



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τηλεπικοινωνιών
Εργαστήριο Ηλεκτρακουστικής & Τηλεοπτικών Συστημάτων

Διπλωματική Εργασία

Τεχνικές Ανάκτησης Μουσικής Πληροφορίας για την Παραγωγή Ρυθμικών Μοτίβων Τυμπάνων

Αλέξανδρος Ηλιάδης
ΑΕΜ 9300

Επίβλεψη:
Χαράλαμπος Δημούλας
Αναπληρωτής Καθηγητής

Ιούλιος 2022

Copyright © 2022 Αλέξανδρος Ηλιάδης



Το παρόν έργο αδειοδοτείται υπό τους όρους της άδειας Creative Commons Αναφορά Δημιουργού - Μη Εμπορική Χρήση - Παρόμοια Διανομή 4.0.

Για να δείτε ένα αντίγραφο αυτής της άδειας επισκεφτείτε τον ιστότοπο:

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.el>

Περίληψη

Τεχνικές Ανάκτησης Μουσικής Πληροφορίας για την Παραγωγή Ρυθμικών Μοτίβων Τυμπάνων

Η παραγωγή μοτίβων τυμπάνων αποτελεί ένα μέρος του ευρύτερου ερευνητικού πεδίου της αλγορίθμικής σύνθεσης το οποίο επικεντρώνεται στο μουσικό στοιχείο του ρυθμού. Η σχετική έρευνα προσφέρει διάφορες μεθοδολογίες, ανάλογα με τη φύση της εκάστοτε επιθυμητής εφαρμογής, ωστόσο υπάρχει έλλειψη προσεγγίσεων που να στηρίζονται σε ηχητικά δεδομένα για να παράγουν ρυθμικές ακολουθίες τυμπάνων. Η παρούσα διπλωματική εργασία σκοπεύει να αντιμετωπίσει αυτό το ζήτημα παρουσιάζοντας ένα σύστημα που λαμβάνει ένα ακατέργαστο μουσικό απόσπασμα και παράγει συμβολικά μοτίβα τυμπάνων τα οποία είναι ρυθμικά συμβατά με αυτό. Βασικό πυλώνα της προτεινόμενης μεθοδολογίας αποτελεί το λεγόμενο Σήμα Ισχύος Onset (OSS), μια χρονοσειρά που αξιοποιείται συχνά σε θέματα Ανάκτησης Μουσικής Πληροφορίας (MIR), ενώ εξάγονται και επιπλέον χαρακτηριστικά που περιέχουν πληροφορία για τη μετρική δομή του μουσικού αποσπάσματος. Το προτεινόμενο σύστημα, ακολουθώντας την καλά καθιερωμένη τάση των τεχνικών μηχανικής μάθησης, βασίζεται σε ένα μοντέλο Seq2Seq, μια αναδρομική αρχιτεκτονική νευρωνικού δικτύου η οποία προέρχεται από το πεδίο της Επεξεργασίας Φυσικής Γλώσσας (NLP) και επομένως μπορεί να θεωρηθεί κατάλληλη για τη μετάφραση μιας μελωδικής ακολουθίας σε μια ρυθμική ακολουθία. Η αξιολόγηση του συστήματος πραγματοποιήθηκε κατά κύριο λόγο μέσω της διεξαγωγής μιας διαδικτυακής έρευνας στην οποία συμμετέχοντες με διαφορετικό μουσικό υπόβαθρο κλήθηκαν να ακούσουν και να βαθμολογήσουν ποιοτικά χαρακτηριστικά ορισμένων ηχητικών δειγμάτων. Τα δείγματα ήταν ομαδοποιημένα σε ζεύγη, καθένα από τα οποία αποτελούνταν από ένα κοινό μουσικό απόσπασμα συνοδευόμενο τόσο από ανθρωπίνως μεταγραμμένα όσο και από τεχνητά παραγόμενα μοτίβα τυμπάνων. Για να αποφευχθούν μεροληπτικές βαθμολογήσεις η προέλευση των μοτίβων τυμπάνων κάθε δείγματος παρέμεινε κρυφή. Η στατιστική ανάλυση των αποτελεσμάτων έδειξε ότι, κατά μέσο όρο, η μόνη στατιστικά σημαντική διαφορά ανάμεσα στις βαθμολογήσεις των δύο τύπων δειγμάτων αφορούσε τη δημιουργικότητα, στην οποία τα δείγματα των παραγόμενων μοτίβων τυμπάνων υπερτέρησαν. Επιπρόσθετα, οι συμμετέχοντες αντιμετώπισαν ιδιαίτερη δυσκολία στο να αναγνωρίσουν ορθά τα δείγματα των παραγόμενων μοτίβων τυμπάνων, επικυρώνοντας έτσι την αποτελεσματικότητα του συστήματος.

Λέξεις-Κλειδιά: Αλγορίθμική Σύνθεση, Παραγωγή Μοτίβων Τυμπάνων, Ανάκτηση Μουσικής Πληροφορίας, Σήμα Ισχύος Onset, Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Seq2Seq

Abstract

Music Information Retrieval Techniques for Rhythmic Drum Pattern Generation

Drum pattern generation is a part of the wider research field of algorithmic composition which focuses on the music element of the rhythm. Related research offers various methodologies, depending on the nature of the respective desired application, however there is a lack of approaches relying on audio data to generate rhythmic drum sequences. The current thesis intends to address this issue by introducing a system which receives a raw musical excerpt and generates symbolic drum patterns which are rhythmically compatible with it. Core pillar of the proposed methodology is the so-called Onset Strength Signal (OSS), a time series which is often utilized on topics of Music Information Retrieval (MIR), while additional features containing information on the metrical structure of the musical excerpt are also extracted. The proposed system, following the well-established trend of machine learning techniques, is based on a Seq2Seq model, a recurrent neural network architecture which originates from the Natural Language Processing (NLP) field and therefore can be deemed suitable for translating a melodic sequence into a rhythmic sequence. The system evaluation was carried out primarily by conducting an online survey in which participants of varying musical background were asked to listen to and rate qualitative characteristics of certain audio samples. The samples were grouped in pairs, each one of them consisting of a mutual musical excerpt accompanied by both humanly-transcribed and artificially-generated drum patterns. To avoid biased ratings the source of each sample's drum pattern remained hidden. Statistical analysis of the results showed that, on average, the only statistically significant difference between the ratings of the two sample types was in regard to creativity, in which the samples of the generated drum patterns excelled. Furthermore, the participants faced particular difficulty in correctly identifying the samples of the generated drum patterns, thus validating the effectiveness of the system.

Keywords: Algorithmic Composition, Drum Pattern Generation,
Music Information Retrieval, Onset Strength Signal,
Machine Learning, Neural Networks, Seq2Seq

Πρόλογος

Η παρούσα διπλωματική εργασία ολοκληρώνει τις φοιτητικές μου υποχρεώσεις στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης. Ένας ξεχωριστός κύκλος 6 χρόνων σπουδών κλείνει με την πραγματοποίηση μιας ιδέας η οποία γεννήθηκε πριν από πολύ καιρό, μιας ιδέας που κάποτε πίστευα πως θα ήταν αδύνατο να υλοποιηθεί ποτέ στην πράξη. Η εμπειρία και οι γνώσεις που απέκτησα με τα χρόνια μέσα από την ακαδημαϊκή μου πορεία μου έδωσαν το κίνητρο και τα εφόδια να προκαλέσω τον εαυτό μου και να ξεκινήσω πριν 9 μήνες ένα ταξίδι που δεν θα μπορούσε ποτέ να έχει πραγματοποιηθεί χωρίς τη συμβολή κάποιων ανθρώπων τους οποίους οφείλω να αναγνωρίσω.

Αρχικά, θα ήθελα να ευχαριστήσω τον υπεύθυνο καθηγητή μου, κύριο Χαράλαμπο Δημούλα, για την εμπιστοσύνη που έδειξε στο πρόσωπο μου και την ελευθερία που μου έδωσε στην επιλογή του θέματος της εργασίας και την ανάπτυξη της προτεινόμενης μεθοδολογίας. Επίσης, θέλω να ευχαριστήσω τους κυρίους Ιορδάνη Θωίδη, Νικόλαο Βρύζα και Λάζαρο Βρύση για τις προτάσεις τους, τα σχόλια τους και τη συμμετοχή τους στην επίβλεψη της εργασίας. Οφείλω να δώσω επιπλέον ευχαριστίες στους κυρίους Δημούλα και Θωίδη για τη διδασκαλία του μαθήματος «Τεχνολογία του Ήχου και της Εικόνας», το οποίο έθεσε τις βάσεις για την εκπόνηση αυτής της εργασίας. Ακόμα, θα ήθελα να ευχαριστήσω τους κυρίους Γεώργιο Παπανικολάου και Χρήστο Σεβαστιάδη για τη διδασκαλία των μαθημάτων «Ηλεκτρακουστική I» και «Ηλεκτρακουστική II», καθώς ήταν τα πρώτα μαθήματα που μου έδειξαν ότι η επιστήμη του Ηλεκτρολόγου Μηχανικού μπορεί να συνδυαστεί με τον ήχο και τη μουσική, στρώνοντας έτσι τον δρόμο για την μετέπειτα πορεία μου.

Ιδιαίτερες ευχαριστίες νιώθω την ανάγκη να δώσω στον καλό φίλο και πρώην συμφοιτητή μου Αθανάσιο Παρασκευά, ο οποίος με την εμπειρία του στον τομέα της μηχανικής μάθησης με συμβούλευε διαρκώς και με βοήθησε να ξεπεράσω πολλά εμπόδια που εμφανίστηκαν κατά τη διάρκεια εκπόνησης της εργασίας.

Τέλος, οφείλω να ευχαριστήσω την οικογένεια, τους φίλους, τις φίλες και τους μουσικούς οι οποίοι βρίσκονταν φυσικά και νοητά δίπλα μου όλα αυτά τα χρόνια και καθόρισαν, ο καθένας με τον τρόπο του, το άτομο που είμαι σήμερα.

Αλέξανδρος Ηλιάδης
Θεσσαλονίκη, Ιούλιος 2022

Περιεχόμενα

Περίληψη	i
Abstract	ii
Πρόλογος	iii
Περιεχόμενα	iv
Λίστα Εικόνων	vi
Λίστα Πινάκων	viii
Λίστα Συντομογραφιών	ix
Εισαγωγή	x
1 Θεώρηση του Προβλήματος	1
1.1 Παραγωγή Μοτίβων Τυμπάνων	1
1.2 Βασικές Προσεγγίσεις	3
1.3 Εφαρμογές	3
1.3.1 Logic Pro Drummer	3
1.3.2 Magenta Studio	4
1.4 Στοχευμένη Καινοτομία	5
2 Θεωρητικό Υπόβαθρο	7
2.1 Στοιχεία Μουσικής Θεωρίας	7
2.2 Ψηφιακός Ήχος	9
2.2.1 Ψηφιακή Επεξεργασία Σήματος	10
2.2.2 Ψηφιακή Διασύνδεση Μουσικών Οργάνων	13
2.3 Νευρωνικά Δίκτυα	14
2.3.1 Βασικές Αρχιτεκτονικές	15
2.3.2 Προχωρημένες Αρχιτεκτονικές	17
2.4 Ερευνητικά Έργα	19
3 Μεθοδολογία και Υλοποίηση	21
3.1 Διαδικασία	21
3.2 Σύνολο Δεδομένων	22
3.3 Προεπεξεργασία	23
3.3.1 Μετρική Δομή	23

3.3.2 Αναπαράσταση Μοτίβων Τυμπάνων	23
3.3.3 Εξαγωγή Χαρακτηριστικών	26
3.4 Νευρωνικό Δίκτυο	32
3.5 Μετεπεξεργασία	32
4 Αξιολόγηση	35
4.1 Διαδικασίες Αξιολόγησης	35
4.1.1 Αντικειμενική Αξιολόγηση	35
4.1.2 Υποκειμενική Αξιολόγηση	36
4.2 Αποτελέσματα Αξιολόγησης	37
4.2.1 Αποτελέσματα Εκπαίδευσης	37
4.2.2 Αποτελέσματα Έρευνας	41
4.3 Οπτικοποίηση Δειγμάτων	52
5 Συμπεράσματα	55
5.1 Σύνοψη	55
5.2 Συνεισφορά	55
5.3 Μελλοντικές Επεκτάσεις	55
Βιβλιογραφικές Αναφορές	57
Παράρτημα: Κώδικας & Εργαλεία	61

Λίστα Εικόνων

Εικόνα 1.1:	Διαδικασία παραγωγής μοτίβων τυμπάνων	1
Εικόνα 1.2:	Τα μέρη ενός τυπικού συνόλου τυμπάνων	2
Εικόνα 1.3:	Το γραφικό περιβάλλον του <i>Drummer</i>	4
Εικόνα 1.4:	Τα γραφικά περιβάλλοντα των εφαρμογών του <i>Magenta Studio</i>	4
Εικόνα 2.1:	Οι βασικότερες αξίες στη μουσική	7
Εικόνα 2.2:	Παρεστιγμένες και συζευγμένες νότες	8
Εικόνα 2.3:	Τρίγχα και ισοδύναμες αξίες	8
Εικόνα 2.4:	Παρτιτούρα μελωδίας διάρκειας δύο μέτρων 4/4	9
Εικόνα 2.5:	Ηχητικό σήμα στο πεδίο του χρόνου	10
Εικόνα 2.6:	Ηχητικό σήμα στο πεδίο της συχνότητας	11
Εικόνα 2.7:	Ηχητικό σήμα στο πεδίο του χρόνου και της συχνότητας	12
Εικόνα 2.8:	Σχηματική αναπαράσταση νευρώνα	14
Εικόνα 2.9:	Σχηματική αναπαράσταση δικτύου MLP	15
Εικόνα 2.10:	Σχηματική αναπαράσταση δικτύου CNN	16
Εικόνα 2.11:	Σχηματική αναπαράσταση κελιού RNN	17
Εικόνα 2.12:	Σχηματική αναπαράσταση κελιού LSTM	17
Εικόνα 2.13:	Σχηματική αναπαράσταση μοντέλου GAN	18
Εικόνα 2.14:	Σχηματική αναπαράσταση μοντέλου VAE	18
Εικόνα 2.15:	Σχηματική αναπαράσταση μοντέλου Seq2Seq	19
Εικόνα 3.1:	Διαδικασία υλοποίησης του προτεινόμενου συστήματος	21
Εικόνα 3.2:	Διαδικασία χρήσης του προτεινόμενου συστήματος	22
Εικόνα 3.3:	Κατανομή των νοτών τυμπάνων στο σύνολο δεδομένων	26
Εικόνα 3.4:	Κατανομή των νοτών τυμπάνων στα υποσύνολα δεδομένων	26
Εικόνα 3.5:	Διαχωρισμός του ηχητικού σήματος σε μουσικά μέτρα	27
Εικόνα 3.6:	Κυματομορφή ηχητικού σήματος	28
Εικόνα 3.7:	Φασματογράφημα ηχητικού σήματος	28
Εικόνα 3.8:	OSS χωρίς εξομάλυνση	29
Εικόνα 3.9:	OSS με εξομάλυνση	29
Εικόνα 3.10:	Χαρακτηριστικό 1: Σήμα Ισχύος Onset	30
Εικόνα 3.11:	Χαρακτηριστικό 2: Εγγύτητα Beat	31
Εικόνα 3.12:	Χαρακτηριστικό 3: Πρόοδος Μέτρου	31
Εικόνα 3.13:	Μοντέλο Seq2Seq του προτεινόμενου συστήματος	32
Εικόνα 4.1:	Πίνακας σύγχυσης δυαδικής ταξινόμησης	35
Εικόνα 4.2:	Συνάρτηση κόστους στα σύνολα εκπαίδευσης και επικύρωσης	38
Εικόνα 4.3:	Θεμελιώδεις μετρικές ταξινόμησης στο σύνολο εκπαίδευσης	39
Εικόνα 4.4:	Παράγωγες μετρικές ταξινόμησης στο σύνολο εκπαίδευσης	39

Εικόνα 4.5: Θεμελιώδεις μετρικές ταξινόμησης στο σύνολο επικύρωσης	40
Εικόνα 4.6: Παράγωγες μετρικές ταξινόμησης στο σύνολο επικύρωσης	40
Εικόνα 4.7: Δείκτης Jaccard στα σύνολα εκπαίδευσης και επικύρωσης	41
Εικόνα 4.8: Δημογραφικά στοιχεία έρευνας (I)	42
Εικόνα 4.9: Δημογραφικά στοιχεία έρευνας (II)	43
Εικόνα 4.10: Δημογραφικά στοιχεία έρευνας (III)	43
Εικόνα 4.11: Δημογραφικά στοιχεία έρευνας (IV)	44
Εικόνα 4.12: Δημογραφικά στοιχεία έρευνας (V)	44
Εικόνα 4.13: Ομάδα 1 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων	45
Εικόνα 4.14: Ομάδα 1 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	45
Εικόνα 4.15: Ομάδα 2 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων	46
Εικόνα 4.16: Ομάδα 2 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	46
Εικόνα 4.17: Ομάδα 3 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων	47
Εικόνα 4.18: Ομάδα 3 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	47
Εικόνα 4.19: Ομάδα 4 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων	48
Εικόνα 4.20: Ομάδα 4 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	48
Εικόνα 4.21: Ομάδα 5 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων	49
Εικόνα 4.22: Ομάδα 5 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	49
Εικόνα 4.23: Ομάδα 6 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων	50
Εικόνα 4.24: Ομάδα 6 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	50
Εικόνα 4.25: Κυματομορφή τεχνητού δείγματος	53
Εικόνα 4.26: Κυματομορφή ανθρώπινου δείγματος	53
Εικόνα 4.27: Φασματογράφημα τεχνητού δείγματος	54
Εικόνα 4.28: Φασματογράφημα ανθρώπινου δείγματος	54
Εικόνα Π.1: Λογότυπα εργαλείων λογισμικού	62

Λίστα Πινάκων

Πίνακας 2.1: Κατηγορίες προγραμμάτων MIDI	13
Πίνακας 3.1: Αντιστοίχιση των νοτών MIDI στα κρουστά του συστήματος	24
Πίνακας 3.2: Ανάθεση νοτών MIDI στα κρουστά του συστήματος	24
Πίνακας 3.3: Διαχωρισμός του συνόλου δεδομένων	25
Πίνακας 4.1: Αποτελέσματα εκπαίδευσης νευρωνικού δικτύου	38
Πίνακας 4.2: Αποτελέσματα μέσων βαθμολογιών δειγμάτων	51
Πίνακας 4.3: Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων	51
Πίνακας 4.4: Αποτελέσματα t-test (p-τιμές)	52
Πίνακας Π.1: Βιβλιοθήκες της Python	61

Λίστα Συντομογραφιών

ADAM	Adaptive Moment Estimation
ANN	Artificial Neural Network
BCE	Binary Cross-Entropy
BiLSTM	Bidirectional Long Short-Term Memory
BiRNN	Bidirectional Recurrent Neural Network
BPM	Beats Per Minute
CNN	Convolutional Neural Network
CQT	Constant Q Transform
DAW	Digital Audio Workstation
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
FFNN	Feedforward Neural Network
FN	False Negatives
FP	False Positives
GAN	Generative Adversarial Network
LSTM	Long Short-Term Memory
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MLP	Multilayer Perceptron
NLP	Natural Language Processing
OSS	Onset Strength Signal
RNN	Recurrent Neural Network
Seq2Seq	Sequence-to-Sequence
STFT	Short-Time Fourier Transform
TN	True Negatives
TP	True Positives
TPB	Ticks Per Beat
USPB	Microseconds Per Beat
VAE	Variational Autoencoder

Εισαγωγή

Η μουσική είναι μια μορφή τέχνης που αποτελεί αναπόσπαστο στοιχείο της ανθρώπινης ιστορίας. Η σύνθεση της μουσικής έχει τις ρίζες της βαθιά μέσα στην ανάγκη έκφρασης των ανθρωπίνων συναισθημάτων και αποτελεί προϊόν τόσο έμπινευσης όσο και ικανότητας. Ωστόσο, η στενή σχέση της μουσικής με τα μαθηματικά ώθησε πολλούς να αναζητήσουν αντισυμβατικές μεθόδους σύνθεσης μουσικής βασισμένες στη λογική και τις πιθανότητες. Έτσι, με το πέρασμα των χρόνων, γεννήθηκε και εδραιώθηκε το αντικείμενο της αλγορίθμικής σύνθεσης, δηλαδή της σύνθεσης πρωτότυπης μουσικής με τη χρήση αλγορίθμων.⁰

Σήμερα, η ψηφιακή επανάσταση έχει φέρει σαρωτικές αλλαγές στις μεθόδους καταγραφής, επεξεργασίας, αποθήκευσης, αναπαραγωγής και μετάδοσης του ήχου, καθιστώντας τη μουσική πιο προσβάσιμη από ποτέ. Ταυτόχρονα, η εκθετική ανάπτυξη της υπολογιστικής ισχύος και η άνοδος της τεχνητής νοημοσύνης, μέσω της καθιέρωσης τεχνικών μηχανικής μάθησης βασισμένων στα τεχνητά νευρωνικά δίκτυα, πυροδότησε ένα νέο ενδιαφέρον για το ερευνητικό πεδίο της αλγορίθμικής σύνθεσης. Έτσι, οι υπολογιστές σήμερα έχουν πλέον τη δυνατότητα να συνθέτουν μουσική τέτοιας ποιότητας και φυσικότητας ώστε αρκετές φορές να καθίσταται δύσκολη η διάκρισή της από μια ανθρώπινη μουσική δημιουργία.

Ένα από τα θεμελιώδη στοιχεία που απαρτίζουν τη μουσική και καλείται να κατανοήσει ένα σύστημα τεχνητής νοημοσύνης είναι ο ρυθμός. Αυτό που κάνει τον ρυθμό να ξεχωρίζει από άλλα μουσικά στοιχεία, όπως η μελωδία και η αρμονία, είναι η επίδραση που μπορεί να έχει στο ανθρώπινο σώμα, καθώς μπορεί να οδηγήσει στη μετατροπή των ηχητικών ερεθισμάτων σε κίνηση, φανερώνοντας και ενισχύοντας έτσι ακόμα περισσότερο την απόλαυση της μουσικής.

Στη σύγχρονη μουσική το πρωταρχικό ρυθμικό όργανο είναι τα τύμπανα, κοινώς ντραμς, ένα σύνολο από άτονα μεμβρανόφωνα και ιδιόφωνα κρουστά όργανα. Πρόκειται για το όργανο το οποίο αποτελεί τη ραχοκοκαλιά ενός μουσικού συνόλου, καθώς είναι αυτό που οδηγεί τον ρυθμό τον οποίο τα υπόλοιπα όργανα ακολουθούν. Έτσι, δεδομένου του ρόλου που έχουν τα τύμπανα σε μια μουσική εκτέλεση, τα τελευταία χρόνια έχει δοθεί ιδιαίτερη έμφαση σε ένα πολύ συγκεκριμένο κομμάτι του ευρύτερου πεδίου της αλγορίθμικής σύνθεσης, αυτό της αυτόματης παραγωγής ρυθμικών μοτίβων τυμπάνων.

⁰A Brief History of Algorithmic Composition:
<https://ccrma.stanford.edu/~blackrse/algorithm.html>

Η παρούσα εργασία επικεντρώνεται ακριβώς σε αυτό το αντικείμενο και έχει σκοπό να συνεισφέρει σημαντικά στη σχετική έρευνα. Η διάρθρωση των κεφαλαίων της έχει ως εξής:

- Στο Κεφάλαιο 1 γίνεται μία περιεκτική θεώρηση του προβλήματος της παραγωγής μοτίβων τυμπάνων, παρουσιάζονται κάποιες βασικές προσεγγίσεις και σχετικές υλοποιημένες εφαρμογές, ενώ καθορίζεται με σαφήνεια η στοχευμένη καινοτομία της εργασίας.
- Στο Κεφάλαιο 2 προσφέρεται το αναγκαίο θεωρητικό υπόβαθρο για την πλήρη κατανόηση της παρούσας μελέτης. Αυτό περιλαμβάνει μια σύντομη εισαγωγή στη θεωρία της μουσικής, του ψηφιακού ήχου και των τεχνητών νευρωνικών δικτύων, ενώ επίσης γίνεται μια αναφορά στα σημαντικότερα ερευνητικά έργα γύρω από το αντικείμενο της εργασίας.
- Στο Κεφάλαιο 3 αναλύεται η προτεινόμενη μεθοδολογία της εργασίας και το σύστημα που υλοποιήθηκε, περιγράφοντας τα δεδομένα που συλλέχθηκαν και τις μεθόδους που χρησιμοποιήθηκαν στα στάδια της προεπεξεργασίας, του νευρωνικού δικτύου, και της μετεπεξεργασίας.
- Στο Κεφάλαιο 4 πραγματοποιείται η αξιολόγηση του υλοποιημένου συστήματος, τόσο ποσοτικά με τη χρήση μετρικών αξιολόγησης, όσο και ποιοτικά μέσω της στατιστικής ανάλυσης αποτελεσμάτων που προέκυψαν από τη σύνταξη σχετικού ερωτηματολογίου.
- Στο Κεφάλαιο 5 καταγράφονται τα συμπεράσματα που προέκυψαν όσον αφορά τα αποτελέσματα και τη συνεισφορά της εργασίας, ενώ προτείνονται κάποιες μελλοντικές επεκτάσεις με στόχο τη βελτίωση της αποτελεσματικότητας και της λειτουργικότητας του προτεινόμενου συστήματος.

Τέλος, στην ενότητα Βιβλιογραφικές Αναφορές συγκεντρώνονται όλες οι παραπομπές σε πηγές που συνέβαλαν στη συγγραφή του παρόντος κειμένου, ενώ στο Παράρτημα παρατίθεται ο κώδικας που αναπτύχθηκε και παρουσιάζονται συνοπτικά τα εργαλεία λογισμικού που αξιοποιήθηκαν για τους σκοπούς της εργασίας.

Κεφάλαιο 1

Θεώρηση του Προβλήματος

1.1 Παραγωγή Μοτίβων Τυμπάνων

Η διαδικασία της παραγωγής μοτίβων τυμπάνων μπορεί να αναπαρασταθεί από ένα σύστημα το οποίο δέχεται κάποια πληροφορία στην είσοδο του και παράγει ένα μοτίβο τυμπάνων στην έξοδο του, όπως φαίνεται στην Εικόνα 1.1:



Εικόνα 1.1: Διαδικασία παραγωγής μοτίβων τυμπάνων

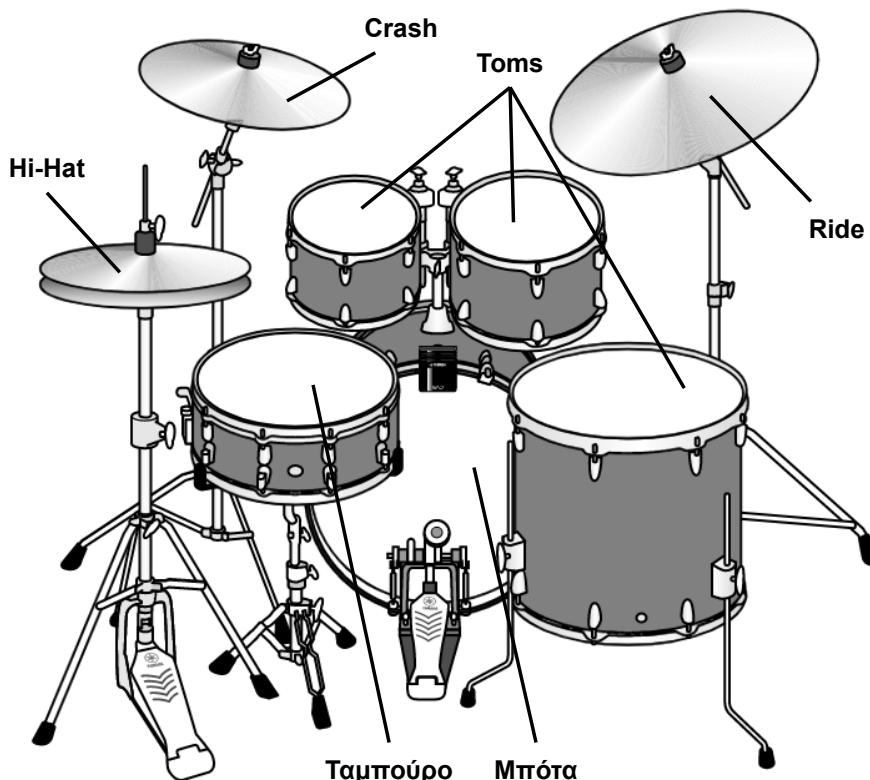
Η εισαγόμενη πληροφορία μπορεί να αναπαρασταθεί με διάφορους τρόπους, ανάλογα με την προσέγγιση του προβλήματος, ενώ το εξαγόμενο μοτίβο τυμπάνων αναπαρίσταται κατά κύριο λόγο συμβολικά. Αυτό συμβαίνει διότι το αντικείμενο της παραγωγής μοτίβων τυμπάνων ενδιαφέρεται για τα δεδομένα της μουσικής εκτέλεσης των τυμπάνων και όχι για το τελικό ηχητικό απόσπασμα που μπορεί να προκύψει μέσω ηχογράφησης ή ηχητικής σύνθεσης των παραγόμενων μοτίβων.

Για κάθε σύστημα παραγωγής μοτίβων τυμπάνων είναι απαραίτητο αρχικά να προσδιοριστούν τα επιμέρους κρουστά όργανα τα οποία θα συμμετέχουν στη μουσική εκτέλεση. Τα μέρη ενός τυπικού συνόλου τυμπάνων παρουσιάζονται στην Εικόνα 1.2 και είναι τα εξής:

- Μπότα (Kick): Το μεγαλύτερο τύμπανο του συνόλου, στηρίζεται στο έδαφος και παίζεται με τη χρήση πεταλιού στο δεξί πόδι. Χαρακτηρίζεται από κοφτό και βαθύ ήχο.
- Ταμπούρο (Snare): Τύμπανο με μεγάλο δυναμικό εύρος και ευελιξία στη χρήση του. Χαρακτηρίζεται από κοφτό και διαπεραστικό ήχο.
- Tom: Τύμπανα διαφορετικού μεγέθους τα οποία χρησιμοποιούνται για να προσδώσουν ενδιαφέρον στο παιχνίδι. Χαρακτηρίζονται από ήχο μικρής διάρκειας, το βάθος του οποίου είναι ανάλογο με το μέγεθος του τυμπάνου.

- Hi-Hat: Αποτελείται από δύο όμοια και αντικριστά πιατίνια η απόσταση των οποίων ελέγχεται με τη χρήση πεταλιού στο αριστερό πόδι. Χρησιμοποιείται για τη διατήρηση ενός σταθερού ρυθμού και μπορεί να παιχτεί σε κλειστή, ανοιχτή ή κάποια ενδιάμεση θέση. Χαρακτηρίζεται από συριστικό ήχο ο οποίος είναι κοφτός σε κλειστή θέση και διαρκής σε ανοιχτή θέση. Μπορεί επίσης να παιχτεί ανοιγοκλείνοντας το γρήγορα με το πετάλι, παράγοντας έτσι έναν κοφτό ήχο παρόμοιο με αυτόν της κλειστής θέσης.
- Ride: Πιατίνι το οποίο χρησιμοποιείται για τη διατήρηση ενός σταθερού ρυθμού. Χαρακτηρίζεται από ήχο μεσαίας διάρκειας που θυμίζει καμπάνα.
- Crash: Πιατίνι το οποίο χρησιμοποιείται αραιά για να προσδώσει έμφαση και δυναμισμό στο παίξιμο. Χαρακτηρίζεται από οξύ και διαπεραστικό ήχο μεγάλης διάρκειας.

Αξίζει να σημειωθεί ότι το πλήθος καθενός από τα παραπάνω όργανα ποικίλει από σύνολο σε σύνολο, ενώ συχνά συμπεριλαμβάνονται και άλλα πιατίνια όπως το splash και το china ή ακόμα και άλλα κρουστά όργανα.



Εικόνα 1.2: Τα μέρη ενός τυπικού συνόλου τυμπάνων¹

¹Πηγή πρωτότυπης εικόνας:

https://manual.yamaha.com/mi/bo/ead10/en/05_sad_en.html

1.2 Βασικές Προσεγγίσεις

Ανάλογα με τη φύση της επιθυμητής εφαρμογής, το πρόβλημα της παραγωγής μοτίβων τυμπάνων μπορεί να αντιμετωπιστεί με διάφορες προσεγγίσεις. Αυτές οι προσεγγίσεις εξαρτώνται σε πρώτο επίπεδο από την αναπαράσταση της πληροφορίας που δέχεται στην είσοδο του το σύστημα. Έτσι, οι βασικότεροι τύποι εισόδου, οι οποίοι μάλιστα συχνά συνδυάζονται μεταξύ τους, είναι οι εξής:

- Τυχαίος Θόρυβος: [1], [2]

Πρόκειται για τυχαία διανύσματα τα οποία έχουν ως αποτέλεσμα την παραγωγή τυχαιοποιημένων μοτίβων τυμπάνων.

- Παράμετρος Ελέγχου: [1]–[5]

Πρόκειται για παραμέτρους που ρυθμίζονται από τον χρήστη της εφαρμογής και αφορούν χαρακτηριστικά όπως το επιθυμητό μουσικό είδος, η ηχηρότητα και η πολυπλοκότητα των παραγόμενων μοτίβων, η πυκνότητα των επιμέρους οργάνων, κ.ά.

- Ρυθμική Ακολουθία: [4]–[11]

Πρόκειται για ακολουθίες που μπορούν να αναπαριστούν ρυθμικές ώσεις, ημιτελή ή ακόμα και ολοκληρωμένα μοτίβα τυμπάνων, και λειτουργούν ως *σπόροι* (seeds) για την παραγωγή νέων ρυθμικών μοτίβων.

- Συμβολικά Δεδομένα: [8]–[12]

Πρόκειται για δεδομένα που μπορούν να περιλαμβάνουν πληροφορία για τη θέση και τη διάρκεια μουσικών νοτών, για την ταχύτητα, τη δομή και την οργάνωση ενός μουσικού κομματιού, κ.ά.

- Ηχητικά Δεδομένα: [12]

Πρόκειται για δεδομένα ηχητικών σημάτων σε ακατέργαστη μορφή τα οποία επεξεργάζονται με κατάλληλες τεχνικές από το σύστημα.

1.3 Εφαρμογές

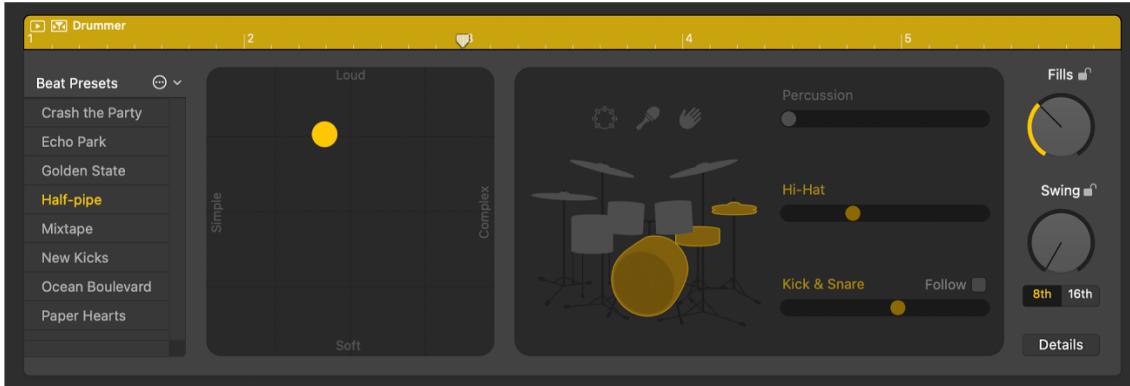
Απότερος σκοπός του αντικειμένου της παραγωγής μοτίβων τυμπάνων είναι η υλοποίηση εφαρμογών οι οποίες θα βοηθούν τους μουσικούς στη διαδικασία της σύνθεσης, παρέχοντας τους χρήσιμα μοτίβα τυμπάνων. Οι εφαρμογές αυτές μπορούν να είναι αυτόνομες (standalone) ή να αποτελούν άρθρωμα (plugin) κάποιου Ψηφιακού Σταθμού Επεξεργασίας Ήχου (Digital Audio Workstation - **DAW**). Στη συνέχεια παρουσιάζονται κάποιες από αυτές τις διαθέσιμες εφαρμογές.

1.3.1 Logic Pro Drummer

Ο *Drummer* αποτελεί αποκλειστικό άρθρωμα του *Logic Pro*, του εμπορικού DAW της *Apple*.² Το γραφικό περιβάλλον του *Drummer* παρουσιάζεται στην Εικόνα 1.3:

²Overview of Drummer in Logic Pro:

<https://support.apple.com/guide/logicpro/overview-lgcp72435777>

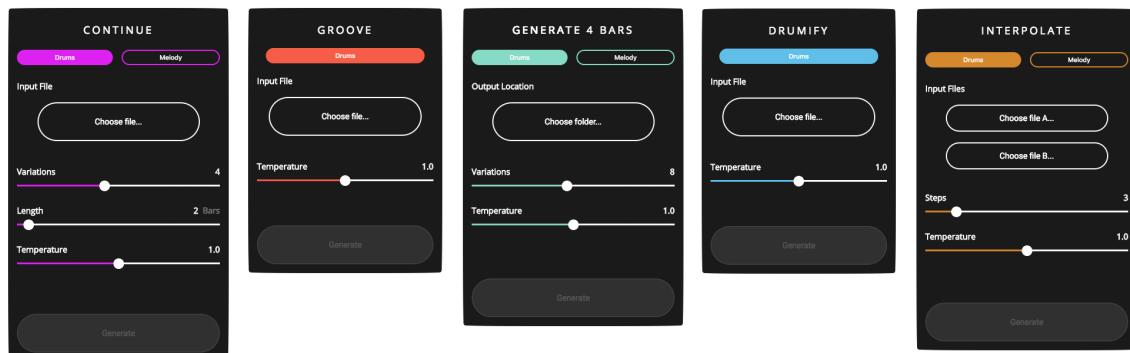


Εικόνα 1.3: Το γραφικό περιβάλλον του *Drummer*³

Όπως φαίνεται, ο *Drummer* δίνει στο χρήστη τη δυνατότητα να επιλέξει κάποια από τις προκαθορισμένες ρυθμίσεις (presets) αλλά και να φτιάξει πολύ εύκολα τις δικές του ρυθμίσεις, ελέγχοντας παραμέτρους και χαρακτηριστικά της μουσικής εκτέλεσης. Επιπλέον, μέσω της επιλογής *Follow*, ο *Drummer* μπορεί να αναλύσει ένα μουσικό απόσπασμα έτσι ώστε να προσαρμοστεί καλύτερα στον ρυθμό του. Έτσι, καθίσταται σαφές πως ο *Drummer* αποτελεί ένα πολύ χρήσιμο εργαλείο, χωρίς όμως να είναι γνωστό αν λειτουργεί με κατάλληλο χειρισμό προεπιλεγμένων μοτίβων τυμπάνων ή με πλήρη αξιοποίηση των δυνατοτήτων της τεχνητής νοημοσύνης.

1.3.2 Magenta Studio

Το *Magenta Studio* αποτελεί μια συλλογή εργαλείων ανοιχτού κώδικα που είναι διαθέσιμα ως αυτόνομες εφαρμογές και ως άρθρωμα του *Ableton Live*, του εμπορικού DAW της *Ableton*.⁴ Τα γραφικά περιβάλλοντα των εφαρμογών του *Magenta Studio* παρουσιάζονται στην Εικόνα 1.4:



Εικόνα 1.4: Τα γραφικά περιβάλλοντα των εφαρμογών του *Magenta Studio*⁵

³Πηγή πρωτότυπης εικόνας:

<https://support.apple.com/guide/logicpro/overview-1gcp72435777>

⁴Magenta Studio:

<https://magenta.tensorflow.org/studio>

⁵Πηγή πρωτότυπης εικόνας:

<https://magenta.tensorflow.org/studio/standalone>

Πρόκειται για πέντε ξεχωριστές εφαρμογές που προσεγγίζουν το πρόβλημα της παραγωγής μοτίβων τυμπάνων υπό διαφορετική σκοπιά η κάθε μία. Αναλυτικότερα, οι εφαρμογές αυτές είναι οι εξής:

- *Continue*: Δέχεται ένα μοτίβο τυμπάνων και παράγει νέα μοτίβα τα οποία θα μπορούσαν να αποτελούν χρονικά μουσική συνέχεια του αρχικού.
- *Groove*: Δέχεται ένα μοτίβο τυμπάνων και το προσαρμόζει έτσι ώστε να ακούγεται πιο αληθινό και φυσικό. Ουσιαστικά αντιμετωπίζεται το πρόβλημα του εξανδρωπισμού (*humanization*) των παραγόμενων μοτίβων τυμπάνων, εισάγοντας ατέλειες που προσομοιώνουν το ανθρώπινο παιξίμο [5], [7].
- *Generate*: Παράγει τυχαιοποιημένα μοτίβα τυμπάνων συγκεκριμένης διάρκειας.
- *Drumify*: Δέχεται μια ρυθμική ακολουθία ή συμβολικά δεδομένα ενός μουσικού αποσπάσματος και παράγει ρυθμικά συμβατά μοτίβα τυμπάνων.
- *Interpolate*: Δέχεται δύο διαφορετικά μοτίβα τυμπάνων και παράγει ενδιάμεσα μοτίβα που έχουν ως σκοπό την ομαλή μετάβαση από το ένα μοτίβο στο άλλο.

1.4 Στοχευμένη Καινοτομία

Έχοντας περιγράψει αναλυτικά το πρόβλημα της παραγωγής μοτίβων τυμπάνων, και έχοντας παρουσιάσει τις βασικές προσεγγίσεις και κάποιες από τις σχετικές υλοποιημένες εφαρμογές, είναι πλέον εφικτό να καθοριστεί με σαφήνεια η προσέγγιση και η στοχευμένη καινοτομία της παρούσας εργασίας. Έτσι λοιπόν, η παρούσα εργασία προσεγγίζει το πρόβλημα της παραγωγής μοτίβων τυμπάνων με την υλοποίηση ενός συστήματος το οποίο θα βασίζεται κυρίως σε ηχητικά δεδομένα και σε κατάλληλες τεχνικές επεξεργασίας τους, μια προσέγγιση η οποία απουσιάζει σε τεράστιο βαθμό από τη σχετική έρευνα. Πρόκειται επομένως για ένα σύστημα εμπνευσμένο από τη λειτουργία *Follow* του *Drummer*, με σημαντική διαφορά όμως πως, σε αντίθεση με τον *Drummer*, το εν λόγω σύστημα θα είναι ικανό να παράγει συνοδευτικά μοτίβα τυμπάνων χωρίς την ανάγκη ρύθμισης παραμέτρων από τον χρήστη. Το κίνητρο για μια τέτοια προσέγγιση είναι η προσομοίωση ενός αληθινού ντράμερ, ο οποίος χρειάζεται απλά να ακούσει ένα μουσικό κομμάτι για να παίξει συνοδευτικά τύμπανα. Ωστόσο για να γίνει αυτή η προσομοίωση εφικτή θα χρειαστεί να αξιοποιηθούν και κάποια συμβολικά δεδομένα τα οποία περιέχουν πληροφορία για τη δομή του μουσικού κομματιού την οποία ένας έμπειρος ντράμερ μπορεί εύκολα να αντιληφθεί. Έτσι, αυτός ο συνδυασμός ηχητικών και συμβολικών δεδομένων για την παραγωγή μοτίβων τυμπάνων, αξιοποιώντας τεχνικές επεξεργασίας από άλλα ερευνητικά πεδία και εκμηδενίζοντας την ανάγκη παρεμβολής από τον χρήστη, συνιστά τη στοχευμένη καινοτομία της εργασίας.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Στοιχεία Μουσικής Θεωρίας

Για την υλοποίηση ενός συστήματος παραγωγής μοτίβων τυμπάνων είναι αρχικά απαραίτητη η κατανόηση της έννοιας του ρυθμού σε Θεωρητικό επίπεδο. Κάθε ρυθμικό μοτίβο αποτελεί μια ακολουθία από νότες οι οποίες έχουν κάποια σχετική διάρκεια που εκφράζεται με την *αξία*. Η αξία αναφοράς η οποία έχει τη μεγαλύτερη διάρκεια ονομάζεται *ολόκληρο* και όλες οι υπόλοιπες αξίες λαμβάνονται αναδρομικά με υποδιπλασιασμό. Έτσι προκύπτει το *μισό*, το *τέταρτο*, το *όγδοο*, κ.ο.κ. Τα σύμβολα των βασικότερων αξιών στη μουσική σημειογραφία παρουσιάζονται στην Εικόνα 2.1:

Αξία	Σύμβολο
Ολόκληρο	○
Μισό	♩
Τέταρτο	♪
Όγδοο	♫
Δέκατο Έκτο	♪♪
Τριακοστό Δεύτερο	♪♪♪
Εξηκοστό Τέταρτο	♪♪♪♪

Εικόνα 2.1: Οι βασικότερες αξίες στη μουσική

Η διάρκεια μιας νότας μπορεί να επεκταθεί με τη *στιγμή διαρκείας* και τη *σύζευξη διαρκείας*. Στην πρώτη περίπτωση η αξία της νότας παρατείνεται κατά το ήμιση της και η νότα ονομάζεται *παρεστιγμένη*, ενώ στη δεύτερη περίπτωση δύο ή περισσότερες νότες συνενώνονται συγχωνεύοντας τις αξίες τους και ονομάζονται *συζευγμένες*.

Με βάση τα παραπάνω, προκύπτει εύκολα το συμπέρασμα ότι μια παρεστιγμένη νότα συνίσταται από δύο συζευγμένες νότες αλλά αυτό δεν σημαίνει πως δύο συζευγμένες νότες συνιστούν πάντα μία παρεστιγμένη νότα. Ο συμβολισμός των παρεστιγμένων και των συζευγμένων νοτών καθώς και η σχέση ισοδυναμίας μεταξύ τους απεικονίζονται στα παραδείγματα της Εικόνας 2.2:

Παρεστιγμένη Νότα	=	Συζευγμένες Νότες
	=	
	=	
	=	
	=	

Εικόνα 2.2: Παρεστιγμένες και συζευγμένες νότες

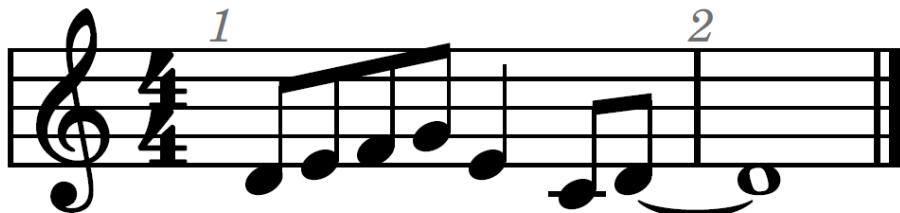
Επιπλέον, είναι δυνατόν μια αξία να διαιρεθεί με παράγοντα διάφορο του 2. Η πιο συνηθισμένη διαίρεση είναι αυτή που χωρίζει την αξία σε τρία ίσα μέρη τα οποία συνιστούν ένα τρίγχο. Ο συμβολισμός των τριήχων καθώς και οι αντίστοιχες ισοδύναμες αξίες τους απεικονίζονται στα παραδείγματα της Εικόνας 2.3:

Τρίηχο	=	Ισοδύναμη Αξία
	=	
	=	
	=	
	=	

Εικόνα 2.3: Τρίηχα και ισοδύναμες αξίες

Η πλήρης απουσία νοτών για κάποια χρονική διάρκεια περιγράφεται με την έννοια της *παύσης* η οποία υποδηλώνει την ύπαρξη σιωπής στη μουσική. Οι αξίες των παύσεων ακολουθούν τους ίδιους ακριβώς κανόνες με τις αξίες των νοτών που περιγράφηκαν παραπάνω. Έτσι, αφού καθοριστούν οι αξίες τους, οι νότες και οι παύσεις ενός μουσικού κομματιού οργανώνονται σε ομάδες που ονομάζονται *μέτρα* και έχουν σκοπό τον διαχωρισμό του κομματιού σε μικρότερα αποσπάσματα.

Κάθε μουσικό μέτρο περιγράφεται από ένα κλάσμα το οποίο καθορίζει τη χρονική διάρκεια του και μπορεί να διαφέρει από μέτρο σε μέτρο. Ο παρονομαστής αυτού του κλάσματος καθορίζει την αξία που θα χρησιμοποιηθεί ως μονάδα μέτρησης του χρόνου ενώ ο αριθμητής καθορίζει πόσοι περιοδικοί χτύποι (beat) αυτής της αξίας χωράνε στο μέτρο. Για παράδειγμα, ένα μέτρο 4/4, το οποίο είναι το συνηθέστερο στη σύγχρονη μουσική, χωράει νότες και παύσεις συνολικής διάρκειας τεσσάρων beat, όπου σε κάθε beat αναλογεί η αξία του τετάρτου. Στην Εικόνα 2.4 παρουσιάζεται μια παρτιτούρα μελωδίας διάρκειας δύο μέτρων 4/4:



Εικόνα 2.4: Παρτιτούρα μελωδίας διάρκειας δύο μέτρων 4/4

Οι αξίες και τα μέτρα χαρακτηρίζονται από μια σχετική διάρκεια η οποία όμως δεν παρέχει κάποια πληροφορία για τον απόλυτο χρόνο που διαρκούν οι διάφορες νότες και παύσεις. Ο χρόνος αυτός καθορίζεται από το *tempo* το οποίο μετριέται σε χτύπους ανα λεπτό (Beats Per Minute - **BPM**) και ορίζει την ταχύτητα του κομματιού. Για παράδειγμα, σε ένα κομμάτι των 60 BPM εμφανίζονται 60 beat σε 60 δευτερόλεπτα και επομένως η διάρκεια του ενός beat είναι 1 δευτερόλεπτο. Ανεξάρτητα από την αξία του beat βάσει της οποίας υπολογίζεται η διάρκεια ενός μέτρου, συνηθίζεται η αξία του beat βάσει της οποίας καθορίζεται το *tempo* να είναι αυτή του τετάρτου. Αυτό σημαίνει ότι, για παράδειγμα, σε ένα μέτρο 7/8 χωράνε επτά beat αξίας ογδόου αλλά αν το αντίστοιχο *tempo* είναι 60 BPM τότε η αξία του τετάρτου είναι αυτή που έχει διάρκεια ενός δευτερολέπτου ενώ αντίστοιχα η αξία του ογδόου θα έχει διάρκεια μισού δευτερολέπτου. Τέλος, η τιμή του *tempo* μπορεί να παραμένει σταθερή σε όλη τη διάρκεια ενός μουσικού κομματιού ή να αλλάζει. Έτσι, είναι δυνατόν το *tempo* να αποκτά νέα τιμή στην αρχή ή στο ενδιάμεσο ενός μέτρου, ή ακόμα και να μεταβάλλεται σταδιακά κατά τη διάρκεια ενός ή περισσότερων μέτρων.

2.2 Ψηφιακός Ήχος

Ο ήχος αποτελεί ένα φυσικό φαινόμενο το οποίο μοντελοποιείται ως ένα σήμα, δηλαδή ως μία συνάρτηση η οποία φέρει κάποια πληροφορία. Σε αντίθεση με τον αναλογικό κόσμο όπου όλα τα σήματα είναι συνεχείς συναρτήσεις, στον ψηφιακό κόσμο των υπολογιστών τα ηχητικά σήματα είναι διακριτά και οι τεχνικές διαχείρισης τους αποτελούν αντικείμενο της Ψηφιακής Επεξεργασίας Σήματος (Digital Signal Processing - **DSP**) [13]. Η μουσική, ως ήχος, μπορεί σαφώς να προσεγγιστεί ως ένα ηχητικό σήμα, αλλά χάρη στη μουσική θεωρία και σημειογραφία είναι δυνατό να αναπαρασταθεί και συμβολικά βάσει κάποιου καθιερωμένου τεχνικού προτύπου, το δημοφιλέστερο εκ των οποίων είναι η Ψηφιακή Διασύνδεση Μουσικών Οργάνων (Musical Instrument Digital Interface - **MIDI**) [14].

2.2.1 Ψηφιακή Επεξεργασία Σήματος

Ένα διακριτό ηχητικό σήμα μπορεί να αναπαρασταθεί με τους εξής τρόπους:

- ως κυματομορφή στο πεδίο του χρόνου
- ως φάσμα στο πεδίο της συχνότητας
- ως φασματογράφημα στο πεδίο του χρόνου και της συχνότητας

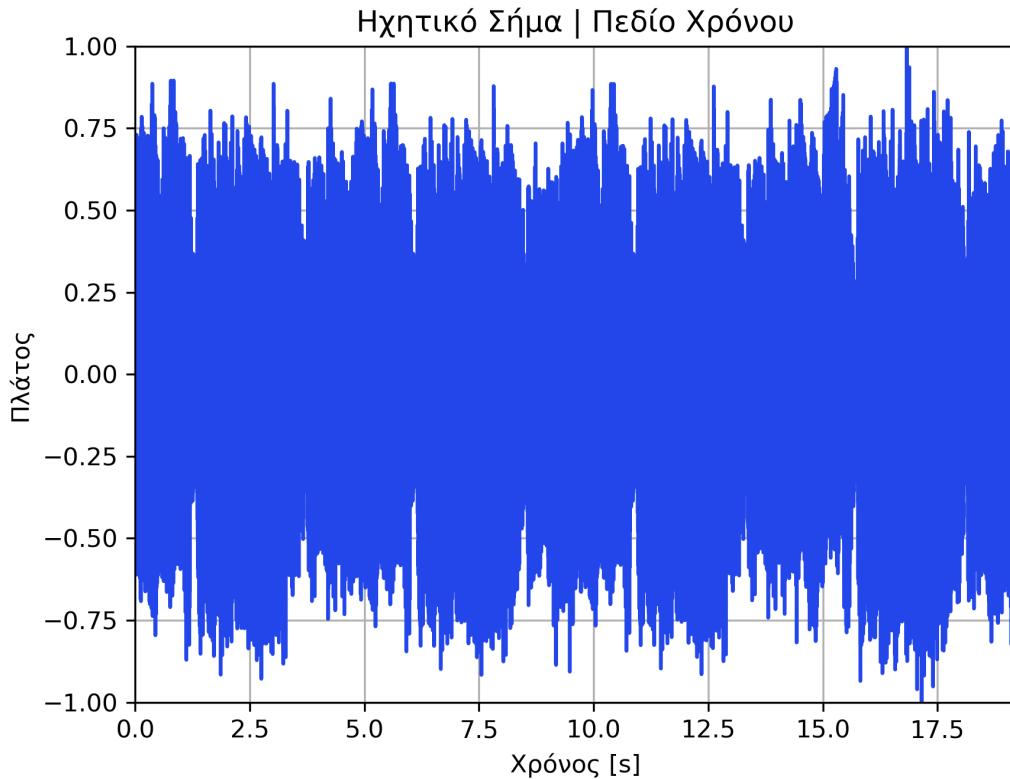
Η αναπαράσταση στο πεδίο του χρόνου περιγράφει τις διακυμάνσεις της έντασης του ηχητικού σήματος και προκύπτει μέσω περιοδικής δειγματοληψίας του αντίστοιχου αναλογικού σήματος $x_a(t)$. Η διαδικασία αυτή περιγράφεται από την Εξίσωση 2.1:

$$x[n] = x_a(nT), \quad 0 \leq n \leq N - 1 \quad (2.1)$$

όπου N είναι ο αριθμός δειγμάτων της διακριτής ακολουθίας $x[n]$, που χωρίς βλάβη της γενικότητας μπορεί να θεωρηθεί άρτιος, και T είναι η περίοδος δειγματοληψίας. Η αντίστοιχη συχνότητα δειγματοληψίας δίνεται από την Εξίσωση 2.2:

$$f_s = \frac{1}{T} \quad (2.2)$$

Ένα ηχητικό σήμα στο πεδίο του χρόνου παρουσιάζεται στην Εικόνα 2.5:

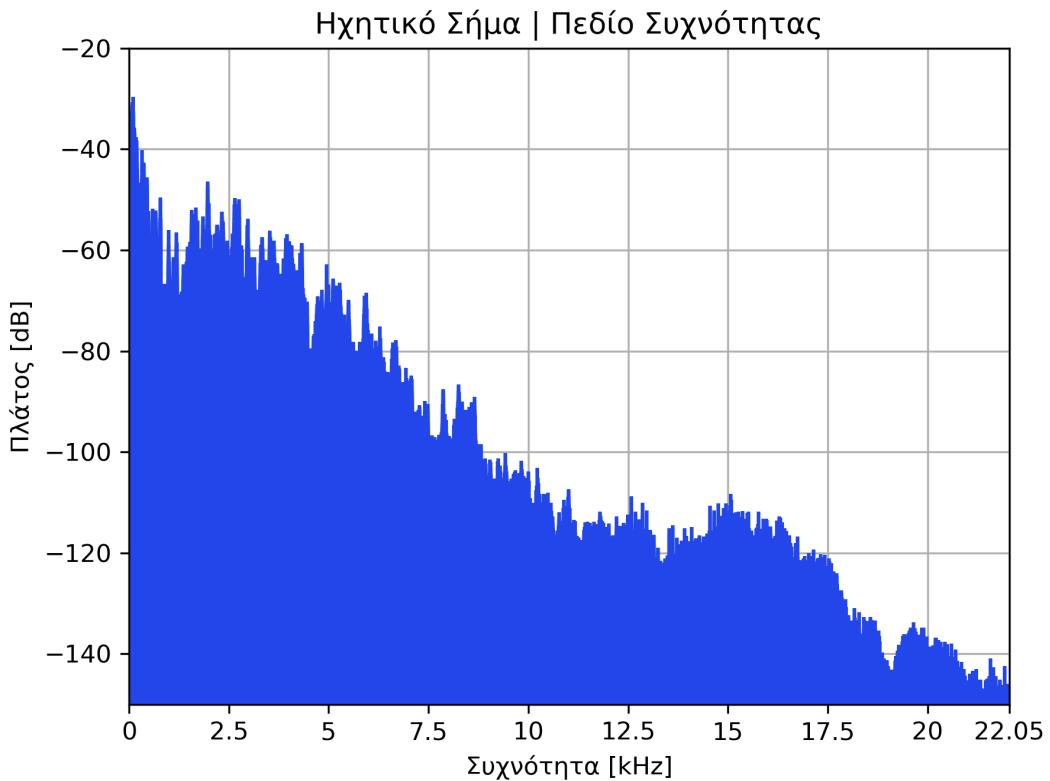


Εικόνα 2.5: Ηχητικό σήμα στο πεδίο του χρόνου

Η αναπαράσταση στο πεδίο της συχνότητας περιγράφει το φασματικό περιεχόμενο του ηχητικού σήματος και προκύπτει εφαρμόζοντας στη διακριτή ακολουθία $x[n]$ τον Διακριτό Μετασχηματισμό Φουριέ (Discrete Fourier Transform - **DFT**). Η διαδικασία αυτή περιγράφεται από την Εξίσωση 2.3:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}, \quad 0 \leq k \leq N-1 \quad (2.3)$$

Η μετασχηματισμένη ακολουθία $X[k]$ είναι μιγαδική, ενώ το πλάτος $|X[k]|$ είναι συμμετρικό γύρω από τον άξονα των συχνοτήτων, καθώς τα ηχητικά σήματα είναι πραγματικά. Συνεπώς, αρκεί να διατηρηθούν μόνο τα δείγματα για $k \in [0, N/2]$ τα οποία αντιστοιχούν στο εύρος συχνοτήτων $[0, f_s/2]$. Προκύπτει έτσι το συμπέρασμα πως η μέγιστη συχνότητα που μπορεί να αναπαρασταθεί στο φάσμα είναι η μισή της συχνότητας δειγματοληψίας. Η παρατήρηση αυτή συμπληρώνεται από το Θεώρημα Nyquist το οποίο ορίζει πως η συχνότητα δειγματοληψίας πρέπει να είναι τουλάχιστον διπλάσια από τη μέγιστη συχνότητα του αναλογικού σήματος. Για το λόγο αυτό, συνηθίζεται ο ήχος να δειγματοληπτείται με συχνότητα $f_s = 44.1\text{kHz}$, καθώς έτσι καλύπτεται πλήρως το εύρος συχνοτήτων $20\text{Hz} - 20\text{kHz}$ που αντιλαμβάνεται το ανθρώπινο αυτό. Επιπλέον, λόγω της λογαριθμικής αντίληψης της έντασης του ήχου, το φάσμα απεικονίζεται σε κλίμακα dB με βάση τον μετασχηματισμό $10 \log_{10} |X[k]|^2$. Ένα ηχητικό σήμα στο πεδίο της συχνότητας παρουσιάζεται στην Εικόνα 2.6:

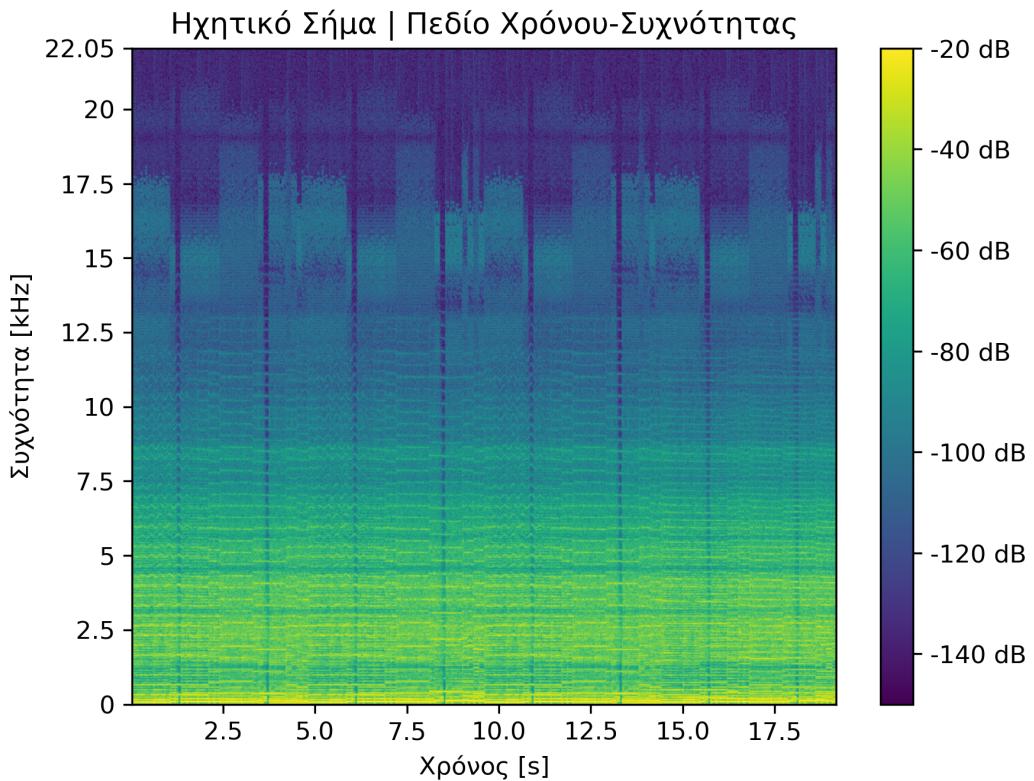


Εικόνα 2.6: Ηχητικό σήμα στο πεδίο της συχνότητας

Η αναπαράσταση στο πεδίο της συχνότητας, αν και περιέχει πληροφορία η οποία είναι χρήσιμη για κάποιες εφαρμογές, έχει το ελάττωμα ότι είναι στατική, δεν περιγράφει δηλαδή τις αλλαγές που υφίσταται το ηχητικό σήμα στη διάρκεια του χρόνου. Τη λύση σε αυτό το πρόβλημα δίνει το φασματογράφημα όπου κομμάτια του σήματος απομονώνονται χρησιμοποιώντας χρονικά κυλιόμενα και επικαλυπτόμενα παράθυρα, υπολογίζοντας για κάθε ξεχωριστό κομμάτι το αντίστοιχο φάσμα. Η αναπαράσταση στο πεδίο του χρόνου και της συχνότητας περιγράφει τις διακυμάνσεις του φασματικού περιεχομένου του ηχητικού σήματος και προκύπτει εφαρμόζοντας στη διακριτή ακολουθία $x[n]$ τον Βραχύχρονο Μετασχηματισμό Φουριέ (Short-Time Fourier Transform - **STFT**). Η διαδικασία αυτή περιγράφεται από την Εξίσωση 2.4:

$$X[n, k] = \sum_{l=0}^{L-1} x[nH + l] w[l] e^{-j\frac{2\pi}{N} kl}, \quad \begin{cases} 0 \leq n \leq \left\lfloor \frac{N-(L-H)}{H} \right\rfloor \\ 0 \leq k \leq N-1 \end{cases} \quad (2.4)$$

όπου L είναι το μήκος και H η μετατόπιση του παραθύρου $w[l]$ σε δείγματα. Παρόμοια με την ανάλυση που έγινε για το πεδίο της συχνότητας, και σε αυτή την περίπτωση αρκεί να διατηρηθούν μόνο τα δείγματα για $k \in [0, N/2]$, τα οποία αντιστοιχούν στο εύρος συχνοτήτων $[0, f_s/2]$, ενώ το φασματογράφημα απεικονίζεται σε κλίμακα dB με βάση τον μετασχηματισμό $10 \log_{10} |X[n, k]|^2$. Ένα ηχητικό σήμα στο πεδίο του χρόνου και της συχνότητας παρουσιάζεται στην Εικόνα 2.7:



Εικόνα 2.7: Ηχητικό σήμα στο πεδίο του χρόνου και της συχνότητας

2.2.2 Ψηφιακή Διασύνδεση Μουσικών Οργάνων

Το πρότυπο MIDI αποτελεί ένα πρωτόκολλο επικοινωνίας μεταξύ ηλεκτρονικών μουσικών οργάνων και υπολογιστών το οποίο συντηρείται και αναπτύσσεται από τη μη κερδοσκοπική οργάνωση The MIDI Association.⁶ Βασικό χαρακτηριστικό του προτύπου είναι ότι η πληροφορία μεταδίδεται μέσω μηνυμάτων τα οποία αναπαριστούν συμβολικά μία μουσική εκτέλεση. Συνεπώς, τα αρχεία MIDI τα οποία έχουν επέκταση .mid δεν περιέχουν ηχητικά δεδομένα, αλλά εντολές σχετικά με τις νότες που παράγουν τα διάφορα μουσικά όργανα, όπως το τονικό ύψος, η ένταση και η διάρκεια τους. Ακόμα, μεταδίδονται μετα-μηνύματα (meta messages) τα οποία μπορούν να περιέχουν γενικότερη πληροφορία για το μουσικό κομμάτι, όπως αλλαγές στο tempo ή στο κλάσμα χρόνου των μέτρων.

Η προδιαγραφή General MIDI 1 του προτύπου ορίζει ότι ένα αρχείο MIDI μπορεί να περιέχει μέχρι 16 διαφορετικά κανάλια οργάνων, εκ των οποίων το κανάλι 10 προορίζεται αποκλειστικά για τύμπανα και κρουστά. Σε κάθε άλλο κανάλι μπορεί να ανατεθεί ένας από τους 128 διαθέσιμους ήχους οι οποίοι αποκαλούνται *προγράμματα* και διακρίνονται στις κατηγορίες που παρουσιάζονται στον Πίνακα 2.1:

Πίνακας 2.1: Κατηγορίες προγραμμάτων MIDI

Αριθμός Προγράμματος	Κατηγορία
1-8	Piano
9-16	Chromatic Percussion
17-24	Organ
25-32	Guitar
33-40	Bass
41-48	Strings
49-56	Ensemble
57-64	Brass
65-72	Reed
73-80	Pipe
81-88	Synth Lead
89-96	Synth Pad
97-104	Synth Effects
105-112	Ethnic
113-120	Percussive
121-128	Sound Effects

Τα προγράμματα δεν διαθέτουν ηχητικά χαρακτηριστικά, αλλά αποτελούν εντολές προς κάποιο εξωτερικό λογισμικό το οποίο μετατρέπει την MIDI πληροφορία σε ήχο.

⁶The MIDI Association: <https://www.midi.org/>

Τα μηνύματα MIDI μεταδίδονται σειριακά με μία χρονική διαφορά η οποία ονομάζεται **χρόνος δέλτα** και μετριέται σε χτύπους (ticks). Έτσι, η διάρκεια μιας νότας σε ticks υπολογίζεται αθροίζοντας τους χρόνους δέλτα που μεσολαβούν μεταξύ των μηνυμάτων "Note On" και "Note Off" για την αντίστοιχη νότα. Τα ticks αποτελούν τη μονάδα μέτρησης χρόνου του προτύπου και μεταφράζονται σε πραγματικό χρόνο μέσω δύο παραμέτρων. Η πρώτη από αυτές είναι σταθερή για κάθε αρχείο MIDI και ορίζει τα ticks ανά beat (Ticks Per Beat - **TPB**), όπου η αξία του beat αναλογεί σε αυτή του τετάρτου. Η δεύτερη είναι το tempo, το οποίο βάσει προτύπου μετριέται σε μια ανά beat (Microseconds Per Beat - **USPB**) και η τιμή του μπορεί εύκολα να μετατραπεί σε τιμή BPM μέσω της Εξίσωσης 2.5:

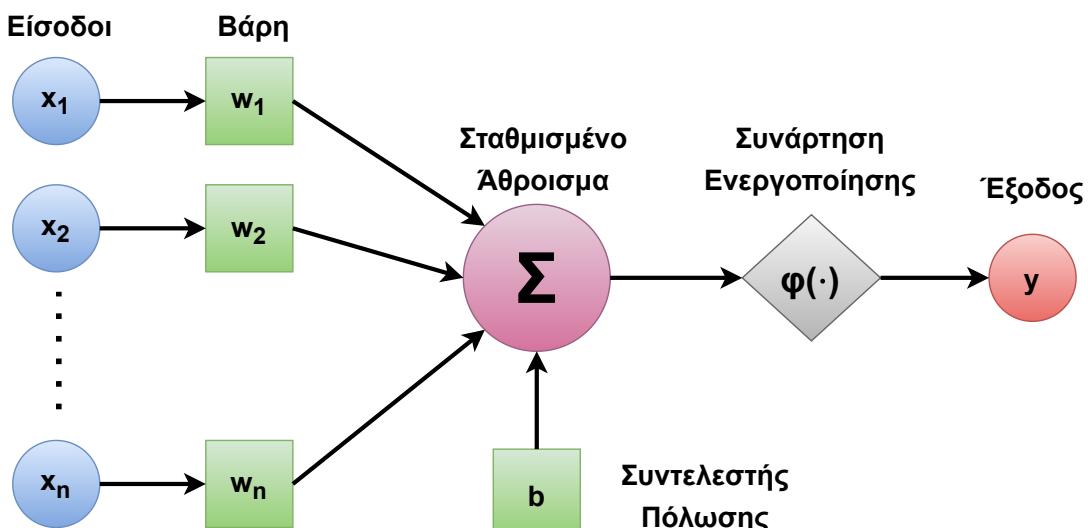
$$BPM = \frac{60 \cdot 10^6}{USPB} \quad (2.5)$$

Έτσι, η μετατροπή των ticks σε δευτερόλεπτα γίνεται βάσει της Εξίσωσης 2.6:

$$seconds = \frac{60}{TPB \cdot BPM} ticks \quad (2.6)$$

2.3 Νευρωνικά Δίκτυα

Ένα τεχνητό νευρωνικό δίκτυο (Artificial Neural Network - **ANN**) [15] αποτελείται από κόμβους οι οποίοι ονομάζονται **νευρώνες** και οργανώνονται σε **επίπεδα** που συνδέονται μεταξύ τους μεταβιβάζοντας πληροφορία από την είσοδο στην έξοδο του δικτύου. Κάθε νευρώνας δέχεται έναν αριθμό εισόδων x_i , οι οποίες σταθμίζονται με βάρη w_i και αθροίζονται, προσθέτοντας επιπλέον έναν συντελεστή πόλωσης b . Στη συνέχεια, το άθροισμα αυτό περνάει μέσα από μία **συνάρτηση ενεργοποίησης** $\phi(\cdot)$, η οποία περιορίζει την τελική έξοδο y σε ένα επιθυμητό εύρος τιμών. Μια σχηματική αναπαράσταση ενός νευρώνα παρουσιάζεται στην Εικόνα 2.8:



Εικόνα 2.8: Σχηματική αναπαράσταση νευρώνα

Η έξοδος ενός νευρώνα συναρτήσει των εισόδων του, των βαρών και του συντελεστή πόλωσης περιγράφεται από την Εξίσωση 2.7:

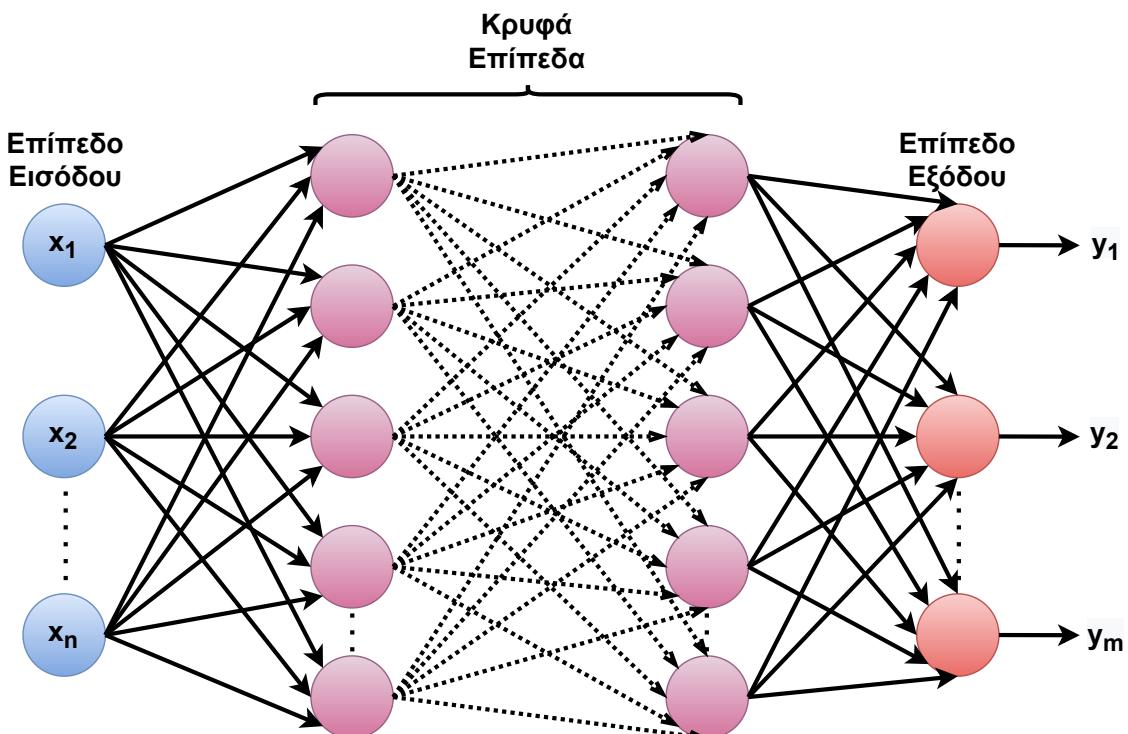
$$y = \phi\left(\sum_{i=0}^n w_i x_i + b\right) \quad (2.7)$$

Ένα νευρωνικό δίκτυο εκπαιδεύεται με κάποιο σύνολο δεδομένων ανανεώνοντας συνεχώς τα βάρη και τους συντελεστές πόλωσης των νευρώνων του σε στάδια που ονομάζονται **εποχές**, έτσι ώστε να ελαχιστοποιηθεί κάποια συνάρτηση κόστους $L(\cdot)$ η οποία περιγράφει την απόκλιση των προβλεπόμενων εξόδων από τις επιθυμητές.

2.3.1 Βασικές Αρχιτεκτονικές

Οι δύο κύριες κατηγορίες νευρωνικών δικτύων είναι τα Νευρωνικά Δίκτυα Πρόσθιας Τροφοδότησης (Feedforward Neural Networks - **FFNNs**) και τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks - **RNNs**). Η ειδοποιός διαφορά τους είναι ότι στα FFNN η πληροφορία κατευθύνεται με μοναδική κατεύθυνση από την είσοδο προς την έξοδο, ενώ στα RNN οι νευρώνες διαθέτουν κλάδους ανάδρασης.

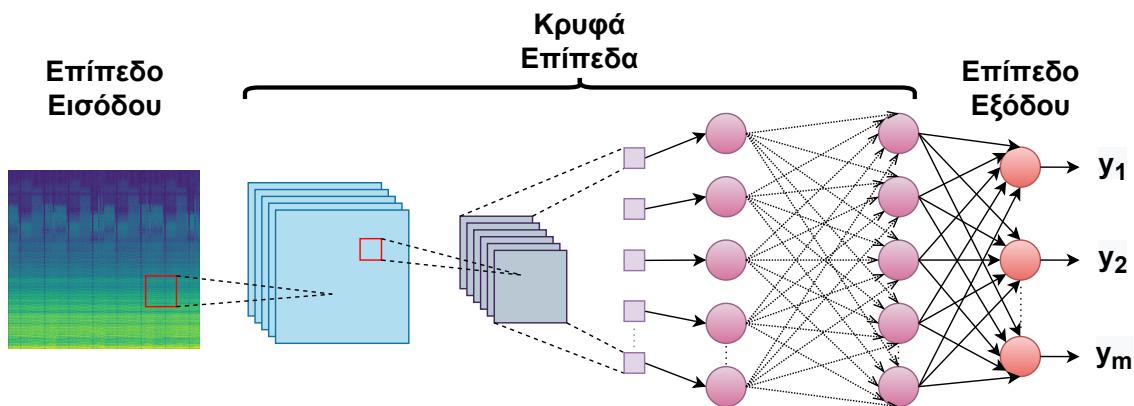
Η βασικότερη αρχιτεκτονική FFNN είναι το Πολυεπίπεδο Perceptron (Multilayer Perceptron - **MLP**), το οποίο αποτελείται από ένα επίπεδο εισόδου, ένα επίπεδο εξόδου και ενδιάμεσα κρυφά επίπεδα. Η σχηματική αναπαράσταση ενός δικτύου MLP παρουσιάζεται στην Εικόνα 2.9:



Εικόνα 2.9: Σχηματική αναπαράσταση δικτύου MLP

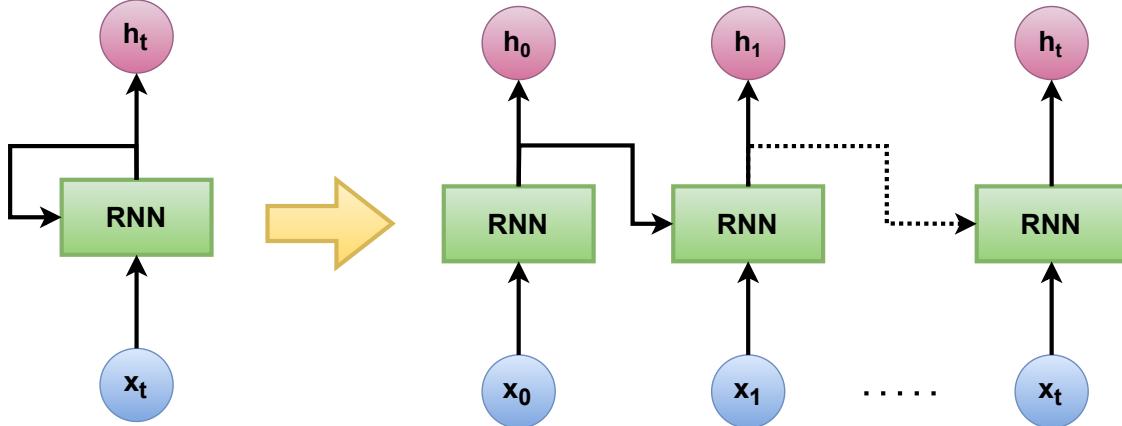
Χαρακτηριστικό του MLP είναι ότι πρόκειται για ένα πλήρως συνδεδεμένο δίκτυο, δηλαδή όλοι οι κόμβοι ενός επιπέδου συνδέονται με όλους τους κόμβους του προηγούμενου επιπέδου. Το επίπεδο εισόδου δέχεται τα δεδομένα, τα οποία μπορεί να έχουν περάσει από κάποιο στάδιο προεπεξεργασίας έτσι ώστε να εξαχθούν χαρακτηριστικά (features), και τα περνάει απευθείας στις εισόδους των νευρώνων του πρώτου κρυφού επιπέδου. Από εκεί, και αφού περάσουν από όλα τα κρυφά επίπεδα, τα δεδομένα καταλήγουν στο επίπεδο εξόδου όπου η τελική έξοδος λαμβάνεται μέσω της επιλεγμένης συνάρτηση ενεργοποίησης.

Μία άλλη συνηθισμένη FFNN αρχιτεκτονική είναι τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - **CNNs**) [16], τα οποία είναι ειδικά σχεδιασμένα ώστε να επεξεργάζονται δεδομένα που χαρακτηρίζονται από κάποια τοπολογία, όπως είναι τα φασματογραφήματα. Η βασική διαφορά τους από τα δίκτυα MLP είναι ότι βασίζονται σε κρυφά επίπεδα τα οποία δεν είναι πλήρως συνδεδεμένα και χρησιμοποιούν πυρήνες για να εκτελούν πράξεις συνέλιξης με σόχο την εξαγωγή χαρακτηριστικών από τα δεδομένα. Η σχηματική αναπαράσταση ενός δικτύου CNN παρουσιάζεται στην Εικόνα 2.10:

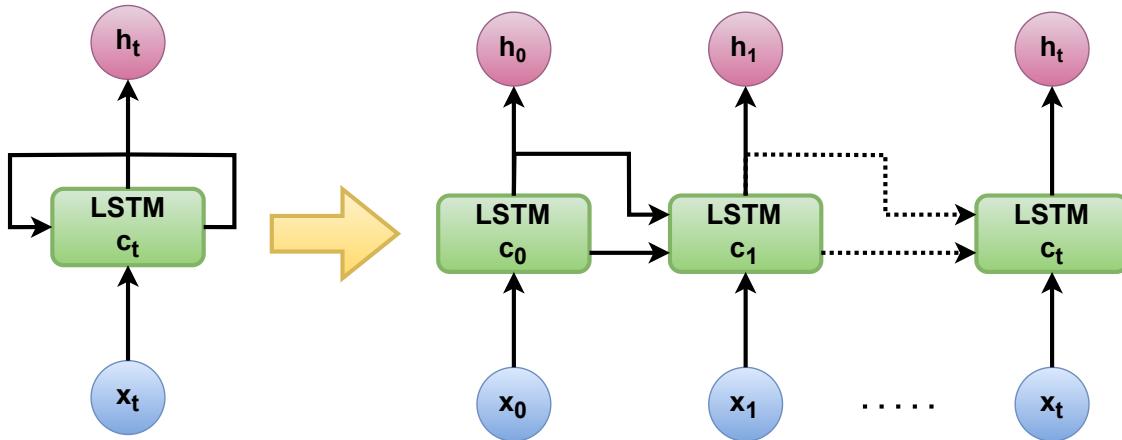


Εικόνα 2.10: Σχηματική αναπαράσταση δικτύου CNN

Παρόμοια με τα δίκτυα MLP είναι και τα δίκτυα RNN, με τη διαφορά πως οι νευρώνες των κρυφών επιπέδων τους αποτελούν κελιά η έξοδος των οποίων αποκαλείται *κρυφή κατάσταση* (hidden state) και εισάγεται σε αυτά μέσω αναδρομής, μοντελοποιώντας έτσι κάποια μορφή μνήμης. Αυτή η αρχή λειτουργίας καθιστά τα δίκτυα RNN ιδανικά για χειρισμό δεδομένων που χαρακτηρίζονται από χρονικές εξαρτήσεις, όπως είναι οι χρονοσειρές. Η σχηματική αναπαράσταση ενός κελιού RNN το οποίο έχει ξεδιπλωθεί στο χρόνο παρουσιάζεται στην Εικόνα 2.11. Μια βελτιωμένη εκδοχή των κλασικών RNN είναι τα αναδρομικά δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory - **LSTM**) [17], τα οποία διαθέτουν μια εσωτερική κατάσταση κελιού (cell state) που τους επιτρέπει να χειριστούν δεδομένα με χρονικές εξαρτήσεις μακράς διαρκείας. Η σχηματική αναπαράσταση ενός κελιού LSTM το οποίο έχει ξεδιπλωθεί στο χρόνο παρουσιάζεται στην Εικόνα 2.12. Τόσο τα κλασικά RNN όσο και τα LSTM μπορούν να υλοποιηθούν έτσι ώστε να λαμβάνουν υπόψιν παρελθοντικές αλλά και μελλοντικές τιμές των δεδομένων. Στην περίπτωση αυτή τα δίκτυα ονομάζονται αμφίδρομα (Bidirectional RNN/LSTM - **BiRNN/BiLSTM**).



Εικόνα 2.11: Σχηματική αναπαράσταση κελιού RNN



Εικόνα 2.12: Σχηματική αναπαράσταση κελιού LSTM

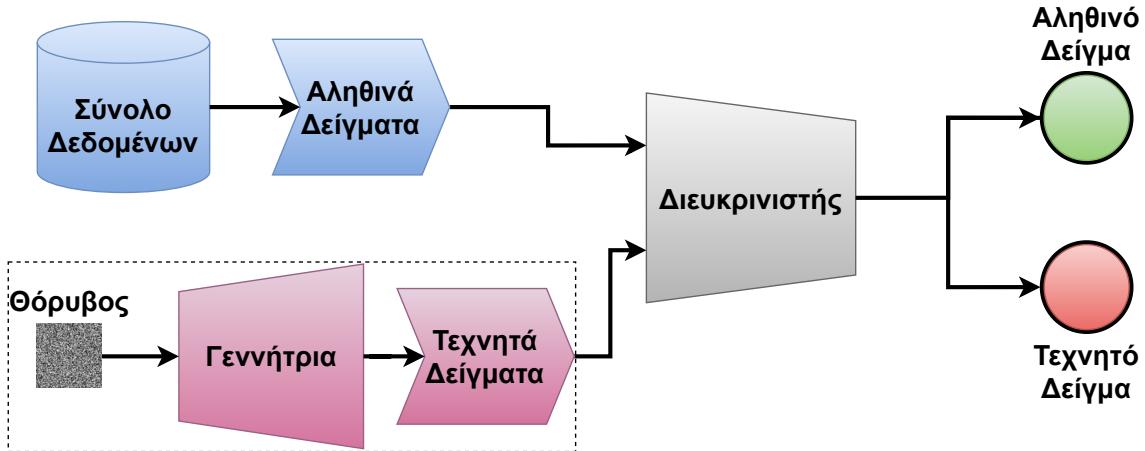
2.3.2 Προχωρημένες Αρχιτεκτονικές

Οι αρχιτεκτονικές νευρωνικών δικτύων που παρουσιάστηκαν μέχρι αυτό το σημείο είναι κατάλληλες για προβλήματα *ταξινόμησης* (classification) και *παθινδρόμησης* (regression), αλλά συνήθως δεν επαρκούν για προβλήματα αλγορίθμικής σύνθεσης όπου στόχος είναι η παραγωγή δεδομένων. Για το λόγο αυτό, οι βασικές αυτές αρχιτεκτονικές χρησιμοποιούνται για την υλοποίηση πιο προχωρημένων μοντέλων νευρωνικών δικτύων, τα σημαντικότερα από τα οποία είναι τα παρακάτω:

- Γενετικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks - **GANs**) [18]
- Μεταβλητοί Αυτοκωδικοποιητές (Variational Autoencoders - **VAEs**) [19]
- Μοντέλα Ακολουθίας σε Ακολουθία (Sequence-to-Sequence - **Seq2Seq**) [20]

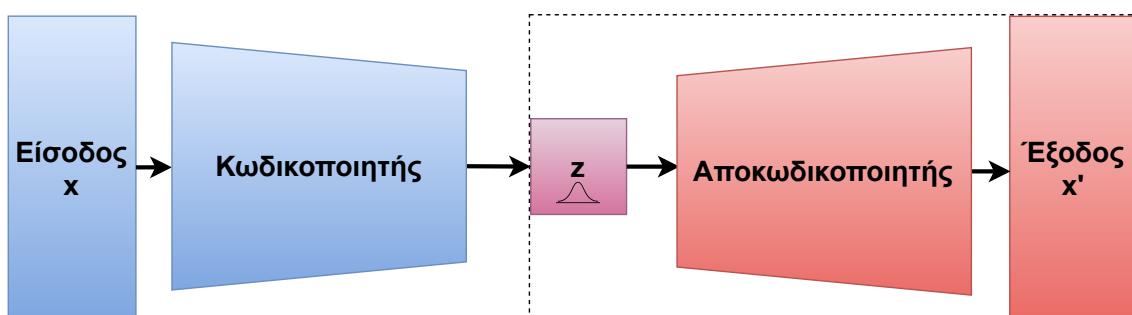
Τα παραπάνω μοντέλα επιδέχονται αρκετές παραλλαγές ενώ μπορούν ακόμα και να συνδυαστούν μεταξύ τους, υλοποιώντας έτσι ακόμα πιο σύνθετες αρχιτεκτονικές νευρωνικών δικτύων. Οι βασικές υλοποιήσεις τους περιγράφονται στη συνέχεια.

Τα μοντέλα GAN χρησιμοποιούνται για την παραγωγή τυχαίων δειγμάτων. Η αρχιτεκτονική τους περιλαμβάνει δύο νευρωνικά δίκτυα τα οποία αποτελούν τη γεννήτρια (generator) και τον διευκρινιστή (discriminator) και εκπαιδεύονται με τη χρήση ενός συνόλου δεδομένων που αποτελείται από αληθινά δείγματα. Στόχος της γεννήτριας είναι να μάθει να παράγει δείγματα ώστε να παραπλανά τον διευκρινιστή, ο οποίος ταυτόχρονα μαθαίνει να ξεχωρίζει αν τα δείγματα της γεννήτριας είναι αληθινά ή τεχνητά. Μετά το πέρας της εκπαίδευσης διατηρείται μόνο το δίκτυο της γεννήτριας, η οποία βρίσκεται σε θέση να παράγει αληθοφανή τεχνητά δείγματα. Η σχηματική αναπαράσταση ενός μοντέλου GAN παρουσιάζεται στην Εικόνα 2.13:



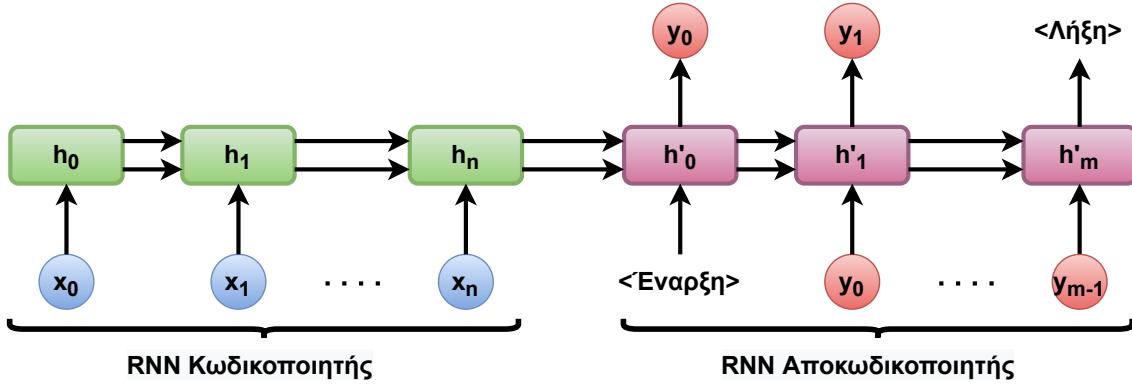
Εικόνα 2.13: Σχηματική αναπαράσταση μοντέλου GAN

Τα μοντέλα VAE χρησιμοποιούνται επίσης για την παραγωγή τυχαίων δειγμάτων, αλλά η αρχή λειτουργίας τους διαφέρει σημαντικά από αυτή των GANs. Κατά την εκπαίδευση τους, τα μοντέλα VAE μαθαίνουν να ανακατασκευάζουν στην έξοδο τους τα δεδομένα που δέχονται στην είσοδο τους. Αυτό επιτυγχάνεται χρησιμοποιώντας ένα δίκτυο κωδικοποιητή (encoder), ο οποίος συμπιέζει την πληροφορία σε έναν λανθάνων χώρο (latent space) που περιγράφεται από την κανονική κατανομή, και ένα δίκτυο αποκωδικοποιητή (decoder), ο οποίος εκτελεί την αντίστροφη διαδικασία. Μετά το πέρας της εκπαίδευσης, διατηρείται μόνο ο αποκωδικοποιητής και η επιλογή ενός τυχαίου σημείου στον λανθάνων χώρο έχει ως αποτέλεσμα την παραγωγή ενός νέου δείγματος που διατηρεί παρόμοια χαρακτηριστικά με αυτά των δεδομένων της εκπαίδευσης. Η σχηματική αναπαράσταση ενός μοντέλου VAE παρουσιάζεται στην Εικόνα 2.14:



Εικόνα 2.14: Σχηματική αναπαράσταση μοντέλου VAE

Τέλος, τα μοντέλα Seq2Seq προέρχονται από το πεδίο της Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing - **NLP**), όπου χρησιμοποιούνται κυρίως για μεταφράσεις, και η αρχιτεκτονική τους αποτελείται αποκλειστικά από αναδρομικά δίκτυα τα οποία υλοποιούν έναν κωδικοποιητή και έναν αποκωδικοποιητή. Ο κωδικοποιητής δέχεται μία ακολουθία δεδομένων και συμπιέζει την πληροφορία της σε ένα διάνυσμα πλαισίου (context vector) το οποίο αποτελείται από τις κρυφές καταστάσεις των αναδρομικών κελιών του δικτύου. Το διάνυσμα αυτό μεταβιβάζεται στον αποκωδικοποιητή ο οποίος δέχεται επιπλέον μια ένδειξη έναρξης και ξεκινάει να παράγει αναδρομικά μια νέα ακολουθία μέχρις ότου προβλέψει την ένδειξη για τη λήξη της διαδικασίας. Η σχηματική αναπαράσταση ενός μοντέλου Seq2Seq παρουσιάζεται στην Εικόνα 2.15:



Εικόνα 2.15: Σχηματική αναπαράσταση μοντέλου Seq2Seq

2.4 Ερευνητικά Έργα

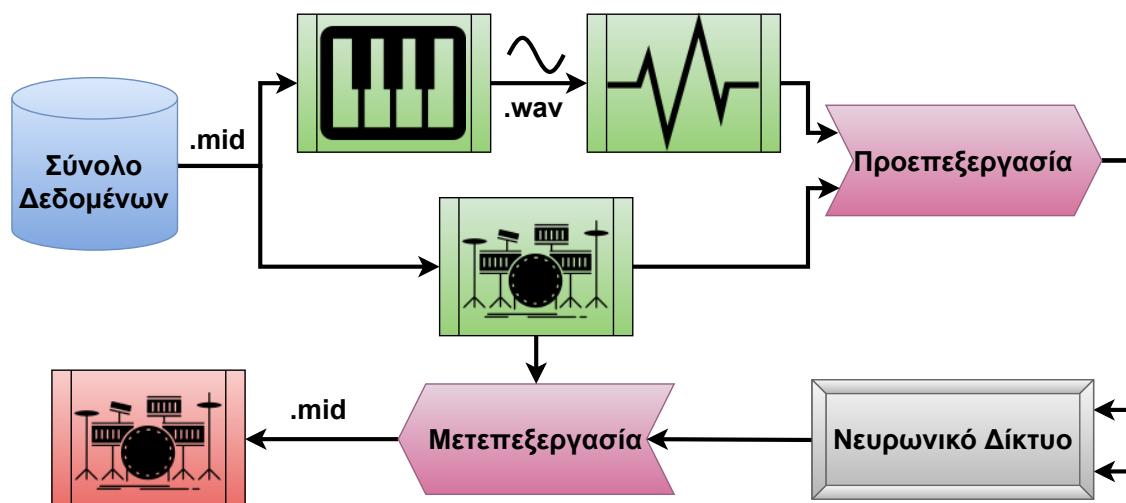
Τη σημαντικότερη συνεισφορά στο πρόβλημα της παραγωγής μοτίβων τυμπάνων με βάση ηχητικά δεδομένα αποτελεί το έργο των Wei et al. [12]. Σε αυτό, οι συγγραφείς αξιοποιούν μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή με χρήση δικτύων CNN, εμπνευσμένη από τα μοντέλα VAE. Το υλοποιημένο σύστημα τροφοδοτείται με μία παραλλαγή του STFT ενός ηχητικού σήματος που ονομάζεται Μετασχηματισμός Σταθερού Q (Constant Q Transform - **CQT**) και παράγει τα αντίστοιχα μοτίβα τυμπάνων για τη συνοδεία του μουσικού αποστάσματος. Αξιόλογα είναι επίσης τα έργα των Makris et al. [8]-[11] οι οποίοι υλοποιούν αρχιτεκτονικές που συνδυάζουν δίκτυα FFNN και LSTM, αλλά και μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή εμπνευσμένη από τα μοντέλα Seq2Seq. Με αυτές τις αρχιτεκτονικές προσεγγίζουν το πρόβλημα της παραγωγής μοτίβων τυμπάνων αξιοποιώντας συμβολικά δεδομένα εκτέλεσης κιθάρας και μπάσου, καθώς επίσης και δεδομένα για το tempo, τα μέτρα και την οργάνωση ενός μουσικού κομματιού. Τέλος, αξίζει να σημειωθεί το έργο των Hakí και Jordà [21], οι οποίοι χρησιμοποιούν ένα μοντέλο Seq2Seq βασισμένο σε δίκτυα LSTM για να υλοποίησουν ένα σύστημα που παράγει συμβολικές μπασογραμμές αξιοποιώντας δεδομένα μοτίβων τυμπάνων. Πρόκειται για την επίλυση ενός προβλήματος που είναι αντίστροφο από αυτό της παρούσας εργασίας, το οποίο ωστόσο επιδεικνύει την ικανότητα των μοντέλων Seq2Seq να μεταφράζουν τα μουσικά μοτίβα ενός οργάνου σε ένα άλλο οργανό.

Κεφάλαιο 3

Μεθοδολογία και Υλοποίηση

3.1 Διαδικασία

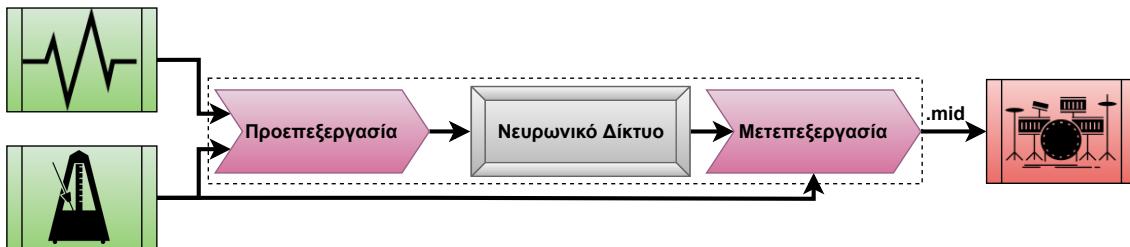
Η γενικότερη διαδικασία που ακολουθείται για την υλοποίηση του προτεινόμενου συστήματος παρουσιάζεται συνοπτικά στην Εικόνα 3.1:



Εικόνα 3.1: Διαδικασία υλοποίησης του προτεινόμενου συστήματος

Αρχικά, συγκροτείται ένα σύνολο δεδομένων από αρχεία μουσικών κομματιών με πληροφορία MIDI. Για κάθε αρχείο διαχωρίζεται το κανάλι των τυμπάνων από τα κανάλια των υπόλοιπων οργάνων, η πληροφορία των οποίων μετατρέπεται σε ήχο μέσω ηχητικής σύνθεσης. Έτσι προκύπτει ένα αρχείο .wav το οποίο περιέχει τα δεδομένα του ηχητικού σήματος και ένα αρχείο .mid το οποίο περιέχει τα δεδομένα των μοτίβων τυμπάνων καθώς επίσης και όλη την πληροφορία για τη μετρική δομή του μουσικού κομματιού. Τα δεδομένα των δύο αρχείων περνάνε από ένα στάδιο προεπεξεργασίας και τροφοδοτούνται ανά μουσικό μέτρο στο νευρωνικό δίκτυο, το οποίο μαθαίνει να παράγει τα επιθυμητά μοτίβα τυμπάνων βάσει των χαρακτηριστικών που εξάγονται από το σήμα ήχου. Τέλος, η έξοδος του νευρωνικού δικτύου περνάει από ένα στάδιο μετεπεξεργασίας έτσι ώστε να προκύψει το τελικό αρχείο MIDI που περιέχει όλη την πληροφορία των παραγόμενων μοτίβων τυμπάνων.

Εφόσον υλοποιηθεί, εκπαιδευτεί και αξιολογηθεί το σύστημα, η χρήση του για την παραγωγή μοτίβων τυμπάνων ακολουθεί τη διαδικασία που περιγράφεται συνοπτικά από την Εικόνα 3.2:



Εικόνα 3.2: Διαδικασία χρήσης του προτεινόμενου συστήματος

Συγκεκριμένα, το σύστημα χρειάζεται ένα μουσικό απόσπασμα σε μορφή ηχητικού σήματος καθώς και την πληροφορία της μετρικής δομής του. Η πληροφορία αυτή μπορεί να αποτελεί μέρος κάποιου αρχείου MIDI, ή να παρέχεται από κάποιο DAW, και χρησιμοποιείται για το διαχωρισμό του μουσικού αποσπάσματος σε μέτρα, βάσει των οποίων γίνεται η εξαγωγή των χαρακτηριστικών που επεξεργάζεται το νευρωνικό δίκτυο για την παραγωγή μοτίβων τυμπάνων ανά μουσικό μέτρο.

3.2 Σύνολο Δεδομένων

Για το σύνολο δεδομένων συλλέχθηκαν από τον ιστότοπο Ultimate Guitar αρχεία με συμβολικά δεδομένα 150 κομματιών από τη δισκογραφία του βρετανικού heavy metal συγκροτήματος Iron Maiden.^{7,8} Για τη διασφάλιση της ποιότητας του συνόλου δεδομένων κάθε αρχείο του επιμελήθηκε ξεχωριστά ακολουθώντας τα εξής βήματα:

- Αφαιρέθηκαν μέτρα με κακή πληροφορία για την εκπαίδευση του δικτύου, όπως τα μέτρα με μη ολοκληρωμένα ή χωρίς καθόλου μοτίβα τυμπάνων.
- Καταργήθηκαν αυτοματισμοί που μετέβαλαν το tempo σε ενδιάμεσα σημεία των μέτρων, επιβάλλοντας έτσι σταθερό tempo σε κάθε μέτρο.
- Διορθώθηκαν λάθη σε κλάσματα χρόνου, θέσεις νοτών και αξίες έτσι ώστε τα beat και οι νότες να βρίσκονται στις σωστές θέσεις των μέτρων.
- Πραγματοποιήθηκε μίξη των οργάνων χρησιμοποιώντας ρεαλιστικούς ήχους, έτσι ώστε το συνθετικό ηχητικό σήμα που θα προκύψει από τη μετατροπή της MIDI πληροφορίας σε ήχο να έχει ηχόχρωμα που θα προσομοιώνει αυτό της πραγματικής μουσικής.

Από την εκκαθάριση του συνόλου δεδομένων απέμειναν συνολικά 24774 μέτρα 16 διαφορετικών κλασμάτων χρόνου, με χονδρικά το 95% αυτών να πρόκειται για μέτρα 4/4. Μετά τον διαχώρισμό των τυμπάνων από τα υπόλοιπα όργανα προκύπτουν 150 αρχεία .mid και 150 αρχεία .wav, τα οποία συντελούν το τελικό σύνολο δεδομένων που χρησιμοποιείται άμεσα για την εκτέλεση της προτεινόμενης μεθοδολογίας.

⁷Ultimate Guitar: <https://www.ultimate-guitar.com>

⁸Iron Maiden: <https://www.ironmaiden.com>

3.3 Προεπεξεργασία

Το στάδιο της προεπεξεργασίας μπορεί να χωριστεί σε τρία υποστάδια τα οποία είναι:

- Η ανάκτηση της πληροφορίας της μετρικής δομής του μουσικού κομματιού για τον διαχωρισμό του σε μέτρα.
- Η μετατροπή των συμβολικών δεδομένων MIDI των μοτίβων τυμπάνων σε κάποια αναπαράσταση κατάλληλη για χειρισμό από το σύστημα.
- Η επεξεργασία του ηχητικού σήματος για την εξαγωγή των χαρακτηριστικών που θα τροφοδοτηθούν στο νευρωνικό δίκτυο.

3.3.1 Μετρική Δομή

Η μετρική δομή ενός μουσικού κομματιού περιλαμβάνει πληροφορία όπως το tempo και τα κλάσματα χρόνου των μέτρων του. Ένα αρχείο MIDI δεν παρέχει άμεσα την πληροφορία για τον διαχωρισμό ενός μουσικού κομματιού σε μέτρα, αλλά ο διαχωρισμός αυτός μπορεί να επιτευχθεί ανακτώντας τα μετα-μηνύματα "Set Tempo" και "Time Signature", τα οποία όριζουν αντίστοιχα μια νέα τιμή tempo και ένα νέο κλάσμα χρόνου σε κάποια χρονική στιγμή. Ετοι, γνωρίζοντας το κλάσμα χρόνου ενός μέτρου η διάρκεια του σε ticks μπορεί να υπολογιστεί από την Εξίσωση 3.1:

$$\text{ticks} = 4 \cdot \text{TPB} \cdot \frac{N}{D} \quad (3.1)$$

όπου N, D είναι αντίστοιχα ο αριθμητής και ο παρονομαστής του κλάσματος χρόνου και TPB είναι ο σταθερός αριθμός των ticks ανά beat που ορίζει το αρχείο MIDI. Ο διαχωρισμός του μουσικού κομματιού σε μέτρα στο συμβολικό πεδίο μπορεί να επιτευχθεί εντοπίζοντας τις χρονικές θέσεις των μετα-μηνυμάτων "Time Signature" και υπολογίζοντας την διάρκεια κάθε μέτρου βάσει της Εξίσωσης 3.1, ξεκινώντας από την αρχή και φτάνοντας μέχρι το τέλος των μηνυμάτων MIDI. Στη συνέχεια, εντοπίζοντας τις χρονικές θέσεις των μετα-μηνυμάτων "Set Tempo" και αξιοποιώντας τις Εξισώσεις 2.5 και 2.6 για τη μετατροπή των ticks σε δευτερόλεπτα, επιτυγχάνεται ο διαχωρισμός του μουσικού κομματιού σε μέτρα στο πεδίο του ήχου. Η διαδικασία αυτή διευκολύνεται κατά πολύ όταν το tempo κάθε μέτρου είναι σταθερό και αυτός είναι ένας από τους λόγους για τους οποίους καταργήθηκαν όλες οι ενδιάμεσες μεταβολές του tempo κατά την επιμέλεια των αρχείων του συνόλου δεδομένων.

3.3.2 Αναπαράσταση Μοτίβων Τυμπάνων

Η αναπαράσταση των μοτίβων τυμπάνων μέσω μηνυμάτων MIDI δυσχεραίνει τον χειρισμό τους από το σύστημα, καθώς δεν είναι κατάλληλη για την εκτέλεση πράξεων και υπολογισμών. Για τον λόγο αυτό, τα συμβολικά δεδομένα χρειάζεται να αναπαρασταθούν με κάποιον πιο μαθηματικό τρόπο ο οποίος αρχικά προϋποθέτει να επιλεχθούν τα επιμέρους κρουστά όργανα του συνόλου τυμπάνων τα οποία θα μάθει να παράγει το νευρωνικό δίκτυο. Για το σύστημα που υλοποιήθηκε επιλέχθηκαν συνολικά εννιά κρουστά, τα οποία ήταν η μπότα, το ταμπούρο, τα πιατίνια crash, ride και hi-hat (σε ανοιχτή και κλειστή θέση) και τρία toms (χαμηλό/μεσαίο/υψηλό).

Σύμφωνα με την προδιαγραφή General MIDI 1, το κανάλι 10 διαθέτει 47 διαφορετικούς ήχους τυμπάνων καθένας από τους οποίους λαμβάνει έναν μοναδικό αριθμό νότας MIDI από το 35 έως το 81. Η αντιστοίχιση αυτών των νοτών MIDI στα επιλεγμένα κρουστά του συστήματος γίνεται σύμφωνα με τον Πίνακα 3.1:

Πίνακας 3.1: Αντιστοίχιση των νοτών MIDI στα κρουστά του συστήματος

Νότα MIDI	Κρουστό	Νότα MIDI	Κρουστό
(35) Acoustic Bass Drum	Μπότα	(48) Hi-Mid Tom	Υψηλό Tom
(36) Bass Drum 1	Μπότα	(49) Crash Cymbal 1	Crash
(37) Side Stick	Ταμπούρο	(50) High Tom	Υψηλό Tom
(38) Acoustic Snare	Ταμπούρο	(51) Ride Cymbal 1	Ride
(39) Hand Clap	Ταμπούρο	(52) Chinese Cymbal	Crash
(40) Electric Snare	Ταμπούρο	(53) Ride Bell	Ride
(41) Low Floor Tom	Χαμηλό Tom	(54) Tambourine	Κλειστό Hi-Hat
(42) Closed Hi-Hat	Κλειστό Hi-Hat	(55) Splash Cymbal	Crash
(43) High Floor Tom	Χαμηλό Tom	(56) Cowbell	Ride
(44) Pedal Hi-Hat	Κλειστό Hi-Hat	(57) Crash Cymbal 2	Crash
(45) Low Tom	Μεσαίο Tom	(58) Vibraphone	Crash
(46) Open Hi-Hat	Ανοιχτό Hi-Hat	(59) Ride Cymbal 2	Ride
(47) Low-Mid Tom	Μεσαίο Tom	(60) — (81)	—

Φυσικά, σε καθένα από τα επιλεγμένα κρουστά του συστήματος θα πρέπει ούτως ή άλλως να ανατεθεί κάποια συγκεκριμένη νότα MIDI για τη συγγραφή των μηνυμάτων του τελικού αρχείου MIDI που θα προκύπτει από το σύστημα. Η ανάθεση αυτή γίνεται σύμφωνα με τον Πίνακα 3.2:

Πίνακας 3.2: Ανάθεση νοτών MIDI στα κρουστά του συστήματος

Κρουστό	Νότα MIDI
Μπότα	(36) Bass Drum 1
Ταμπούρο	(38) Acoustic Snare
Crash	(49) Crash Cymbal 1
Ride	(51) Ride Cymbal 1
Ανοιχτό Hi-Hat	(46) Open Hi-Hat
Κλειστό Hi-Hat	(42) Closed Hi-Hat
Χαμηλό Tom	(43) High Floor Tom
Μεσαίο Tom	(47) Low-Mid Tom
Υψηλό Tom	(48) Hi-Mid Tom

Επομένως, συνδυάζοντας τους Πίνακες 3.1 και 3.2 προκύπτουν οι μετατροπές των νοτών MIDI που γίνονται κατά το στάδιο προεπεξεργασίας του συνόλου δεδομένων.

Μετά την μετατροπή των νοτών MIDI σύμφωνα με τα παραπάνω, εντοπίζονται οι χρονικές θέσεις όλων των μηνυμάτων "Note On" και κάθε τέτοια χρονική θέση συνοδεύεται από ένα δυαδικό διάνυσμα διάστασης 1×9 το οποίο συμβολίζει ποια κρουστά παίζονται τη συγκεκριμένη χρονική στιγμή. Για παράδειγμα, το διάνυσμα $[1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$ περιγράφει την ταυτόχρονη κρούση της μπότας και του crash κάποια χρονική στιγμή, ενώ παρόμοια το διάνυσμα $[0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0]$ υποδηλώνει την ταυτόχρονη κρούση του ταμπούρου και του χαμηλού tom. Αφού συγκεντρωθούν όλα αυτά τα διανύσματα για ένα μουσικό κομμάτι, ακολουθεί η ομαδοποίηση τους σε μέτρα σύμφωνα με την πληροφορία της μετρικής δομής που έχει ανακτηθεί προηγουμένως. Ειδικότερα, κάθε μέτρο διαμερίζεται σε διακριτά και ισαπέχοντα σημεία, η απόσταση των οποίων αντιστοιχεί στη διάρκεια κάποιας ελάχιστης αξίας νότας που επιλέγεται, και στη συνέχεια οι χρονικές θέσεις των δυαδικών διανυσμάτων κβαντίζονται στις θέσεις αυτών των σημείων. Το πλήθος K των σημείων αυτών ταυτίζεται ουσιαστικά με το μέγιστο πλήθος νοτών της ελάχιστης αξίας V_{min} που χωράνε στο μέτρο και δίνεται από την Εξίσωση 3.2:

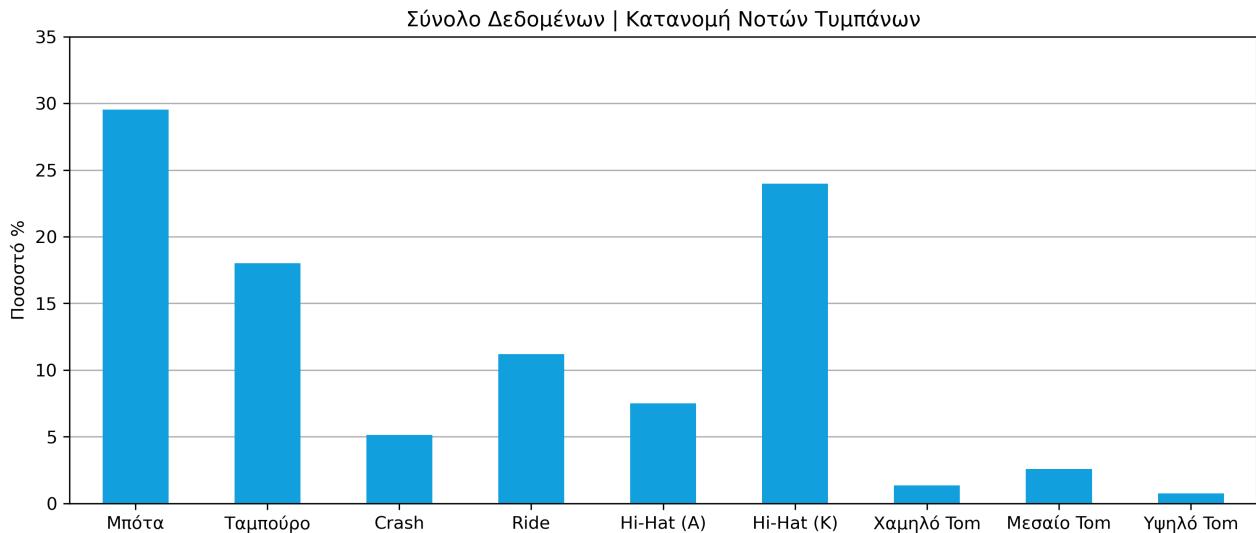
$$K = \frac{1}{V_{min}} \cdot \frac{N}{D} \quad (3.2)$$

Στην περίπτωση που είναι επιθυμητή η υποδιαιρεση της ελάχιστης αξίας σε τρίηχα θα πρέπει το νούμερο που προκύπτει από την παραπάνω εξίσωση να τριπλασιαστεί. Έτσι, τα μοτίβα τυμπάνων σε κάθε μέτρο αναπαρίστανται από έναν δυαδικό πίνακα διάστασης $K \times 9$ ο οποίος είναι και αραιός (sparse), καθώς όλες οι θέσεις στις οποίες δεν υπάρχουν νότες τυμπάνων περιγράφονται από το διάνυσμα $[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$. Για το σύστημα που υλοποιήθηκε επιλέχθηκε ως ελάχιστη αξία νότας αυτή του δεκάτου έκτου ($V_{min} = 1/16$) με δυνατότητα υποδιαιρεσης τριήχου, και επομένως κάθε μέτρο 4/4 διαμερίζεται σε $K = 48$ σημεία. Συνεπώς, το μοτίβο τυμπάνων ενός μέτρου 4/4 θα αναπαρίσταται από έναν αραιό δυαδικό πίνακα διάστασης 48×9 .

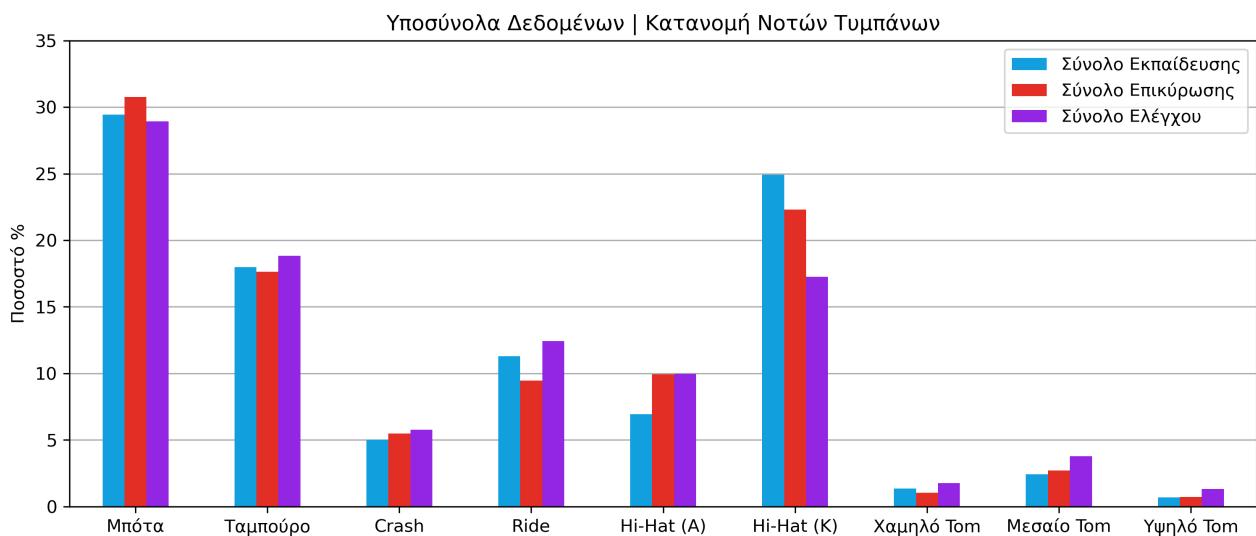
Αφού τα μοτίβα τυμπάνων αναπαρασταθούν σύμφωνα με τον παραπάνω τρόπο, ο ενιαίος αραιός πίνακας που περιέχει όλα τα μέτρα του συνόλου δεδομένων έχει πυκνότητα 3.12% και η σχετική κατανομή των νοτών τυμπάνων του συνόλου παρουσιάζεται στην Εικόνα 3.3. Στη συνέχεια, το σύνολο δεδομένων χωρίζεται σε υποσύνολα εκπαίδευσης (training), επικύρωσης (validation) και ελέγχου (testing) και η σχετική κατανομή των νοτών τυμπάνων κάθε υποσυνόλου παρουσιάζεται στην Εικόνα 3.4. Ο διαχωρισμός του συνόλου δεδομένων στα υποσύνολα εκπαίδευσης, επικύρωσης και ελέγχου του νευρωνικού δικτύου περιγράφεται από τον Πίνακα 3.3:

Πίνακας 3.3: Διαχωρισμός του συνόλου δεδομένων

	Σύνολο Εκπαίδευσης	Σύνολο Επικύρωσης	Σύνολο Ελέγχου
Ποσοστό	80%	10%	10%
Κομμάτια	120	15	15
Μέτρα	19821	2602	2351
Πυκνότητα	3.16%	3.03%	2.91%



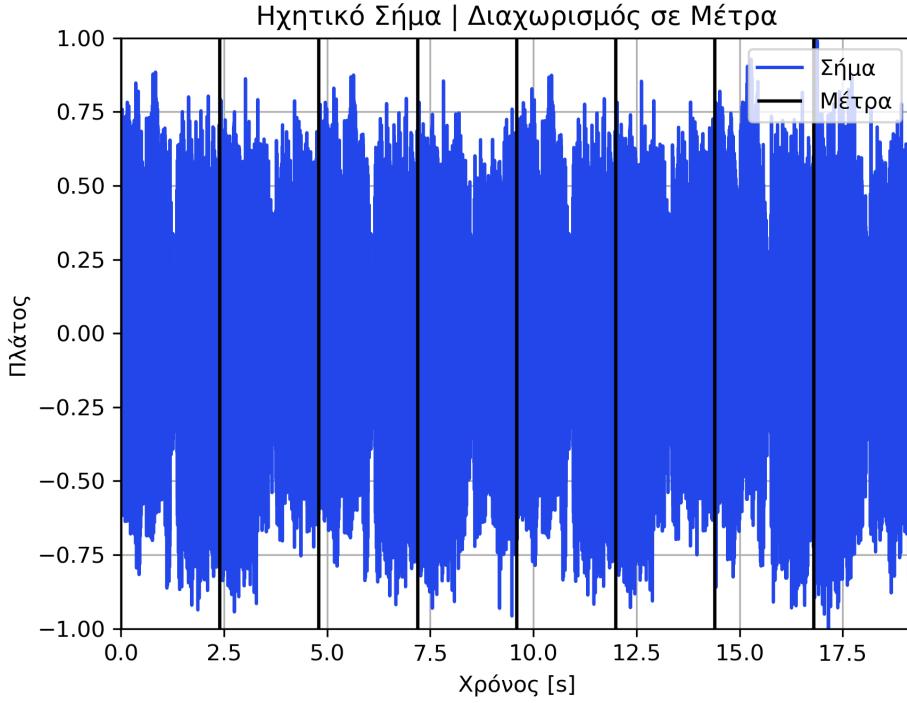
Εικόνα 3.3: Κατανομή των νοτών τυμπάνων στο σύνολο δεδομένων



Εικόνα 3.4: Κατανομή των νοτών τυμπάνων στα υποσύνολα δεδομένων

3.3.3 Εξαγωγή Χαρακτηριστικών

Το πρώτο βήμα για την εξαγωγή χαρακτηριστικών από το ηχητικό σήμα είναι η υποδειγματοληψία του σήματος με συχνότητα $f_s = 8kHz$. Αυτό έχει ως αποτέλεσμα τη διατήρηση του φασματικού περιεχομένου του ήχου μέχρι τα $4kHz$, όπου περιέχεται το σύνολο της μουσικής πληροφορίας, και την απόρριψη των υψηλότερων συχνοτήτων που αντιμετωπίζονται σαν θόρυβος. Παράλληλα, ο ήχος μετατρέπεται από στερεοφωνικός σε μονοφωνικός και κανονικοποιείται έτσι ώστε το πλάτος της κυματομορφής του ηχητικού σήματος να βρίσκεται στο διάστημα $[-1, 1]$. Ακολούθως, το ηχητικό σήμα χωρίζεται σε μουσικά μέτρα αξιοποιώντας την πληροφορία της μετρικής δομής που έχει ανακτηθεί νωρίτερα, όπως φαίνεται στην Εικόνα 3.5:



Εικόνα 3.5: Διαχωρισμός του ηχητικού σήματος σε μουσικά μέτρα

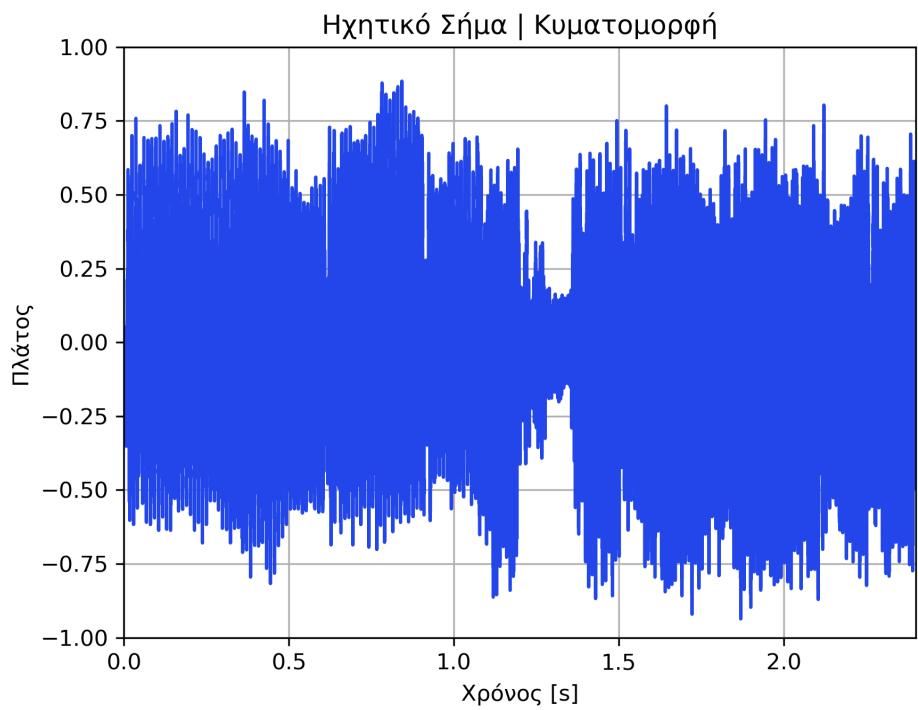
Στη συνέχεια, λαμβάνεται η κυματομορφή του ηχητικού σήματος για κάθε μέτρο ξεχωριστά και υπολογίζεται το αντίστοιχο φασματογράφημα βάσει της Εξίσωσης 2.4, όπως φαίνεται στις Εικόνες 3.6 και 3.7. Τα παράθυρα που χρησιμοποιούνται για τον υπολογισμό του STFT είναι τύπου *Hannning* [13] με επικάλυψη 2/3 και το μήκος τους αντιστοιχίζεται στη διάρκεια της ελάχιστης αξίας που έχει επιλεχθεί για τη διαμέριση κάθε μέτρου στα σημεία που αποτελούν τις θέσεις νοτών των τυμπάνων. Η διάρκεια αυτή είναι αντιστρόφως ανάλογη του tempo και δίνεται από την Εξίσωση 3.3:

$$D = 4 \cdot V_{min} \cdot \frac{60}{BPM} \quad (3.3)$$

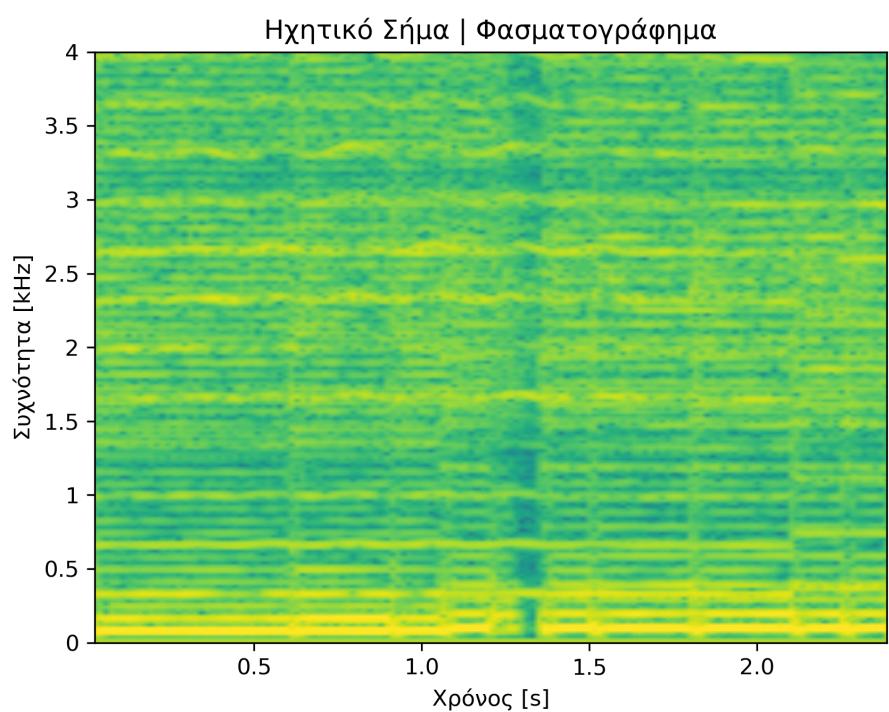
όπου, εφόσον επιτρέπεται η υποδιαιρεση της ελάχιστης αξίας σε τρίτη, το νούμερο που προκύπτει από την παραπάνω εξίσωση θα πρέπει να υποτριπλασιαστεί. Τέλος, υπολογίζεται το Σήμα Ισχύος Onset (Onset Strength Signal - **OSS**), μια χρονοσειρά που συναντάται στο πεδίο της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval - **MIR**) [22] και περιγράφεται από την Εξίσωση 3.4:

$$OSS[n] = \sum_{k=0}^{N-1} (|X[n+1, k]| - |X[n, k]|) \quad (3.4)$$

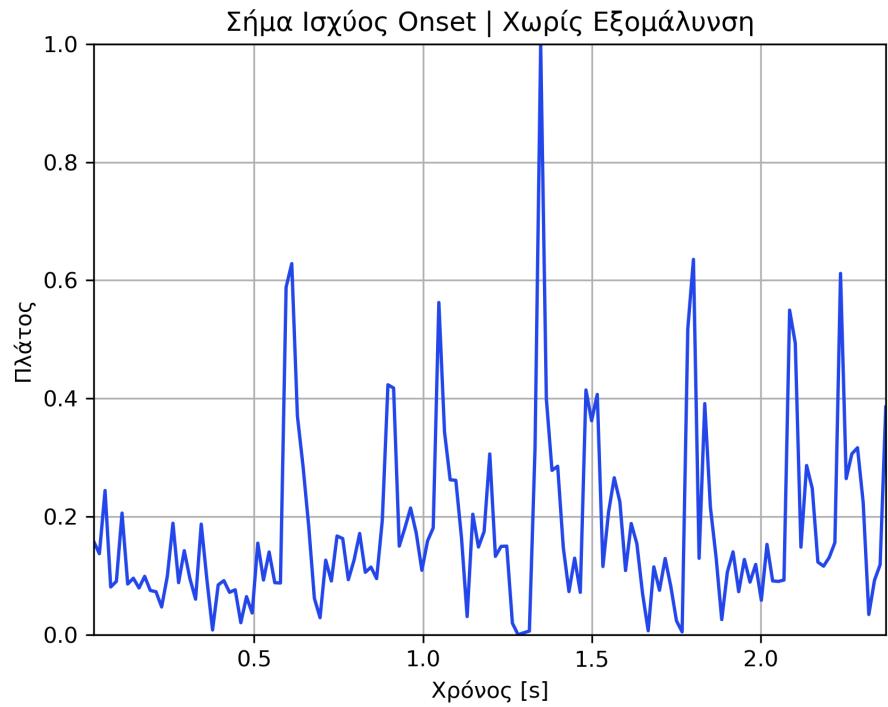
όπου στο παραπάνω άθροισμα συμμετέχουν μόνο τα k για τα οποία ισχύει $|X[n+1, k]| - |X[n, k]| > 0$ και επομένως πρόκειται για το άθροισμα των θετικών πρώτων διαφορών ως προς τον άξονα της συχνότητας του STFT. Η επεξεργασία του σήματος ήχου ολοκληρώνεται εξομαλύνοντας το OSS με φίλτρο κινούμενου μέσου μήκους 3 σημείων [13], η επίδραση του οποίου φαίνεται στις Εικόνες 3.8 και 3.9.



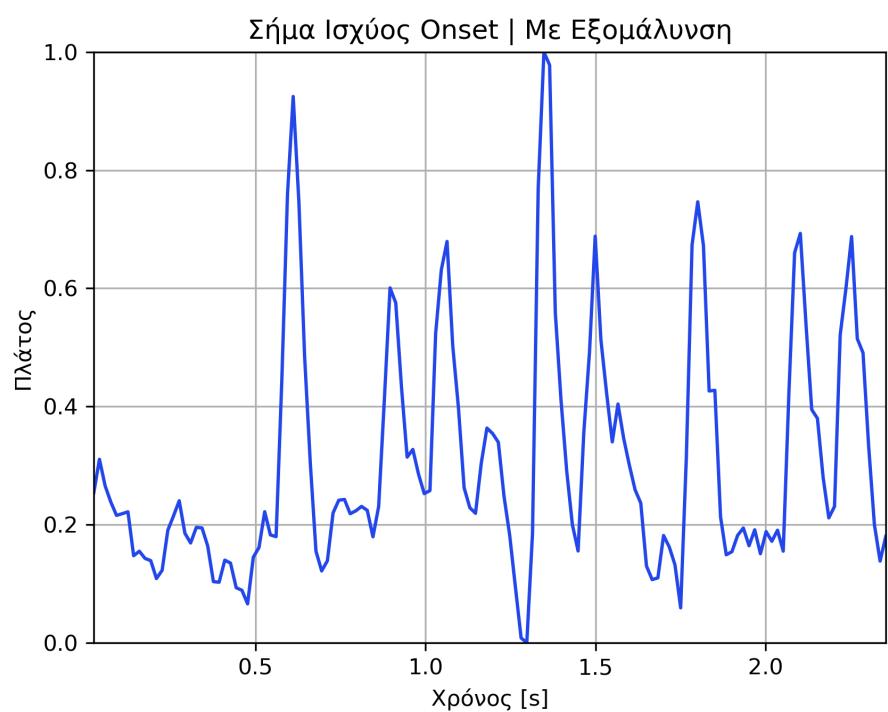
Εικόνα 3.6: Κυματομορφή ηχητικού σήματος



Εικόνα 3.7: Φασματογράφημα ηχητικού σήματος

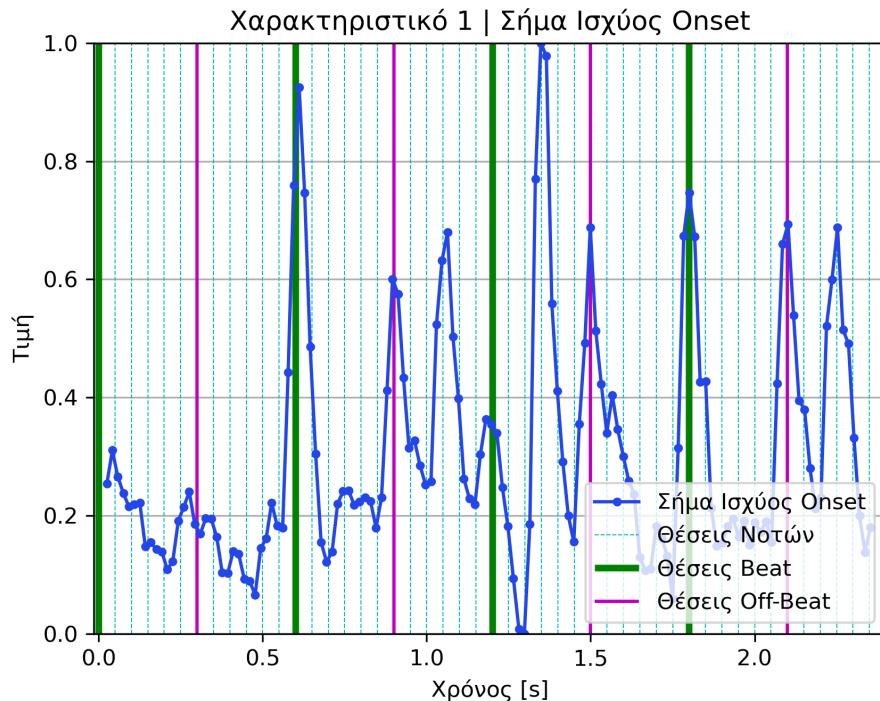


Εικόνα 3.8: OSS χωρίς εξομάλυνση

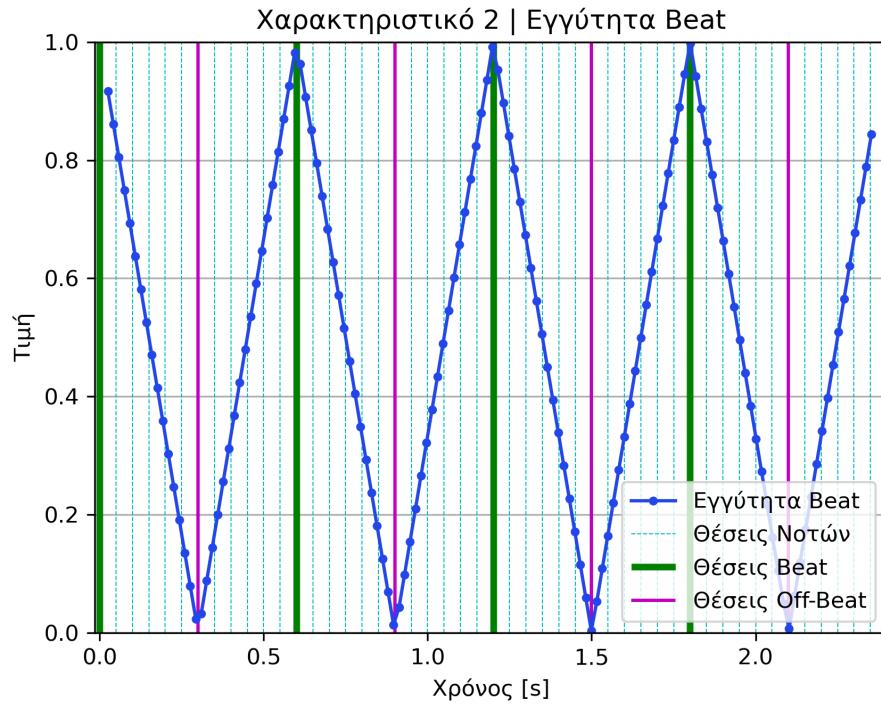


Εικόνα 3.9: OSS με εξομάλυνση

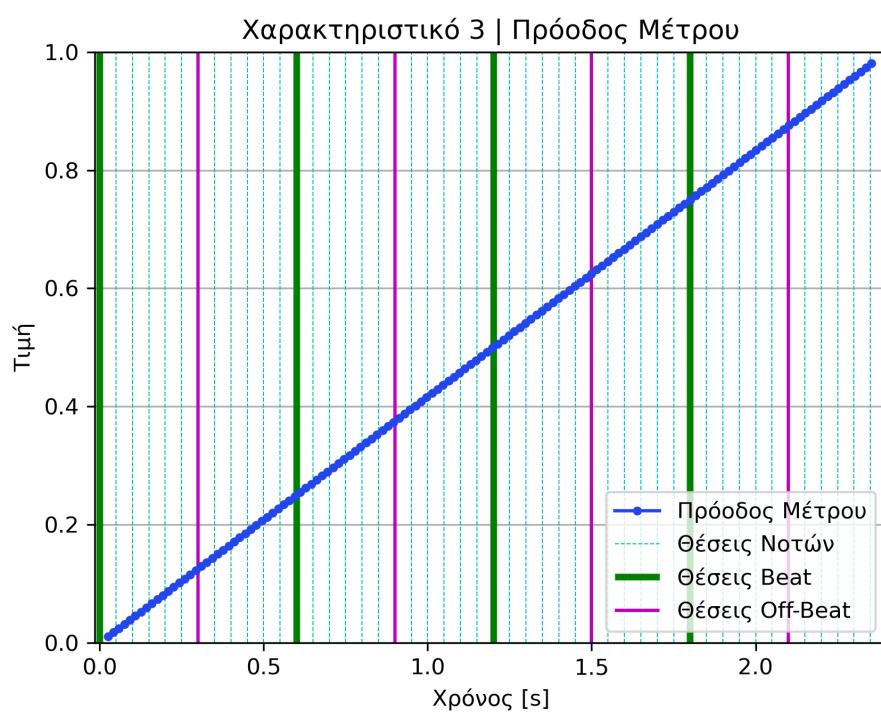
Το OSS αναπαριστά την αύξηση της ενέργειας ενός ηχητικού σήματος, η οποία προκαλείται από την έναρξη ηχητικών συμβάντων στον χρόνο (onsets). Συνεπώς, το παίξιμο νοτών σε μια μουσική εκτέλεση έχει ως αποτέλεσμα την εμφάνιση μιας ακολουθίας κορυφών στο OSS, η οποία μπορεί να αποτελέσει έναν πολύ καλό οδηγό για την αναγνώριση του ρυθμικού μοτίβου από ένα νευρωνικό δίκτυο. Επιπλέον, η επιλογή μήκους παραθύρου ίσου με τη διάρκεια της ελάχιστης αξίας του μέτρου με επικάλυψη 2/3 για τον υπολογισμό του STFT έχει ως αποτέλεσμα των υπολογισμό τριών σημείων του OSS γύρω από κάθε θέση νότας τυμπάνων, ευνοώντας έτσι την εμφάνιση κορυφών γύρω από αυτές τις θέσεις. Από την άλλη, η μεταβολή του μήκους του παραθύρου βάσει του tempo κάθε μέτρου έχει ως αποτέλεσμα σταθερή χρονική αναλυτικότητα σε κάθε μέτρο, γεγονός που επιτρέπει στο προτεινόμενο σύστημα να προσαρμόζεται απόλυτα στη μετρική δομή κάθε μουσικού κομματιού. Έτσι, το OSS αποτελεί το βασικό χαρακτηριστικό το οποίο αξιοποιείται για την παραγωγή μοτίβων τυμπάνων από το νευρωνικό δίκτυο, ωστόσο από μόνο του αποτελεί απλώς μία ακολουθία που δεν περιέχει κάποια πληροφορία για τη μετρική δομή. Για το λόγο αυτό, κατασκευάζονται δύο επιπλέον χαρακτηριστικά με βάση τα χρονικά σημεία στα οποία έχει υπολογιστεί το OSS. Το πρώτο αποκαλείται *Εγγύτητη Beat* και έχει τη μορφή τριγωνικής συνάρτησης που περιγράφει την απόσταση κάθε σημείου από τη θέση του πλησιέστερου beat, ενώ το δεύτερο αποκαλείται *Πρόοδος Μέτρου* και έχει τη μορφή γραμμικής συνάρτησης που περιγράφει τη γενικότερη θέση κάθε σημείου μέσα στο μέτρο. Τα τρία χαρακτηριστικά, τα οποία είναι κανονικοποιημένα στο διάστημα [0,1], απεικονίζονται στις Εικόνες 3.10, 3.11 και 3.12 για ένα μέτρο 4/4, μαζί με τις θέσεις των νοτών τυμπάνων, τις θέσεις των beat και τις ενδιάμεσες θέσεις των off-beat που είναι ιδιαίτερης σημασίας για πολλά ρυθμικά μοτίβα.



Εικόνα 3.10: Χαρακτηριστικό 1: Σήμα Ισχύος Onset



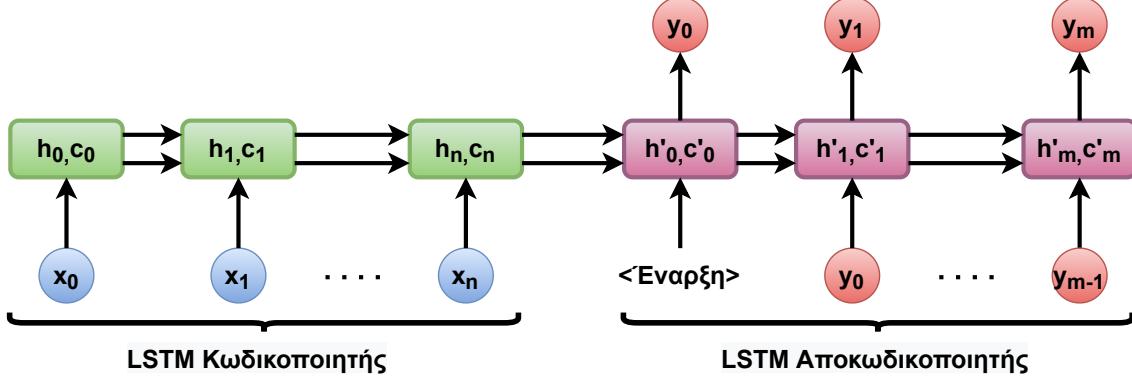
Εικόνα 3.11: Χαρακτηριστικό 2: Εγγύτητα Beat



Εικόνα 3.12: Χαρακτηριστικό 3: Πρόοδος Μέτρου

3.4 Νευρωνικό Δίκτυο

Για το νευρωνικό δίκτυο του προτεινόμενου συστήματος υλοποιήθηκε ένα μοντέλο Seq2Seq με χρήση δικτύων LSTM το οποίο απεικονίζεται στην Εικόνα 3.13:



Εικόνα 3.13: Μοντέλο Seq2Seq του προτεινόμενου συστήματος

Το επίπεδο εισόδου του δικτύου λαμβάνει τις ακολουθίες των τριών χαρακτηριστικών που έχουν εξαχθεί για ένα μέτρο κατά το στάδιο της προεπεξεργασίας και τις μεταβιβάζει στον κωδικοποιητή, όπου η πληροφορία κωδικοποιείται για να περάσει στη συνέχεια στον αποκωδικοποιητή. Τόσο ο κωδικοποιητής όσο και ο αποκωδικοποιητής αποτελούνται από ένα αναδρομικό επίπεδο 512 κελιών LSTM με παράμετρο *dropout* 0.25 η οποία βελτιώνει την ικανότητα γενίκευσης του νευρωνικού δικτύου, ώστε να μπορεί να χειρίζεται αποτελεσματικά δεδομένα που δεν βρίσκονται στο σύνολο εκπαίδευσης [15]. Ο αποκωδικοποιητής λαμβάνει την ένδειξη έναρξης και ξεκινάει να παράγει αναδρομικά στην έξοδο του την αντίστοιχη ακολουθία των μοτίβων τυμπάνων, μέχρις ότου το μήκος αυτής φτάσει το μέγιστο πλήθος νοτών που χωράνε στο μέτρο. Το επίπεδο εξόδου αποτελείται από έναν κόμβο για κάθε ένα από τα εννιά επιμέρους κρουστά όργανα που έχουν επιλεχθεί για να συμμετέχουν στα παραγόμενα μοτίβα τυμπάνων. Λόγω της δυαδικής αναπαράστασης των μοτίβων τυμπάνων, για συνάρτηση ενεργοποίησης στο επίπεδο εξόδου χρησιμοποιήθηκε η σιγμοειδής (sigmoid) [23] η οποία μετατρέπει την τιμή κάθε εξόδου ανεξάρτητα σε μια πιθανότητα στο διάστημα [0, 1]. Επιπλέον, για συνάρτηση κόστους χρησιμοποιήθηκε η Δυαδική Διασταυρωμένη Εντροπία (Binary Cross-Entropy - **BCE**) [24], για την ελαχιστοποίηση της οποίας χρησιμοποιήθηκε ο βελτιστοποιητής Προσαρμοστικής Εκτίμησης Ροπών (Adaptive Moment Estimation - **ADAM**) [25].

3.5 Μετεπεξεργασία

Κύριο και απαραίτητο βήμα του σταδίου μετεπεξεργασίας είναι η μετατροπή των πιθανοτήτων που προκύπτουν στην έξοδο του νευρωνικού δικτύου από τη σιγμοειδή συνάρτηση σε μία δυαδική τιμή, η οποία αναπαριστά την ύπαρξη ή όχι νότας MIDI σε κάποια χρονική στιγμή. Αυτό επιτυγχάνεται πολύ εύκολα με στρογγυλοποίηση των πιθανοτήτων στον κοντινότερο ακέραιο, δηλαδή σε κάποια από τις τιμές 0 και 1.

Το επόμενο βήμα της μετεπεξεργασίας είναι προαιρετικό, αλλά πολύ σημαντικό, και αποσκοπεί στο φιλτράρισμα των παραγόμενων μοτίβων τυμπάνων, έτσι ώστε αυτά να ακούγονται πιο φυσικά και να μην είναι αδύνατο να παιχτούν από άνθρωπο. Το πρόβλημα πηγάζει από το γεγονός ότι το νευρωνικό έχει μάθει να παράγει ρυθμικές ακολουθίες για κάθε κρουστό όργανο του συνόλου τυμπάνων ξεχωριστά, καθιστώντας έτσι πιθανό σε κάποιες χρονικές στιγμές να εμφανίζονται ασυνήθιστοι ή ακόμα και αδύνατοι συνδυασμοί δύο ή περισσότερων κρουστών. Για το λόγο αυτό, χρειάζεται να εισαχθούν κάποιοι λογικοί και φυσικοί περιορισμοί οι οποίοι θα αντιμετωπίζουν αυτό το πρόβλημα. Παρακάτω ορίζονται οι βασικοί κανόνες οι οποίοι ακολουθούνται για να αντιμετωπίσουν το ταυτόχρονο παίξιμο συγκεκριμένων συνδυασμών δύο κρουστών οργάνων, διατηρώντας μόνο αυτό που κρίνεται σημαντικότερο:

- Μπότα + Ταμπούρο → Ταμπούρο
- Crash + Ride → Crash
- Crash + Ανοιχτό Hi-Hat → Crash
- Crash + Κλειστό Hi-Hat → Crash
- Ride + Ανοιχτό Hi-Hat → Ride
- Ride + Κλειστό Hi-Hat → Ride
- Ανοιχτό Hi-Hat + Κλειστό Hi-Hat → Ανοιχτό Hi-Hat

Επιπλέον, λαμβάνεται υπόψιν το γεγονός πως ένας άνθρωπος μπορεί να παίξει μέχρι δύο κρουστά ταυτόχρονα με τα χέρια του ή μέχρι τρία κρουστά συνολικά αν συμπεριληφθεί και η μπότα η οποία παίζεται με το πόδι. Επομένως, στις περιπτώσεις όπου εμφανίζονται ταυτόχρονα περισσότερα από δύο ή τρία κρουστά αντίστοιχα, τα όργανα που θα συμπεριληφθούν τελικά στη μουσική εκτέλεση ακολουθούν μία σειρά προτεραιότητας η οποία ορίζεται από την παρακάτω λίστα:

1. Μπότα
2. Ταμπούρο
3. Crash
4. Ride
5. Ανοιχτό Hi-Hat
6. Κλειστό Hi-Hat
7. Χαμηλό Tom
8. Μεσαίο Tom
9. Υψηλό Tom

Το τελευταίο βήμα του σταδίου της μετεπεξεργασίας είναι η μετατροπή της δυαδικής αναπαράστασης των παραγόμενων μοτίβων τυμπάνων σε μηνύματα MIDI. Τα μηνύματα αυτά συνδυάζονται με τα μετα-μηνύματα που περιέχουν την πληροφορία για τη μετρική δομή του μουσικού κομματιού και είχαν προηγουμένως αξιοποιηθεί στο στάδιο της προεπεξεργασίας, δημιουργώντας έτσι το τελικό αρχείο MIDI που περιέχει όλη τη συμβολική πληροφορία για τη μουσική εκτέλεση των τυμπάνων.

Κεφάλαιο 4

Αξιολόγηση

4.1 Διαδικασίες Αξιολόγησης

Η αξιολόγηση του υλοποιημένου συστήματος μπορεί να προσεγγιστεί τόσο με αντικειμενικά όσο και με υποκειμενικά κριτήρια. Η πρώτη περίπτωση αφορά την ποσοτική αξιολόγηση του με τη χρήση μετρικών, ενώ η δεύτερη περίπτωση αφορά την ποιοτική αξιολόγηση του μέσω της συλλογικής ακρόασης και βαθμολόγησης των παραγόμενων μοτίβων τυμπάνων.

4.1.1 Αντικειμενική Αξιολόγηση

Η δυαδική αναπαράσταση των μοτίβων τυμπάνων καθιστά δυνατή τη χρήση μετρικών δυαδικής ταξινόμησης για την αντικειμενική αξιολόγηση του συστήματος [26]. Θεωρώντας την κλάση 1 ως θετική και την κλάση 0 ως αρνητική, μπορούν να οριστούν οι θεμελιώδεις μετρικές δυαδικής ταξινόμησης *True Positives (TP)*, *True Negatives (TN)*, *False Positives (FP)* και *False Negatives (FN)* που περιγράφονται από τον πίνακα σύγχυσης (confusion matrix) της Εικόνας 4.1:

		Προβλεπόμενη Κλάση	
		1	0
Πραγματική Κλάση	1	True Positives (TP)	False Negatives (FN)
	0	False Positives (FP)	True Negatives (TN)

Εικόνα 4.1: Πίνακας σύγχυσης δυαδικής ταξινόμησης

Οι παραπάνω μετρικές περιγράφουν το πλήθος των σωστών και λανθασμένων ταξινομήσεων αλλά, καθώς οι τιμές τους εξαρτώνται από το μέγεθος του συνόλου δεδομένων, δεν μπορούν να αποτελέσουν σημείο αναφοράς για την αξιολόγηση του συστήματος. Για το λόγο αυτό ορίζονται οι παράγωγες μετρικές *Recall*, *Precision* και *F1-Score* οι οποίες παίρνουν τιμές στο διάστημα [0, 1] και περιγράφονται από τις Εξισώσεις 4.1, 4.2 και 4.3:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.2)$$

$$F1\text{-Score} = 2 \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4.3)$$

Πέρα από τις παραπάνω κλασικές μετρικές ταξινόμησης, μπορεί να αξιοποιηθεί επιπλέον ο δείκτης ομοιότητας *Jaccard* [27]. Η μετρική αυτή παίρνει επίσης τιμές στο διάστημα [0, 1] και μπορεί να περιγράψει την ομοιότητα μεταξύ των επιθυμητών και των παραγόμενων μοτίβων τυμπάνων βάσει της Εξίσωσης 4.4:

$$\text{Jaccard} = \frac{TP}{TP + FP + FN} \quad (4.4)$$

4.1.2 Υποκειμενική Αξιολόγηση

Η υποκειμενική αξιολόγηση του συστήματος πραγματοποιήθηκε διεξάγοντας μία διαδικτυακή έρευνα στην οποία οι συμμετέχοντες κλήθηκαν να ακούσουν και να βαθμολογήσουν ηχητικά δείγματα τα οποία προήλθαν από μουσικά κομμάτια που δεν είχαν χρησιμοποιηθεί για την εκπαίδευση του συστήματος. Αρχικά, οι συμμετέχοντες κλήθηκαν να απαντήσουν σε ορισμένες δημογραφικές ερωτήσεις, προκειμένου να συγκεντρωθούν στατιστικά δεδομένα για το μουσικό υπόβαθρο τους. Οι ερωτήσεις αυτές ήταν οι εξής:

- I. Ποια είδη μουσικής ακούτε;
- II. Είστε εξοικειωμένοι με τη μουσική του συγκροτήματος *Iron Maiden*;
- III. Ξέρετε να παίζετε τύμπανα;
- IV. Ξέρετε να παίζετε κάποιο άλλο μουσικό όργανο;
- V. Έχετε κάποια άλλη μουσική ενασχόληση;

Τα ηχητικά δείγματα που παρουσιάστηκαν στους συμμετέχοντες για βαθμολόγηση ήταν ομαδοποιημένα σε 20 ζεύγη. Κάθε ένα από αυτά τα ζεύγη αποτελούνταν από ένα κοινό μουσικό απόσπασμα το οποίο συνοδευόταν αφενός από μοτίβα τυμπάνων που προέκυψαν από ανθρώπινη μεταγραφή του αυθεντικού μουσικού κομματιού σε MIDI και αφετέρου από μοτίβα τυμπάνων που παρήχθησαν τεχνητά από το σύστημα.

Για την αποφυγή μεροληπτικών βαθμολογήσεων η προέλευση των μοτίβων τυμπάνων κάθε δείγματος παρέμεινε κρυφή από τους συμμετέχοντες, ενώ η σειρά ακρόασης των δειγμάτων τυχαιοποιήθηκε έτσι ώστε στα μισά ζεύγη να προηγούνται τα ανθρώπινα δείγματα και στα άλλα μισά ζεύγη να προηγούνται τα τεχνητά δείγματα. Παράλληλα, η ακουστότητα όλων των δειγμάτων βαθμονομήθηκε στο ίδιο επίπεδο. Τα ποιοτικά χαρακτηριστικά των δειγμάτων που κλήθηκαν να βαθμολογήσουν οι συμμετέχοντες ήταν τα εξής:

- *Αισθητική*: πόσο ικανοποιητικό είναι ηχητικά το παιξιμό των τυμπάνων.
- *Προσαρμοστικότητα*: πόσο ταιριάζει με τον ρυθμό της μουσικής το παιξιμό των τυμπάνων.
- *Δημιουργικότητα*: πόσο ενδιαφέρον, εναλλακτικό ή απρόβλεπτο είναι το παιξιμό των τυμπάνων.
- *Φυσικότητα*: πόσο ανθρώπινο ακούγεται το παιξιμό των τυμπάνων.

Επιπλέον, για κάθε ζεύγος ζητήθηκε από τους συμμετέχοντες να βαθμολογήσουν την ομοιότητα των δύο δειγμάτων του, καθώς επίσης και να προσπαθήσουν να αναγνωρίσουν ποιο από τα δύο δείγματα ήταν αυτό που παρήχθη τεχνητά από το σύστημα. Η βαθμολόγηση των δειγμάτων έγινε σε μια γραμμική κλίμακα από το 0 έως το 10, ενώ στην ερώτηση «Ποιο από τα δύο δείγματα πιστεύετε ότι δημιουργήθηκε τεχνητά;» οι συμμετέχοντες μπορούσαν να επιλέξουν ένα από τα δύο δείγματα ή την απάντηση «Δεν μπορώ να πω με σιγουριά». Κατά τη στατιστική ανάλυση των αποτελεσμάτων, η ύπαρξη στατιστικής σημαντικότητας στη διαφορά των μέσων βαθμολογιών μεταξύ των ανθρώπινων και των τεχνητών δειγμάτων ελέγχεται με το *t-test* του Welch [28] για επίπεδο σημαντικότητας 5%. Έτσι, στις περιπτώσεις όπου οι *p*-τιμές που προκύπτουν από τον έλεγχο είναι μικρότερες από 0.05 υπάρχει πράγματι στατιστικά σημαντική διαφορά στα αποτελέσματα.

4.2 Αποτελέσματα Αξιολόγησης

Το νευρωνικό δίκτυο του συστήματος εκπαιδεύτηκε συνολικά για 300 εποχές, θέτοντας σημείο ελέγχου έτσι ώστε να αποθηκεύεται ένα στιγμιότυπο του ανά 25 εποχές. Η ακρόαση των παραγόμενων μοτίβων τυμπάνων πάνω σε μουσικά κομμάτια που δεν είχαν χρησιμοποιηθεί για την εκπαίδευση του οδήγησε στο συμπέρασμα πως τα καλύτερα αποτελέσματα προέκυπταν όταν αυτό είχε εκπαίδευτεί για 225 εποχές. Συνεπώς, τα αποτελέσματα της εκπαίδευσης και της έρευνας που παρουσιάζονται στη συνέχεια αφορούν το νευρωνικό δίκτυο ύστερα από εκπαίδευση 225 εποχών.

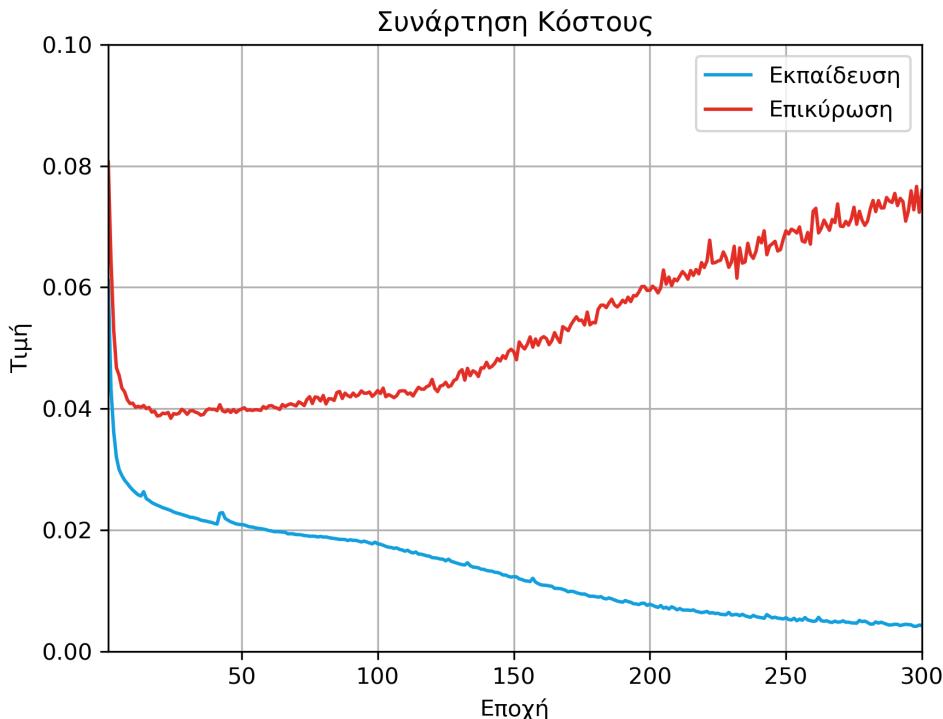
4.2.1 Αποτελέσματα Εκπαίδευσης

Οι τιμές της συνάρτησης κόστους και των μετρικών αξιολόγησης του νευρωνικού δικτύου στα σύνολα εκπαίδευσης, επικύρωσης και ελέγχου παρουσιάζονται παρακάτω στον Πίνακα 4.1. Όπως είναι αναμενόμενο, η απόδοση στο σύνολο εκπαίδευσης υπερτερεί των αποδόσεων στα σύνολα επικύρωσης και ελέγχου.

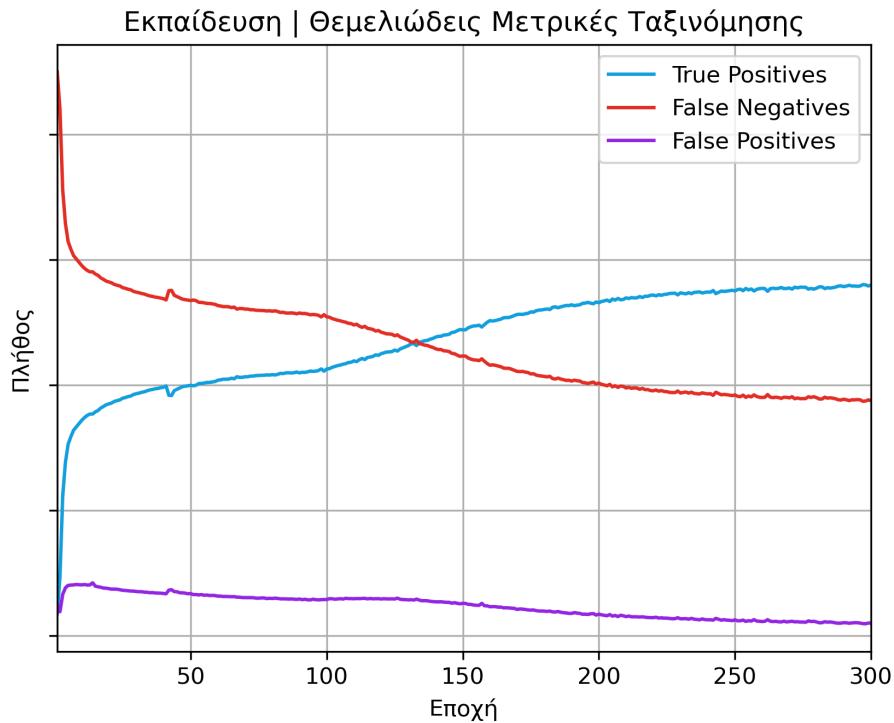
Πίνακας 4.1: Αποτελέσματα εκπαίδευσης νευρωνικού δικτύου

	Σύνολο Εκπαίδευσης	Σύνολο Επικύρωσης	Σύνολο Ελέγχου
Συνάρτηση Κόστους	0.0061	0.0658	0.0423
Recall	58.47%	61.99%	37.28%
Precision	95.23%	67.94%	66.56%
F1-Score	72.45%	64.83%	47.79%
Δείκτης Jaccard	56.80%	47.96%	31.40%

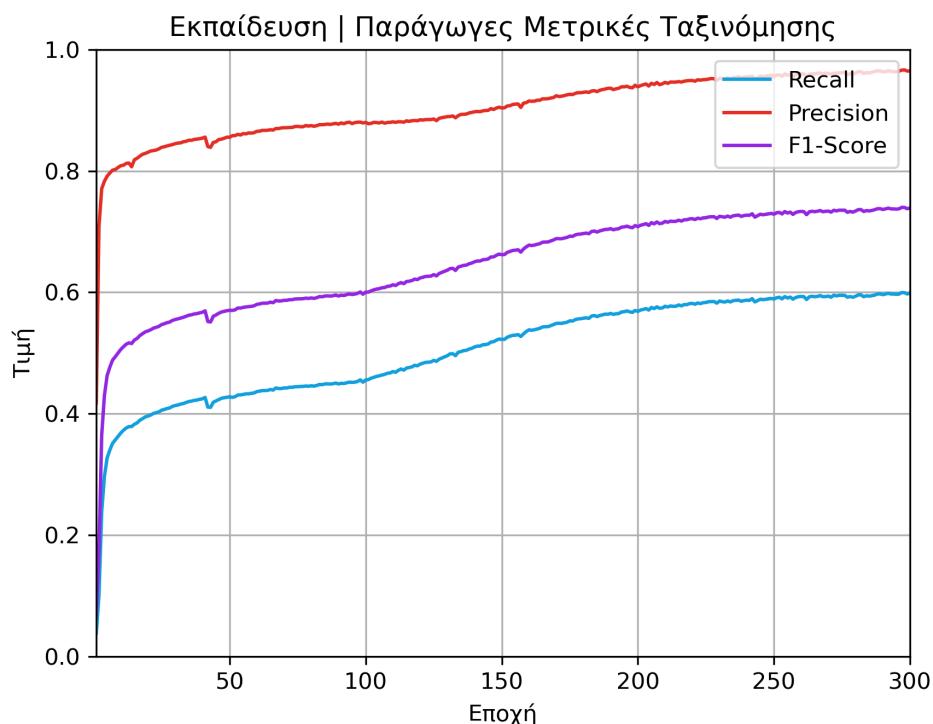
Σημαντικό είναι να παρατηρηθεί, ωστόσο, πως η απόδοση στο σύνολο ελέγχου υστερεί σημαντικά της απόδοσης στο σύνολο επικύρωσης, πράγμα που μπορεί να εγείρει ερωτήματα για την ικανότητα γενίκευσης του συστήματος. Η εξήγηση σε αυτό προκύπτει από το γεγονός ότι η φύση των δεδομένων είναι τέτοια που είναι αδύνατον να διαχωριστεί το σύνολο δεδομένων σε υποσύνολα εκπαίδευσης, επικύρωσης και ελέγχου παρόμοιων κατανομών. Έτσι, είναι πιθανό το σύνολο ελέγχου να περιέχει περισσότερες άγνωστες προς το σύστημα ρυθμικές ακολουθίες απ'ότι το σύνολο επικύρωσης, δικαιολογώντας έτσι τη χαμηλότερη απόδοση. Παρ'όλα αυτά, η εκπαίδευση του νευρωνικού δικτύου έδειξε πως φαινομενικά κακή απόδοση δεν συνεπάγεται και κακά ηχητικά αποτελέσματα. Το πόρισμα αυτό μπορεί να γίνει πιο κατανοητό αναλύοντας το πλήρες ιστορικό του νευρωνικού δικτύου για τα σύνολα εκπαίδευσης και επικύρωσης το οποίο παρουσιάζεται στις Εικόνες 4.2–4.7:



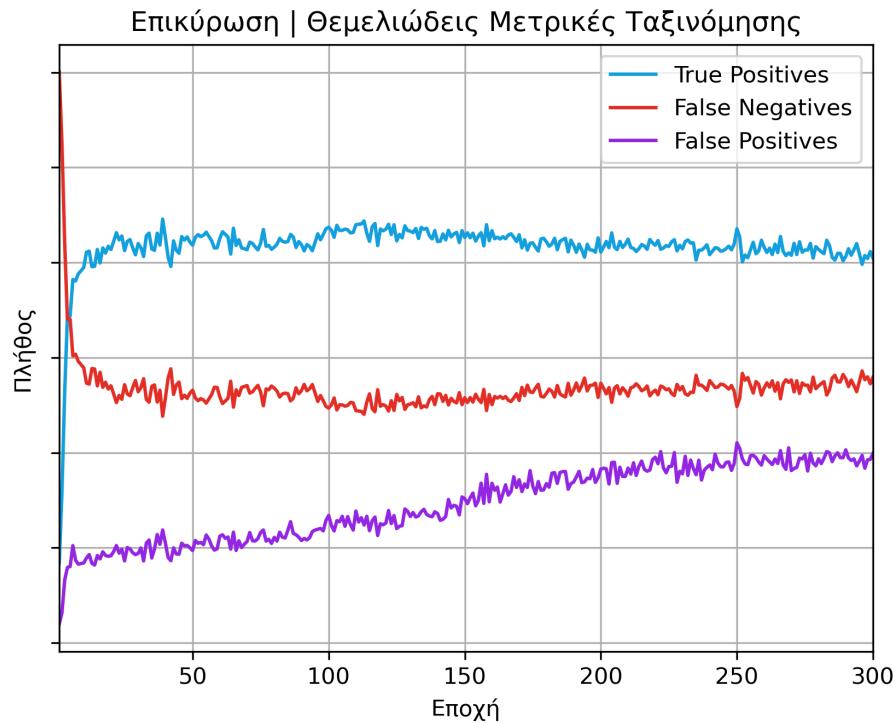
Εικόνα 4.2: Συνάρτηση κόστους στα σύνολα εκπαίδευσης και επικύρωσης



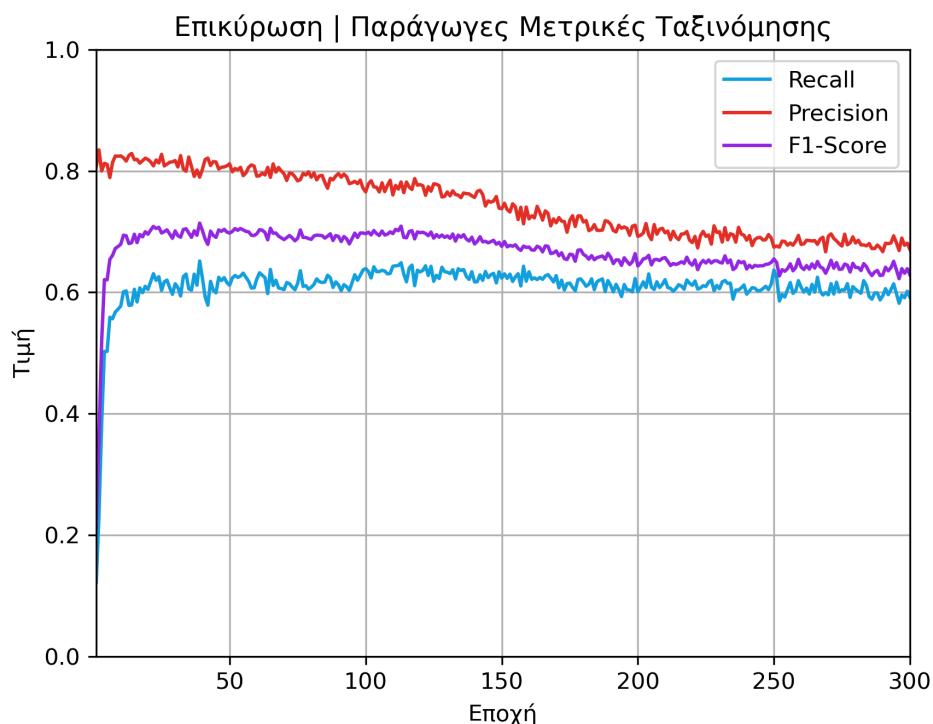
Εικόνα 4.3: Θεμελιώδεις μετρικές ταξινόμησης στο σύνολο εκπαίδευσης



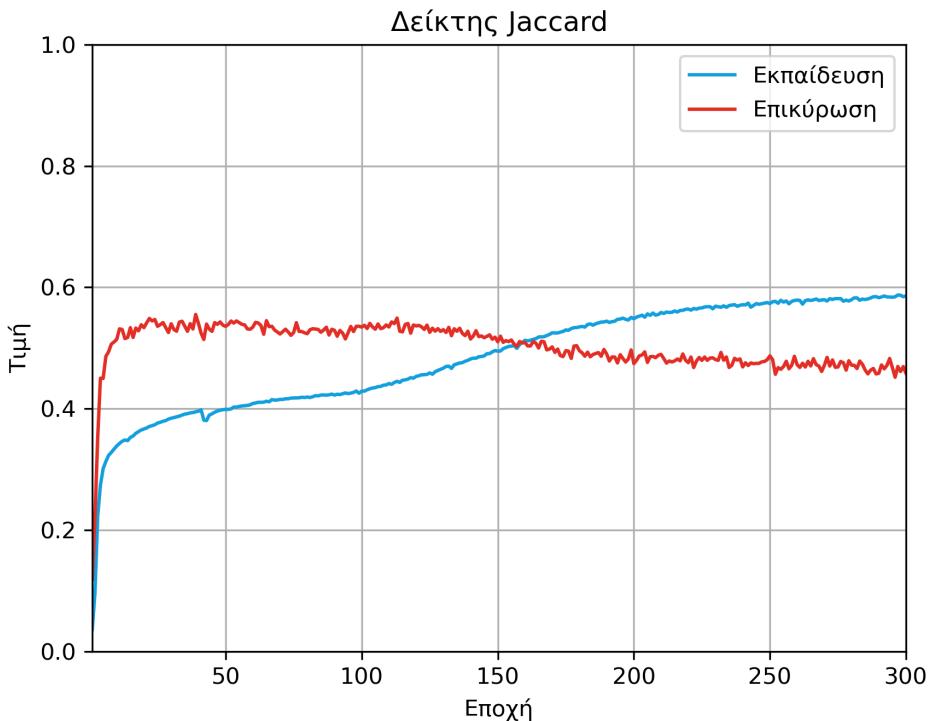
Εικόνα 4.4: Παράγωγες μετρικές ταξινόμησης στο σύνολο εκπαίδευσης



Εικόνα 4.5: Θεμελιώδεις μετρικές ταξινόμησης στο σύνολο επικύρωσης



Εικόνα 4.6: Παράγωγες μετρικές ταξινόμησης στο σύνολο επικύρωσης



Εικόνα 4.7: Δείκτης Jaccard στα σύνολα εκπαίδευσης και επικύρωσης

Όπως είναι λογικό, στο σύνολο εκπαίδευσης τόσο η συνάρτηση κόστους όσο και οι μετρικές αξιολόγησης παρουσιάζουν μια αργή αλλά πρακτικά σταθερή βελτίωση. Αντιθέτως, στο σύνολο επικύρωσης φαίνεται πως μετά από περίπου 100 εποχές η απόδοση του συστήματος αρχίζει να πέφτει σημαντικά. Αυτό που συμβαίνει στην πραγματικότητα είναι πως στις αρχές της εκπαίδευσης το νευρωνικό δίκτυο έχει μάθει να παράγει μόνο πολύ απλά και βασικά μοτίβα τυμπάνων, τα οποία ωστόσο ακούγονται αρκετά αδιάφορα και επαναλαμβανόμενα. Καθώς η εκπαίδευση του δικτύου προχωράει το πλήθος των False Positives αυξάνει, με αποτέλεσμα τα παραγόμενα μοτίβα τυμπάνων να παρεκκλίνουν από τα επιθυμητά αλλά ταυτόχρονα να γίνονται πολύ πιο δημιουργικά και να προσαρμόζονται καλύτερα στους διάφορους ρυθμούς. Έτσι, καθίσταται σαφές πως η φύση του προβλήματος απαιτεί το νευρωνικό δίκτυο να περάσει σε φάση υπερεκπαίδευσης, προσέχοντας ωστόσο έτοι ώστε να διατηρηθεί η ικανότητα γενίκευσης. Επομένως, οι μετρικές που αξιοποιήθηκαν αφενός είναι χρήσιμες για την παρακολούθηση της εκπαίδευσης του νευρωνικού δικτύου, αφετέρου δεν επαρκούν για την αντικειμενική αξιολόγηση του συστήματος.

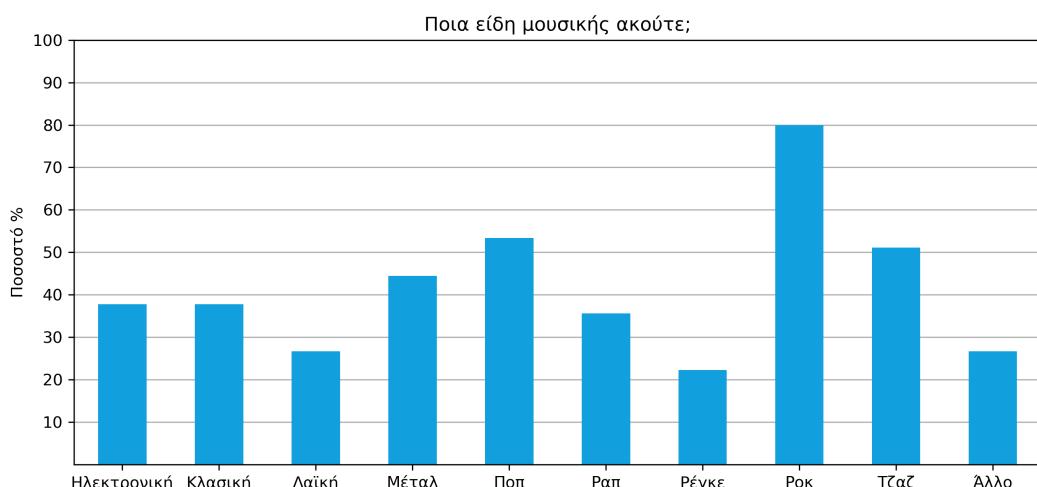
4.2.2 Αποτελέσματα 'Έρευνας'

Στην έρευνα που διεξήχθη έλαβαν μέρος συνολικά 45 άτομα τα δημογραφικά στοιχεία των οποίων παρουσιάζονται στις Εικόνες 4.8–4.12. Όπως φαίνεται, οι προτιμήσεις των συμμετεχόντων καλύπτουν ένα μεγάλο εύρος μουσικών ειδών, με ένα σημαντικό ποσοστό 80% των συμμετεχόντων να ακούει το είδος της rock μουσικής, το οποίο είναι αρκετά κοντά στη μουσική του συγκροτήματος Iron Maiden.

Ειδικότερα με τη μουσική αυτή φαίνεται πως είναι εξοικειωμένη η πλειοψηφία των συμμετεχόντων με ένα ποσοστό 64.44%. Όσον αφορά τη μουσική εμπειρία, ένα αξιόλογο ποσοστό 26.67% των συμμετεχόντων ξέρει να παίζει τύμπανα, ενώ ένα μεγαλύτερο ποσοστό τους 57.78% ξέρει να παίζει κάποιο άλλο όργανο. Τέλος, το 46.67% των συμμετεχόντων έχει κάποια άλλη μουσική ενασχόληση, όπως μπορεί να είναι το τραγούδι, η ηχοληψία, το DJing κ.ά. Βάσει των παραπάνω δημογραφικών στοιχείων οι συμμετέχοντες της έρευνας χωρίστηκαν σε έξι διαφορετικές ομάδες, για κάθε μία από τις οποίες πραγματοποιήθηκε ξεχωριστή στατιστική ανάλυση των αποτελεσμάτων. Οι ομάδες αυτές συντελέστηκαν ως εξής:

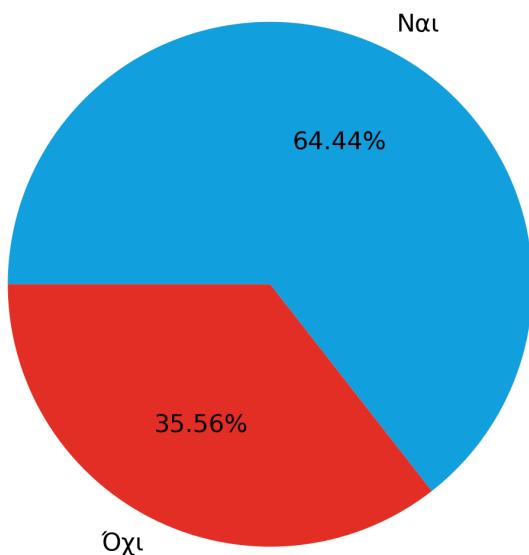
- 1) 'Όλοι οι συμμετέχοντες.
Συμμετοχή: 45/45 (100%)
- 2) 'Οσοι είναι εξοικειωμένοι με τη μουσική του συγκροτήματος Iron Maiden.
Συμμετοχή: 29/45 (64.44%)
- 3) 'Οσοι δεν είναι εξοικειωμένοι με τη μουσική του συγκροτήματος Iron Maiden.
Συμμετοχή: 16/45 (35.56%)
- 4) 'Οσοι ξέρουν να παίζουν τύμπανα.
Συμμετοχή: 12/45 (26.67%)
- 5) 'Οσοι δεν ξέρουν να παίζουν τύμπανα, αλλά ξέρουν να παίζουν κάποιο άλλο όργανο ή έχουν κάποια άλλη μουσική ενασχόληση.
Συμμετοχή: 23/45 (51.11%)
- 6) 'Οσοι δεν ξέρουν να παίζουν τύμπανα, δεν ξέρουν να παίζουν κάποιο άλλο μουσικό όργανο και δεν έχουν κάποια άλλη μουσική ενασχόληση.
Συμμετοχή: 10/45 (22.22%)

Τα αποτελέσματα των μέσων βαθμολογιών για τα χαρακτηριστικά των ανθρώπινων και τεχνητών δειγμάτων καθώς επίσης και των συνολικών ποσοστών της αναγνώρισης των τεχνητών δειγμάτων για κάθε ομάδα παρουσιάζονται στις Εικόνες 4.13—4.24.



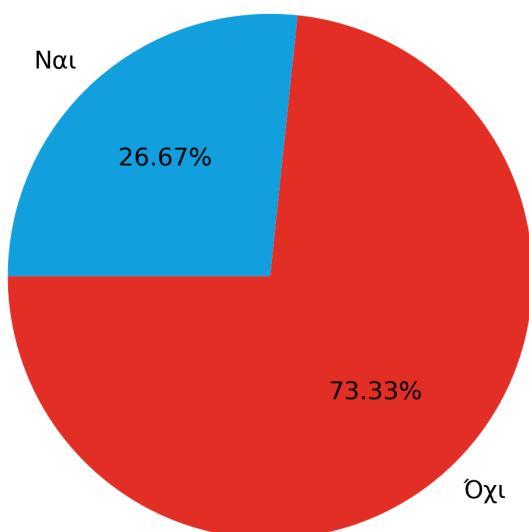
Εικόνα 4.8: Δημογραφικά στοιχεία έρευνας (I)

Είστε εξοικειωμένοι με τη μουσική του συγκροτήματος Iron Maiden;



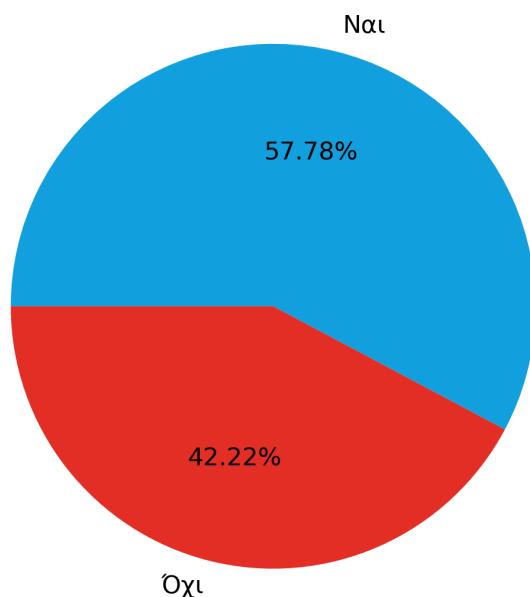
Εικόνα 4.9: Δημογραφικά στοιχεία έρευνας (II)

Ξέρετε να παίζετε τύμπανα;



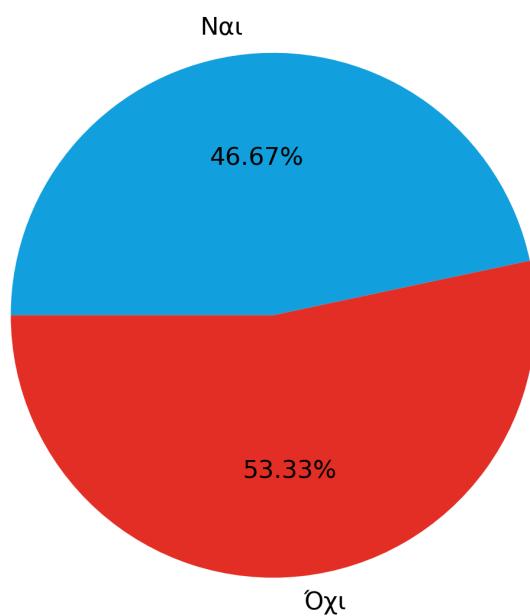
Εικόνα 4.10: Δημογραφικά στοιχεία έρευνας (III)

Ξέρετε να παίζετε κάποιο άλλο μουσικό όργανο;

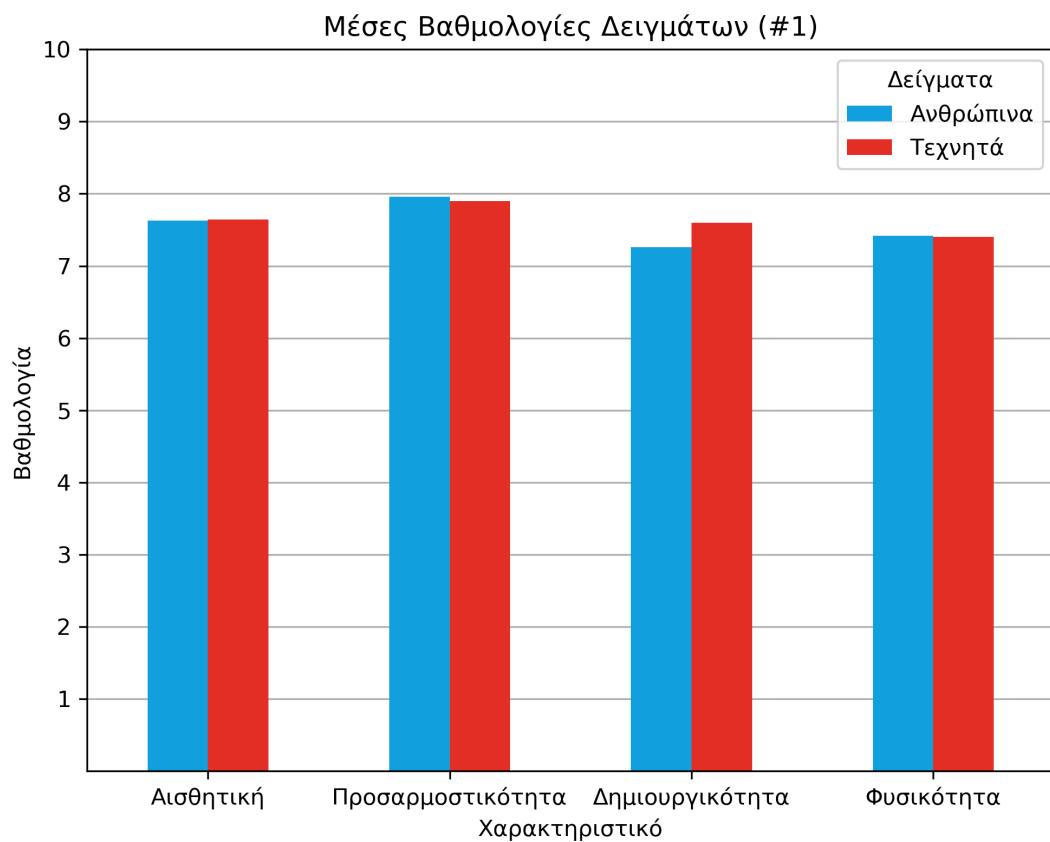


Εικόνα 4.11: Δημογραφικά στοιχεία έρευνας (IV)

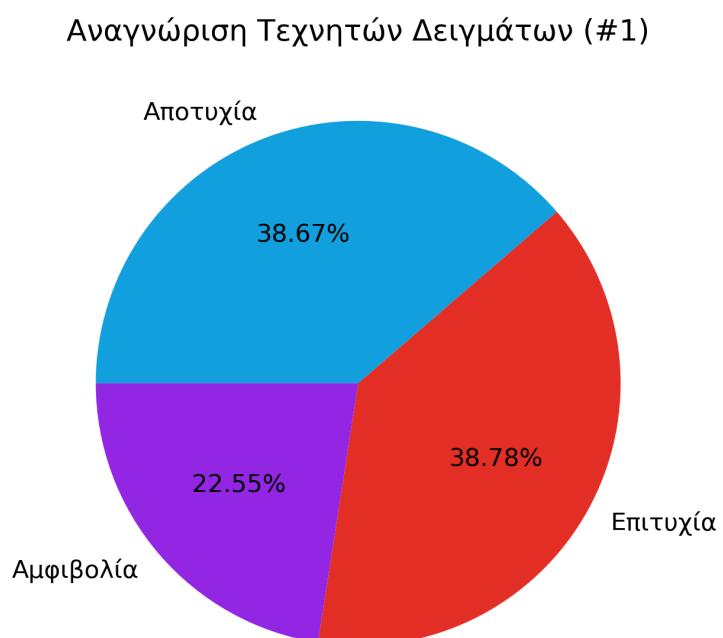
Έχετε κάποια άλλη μουσική ενασχόληση;



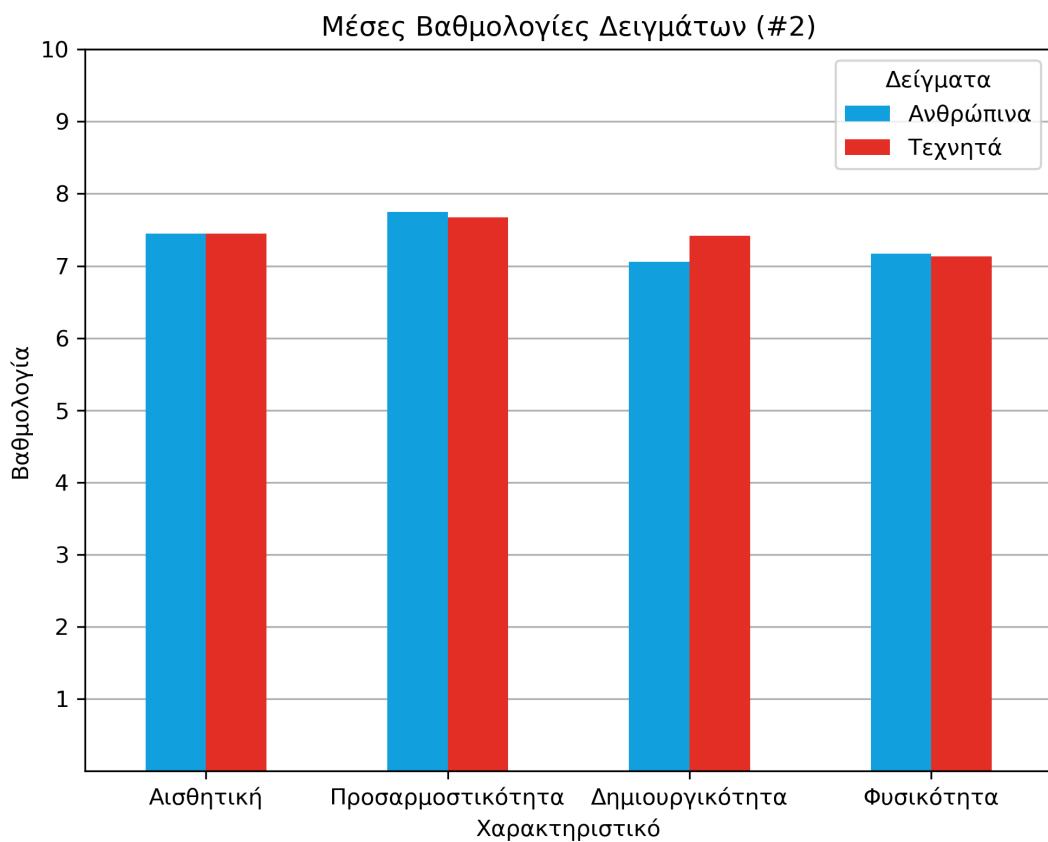
Εικόνα 4.12: Δημογραφικά στοιχεία έρευνας (V)



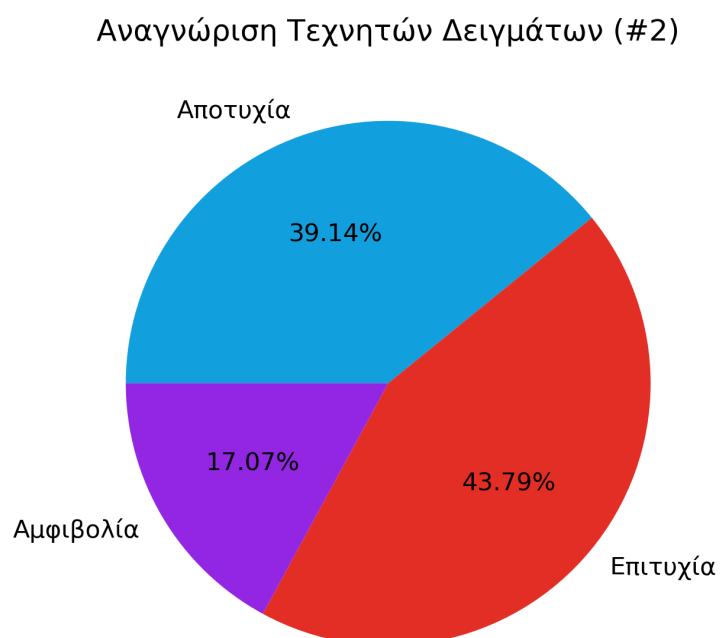
Εικόνα 4.13: Ομάδα 1 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων



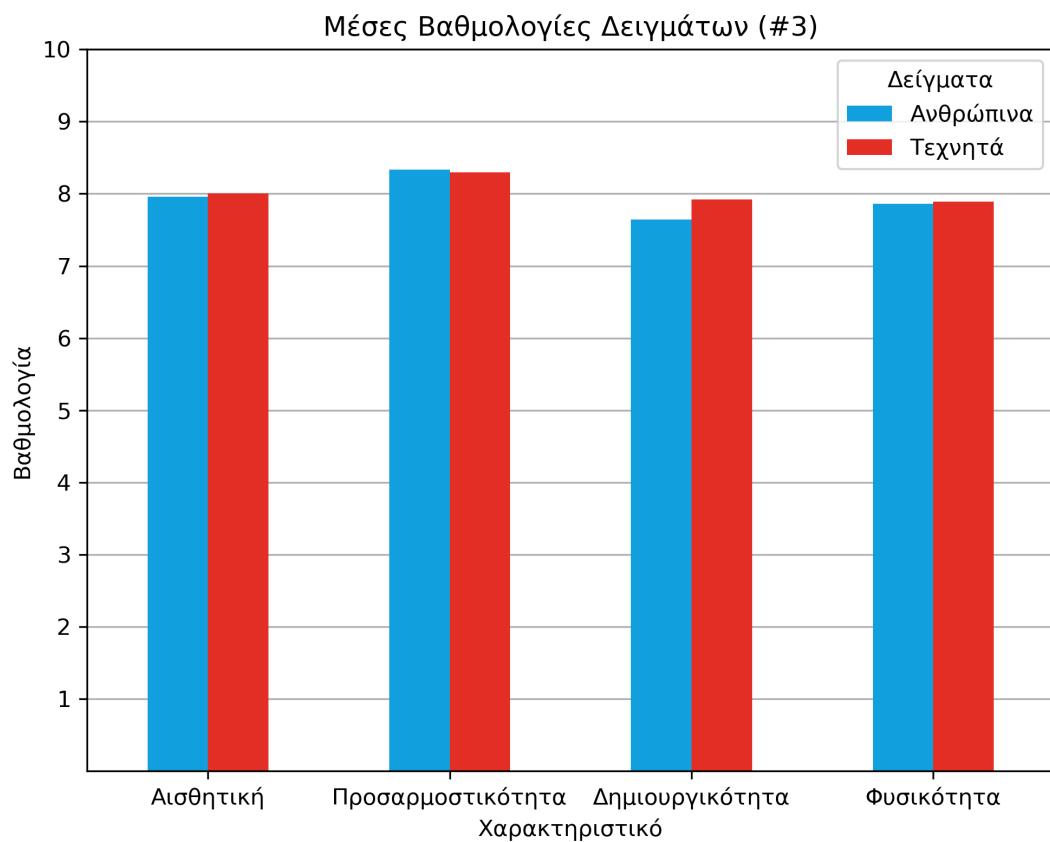
Εικόνα 4.14: Ομάδα 1 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων



Εικόνα 4.15: Ομάδα 2 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων

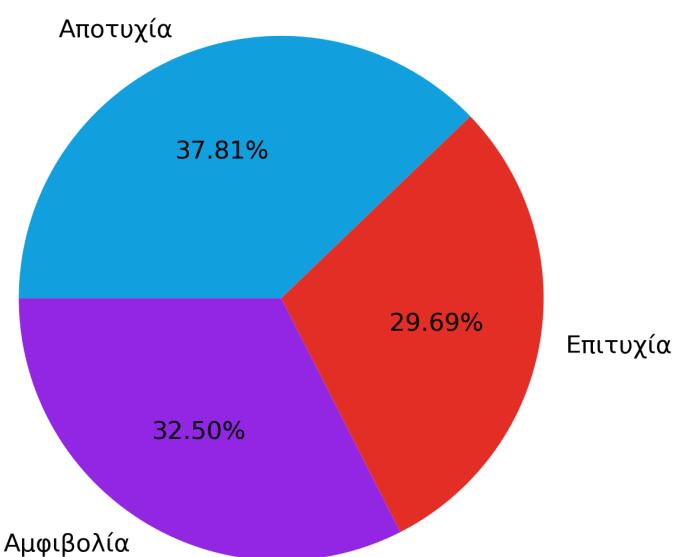


Εικόνα 4.16: Ομάδα 2 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων

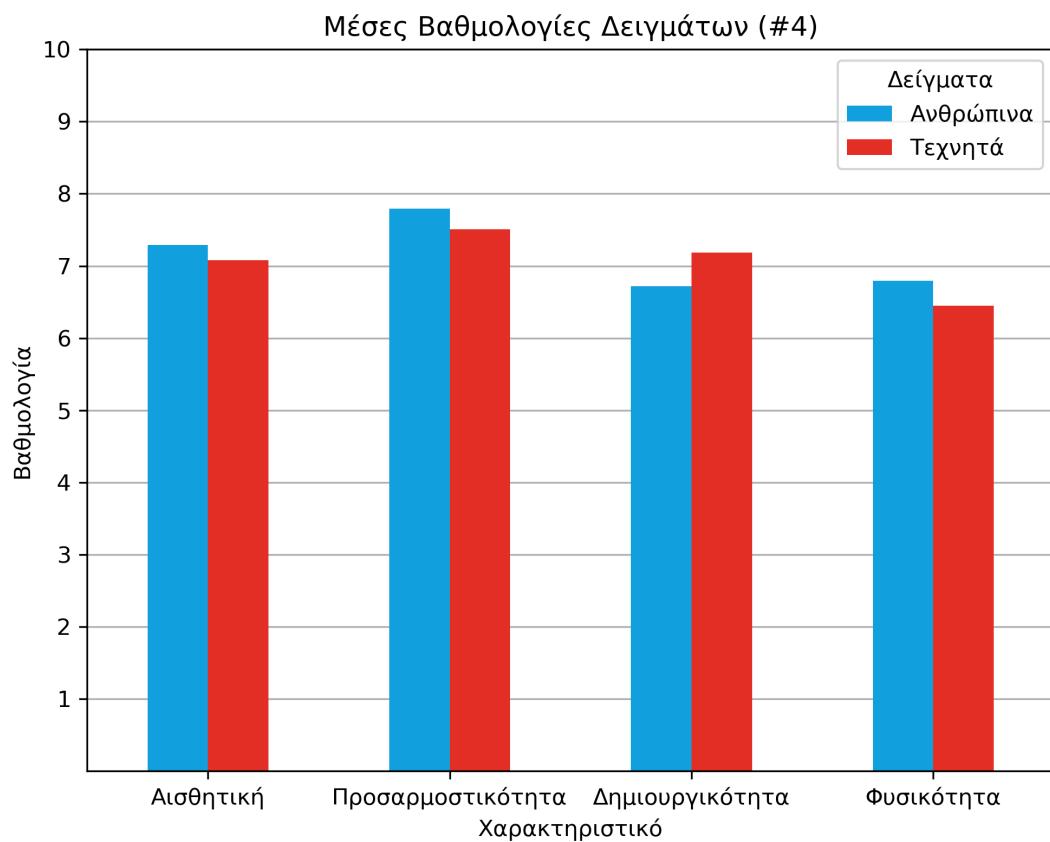


Εικόνα 4.17: Ομάδα 3 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων

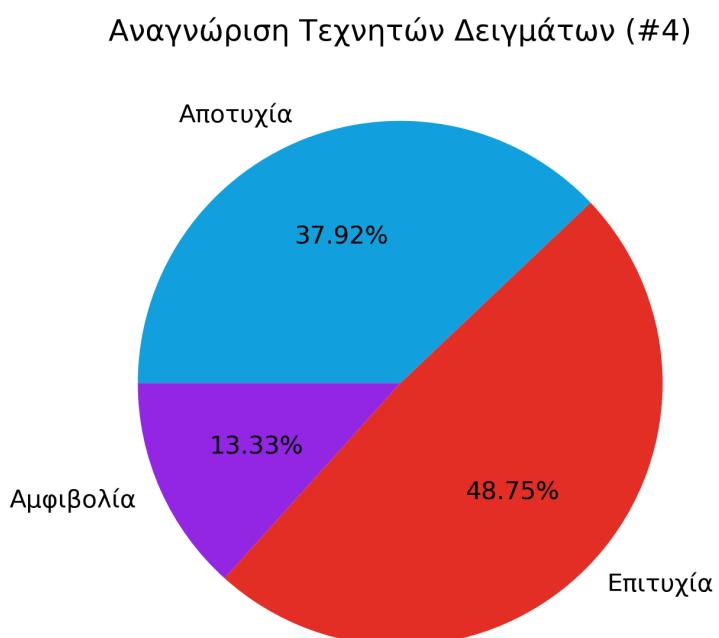
Αναγνώριση Τεχνητών Δειγμάτων (#3)



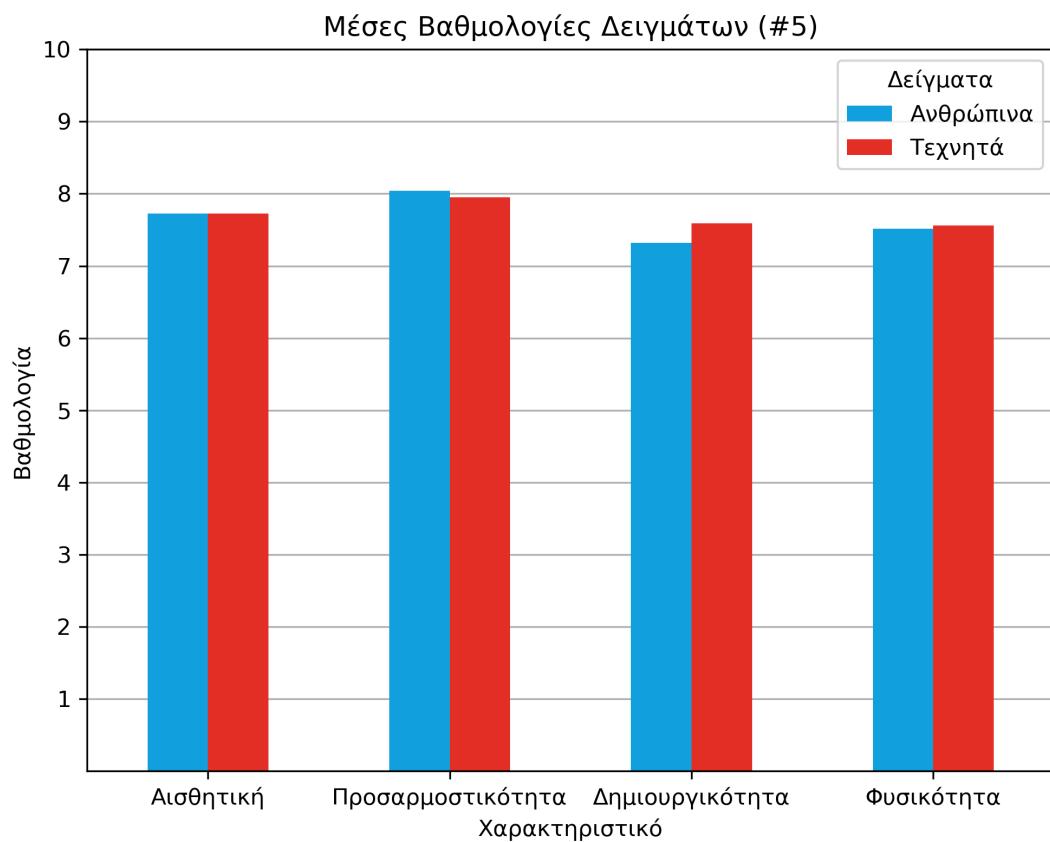
Εικόνα 4.18: Ομάδα 3 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων



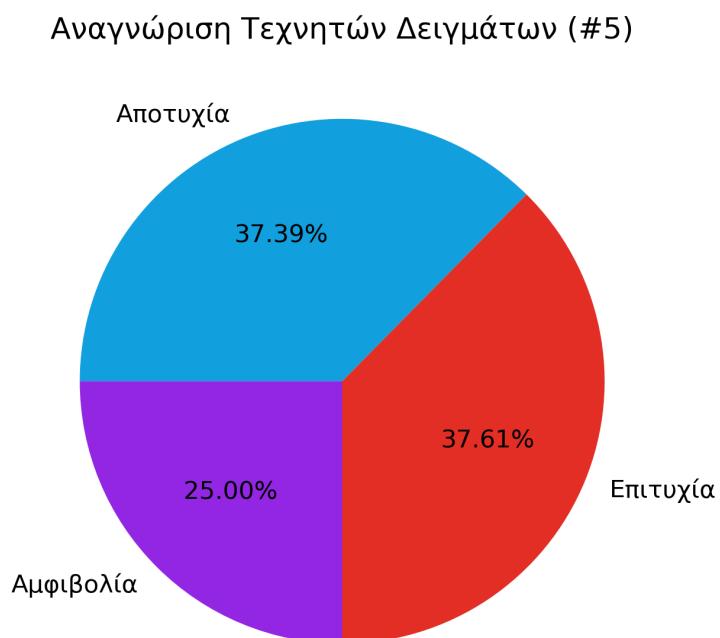
Εικόνα 4.19: Ομάδα 4 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων



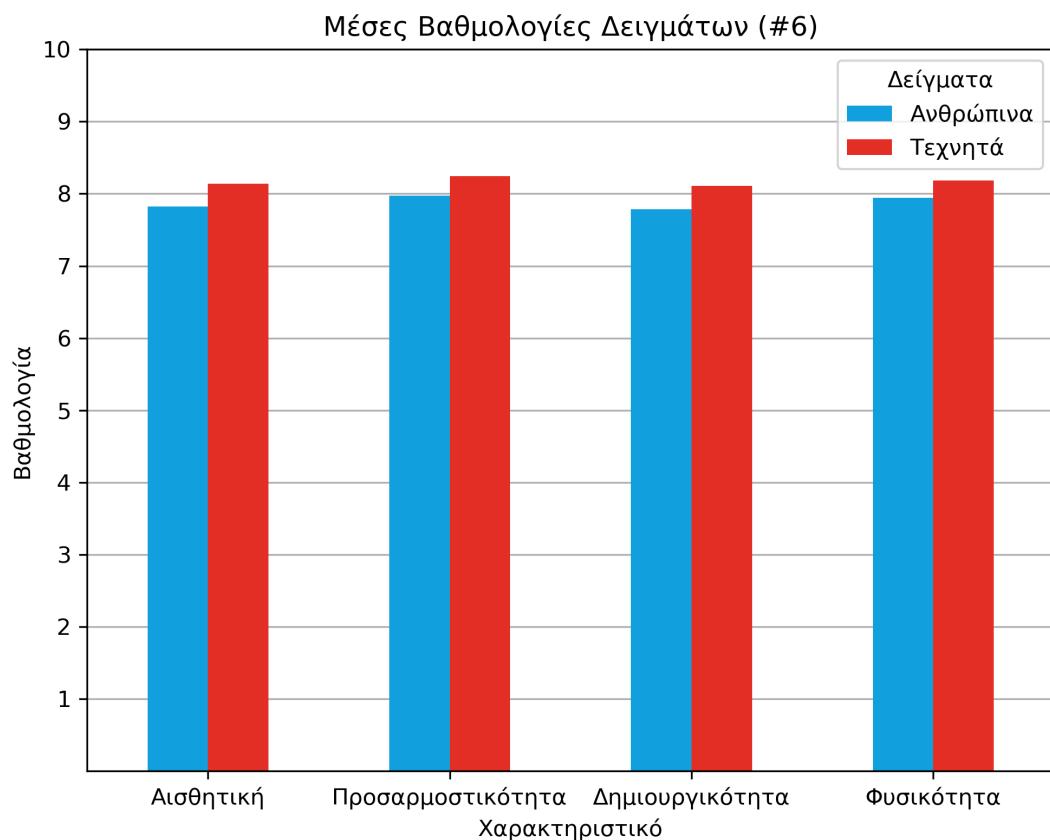
Εικόνα 4.20: Ομάδα 4 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων



Εικόνα 4.21: Ομάδα 5 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων

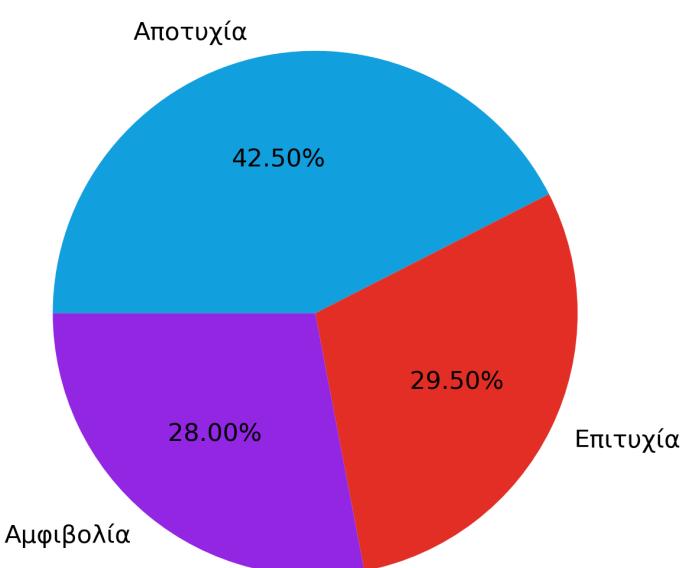


Εικόνα 4.22: Ομάδα 5 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων



Εικόνα 4.23: Ομάδα 6 - Αποτελέσματα μέσων βαθμολογιών δειγμάτων

Αναγνώριση Τεχνητών Δειγμάτων (#6)



Εικόνα 4.24: Ομάδα 6 - Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων

Τα συγκεντρωτικά αποτελέσματα της έρευνας παρουσιάζονται αναλυτικότερα στους Πίνακες 4.2 και 4.3, από όπου προκύπτουν και τα παρακάτω γενικά συμπεράσματα:

- Οι ομάδες {2,4} δείχνουν μεγαλύτερη προτίμηση προς τα ανθρώπινα δείγματα, παρατηρούν μικρότερη ομοιότητα ανάμεσα στους δύο τύπους δειγμάτων και έχουν μεγαλύτερα ποσοστά επιτυχίας αναγνώρισης των τεχνητών δειγμάτων.
- Οι ομάδες {3,6} δείχνουν μεγαλύτερη προτίμηση προς τα τεχνητά δείγματα, παρατηρούν μεγαλύτερη ομοιότητα ανάμεσα στους δύο τύπους δειγμάτων και έχουν μικρότερα ποσοστά επιτυχίας αναγνώρισης των τεχνητών δειγμάτων.
- Οι ομάδες {1,5} βρίσκονται στο ενδιάμεσο, χωρίς να δείχνουν κάποια ιδιαίτερη προτίμηση προς κάποιον από τους δύο τύπους δειγμάτων, παρατηρώντας μια σχετική ομοιότητα ανάμεσα τους και έχοντας παραπλήσια ποσοστά επιτυχίας και αποτυχίας αναγνώρισης των τεχνητών δειγμάτων.

Πίνακας 4.2: Αποτελέσματα μέσων βαθμολογιών δειγμάτων

	Αισθητική		Προσαρμοστικότητα		Δημιουργικότητα		Φυσικότητα		Ομοιότητα
	A	T	A	T	A	T	A	T	
Ομάδα 1	7.63	<u>7.65</u>	<u>7.96</u>	7.90	7.26	<u>7.60</u>	<u>7.42</u>	7.40	6.26
Ομάδα 2	<u>7.45</u>	7.44	<u>7.75</u>	7.68	7.05	<u>7.42</u>	<u>7.17</u>	7.13	5.83
Ομάδα 3	7.96	<u>8.00</u>	<u>8.34</u>	8.30	7.64	<u>7.92</u>	7.86	<u>7.89</u>	7.01
Ομάδα 4	<u>7.29</u>	7.08	<u>7.79</u>	7.50	6.72	<u>7.18</u>	<u>6.80</u>	6.45	5.24
Ομάδα 5	7.72	7.72	<u>8.04</u>	7.95	7.32	<u>7.59</u>	7.51	<u>7.56</u>	6.31
Ομάδα 6	7.82	<u>8.14</u>	7.97	<u>8.24</u>	7.78	<u>8.11</u>	7.94	<u>8.18</u>	7.34

(A: Ανθρώπινα δείγματα | T: Τεχνητά δείγματα)

Πίνακας 4.3: Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων

	Επιτυχία	Αποτυχία	Αμφιβολία
Ομάδα 1	<u>38.78%</u>	38.67%	22.55%
Ομάδα 2	<u>43.79%</u>	39.14%	17.07%
Ομάδα 3	29.69%	<u>37.81%</u>	32.50%
Ομάδα 4	<u>48.75%</u>	37.92%	13.33%
Ομάδα 5	<u>37.61%</u>	37.39%	25.00%
Ομάδα 6	29.50%	<u>42.50%</u>	28.00%

Οι παραπάνω ισχυρισμοί που αφορούν την προτίμηση κάποιου από τους δύο τύπους δειγμάτων μπορούν να επιβεβαιωθούν ή να απορριφθούν σε επίπεδο στατιστικής σημαντικότητας εφαρμόζοντας έλεγχο υπόθεσης για τη διαφορά των μέσων βαθμολογιών κάθε χαρακτηριστικού μεταξύ των ανθρώπινων και των τεχνητών δειγμάτων. Τα αποτελέσματα των t-test παρουσιάζονται στον Πίνακα 4.4.

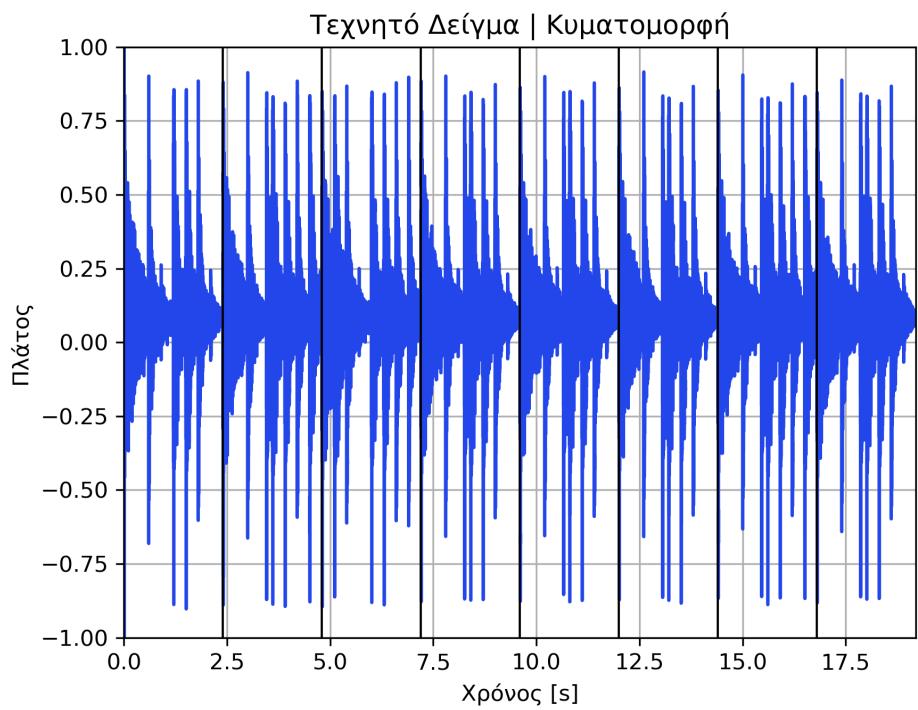
Πίνακας 4.4: Αποτελέσματα t-test (p-τιμές)

Αισθητική	Προσαρμοστικότητα	Δημιουργικότητα	Φυσικότητα
Ομάδα 1	0.8644	0.4471	<u>0.0002</u>
Ομάδα 2	0.9742	0.4566	<u>0.0008</u>
Ομάδα 3	0.7283	0.7741	0.0612
Ομάδα 4	0.1992	0.0506	<u>0.0056</u>
Ομάδα 5	1.0000	0.4828	<u>0.0427</u>
Ομάδα 6	<u>0.0155</u>	<u>0.0382</u>	<u>0.0140</u>
			0.0624

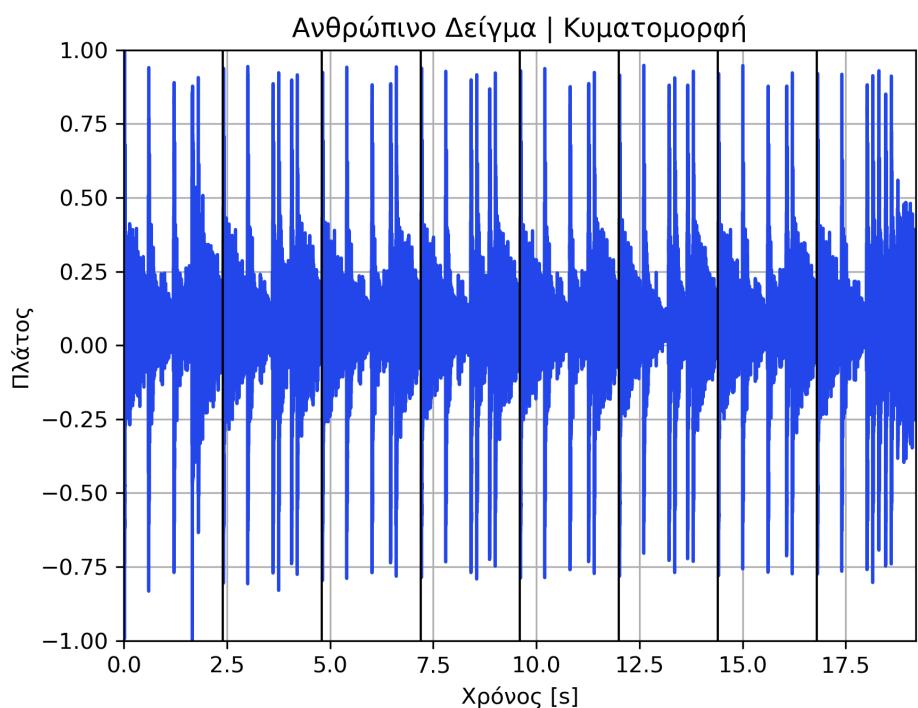
Αποδεικνύεται επομένως πως όλες οι ομάδες, πλην της ομάδας 3, δείχνουν στατιστικά σημαντική προτίμηση ως προς τη δημιουργικότητα των τεχνητών δειγμάτων, ενώ η ομάδα 6 δείχνει επιπλέον στατιστικά σημαντική προτίμηση ως προς την αισθητική και την προσαρμοστικότητα των τεχνητών δειγμάτων. Συνεπώς, τα παραγόμενα μοτίβα τυμπάνων του συστήματος βρίσκονται χωρίς αμφιθολία τουλάχιστον στο ίδιο επίπεδο με τα αντίστοιχα μοτίβα τυμπάνων που θα έπαιζε ένας άνθρωπος, επικυρώνοντας έτσι την αποτελεσματικότητα του συστήματος.

4.3 Οπτικοποίηση Δειγμάτων

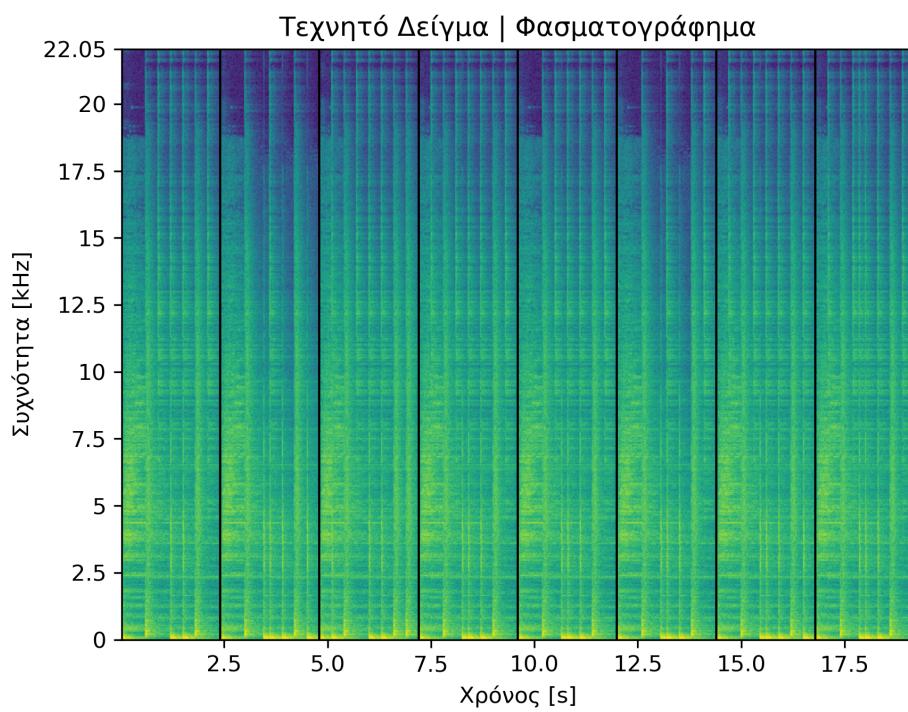
Ένας άλλος τρόπος με τον οποίο μπορεί να πραγματοποιηθεί ένας ποιοτικός έλεγχος των παραγόμενων μοτίβων τυμπάνων είναι μέσω της οπτικοποίησης των ηχητικών δειγμάτων που προκύπτουν από την ηχητική σύνθεση της συμβολικής πληροφορίας. Η οπτικοποίηση αυτή μπορεί να λάβει χώρα στο πεδίο του χρόνου και στο πεδίο του χρόνου και της συχνότητας. Στις Εικόνες 4.25 και 4.26 παρουσιάζονται οι κυματομορφές ενός τεχνητού και ενός ανθρώπινου δείγματος που αφορούν το ίδιο μουσικό απόσπασμα διάρκειας οκτώ μέτρων 4/4. Οι κορυφές των κυματομορφών αυτών αναπαριστούν τις ρυθμικές εναλλαγές της μπότας και του ταμπούρου που αποτελούν τον κορμό του ρυθμικού μοτίβου. Όπως μπορεί να παρατηρηθεί, οι ρυθμοί των δύο δειγμάτων διαφέρουν μεταξύ τους αλλά φαίνεται πως ο ρυθμός του τεχνητού δείγματος είναι συμπαγής και διαθέτει συνοχή από μέτρο σε μέτρο, όπως συμβαίνει και με τον ρυθμό του ανθρώπινο δείγματος, πράγμα το οποίο είναι θεμιτό. Στις Εικόνες 4.27 και 4.28 παρουσιάζονται τα αντίστοιχα φασματογραφήματα των δύο δειγμάτων. Εδώ, αν και εφικτό, δεν είναι τόσο εύκολο να παρατηρηθούν τα μοτίβα της μπότας και του ταμπούρου καθώς η πληροφορία τους βρίσκεται στις κατώτερες συχνότητες, ωστόσο είναι αρκετά ευδιάκριτη η χρήση των πιατινιών. Συγκεκριμένα, παρατηρώντας γύρω από την προσεγγιστική περιοχή 3kHz – 8kHz μπορεί να φανεί η αύξηση της ενέργειας που προκαλείται από το χτύπημα του crash. Στο τεχνητό δείγμα αυτό συμβαίνει στην αρχή κάθε μέτρου για έμφαση, ενώ στο ανθρώπινο δείγμα συμβαίνει τόσο στην αρχή όσο και προς το τέλος των μέτρων. Τέλος, οι κάθετες ισαπέχουσες γραμμές που διακρίνονται στις υψηλές συχνότητες φανερώνουν το χτύπημα κάποιου πιατινιού για τη διατήρηση ενός σταθερού ρυθμού. Στο τεχνητό δείγμα αυτό γίνεται με χρήση του ride, ενώ στο ανθρώπινο δείγμα γίνεται κατά κύριο λόγο με χτύπημα του hi-hat σε ανοιχτή θέση.



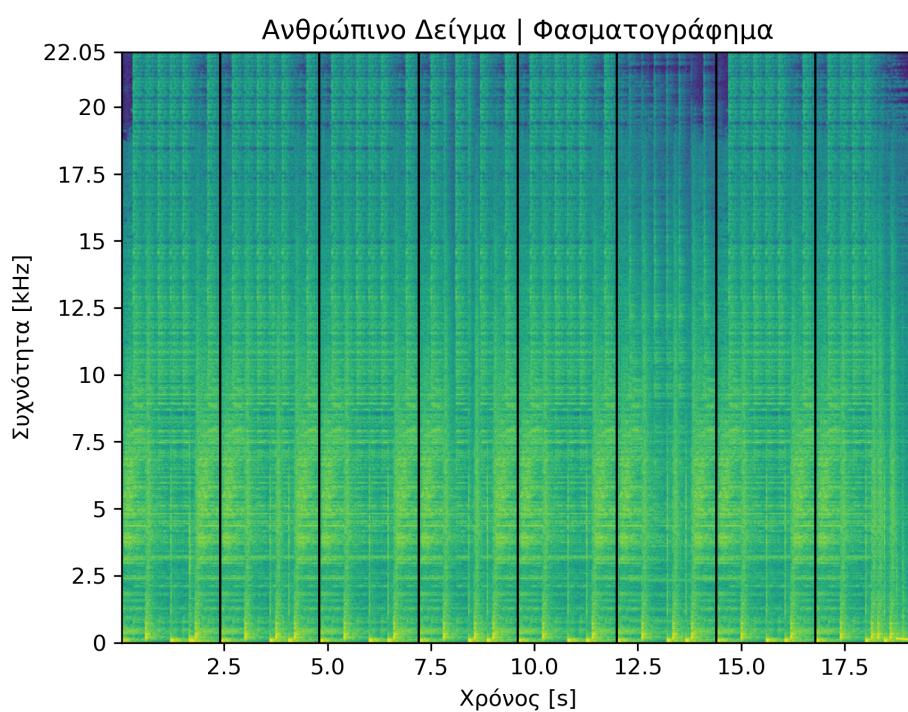
Εικόνα 4.25: Κυματομορφή τεχνητού δείγματος



Εικόνα 4.26: Κυματομορφή ανθρώπινου δείγματος



Εικόνα 4.27: Φασματογράφημα τεχνητού δείγματος



Εικόνα 4.28: Φασματογράφημα ανθρώπινου δείγματος

Κεφάλαιο 5

Συμπεράσματα

5.1 Σύνοψη

Τα αποτελέσματα που προέκυψαν από τη διαδικασία αξιολόγησης έδειξαν πως το προτεινόμενο σύστημα είναι ικανό να παράγει μοτίβα τυμπάνων τα οποία είναι αισθητικά ικανοποιητικά και προσαρμόζονται καλά στον ρυθμό της μουσικής, ενώ είναι επίσης ιδιαίτερα δημιουργικά και η φυσικότητα τους βρίσκεται σε ανθρώπινο επίπεδο. Το μουσικό υπόβαθρο των ακροατών φαίνεται πως παίζει σημαντικό ρόλο στην αντίληψη αυτών των χαρακτηριστικών, αλλά αποδείχθηκε πως ακόμα και οι πιο έμπειροι μουσικοί μπορούν να παραπλανηθούν από την ποιότητα των τεχνητών μοτίβων. Από την άλλη μεριά, η φύση του προβλήματος είναι τέτοια που καθιστά ανεπαρκή τη χρήση μετρικών για την αξιολόγηση ενός τέτοιου συστήματος και η εύρεση κατάλληλων αντικειμενικών κριτηρίων παραμένει ένα ζήτημα προς διερεύνηση. Σε κάθε περίπτωση, το μόνο σίγουρο είναι πως όταν πρόκειται για τέχνη ο καλύτερος κριτής μπορεί να είναι μόνο ο άνθρωπος.

5.2 Συνεισφορά

Η προσέγγιση της παρούσας εργασίας αποτελεί από μόνη της μεγάλη συνεισφορά στο πεδίο της παραγωγής μοτίβων τυμπάνων, καθώς αντίστοιχες μεθοδολογίες που βασίζονται σε ηχητικά δεδομένα για την επίλυση του προβλήματος βρίσκονται σε έλλειψη. Η χρήση του OSS για τον σκοπό αυτό αποτελεί καινοτόμο στοιχείο και μπορεί να αποτελέσει σημαντικό θεμέλιο σε παρόμοιες μελλοντικές προσεγγίσεις. Επιπλέον, η ικανότητα του προτεινόμενου συστήματος να παράγει μοτίβα τυμπάνων τα οποία είναι προσαρμοσμένα στη μετρική δομή του μουσικού κομματιού αποτελεί ιδιαίτερο χαρακτηριστικό, το οποίο μπορεί να δώσει το έναυσμα για την υλοποίηση παρόμοιων συστημάτων.

5.3 Μελλοντικές Επεκτάσεις

Το εν λόγω σύστημα υλοποιήθηκε με γνώμονα την ικανοποιητική επίλυση του προβλήματος της παραγωγής μοτίβων τυμπάνων με τον απλούστερο δυνατό τρόπο.

Συνεπώς, πρόκειται ήδη για μία πολύ γερή βάση η οποία ωστόσο επιδέχεται πολλές βελτιώσεις, κάποιες από τις οποίες περιγράφονται παρακάτω:

- Ο περαιτέρω πειραματισμός με τις διάφορες παραμέτρους του συστήματος και την αρχιτεκτονική του νευρωνικού δικτύου μπορεί να φέρει στο προσκήνιο κάποια υλοποίηση με ακόμα πιο ικανοποιητικά αποτελέσματα.
- Η εκπαίδευση του νευρωνικού δικτύου με περισσότερα δεδομένα και ιδιαίτερα με δεδομένα τραγουδιών διαφορετικού μουσικού είδους θα εντείνει την προσαρμοστικότητα του συστήματος σε πολλούς και διαφορετικούς ρυθμούς.
- Η εισαγωγή παραμέτρων ελέγχου θα δώσει στους χρήστες τη δυνατότητα να ρυθμίζουν τη διαδικασία της παραγωγής μοτίβων τυμπάνων με βάση τις προτιμήσεις τους.
- Ο εξανθρωπισμός των παραγόμενων μοτίβων τυμπάνων θα προσδώσει μεγαλύτερη φυσικότητα στον ήχο που προκύπτει από τη συνθετική μετατροπή της συμβολικής πληροφορίας.
- Η υλοποίηση μιας ολοκληρωμένης εφαρμογής με γραφικό περιβάλλον φιλικό προς τον χρήστη θα επιτελέσει τον απότερο σκοπό του αντικειμένου της παραγωγής μοτίβων τυμπάνων, ο οποίος φυσικά δεν είναι η αντικατάσταση των αληθινών ντράμερ από τις μηχανές, αλλά η ανάπτυξη χρήσιμων εργαλείων που θα υποβοηθούν ερασιτέχνες και πιθανόν επαγγελματίες μουσικούς στη διαδικασία της σύνθεσης.

Βιβλιογραφικές Αναφορές

- [1] N. Tokui, “Can GAN Originate New Electronic Dance Music Genres? - Generating Novel Rhythm Patterns Using GAN with Genre Ambiguity Loss”, 2020. [Online]. Διαθέσιμο: <https://arxiv.org/abs/2011.13062>.
- [2] R. Vogl, H. Eghbal-Zadeh, and P. Knees, “An Automatic Drum Machine with Touch UI Based on a Generative Neural Network”, in *24th International Conference on Intelligent User Interfaces (IUI'19 Companion)*, Marina del Rey, CA, USA: Association for Computing Machinery, NY, USA, 2019, pp. 91–92. doi: 10.1145/3308557.3308673.
- [3] N. Tokui, “Towards Democratizing Music Production with AI - Design of Variational Autoencoder-Based Rhythm Generator as a DAW Plugin”, 2020. [Online]. Διαθέσιμο: <https://arxiv.org/abs/2004.01525>.
- [4] F. Bruford, S. McDonald, and M. Sandler, “*jaki*: User-Controllable Generation of Drum Patterns using LSTM Encoder-Decoder and Deep Reinforcement Learning”, in *1st Joint Conference on AI Music Creativity*, Stockholm, Sweden: AIMC, 2020, pp. 212–220. doi: 10.5281/zenodo.4285414.
- [5] K. Lee, W. Kim, and J. Nam, “PocketVAE: A Two-step Model for Groove Generation and Control”, 2021. [Online]. Διαθέσιμο: <https://arxiv.org/abs/2107.05009>.
- [6] P. Hutchings, “Talking Drums: Generating Drum Grooves with Neural Networks”, in *1st International Workshop on Deep Learning and Music joint with IJCNN*, Anchorage, AK, USA, 2017, pp. 43–47. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1706.09558>.
- [7] J. Gillick, A. Roberts, J. Engel, D. Eck, and D. Bamman, “Learning to Groove with Inverse Sequence Transformations”, in *36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 97, PMLR, 2019, pp. 2269–2279. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1905.06118>.
- [8] D. Makris, M. Kaliakatsos-Papakostas, I. Karydis, and K. L. Kermanidis, “Combining LSTM and Feed Forward Neural Networks for Conditional Rhythm Composition”, in *EANN 2017: Engineering Applications of Neural Networks*, ser. Communications in Computer and Information Science, vol. 744, Springer, 2017, pp. 570–582. doi: 10.1007/978-3-319-65172-9_48.

- [9] D. Makris, M. Kaliakatsos-Papakostas, and K. L. Kermanidis, “DeepDrum: An Adaptive Conditional Neural Network for Generating Drum Rhythms”, 2018. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1809.06127>.
- [10] D. Makris, M. Kaliakatsos-Papakostas, I. Karydis, and K. L. Kermanidis, “Conditional Neural Sequence Learners for Generating Drums’ Rhythms”, *Neural Computing and Applications*, vol. 31, no. 6, pp. 1793–1804, 2019. doi: [10.1007/s00521-018-3708-6](https://doi.org/10.1007/s00521-018-3708-6).
- [11] D. Makris, G. Zixun, M. Kaliakatsos-Papakostas, and D. Herremans, “Conditional Drums Generation using Compound Word Representations”, in *11th International Conference on Artificial Intelligence in Music, Sound, Art and Design*, Madrid, Spain: EvoMUSART, Apr. 2022, pp. 179–194. doi: [10.1007/978-3-031-03789-4_12](https://doi.org/10.1007/978-3-031-03789-4_12).
- [12] I.-C. Wei, C.-W. Wu, and L. Su, “Generating Structured Drum Pattern Using Variational Autoencoder and Self-Similarity Matrix”, in *20th International Society for Music Information Retrieval Conference*, Delft, The Netherlands: ISMIR, 2019, pp. 847–854. doi: [10.5281/zenodo.3527946](https://doi.org/10.5281/zenodo.3527946).
- [13] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Prentice Hall, 1999, ISBN: 9780130834430.
- [14] F. Lazarinis, *Πολυμέσα*. Kallipos, Open Academic Editions, 2015, ISBN: 978-960-603-141-0. [Online]. Διαθέσιμο: <http://hdl.handle.net/11419/2045>.
- [15] H. Wang and B. Raj, “On the Origin of Deep Learning”, 2017. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1702.07800>.
- [16] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, “Object Recognition with Gradient-Based Learning”, *Lecture Notes in Computer Science*, vol. 1681, pp. 319–345, Oct. 1999. doi: [10.1007/3-540-46805-6_19](https://doi.org/10.1007/3-540-46805-6_19).
- [17] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory”, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, “Generative Adversarial Nets”, in *Advances in Neural Information Processing Systems*, vol. 27, NIPS, 2014. [Online]. Διαθέσιμο: <https://arxiv.org/abs/2203.00667>.
- [19] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes”, in *2nd International Conference on Learning Representations*, ICLR, 2014. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1312.6114>.
- [20] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks”, in *Advances in Neural Information Processing Systems*, vol. 27, NIPS, 2014. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1409.3215>.
- [21] B. Haki and S. Jordà, “A Bassline Generation System Based on Sequence-to-Sequence Learning”, in *International Conference on New Interfaces for Musical Expression*, Porto Alegre, Brazil: NIME, Jun. 2019, pp. 204–209. doi: [10.5281/zenodo.3672928](https://doi.org/10.5281/zenodo.3672928).

- [22] G. Percival and G. Tzanetakis, “Streamlined Tempo Estimation Based on Autocorrelation and Cross-Correlation with Pulses”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1765–1776, 2014. doi: 10.1109/TASLP.2014.2348916.
- [23] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions: Comparison of Trends in Practice and Research for Deep Learning”, Nov. 2018. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1811.03378>.
- [24] S. Jadon, “A survey of loss functions for semantic segmentation”, in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, 2020. doi: 10.1109/CIBCB48159.2020.9277638.
- [25] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, in *3rd International Conference for Learning Representations*, San Diego, USA: ICLR, 2015. [Online]. Διαθέσιμο: <https://arxiv.org/abs/1412.6980>.
- [26] Željko Đ. Vujošić, “Classification Model Evaluation Metrics”, *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021. doi: 10.14569/IJACSA.2021.0120670.
- [27] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool”, *BMC Medical Imaging*, vol. 15, Aug. 2015. doi: 10.1186/s12880-015-0068-x.
- [28] B. L. Welch, “The Significance of the Difference Between Two Means when the Population Variances are Unequal”, *Biometrika*, vol. 29, no. 3/4, pp. 350–362, 1938. doi: 10.2307/2332010.

Παράρτημα Κώδικας & Εργαλεία

Για τους σκοπούς της εργασίας, και συγκεκριμένα για την υλοποίηση των σταδίων της προεπεξεργασίας, του νευρωνικού δικτύου και της μετεπεξεργασίας, αναπτύχθηκε κώδικας ο οποίος βρίσκεται στο αποθετήριο του ακόλουθου συνδέσμου:

<https://github.com/mogeadis/AIron-Drummer>

Η ανάπτυξη του κώδικα έγινε με τη γενικού σκοπού και υψηλού επιπέδου γλώσσα προγραμματισμού *Python*.⁹ Οι βιβλιοθήκες της Python που αξιοποιήθηκαν για την υλοποίηση του προτεινόμενου συστήματος παρουσιάζονται στον Πίνακα Π.1:

Πίνακας Π.1: Βιβλιοθήκες της Python

Βιβλιοθήκη	Χρήση
Mido ¹⁰	Επεξεργασία δεδομένων MIDI
Pandas ¹¹	Χειρισμός και ανάλυση δεδομένων
Numpy ¹²	Αριθμητικοί υπολογισμοί
Scipy ¹³	Επεξεργασία σημάτων
Keras ¹⁴	Υλοποίηση νευρωνικού δικτύου
Tensorflow ¹⁵	
Matplotlib ¹⁶	Σχεδίαση επιστημονικών διαγραμμάτων

Η συγγραφή του κώδικα Python έγινε με το δωρεάν λογισμικό επεξεργασίας κώδικα *Visual Studio Code* της *Microsoft*, ενώ η εκπαίδευση του νευρωνικού δικτύου πραγματοποιήθηκε στο περιβάλλον του *Google Colaboratory*, το οποίο αποτελεί μια υπηρεσία υπολογιστικού νέφους που δίνει εξ αποστάσεως πρόσθαση σε κάρτες γραφικών για σημαντική επιτάχυνση των αριθμητικών υπολογισμών.^{17,18}

⁹Python: <https://www.python.org>

¹⁰Mido: <https://mido.readthedocs.io>

¹¹Pandas: <https://pandas.pydata.org>

¹²Numpy: <https://numpy.org>

¹³Scipy: <https://scipy.org>

¹⁴Keras: <https://keras.io>

¹⁵Tensorflow: <https://www.tensorflow.org>

¹⁶Matplotlib: <https://matplotlib.org>

¹⁷Visual Studio Code: <https://code.visualstudio.com>

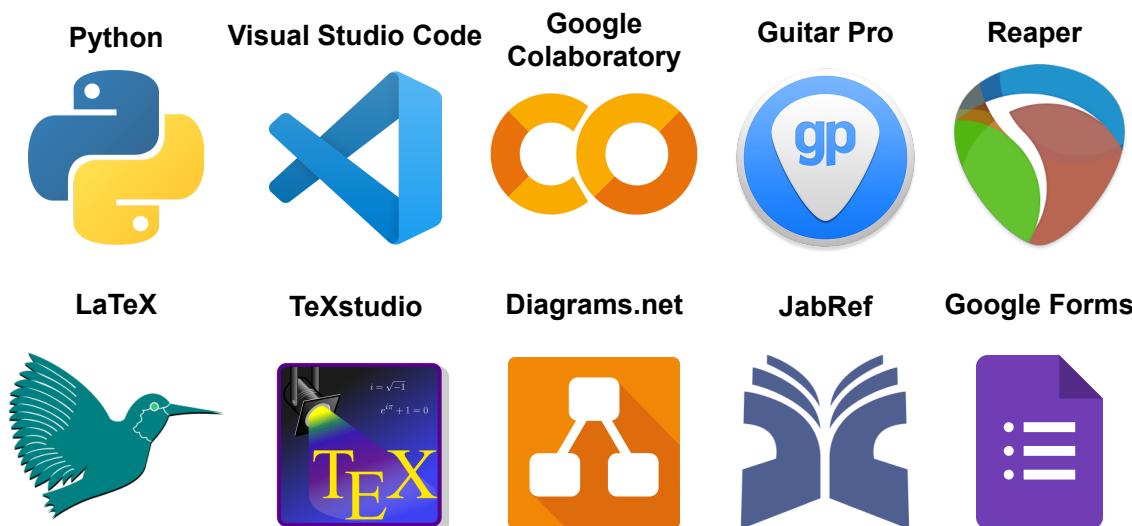
¹⁸Google Colaboratory: <https://colab.research.google.com>

Το συνόλο δεδομένων που συλλέχθηκε για την εκπαίδευση του νευρωνικού δικτύου επιμελήθηκε με το *Guitar Pro*, το εμπορικό λογισμικό μουσικής σημειογραφίας και επεξεργασίας MIDI της *Arobas Music*, ενώ η προετοιμασία των ηχητικών δειγμάτων που συμπεριλήφθηκαν στην έρευνα αξιολόγησης του συστήματος πραγματοποιήθηκε με χρήση του *REAPER*, του εμπορικού DAW της *Cockos*.^{19,20}

Η βιβλιογραφία που συγκεντρώθηκε κατά το στάδιο επισκόπησης του ερευνητικού πεδίου οργανώθηκε στο δωρεάν λογισμικό διαχείρισης βιβλιογραφικών αναφορών *JabRef*, ενώ το ερωτηματολόγιο της έρευνας αξιολόγησης του συστήματος συντάχθηκε στο περιβάλλον διαχείρισης ερευνών *Google Forms*.^{21,22}

Τέλος, η συγγραφή του παρόντος κειμένου πραγματοποιήθηκε με το σύστημα στοιχειοθεσίας *LATEX* στο δωρεάν ολοκληρωμένο περιβάλλον συγγραφής *TeXstudio*, ενώ για τη σχεδίαση των διαγραμμάτων που παρουσιάστηκαν αξιοποιήθηκε το δωρεάν λογισμικό ανοιχτού κώδικα *Diagrams.net*.^{23,24,25}

Τα λογότυπα των εργαλείων λογισμικού που αξιοποιήθηκαν για τους σκοπούς της εργασίας απεικονίζονται στην Εικόνα Π.1:



Εικόνα Π.1: Λογότυπα εργαλείων λογισμικού

¹⁹ *Guitar Pro*: <https://www.guitar-pro.com>

²⁰ *REAPER*: <https://www.reaper.fm>

²¹ *JabRef*: <https://www.jabref.org>

²² *Google Forms*: <https://docs.google.com/forms>

²³ The *LATEX* Project: <https://www.latex-project.org>

²⁴ *TeXstudio*: <https://www.texstudio.org>

²⁵ *Diagrams.net*: <https://diagrams.net>