',0.371)	('F',0.947)
('D#',0.947)	('Dm',0.644)	('D#',0.33)	('F',0.957)
('D#',0.88)	('Dm',0.73)	('D#',0.0)	('F',0.903)
('D#',0.808)	('A',0.773)	('D#',0.018)	('Am',0.906)
('D#',0.661)	('A',0.915)	('D#',0.027)	('Am',0.86)
('NoChord',0)	('A',0.924)	('NoChord',0)	('A',0.827)
('Em',0.64)	('A',0.851)	('Em',0.443)	('A',0.858)
('Em',0.705)	('A',0.799)	('Em',0.474)	('A',0.84)
('Em',0.786)	('A',0.764)	('Em',0.455)	('A',0.836)
('Em',0.791)	('A',0.753)	('Em',0.452)	('A',0.851)
('Em',0.804)	('A',0.766)	('Em',0.458)	('A',0.851)
('Em',0.837)	('A',0.77)	('Em',0.463)	('A',0.84)
('Em',0.846)	('A',0.765)	('Em',0.49)	('A',0.814)
('Em',0.841)	('A',0.748)	('Em',0.501)	('A',0.811)
('Em',0.838)	('A',0.745)	('Em',0.499)	('A',0.798)
('Em',0.836)	('A',0.74)	('E',0.493)	('A',0.797)
('Em',0.846)	('A',0.73)	('E',0.494)	('A',0.796)
('Em',0.862)	('A',0.686)	('E',0.482)	('A',0.775)

Reference : 0 , PCA direction : 8			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',0.983)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('Gm',0.798)	('D#',0.955)	('D#',1.0)	('D#',1.0)
('Fm',0.776)	('D#',0.855)	('C#',0.0)	('D#',1.0)
('Fm',0.989)	('D#',0.757)	('C#',0.009)	('D#',1.0)
('F',0.941)	('A#',0.74)	('C#',0.013)	('D#',1.0)
('F',1.0)	('A#',0.867)	('F',0.013)	('D#',1.0)
('F',0.938)	('A#',0.916)	('F',0.013)	('D#',1.0)
('Am',0.914)	('A#',0.923)	('F',0.012)	('D#',1.0)
('Am',0.922)	('A#',0.957)	('F',0.012)	('D#',1.0)
('A',0.871)	('A#',0.936)	('Am',0.012)	('D#',0.989)
('A',0.836)	('A#',0.907)	('Am',0.054)	('D#',0.988)
('A',0.806)	('A#',0.774)	('Am',0.152)	('D#',0.977)
('A',0.804)	('A#',0.679)	('Am',0.192)	('D#',0.965)
('A',0.818)	('Fm',0.68)	('A',0.0)	('D#',0.936)
('A',0.809)	('Fm',0.711)	('A',0.0)	('D#',0.914)
('A',0.809)	('Fm',0.752)	('A',0.006)	('D#',0.873)
('A',0.802)	('Fm',0.784)	('C#m',0.019)	('D#',0.855)
('A',0.784)	('Fm',0.782)	('A',0.045)	('D#',0.84)
('A',0.757)	('Fm',0.828)	('A',0.064)	('D#',0.83)
('A',0.74)	('Fm',0.847)	('A',0.074)	('D#',0.83)
('A',0.728)	('F',0.827)	('A',0.075)	('D#',0.818)
('A',0.681)	('F',0.863)	('A',0.079)	('D#',0.799)
('A',0.644)	('F',0.893)	('A',0.083)	('D#',0.735)
('C#m',0.601)	('F',0.921)	('C#m',0.094)	('D#',0.634)
('NoChord',0)	('F',0.94)	('NoChord',0)	('B',0.624)
('NoChord',0)	('F',0.964)	('NoChord',0)	('B',0.658)
('NoChord',0)	('F',0.968)	('NoChord',0)	('B',0.653)
('NoChord',0)	('F',0.946)	('NoChord',0)	('B',0.633)
('NoChord',0)	('F',0.929)	('NoChord',0)	('B',0.637)

Reference : 0 , PCA direction : 9			
Bar1	Bar2	Bar3	Bar4
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',1.0)
('D#',1.0)	('D#',1.0)	('D#',1.0)	('D#',0.994)
('D#',0.975)	('D#',1.0)	('D#',1.0)	('D#',0.826)
('D#',0.827)	('D#',0.997)	('D#',1.0)	('Gm',0.785)
('D#',0.633)	('D#',0.993)	('D#',1.0)	('A#',0.662)
('A#',0.791)	('D#',0.99)	('Gm',0.652)	('Fm',0.739)
('A#',0.864)	('D#',0.99)	('Gm',0.647)	('Fm',0.876)
('A#',0.976)	('D#',0.994)	('Gm',0.629)	('Fm',0.816)
('A#',0.996)	('D#',0.99)	('Gm',0.63)	('F',0.901)
('A#',0.978)	('D#',0.99)	('Gm',0.65)	('F',0.9)
('A#',0.939)	('D#',0.928)	('Gm',0.645)	('F',0.871)
('A#',0.914)	('D#',0.819)	('Gm',0.604)	('F',0.78)
('A#',0.892)	('Em',0.752)	('A#',0.471)	('Am',0.849)
('A#',0.868)	('C',0.829)	('A#',0.424)	('Am',0.824)
('A#',0.867)	('C',0.878)	('A#',0.417)	('A',0.803)
('A#',0.795)	('C',0.922)	('Fm',0.534)	('A',0.784)
('F',0.752)	('C',0.938)	('Fm',0.563)	('A',0.763)
('F',0.902)	('C',0.928)	('Fm',0.589)	('A',0.758)
('F',0.957)	('C',0.886)	('Fm',0.584)	('A',0.772)
('F',0.971)	('Em',0.865)	('Fm',0.588)	('A',0.778)
('F',0.969)	('Em',0.876)	('F',0.591)	('A',0.766)
('F',0.97)	('Em',0.875)	('F',0.601)	('A',0.77)
('F',0.974)	('Em',0.864)	('F',0.616)	('A',0.757)
('F',0.972)	('Em',0.833)	('F',0.617)	('A',0.745)
('F',0.975)	('Em',0.796)	('F',0.631)	('A',0.729)
('F',0.966)	('Em',0.779)	('F',0.624)	('A',0.707)
('F',0.899)	('Em',0.776)	('F',0.618)	('A',0.701)
('F',0.833)	('Em',0.764)	('F',0.618)	('A',0.69)
('Am',0.778)	('Em',0.763)	('F',0.611)	('A',0.664)
('Am',0.835)	('Em',0.751)	('F',0.625)	('A',0.646)

Reference : 1 , PCA direction : 0			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.871)	('Em',0.522)	('F',0.866)
('G',1.0)	('F',0.846)	('Em',0.539)	('F',0.8)
('G',1.0)	('F',0.776)	('Em',0.557)	('F',0.761)
('G',1.0)	('C',0.79)	('Em',0.563)	('Am',0.783)
('G',1.0)	('C',0.956)	('Em',0.49)	('Am',0.796)
('G',1.0)	('C',0.996)	('Em',0.494)	('Am',0.761)
('G',1.0)	('C',0.992)	('Em',0.457)	('Am',0.674)
('G',1.0)	('C',0.995)	('Em',0.434)	('D',0.7)
('G',1.0)	('C',0.989)	('C',1.0)	('D',0.811)
('G',1.0)	('C',0.992)	('C',1.0)	('D',0.939)
('G',1.0)	('C',0.982)	('C',1.0)	('D',0.907)
('C',1.0)	('C',0.977)	('C',1.0)	('D',0.868)
('C',1.0)	('C',0.935)	('C',0.99)	('D',0.861)
('C',1.0)	('C',0.881)	('C',0.99)	('D',0.831)
('G',1.0)	('C',0.8)	('G',0.173)	('D',0.873)
('Gm',1.0)	('Am',0.719)	('Gm',0.113)	('D',0.763)
('Gm',0.976)	('A',0.696)	('Gm',0.224)	('Dm',0.689)
('A#',0.955)	('A',0.729)	('A#',0.46)	('Dm',0.629)
('A#',0.935)	('D#',0.649)	('A#',0.629)	('NoChord',0)
('A#',0.885)	('D#',0.796)	('A#',0.692)	('NoChord',0)
('A#',0.806)	('D#',0.841)	('A#m',0.697)	('NoChord',0)
('A#',0.782)	('D#',0.855)	('A#',0.7)	('NoChord',0)
('A#',0.793)	('D#m',0.879)	('A#',0.697)	('NoChord',0)
('A#',0.785)	('D#m',0.874)	('A#',0.683)	('NoChord',0)
('A#',0.779)	('D#m',0.873)	('A#',0.651)	('NoChord',0)
('A#',0.777)	('D#m',0.861)	('A#',0.635)	('NoChord',0)
('A#',0.759)	('D#m',0.858)	('A#',0.615)	('NoChord',0)
('A#',0.736)	('D#m',0.85)	('A#',0.586)	('NoChord',0)
('A#',0.715)	('D#m',0.834)	('A#',0.575)	('NoChord',0)

Reference : 1 , PCA direction : 1			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.977)	('Em',0.447)	('F',0.924)
('G',0.976)	('F',0.987)	('Em',0.432)	('F',0.924)
('G',0.937)	('F',0.988)	('Em',0.427)	('F',0.939)
('G',0.874)	('F',0.988)	('Em',0.422)	('F',0.943)
('Em',0.839)	('F',0.985)	('C',0.886)	('F',0.938)
('Em',0.804)	('F',0.984)	('C',0.856)	('F',0.889)
('C',0.74)	('F',0.987)	('C',0.819)	('F',0.864)
('C',0.77)	('F',0.969)	('C',0.782)	('F',0.735)
('C',0.789)	('F',0.837)	('C',0.602)	('F',0.717)
('C',0.841)	('F',0.988)	('C',0.672)	('Am',0.743)
('C',0.854)	('F',0.991)	('C',0.612)	('Am',0.76)
('C',0.903)	('F',0.976)	('C',0.593)	('Am',0.757)
('C',0.903)	('F',0.967)	('C',0.533)	('Am',0.733)
('C',0.893)	('F',0.955)	('C',0.489)	('Am',0.696)
('C',0.839)	('F',0.938)	('C',0.477)	('Am',0.687)
('C',0.812)	('F',0.932)	('C',0.461)	('C',0.643)
('C',0.789)	('F',0.876)	('C',0.486)	('C',0.623)
('C',0.743)	('F',0.847)	('C',0.505)	('C',0.614)
('C',0.74)	('F',0.826)	('C',0.512)	('C',0.626)
('C',0.768)	('F',0.835)	('C',0.49)	('C',0.634)
('C',0.758)	('F',0.841)	('C',0.457)	('C',0.616)
('C',0.734)	('F',0.806)	('C',0.44)	('NoChord',0)
('C',0.702)	('F',0.778)	('C',0.42)	('NoChord',0)
('Am',0.676)	('F',0.741)	('Am',0.403)	('NoChord',0)
('Am',0.646)	('F',0.701)	('Am',0.397)	('NoChord',0)
('NoChord',0)	('F',0.692)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('F',0.689)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('F',0.684)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('F',0.68)	('NoChord',0)	('NoChord',0)

Reference : 1 , PCA direction : 2			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.91)	('Em',0.492)	('F',0.781)
('G',1.0)	('F',0.906)	('Em',0.51)	('Am',0.698)
('G',1.0)	('F',0.88)	('Em',0.534)	('Am',0.906)
('G',1.0)	('F',0.726)	('Em',0.544)	('Am',0.924)
('G',0.996)	('Am',0.848)	('Em',0.577)	('Am',0.888)
('G',0.997)	('Am',0.877)	('C',1.0)	('Am',0.836)
('G',0.997)	('Am',0.896)	('C',0.996)	('Am',0.815)
('G',0.996)	('Am',0.862)	('C',0.996)	('Am',0.811)
('G',0.97)	('Am',0.848)	('C',1.0)	('Am',0.785)
('G',0.957)	('Am',0.813)	('C',1.0)	('Am',0.748)
('G',0.929)	('Am',0.781)	('C',0.978)	('Am',0.696)
('G',0.88)	('Am',0.797)	('Em',0.532)	('A',0.652)
('G',0.859)	('Am',0.762)	('G',0.489)	('A',0.66)
('G',0.804)	('Am',0.683)	('G',0.52)	('A',0.675)
('G',0.736)	('NoChord',0)	('G',0.539)	('A',0.698)
('G',0.664)	('NoChord',0)	('G',0.571)	('A',0.664)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Gm',0.646)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Gm',0.7)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Gm',0.646)	('NoChord',0)	('NoChord',0)
('Gm',0.632)	('NoChord',0)	('Gm',0.583)	('NoChord',0)
('Gm',0.673)	('NoChord',0)	('Gm',0.565)	('NoChord',0)
('Gm',0.669)	('NoChord',0)	('Gm',0.599)	('NoChord',0)
('Gm',0.641)	('NoChord',0)	('Gm',0.65)	('NoChord',0)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('NoChord',0)	('NoChord',0)	('NoChord',0)

Reference : 1 , PCA direction : 3			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.891)	('Em',0.542)	('F',0.759)
('G',1.0)	('F',0.779)	('Em',0.586)	('Am',0.636)
('G',1.0)	('C',0.858)	('Em',0.587)	('Am',0.713)
('G',1.0)	('C',0.993)	('Em',0.557)	('Am',0.684)
('G',0.986)	('C',0.993)	('Em',0.52)	('D',0.684)
('G',0.978)	('C',0.993)	('Em',0.509)	('D',0.749)
('G',0.98)	('C',0.989)	('Em',0.502)	('D',0.728)
('G',0.973)	('C',0.981)	('Em',0.495)	('D',0.725)
('G',0.97)	('C',0.973)	('Em',0.483)	('D',0.712)
('G',0.935)	('C',0.957)	('Em',0.474)	('D',0.714)
('G',0.909)	('C',0.944)	('Em',0.486)	('D',0.688)
('G',0.894)	('C',0.951)	('Em',0.476)	('G',0.701)
('G',0.875)	('C',0.948)	('Em',0.471)	('G',0.749)
('G',0.861)	('C',0.927)	('Em',0.461)	('G',0.819)
('G',0.849)	('C',0.875)	('Em',0.446)	('G',0.822)
('G',0.847)	('C',0.819)	('Em',0.441)	('G',0.826)
('G',0.846)	('C',0.816)	('Em',0.437)	('G',0.857)
('G',0.806)	('C',0.803)	('G',0.242)	('G',0.857)
('G',0.738)	('C',0.799)	('G',0.226)	('G',0.825)
('G',0.706)	('C',0.795)	('G',0.204)	('G',0.778)
('G',0.682)	('C',0.803)	('G',0.196)	('G',0.75)
('Gm',0.677)	('C',0.799)	('G',0.19)	('G',0.735)
('Gm',0.681)	('C',0.783)	('G',0.171)	('G',0.714)
('Gm',0.691)	('C',0.777)	('Gm',0.187)	('G',0.7)
('Gm',0.694)	('C',0.765)	('Gm',0.197)	('G',0.673)
('Gm',0.692)	('C',0.739)	('Gm',0.199)	('G',0.677)
('Gm',0.685)	('C',0.687)	('Gm',0.211)	('G',0.656)
('Gm',0.681)	('C',0.637)	('Gm',0.213)	('G',0.64)
('Gm',0.675)	('NoChord',0)	('Gm',0.204)	('Gm',0.633)

Reference : 1 , PCA direction :5			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.916)	('Em',0.446)	('F',0.849)
('G',1.0)	('F',0.888)	('Em',0.46)	('F',0.781)
('G',1.0)	('F',0.801)	('Em',0.471)	('F',0.748)
('G',1.0)	('F',0.757)	('Em',0.491)	('Dm',0.721)
('G',1.0)	('C',0.805)	('Em',0.5)	('NoChord',0)
('G',1.0)	('C',0.992)	('Em',0.515)	('NoChord',0)
('G',1.0)	('C',0.985)	('Em',0.488)	('G',0.655)
('G',0.991)	('C',0.902)	('Em',0.481)	('G',0.756)
('G',0.977)	('C',0.761)	('G',0.461)	('G',0.944)
('G',0.972)	('C',0.655)	('G',0.518)	('G',0.955)
('G',0.947)	('C',0.612)	('G',0.536)	('G',0.979)
('G',0.891)	('NoChord',0)	('G',0.561)	('G',0.979)
('G',0.815)	('NoChord',0)	('G',0.555)	('G',0.976)
('Em',0.832)	('NoChord',0)	('G',0.505)	('G',0.975)
('Em',0.836)	('NoChord',0)	('G',0.443)	('G',0.958)
('Em',0.843)	('NoChord',0)	('G',0.422)	('G',0.95)
('Em',0.853)	('NoChord',0)	('G',0.422)	('G',0.953)
('Em',0.862)	('D',0.6)	('G',0.404)	('G',0.945)
('Em',0.869)	('D',0.619)	('G',0.373)	('G',0.908)
('Em',0.868)	('D',0.648)	('G',0.338)	('G',0.898)
('Em',0.883)	('D',0.653)	('G',0.298)	('G',0.882)
('Em',0.902)	('D',0.664)	('G',0.277)	('G',0.875)
('Em',0.911)	('D',0.659)	('G',0.285)	('G',0.851)
('Em',0.917)	('D',0.627)	('G',0.305)	('G',0.824)
('Em',0.916)	('D',0.604)	('G',0.31)	('G',0.762)
('Em',0.91)	('NoChord',0)	('Em',0.148)	('G',0.722)
('Em',0.916)	('NoChord',0)	('Em',0.163)	('G',0.693)
('Em',0.928)	('NoChord',0)	('Em',0.181)	('G',0.663)
('Em',0.932)	('NoChord',0)	('Em',0.194)	('G',0.601)

Reference : 1 , PCA direction : 6			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.99)	('Em',0.413)	('F',0.85)
('G',1.0)	('F',0.986)	('Em',0.379)	('F',0.822)
('G',1.0)	('F',0.983)	('Em',0.344)	('Am',0.795)
('G',1.0)	('F',0.98)	('Em',0.27)	('Am',0.886)
('G',1.0)	('F',0.981)	('G',0.138)	('Am',0.969)
('G',1.0)	('F',0.982)	('G',0.096)	('Am',0.973)
('G',1.0)	('F',0.981)	('G',0.027)	('Am',0.968)
('G',1.0)	('F',0.979)	('G',0.0)	('Am',0.96)
('G',1.0)	('F',0.98)	('G',0.0)	('Am',0.963)
('G',1.0)	('F',0.979)	('G',0.0)	('Am',0.966)
('G',1.0)	('F',0.969)	('G',0.0)	('Am',0.933)
('G',1.0)	('F',0.762)	('Em',0.014)	('Am',0.899)
('G',1.0)	('Am',0.918)	('Em',0.014)	('Am',0.883)
('G',0.985)	('Am',0.89)	('Em',0.006)	('A',0.927)
('G',0.981)	('Am',0.92)	('Em',0.036)	('A',0.937)
('G',0.933)	('Am',0.924)	('Em',0.159)	('A',0.935)
('G',0.846)	('Am',0.897)	('Em',0.272)	('A',0.936)
('Em',0.824)	('Am',0.898)	('Em',0.287)	('A',0.918)
('G',0.821)	('Am',0.847)	('Em',0.277)	('A',0.903)
('G',0.775)	('A',0.889)	('Em',0.273)	('A',0.877)
('G',0.767)	('A',0.884)	('Em',0.27)	('A',0.875)
('Em',0.738)	('A',0.846)	('Em',0.298)	('A',0.865)
('Em',0.724)	('A',0.812)	('Em',0.339)	('A',0.86)
('Em',0.71)	('A',0.826)	('Em',0.32)	('A',0.852)
('Em',0.697)	('A',0.83)	('Em',0.349)	('A',0.842)
('Em',0.688)	('A',0.831)	('Em',0.36)	('A',0.838)
('G',0.698)	('A',0.816)	('Em',0.375)	('A',0.825)
('G',0.733)	('A',0.818)	('Em',0.388)	('A',0.814)
('G',0.717)	('A',0.81)	('Em',0.385)	('A',0.79)

Reference : 1 , PCA direction : 7			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.99)	('Em',0.322)	('F',0.88)
('G',1.0)	('F',0.993)	('Em',0.17)	('F',0.827)
('G',0.995)	('F',0.989)	('G',0.105)	('Am',0.88)
('G',0.995)	('F',0.988)	('G',0.034)	('Am',0.96)
('G',0.97)	('F',0.981)	('G',0.011)	('Am',0.949)
('G',0.938)	('F',0.972)	('G',0.01)	('Am',0.897)
('G',0.925)	('F',0.913)	('G',0.004)	('Am',0.851)
('G',0.91)	('Am',0.892)	('Em',0.008)	('Am',0.786)
('G',0.901)	('Am',0.861)	('Em',0.019)	('A',0.783)
('G',0.87)	('Am',0.845)	('Em',0.109)	('A',0.773)
('G',0.82)	('Am',0.803)	('Em',0.275)	('A',0.797)
('G',0.817)	('Am',0.774)	('Em',0.386)	('A',0.811)
('G',0.787)	('Am',0.721)	('Em',0.429)	('A',0.797)
('G',0.742)	('A',0.778)	('Em',0.462)	('A',0.792)
('G',0.725)	('A',0.817)	('Em',0.547)	('A',0.777)
('Em',0.708)	('A',0.811)	('Em',0.58)	('A',0.739)
('Em',0.723)	('A',0.796)	('Em',0.564)	('A',0.718)
('Em',0.737)	('A',0.793)	('Em',0.542)	('A',0.695)
('Em',0.747)	('A',0.782)	('Em',0.51)	('A',0.678)
('Em',0.741)	('A',0.744)	('Em',0.464)	('A',0.642)
('Em',0.733)	('A',0.71)	('Em',0.438)	('A',0.626)
('Em',0.714)	('A',0.67)	('Em',0.408)	('C#m',0.646)
('Em',0.72)	('A',0.64)	('Em',0.389)	('C#m',0.671)
('Em',0.716)	('C#m',0.631)	('Em',0.359)	('C#m',0.697)
('Em',0.704)	('C#m',0.663)	('Em',0.342)	('C#m',0.709)
('Em',0.678)	('C#m',0.67)	('Em',0.349)	('C#m',0.706)
('Em',0.66)	('C#m',0.676)	('Em',0.347)	('C#m',0.707)
('C',0.633)	('C#m',0.672)	('Em',0.354)	('C#m',0.692)
('C',0.619)	('C#m',0.664)	('Em',0.361)	('C#m',0.674)

Reference : 1 , PCA direction : 8			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',0.926)	('F',0.977)	('Em',0.471)	('F',0.884)
('C',0.849)	('F',0.99)	('C',0.972)	('F',0.892)
('C',1.0)	('F',0.993)	('C',0.976)	('F',0.883)
('C',0.964)	('F',0.993)	('C',0.977)	('F',0.847)
('C',0.872)	('F',0.993)	('C',0.977)	('F',0.814)
('Am',0.795)	('F',0.992)	('C',0.981)	('F',0.795)
('Am',0.964)	('F',0.992)	('Am',0.706)	('F',0.788)
('Am',0.985)	('F',0.992)	('Am',0.705)	('F',0.764)
('Am',0.982)	('F',0.996)	('Am',0.701)	('F',0.704)
('Am',0.933)	('F',0.992)	('Am',0.695)	('F',0.689)
('Am',0.86)	('F',0.992)	('Am',0.678)	('F',0.673)
('A',0.83)	('F',0.988)	('Am',0.68)	('F',0.649)
('A',0.858)	('F',0.992)	('Am',0.685)	('F',0.622)
('A',0.878)	('F',0.988)	('Am',0.703)	('F',0.612)
('A',0.842)	('F',0.976)	('Am',0.684)	('NoChord',
('A',0.804)	('F',0.972)	('A',0.203)	('NoChord',
('A',0.763)	('F',0.94)	('A',0.221)	('NoChord',
('A',0.721)	('F',0.9)	('A',0.22)	('NoChord',
('A',0.676)	('F',0.863)	('A',0.23)	('NoChord',
('A',0.636)	('F',0.837)	('A',0.23)	('NoChord',
('A',0.602)	('F',0.799)	('A',0.222)	('NoChord',
('NoChord',0)	('F',0.758)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Am',0.767)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Am',0.792)	('NoChord',0)	('NoChord',0)
('NoChord',0)	('Am',0.809)	('NoChord',0)	('G',0.616)
('NoChord',0)	('Am',0.81)	('NoChord',0)	('G',0.638)
('NoChord',0)	('Am',0.795)	('NoChord',0)	('G',0.634)
('NoChord',0)	('Am',0.788)	('NoChord',0)	('G',0.623)
('NoChord',0)	('Am',0.783)	('NoChord',0)	('G',0.617)

Reference : 1 , PCA direction : 9			
Bar1	Bar2	Bar3	Bar4
('G',1.0)	('F',0.956)	('Em',0.455)	('F',0.903)
('G',1.0)	('F',0.798)	('Em',0.424)	('F',0.869)
('G',1.0)	('C',0.846)	('Em',0.43)	('Am',0.81)
('G',1.0)	('C',0.987)	('Em',0.413)	('Am',0.935)
('G',1.0)	('C',0.992)	('Em',0.41)	('Am',0.944)
('G',1.0)	('C',0.992)	('Em',0.42)	('Am',0.913)
('G',0.966)	('C',0.992)	('Em',0.418)	('Am',0.836)
('G',0.961)	('C',0.993)	('Em',0.38)	('Am',0.796)
('G',0.916)	('C',0.993)	('Em',0.333)	('A',0.744)
('G',0.88)	('C',0.993)	('G',0.227)	('A',0.719)
('G',0.846)	('C',0.993)	('G',0.199)	('A',0.721)
('G',0.81)	('C',0.993)	('G',0.188)	('A',0.733)
('G',0.77)	('C',0.993)	('G',0.173)	('A',0.724)
('G',0.727)	('C',0.993)	('C',0.6)	('A',0.724)
('C',0.644)	('C',0.976)	('C',0.598)	('A',0.73)
('C',0.601)	('C',0.959)	('C',0.59)	('A',0.72)
('NoChord',0)	('C',0.932)	('NoChord',0)	('A',0.698)
('NoChord',0)	('C',0.907)	('NoChord',0)	('A',0.676)
('NoChord',0)	('C',0.85)	('NoChord',0)	('A',0.67)
('NoChord',0)	('C',0.809)	('NoChord',0)	('A',0.665)
('NoChord',0)	('C',0.781)	('NoChord',0)	('A',0.646)
('Am',0.634)	('Am',0.751)	('Am',0.64)	('A',0.624)
('Am',0.661)	('Am',0.759)	('Am',0.689)	('C#m',0.616)
('Am',0.69)	('Am',0.753)	('Am',0.711)	('C#m',0.616)
('Am',0.702)	('Am',0.755)	('Am',0.737)	('C#m',0.604)
('Am',0.72)	('Am',0.723)	('Am',0.732)	('NoChord',0)
('Am',0.736)	('Am',0.704)	('Am',0.724)	('NoChord',0)
('Am',0.763)	('Am',0.7)	('Am',0.695)	('NoChord',0)
('Am',0.757)	('Am',0.678)	('Am',0.685)	('NoChord',0)
('Am',0.752)	('Am',0.667)	('Am',0.67)	('NoChord',0)

6.2 Πίνακες Αποτελεσμάτων της Chord Quality

Reference : 0 Chord Quality									
PCA0	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.000	0.979	0.991	1.000	1.000	1.000	1.000	0.976	0.996	0.998
1.000	0.910	1.000	1.000	1.000	0.973	0.965	0.763	0.938	0.950
1.000	0.769	0.982	0.888	1.000	0.907	0.893	0.756	0.658	0.902
0.986	0.789	0.888	0.867	1.000	0.759	0.861	0.829	0.689	0.822
0.906	0.770	0.925	0.816	0.989	0.799	0.849	0.818	0.674	0.793
0.878	0.817	0.992	0.863	0.958	0.716	0.851	0.745	0.720	0.844
0.887	0.785	0.988	0.829	0.881	0.753	0.787	0.810	0.717	0.854
0.831	0.836	0.982	0.802	0.813	0.737	0.730	0.835	0.712	0.879
0.890	0.864	0.969	0.796	0.795	0.772	0.666	0.748	0.723	0.880
0.822	0.781	0.951	0.809	0.796	0.744	0.657	0.755	0.702	0.846
0.701	0.769	0.930	0.831	0.863	0.786	0.625	0.743	0.696	0.779
0.939	0.661	0.907	0.851	0.845	0.657	0.624	0.720	0.677	0.741
0.903	0.760	0.886	0.837	0.776	0.763	0.655	0.628	0.660	0.736
0.819	0.778	0.868	0.808	0.717	0.663	0.632	0.626	0.609	0.741
0.807	0.789	0.847	0.793	0.516	0.711	0.594	0.616	0.609	0.759
0.875	0.787	0.843	0.778	0.475	0.741	0.633	0.438	0.610	0.754
0.869	0.774	0.800	0.787	0.188	0.749	0.490	0.698	0.615	0.794
0.835	0.751	0.820	0.718	0.183	0.747	0.604	0.705	0.613	0.800
0.753	0.768	0.865	0.701	0.707	0.758	0.577	0.710	0.620	0.801
0.756	0.767	0.877	0.687	0.724	0.758	0.593	0.712	0.623	0.801
0.758	0.757	0.903	0.677	0.730	0.755	0.635	0.720	0.612	0.804
0.699	0.707	0.883	0.670	0.568	0.748	0.619	0.727	0.606	0.803
0.782	0.722	0.860	0.681	0.544	0.743	0.620	0.729	0.589	0.792
0.901	0.741	0.829	0.690	0.186	0.730	0.605	0.725	0.563	0.783
0.904	0.747	0.786	0.693	0.330	0.724	0.601	0.720	0.391	0.769
0.861	0.737	0.746	0.684	0.323	0.718	0.600	0.717	0.405	0.749
0.825	0.719	0.365	0.668	0.313	0.690	0.599	0.717	0.405	0.726
0.799	0.699	0.356	0.658	0.000	0.680	0.604	0.710	0.395	0.704
0.780	0.661	0.185	0.640	0.000	0.462	0.597	0.701	0.392	0.714

Reference : 1 Chord Quality									
PCA0	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9
0.829	0.829	0.829	0.829	0.829	0.829	0.829	0.829	0.829	0.829
0.815	0.837	0.796	0.798	0.818	0.803	0.813	0.798	0.815	0.773
0.796	0.830	0.779	0.750	0.814	0.782	0.797	0.747	0.926	0.772
0.774	0.823	0.830	0.790	0.814	0.755	0.781	0.742	0.963	0.834
0.784	0.807	0.799	0.809	0.825	0.742	0.784	0.744	0.945	0.837
0.811	0.912	0.827	0.796	0.810	0.576	0.772	0.728	0.914	0.831
0.813	0.883	0.928	0.807	0.755	0.627	0.763	0.704	0.891	0.803
0.781	0.852	0.926	0.800	0.741	0.782	0.744	0.673	0.863	0.783
0.782	0.814	0.916	0.794	0.763	0.783	0.735	0.649	0.861	0.746
0.950	0.736	0.901	0.785	0.791	0.786	0.736	0.641	0.846	0.705
0.983	0.811	0.880	0.770	0.810	0.775	0.736	0.649	0.827	0.690
0.972	0.804	0.846	0.757	0.817	0.769	0.726	0.674	0.801	0.681
0.961	0.807	0.715	0.756	0.817	0.608	0.669	0.697	0.787	0.665
0.946	0.784	0.693	0.761	0.781	0.587	0.704	0.684	0.789	0.761
0.926	0.758	0.671	0.767	0.832	0.578	0.702	0.694	0.795	0.737
0.712	0.735	0.493	0.748	0.738	0.559	0.719	0.717	0.626	0.718
0.649	0.712	0.475	0.733	0.692	0.554	0.738	0.710	0.495	0.408
0.646	0.694	0.000	0.739	0.673	0.557	0.738	0.700	0.481	0.396
0.693	0.677	0.000	0.677	0.331	0.703	0.732	0.692	0.460	0.380
0.553	0.676	0.162	0.647	0.318	0.692	0.712	0.679	0.442	0.369
0.593	0.682	0.175	0.621	0.154	0.688	0.704	0.648	0.426	0.357
0.586	0.668	0.162	0.608	0.155	0.679	0.699	0.627	0.406	0.662
0.584	0.495	0.304	0.600	0.000	0.680	0.687	0.609	0.190	0.681
0.592	0.475	0.310	0.587	0.000	0.677	0.684	0.605	0.192	0.693
0.586	0.455	0.317	0.589	0.000	0.668	0.677	0.601	0.198	0.700
0.576	0.436	0.323	0.582	0.000	0.648	0.680	0.605	0.356	0.544
0.568	0.173	0.000	0.577	0.000	0.445	0.679	0.601	0.362	0.541
0.558	0.172	0.000	0.560	0.000	0.443	0.678	0.598	0.357	0.540
0.543	0.171	0.000	0.543	0.000	0.443	0.688	0.588	0.353	0.530
0.531	0.170	0.000	0.378	0.000	0.432	0.676	0.580	0.350	0.522

Βιβλιογραφία.

- [1] Pypianoroll documentation: <https://salu133445.github.io/pypianoroll/>, 2020
- [2] Nicolas Boulanger-Lewandowski, Yoshua Bengio, and Pascal Vincent. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. arXiv preprint arXiv:1206.6392, 2012
- [3] Gino Brunner, Andres Konrad, Yuyi Wang, and Roger Wattenhofer. Midivae: Modeling dynamics and instrumentation of music with applications to style transfer. arXiv preprint arXiv:1809.07600, 2018.
- [4] Gino Brunner, Yuyi Wang, Roger Wattenhofer, and Sumu Zhao. Symbolic music genre transfer with cyclegan. In 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), pages 786793. IEEE, 2018.
- [5] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Advances in neural information processing systems, pages 21722180, 2016.
- [6] Arden Dertat. Applied deep learning - part 3: Autoencoders, Oct 2017
- [7] Google Developers. Overview of gan structure | generative adversarial networks, 2020.
- [8] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A generative model for music.
- [9] Sander Dieleman. Generating music in the waveform domain, Mar 2020.
- [10] Sander Dieleman, Aaron van den Oord, and Karen Simonyan. The challenge of realistic music generation: modelling raw audio at scale. In Advances in Neural Information Processing Systems, pages 79897999, 2018.
- [11] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. Musegan: Demonstration of a convolutional gan based model for generating multi-track piano-rolls. ISMIR Late Breaking/Demos, 2017.

- [12] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018
- [13] Hao-Wen Dong and Yi-Hsuan Yang. Convolutional generative adversarial networks with binary neurons for polyphonic music generation. arXiv preprint arXiv:1804.09399, 2018.
- [14] Douglas Eck and Juergen Schmidhuber. Finding temporal structure in music: Blues improvisation with lstm recurrent networks. In Proceedings of the 12th IEEE workshop on neural networks for signal processing, pages 747756. IEEE, 2002..
- [15] Felix A Gers, Juergen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. 1999.
- [16] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160, 2016.
- [17] Jack Grundy. Are neural nets close to producing real art?, Apr 2019.
- [18] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the maestro dataset. arXiv preprint arXiv:1810.12247, 2018.
- [19] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in neural information processing systems, pages 66266637, 2017.
- [20] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. arXiv preprint arXiv:1912.04958, 2019.
- [21] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013
- [22] Wonkwang Lee, Donggyun Kim, Seunghoon Hong, and Honglak Lee. High-fidelity synthesis with disentangled representation. arXiv preprint arXiv:2001.04296, 2020.

- [23] Soroush Mehri, Kundan Kumar, Ishaan Gulrajani, Rithesh Kumar, Shubham Jain, Jose Sotelo, Aaron Courville, and Yoshua Bengio. Samplernn: An unconditional end-to-end neural audio generation model. arXiv preprint arXiv:1612.07837, 2016.
- [24] Olof Mogren. C-rnn-gan: Continuous recurrent neural networks with adversarial training. arXiv preprint arXiv:1611.09904, 2016.
- [25] Gerhard Nierhaus. Algorithmic composition: paradigms of automated music generation. Springer Science & Business Media, 2009.
- [26] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499, 2016.
- [27] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. A hierarchical latent vector model for learning long-term structure in music. arXiv preprint arXiv:1803.05428, 2018.
- [28] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 92439252, 2020.
- [29] Aaron van den Oord, Oriol Vinyals, et al. Neural discrete representation learning. In Advances in Neural Information Processing Systems, pages 63066315, 2017.
- [30] Sean Vasquez and Mike Lewis. Melnet: A generative model for audio in the frequency domain. arXiv preprint arXiv:1906.01083, 2019.
- [31] Andrey Voynov and Artem Babenko. Unsupervised discovery of interpretable directions in the gan latent space. arXiv preprint arXiv:2002.03754, 2020.
- [32] Jian Wu, Changran Hu, Yulong Wang, Xiaolin Hu, and Jun Zhu. A hierarchical recurrent neural network for symbolic melody generation. IEEE Transactions on Cybernetics, 50(6):27492757, 2019.
- [33] Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang. Midinet: A convolutional generative adversarial network for symbolic-domain music generation. arXiv preprint arXiv:1703.10847, 2017.

- [34] Erik Härkönen, *GANSpace: Discovering Interpretable GAN Controls*, NeurIPS, 2020.
- [35] Logan Eisenbeiser, *Latent Walking Techniques for Conditioning GAN-Generated Music*, IEEE, 2020.