

Τεχνικές Ανάκτησης Μουσικής Πληροφορίας για την Παραγωγή Ρυθμικών Μοτίβων Τυμπάνων

Αλέξανδρος	Ιορδάνης	Νικόλαος	Λάζαρος	Χαράλαμπος
Ηλιάδης	Θωΐδης	Βρύζας	Βρύσης	Δημούλας
Α.Π.Θ.	Α.Π.Θ.	Α.Π.Θ.	Α.Π.Θ.	Α.Π.Θ.
alexilia	ithoidis	nvryzas	lvrysis	babis
@ece.auth.gr	@auth.gr	@jour.auth.gr	@auth.gr	@eng.auth.gr

ΠΕΡΙΛΗΨΗ

Η παραγωγή μοτίβων τυμπάνων αποτελεί ένα μέρος του ευρύτερου ερευνητικού πεδίου της αλγοριθμικής σύνθεσης το οποίο επικεντρώνεται στο μουσικό στοιχείο του ρυθμού. Η σχετική έρευνα προσφέρει διάφορες μεθοδολογίες, ωστόσο υπάρχει έλλειψη προσεγγίσεων που να στηρίζονται σε ηχητικά δεδομένα για να παράγουν ρυθμικές ακολουθίες τυμπάνων. Η παρούσα εργασία αντιμετωπίζει 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 του προτεινόμενου συστήματος αποτελεί η ανάκτηση του λεγόμενου Σήματος Ισχύος Onset (OSS) και η επεξεργασία του από ένα μοντέλο Seq2Seq. Η αξιολόγηση του συστήματος πραγματοποιήθηκε μέσω της διεξαγωγής μιας διαδικτυακής έρευνας στην οποία συμμετέχοντες κλήθηκαν να ακούσουν και να βαθμολογήσουν ποιοτικά χαρακτηριστικά ορισμένων ηχητικών δειγμάτων. Η στατιστική ανάλυση των αποτελεσμάτων επικύρωσε την αποτελεσματικότητα του συστήματος.

Music Information Retrieval Techniques for Rhythmic Drum Pattern Generation

ABSTRACT

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, however there is a lack of approaches relying on audio data to generate rhythmic drum sequences. The current paper addresses 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 system is the retrieval of the so-called Onset Strength Signal (OSS) and its processing by a Seq2Seq model. The system evaluation was carried out by conducting an online survey in which participants were asked to listen to and rate qualitative characteristics of certain audio samples. Statistical analysis of the results validated the effectiveness of the system.

Εισαγωγή

Η αλγοριθμική σύνθεση μουσικής αποτελεί ένα αντικείμενο με τεράστιο ενδιαφέρον και πολλαπλές εφαρμογές. Η εκθετική ανάπτυξη της υπολογιστικής ισχύος και η υλοποίηση νέων αρχιτεκτονικών νευρωνικών δικτύων έχει δώσει στους υπολογιστές 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. Στη σύγχρονη μουσική το πρωταρχικό ρυθμικό όργανο είναι τα τύμπανα, κοινώς ντραμς, ένα σύνολο από κρουστά όργανα το οποίο αποτελεί τη ραχοκοκαλιά ενός μουσικού συνόλου. 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 αλγοριθμικής σύνθεσης, αυτό της αυτόματης παραγωγής ρυθμικών μοτίβων τυμπάνων.

Το πρόβλημα της παραγωγής μοτίβων τυμπάνων αποτελεί μια διαδικασία κατά την οποία κάποια πληροφορία αξιοποιείται για την παραγωγή μιας ακολουθίας συμβολικών δεδομένων. Η μορφή αυτής της πληροφορίας διαφέρει ανάλογα με τη φύση της εκάστοτε επιθυμητής εφαρμογής και έτσι μπορούν να χρησιμοποιηθούν διανύσματα τυχαίου θορύβου [1]-[2], παράμετροι ελέγχου [1]-[5], ρυθμικές ακολουθίες [4]-[11], συμβολικά δεδομένα [8]-[12] και ηχητικά δεδομένα [12]. Στόχος της παρούσας εργασίας είναι να συνεισφέρει στην περιορισμένη έως τώρα έρευνα για την παραγωγή μοτίβων τυμπάνων με βάση τον ήχο και να παρουσιάσει μια νέα μέθοδο βασισμένη σε τεχνικές Ανάκτησης Μουσικής Πληροφορίας και αξιοποιώντας τα μοντέλα κωδικοποιητή-αποκωδικοποιητή Seq2Seq [13].

1. Σχετική Έρευνα

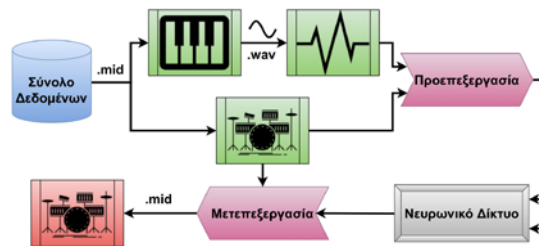
Τη σημαντικότερη συνεισφορά στο πρόβλημα της παραγωγής μοτίβων τυμπάνων με βάση ηχητικά δεδομένα αποτελεί το έργο των Wei et al. [12]. Σε αυτό, οι συγγραφείς αξιοποιούν μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή εμπνευσμένη από τα μοντέλα VAE. Το υλοποιημένο σύστημα παράγει μοτίβα τυμπάνων αξιοποιώντας φασματογραφήματα ενός ηχητικού σήματος το οποίο έχει προηγουμένως διαχωριστεί σε μουσικά μέτρα.

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 Haki και Jordà [14], οι οποίοι χρησιμοποιούν ένα μοντέλο Seq2Seq για να υλοποιήσουν 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 όργανο.

2. Μεθοδολογία και Υλοποίηση

Η γενικότερη διαδικασία που ακολουθείται για την υλοποίηση του προτεινόμενου συστήματος παρουσιάζεται συνοπτικά στο Σχήμα 2.1. 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 καθώς 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.



Σχήμα 2.1 Διαδικασία υλοποίησης του προτεινόμενου συστήματος

2.1 Σύνολο Δεδομένων

Για το σύνολο δεδομένων συλλέχθηκαν από το διαδίκτυο αρχεία με συμβολικά δεδομένα 150 κομματιών από τη δισκογραφία του συγκροτήματος Iron Maiden. Για τη διασφάλιση της ποιότητας του συνόλου δεδομένων κάθε αρχείο επιμελήθηκε ξεχωριστά σε λογισμικό επεξεργασίας MIDI αφαιρώντας μέτρα με κακή πληροφορία, καταργώντας αυτοματισμούς που μετέβαλαν το tempo σε ενδιάμεσα σημεία των μέτρων και πραγματοποιώντας γενικότερες διορθώσεις σε θέσεις νοτών, μουσικές αξίες και κλάσματα χρόνου.

2.2 Προεπεξεργασία

Το στάδιο της προεπεξεργασίας περιλαμβάνει τρία υποστάδια, το πρώτο από τα οποία είναι η ανάκτηση της πληροφορίας της μετρικής δομής του μουσικού κομματιού. Αυτή πραγματοποιείται αξιοποιώντας τη συμβολική πληροφορία που περιέχεται στο αρχείο MIDI και έχει ως αποτέλεσμα τον διαχωρισμό του μουσικού κομματιού σε μέτρα, όπου κάθε μέτρο χαρακτηρίζεται από ένα κλάσμα χρόνου και μια σταθερή τιμή tempo.

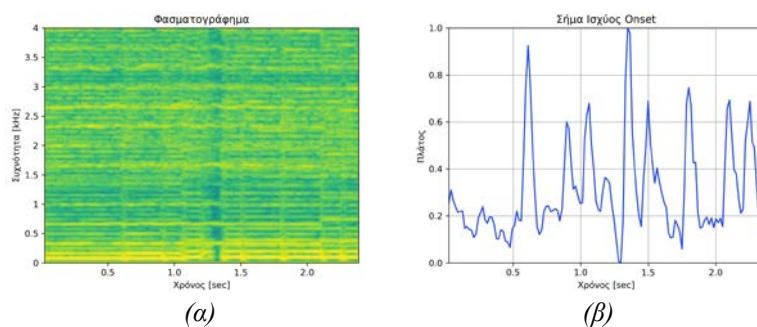
Το δεύτερο υποστάδιο αφορά τη μετατροπή των συμβολικών δεδομένων MIDI των μοτίβων τυμπάνων σε κάποια αναπαράσταση κατάλληλη για χειρισμό από το σύστημα. Αρχικά, επιλέχθηκαν για το σύστημα εννιά κρουστά προς αναπαράσταση τα οποία ήταν η μπότα, το ταμπούρο, τα πιατίνια crash, ride και hi-hat (σε ανοιχτή και κλειστή θέση) και τρία toms (χαμηλό/μεσαίο/υψηλό). Παράλληλα, επιλέχθηκε ως ελάχιστη μουσική αξία για την υποδιαίρεση των μέτρων σε διακριτές θέσεις νοτών το τρίγχο τριακοστού δευτέρου. Έτσι, για κάθε μέτρο προκύπτει ένας αραιός δυαδικός πίνακας διάστασης $K \times 9$ όπου οι τιμές 1 και 0 σε κάποιο κελί του αναπαριστούν την ύπαρξη ή απουσία της αντίστοιχης νότας στην αντίστοιχη χρονική στιγμή. Η μεταβλητή K εξαρτάται από την ελάχιστη επιλεγμένη αξία και το κλάσμα χρόνου του εκάστοτε μέτρου και έτσι για παράδειγμα το μοτίβο τυμπάνων ενός μέτρου $4/4$ θα αναπαρίσταται από έναν αραιό δυαδικό πίνακα διάστασης 48×9 .

Το τρίτο και τελευταίο υποστάδιο αποτελεί η επεξεργασία του ηχητικού σήματος για την εξαγωγή των χαρακτηριστικών που θα τροφοδοτηθούν στο νευρωνικό δίκτυο. Κατ'αρχάς το ηχητικό σήμα υποδειγματοληπείται στα 8kHz έτσι ώστε να αγνοηθούν οι θορυβώδεις υψηλές συχνότητες που δεν περιέχουν χρήσιμη πληροφορία και να βελτιωθεί σημαντικά η ταχύτητα των υπολογισμών. Στη συνέχεια, το ηχητικό σήμα χωρίζεται σε μουσικά μέτρα αξιοποιώντας την πληροφορία της μετρικής δομής που έχει ανακτηθεί νωρίτερα και για κάθε μέτρο ξεχωριστά υπολογίζεται το αντίστοιχο φασματογράφημα με χρήση του Βραχυπρόθεσμου Μετασχηματισμού Φουριέ, όπως φαίνεται στο Σχήμα 2.2 (α). Τα παράθυρα που χρησιμοποιούνται για τον υπολογισμό αυτό είναι τύπου Hanning με επικάλυψη $2/3$ και το μήκος τους αντιστοιχίζεται στη διάρκεια της ελάχιστης αξίας που έχει επιλεγεί. Η διάρκεια αυτή είναι αντιστρόφως ανάλογη της τιμής του tempo και έτσι η χρήση παραθύρων μεταβλητού μήκους εξασφαλίζει σταθερή χρονική αναλυτικότητα σε κάθε μέτρο. Τέλος, αθροίζοντας τις θετικές πρώτες διαφορές ως προς τον άξονα της συχνότητας σε κάθε χρονικό σημείο του φασματογραφήματος λαμβάνεται το λεγόμενο Σήμα Ισχύος Onset (Onset Strength Signal - OSS) [15], το οποίο επιπλέον εξομαλύνεται με χρήση φίλτρου κινούμενου μέσου μήκους 3 σημείων.

Το OSS που προκύπτει από το φασματογράφημα απεικονίζεται στο Σχήμα 2.2 (β) και αποτελεί το πρώτο και κυριότερο χαρακτηριστικό που αξιοποιείται από το νευρωνικό δίκτυο. Παράλληλα, κατασκευάζονται δύο επιπλέον χαρακτηριστικά που περιέχουν πληροφορία για τη δομή του μουσικού μέτρου, ένα σήμα με μορφή τριγωνικής συνάρτησης που περιγράφει την απόσταση κάθε σημείου του OSS από τη θέση του πλησιέστερου beat και ένα σήμα με μορφή γραμμικής συνάρτησης που περιγράφει τη γενικότερη θέση κάθε σημείου του OSS μέσα στο μέτρο.

2.3 Νευρωνικό Δίκτυο

Για το νευρωνικό δίκτυο του προτεινόμενου συστήματος υλοποιήθηκε ένα απλό μοντέλο Seq2Seq. Τόσο ο κωδικοποιητής όσο και ο αποκωδικοποιητής αποτελούνται από ένα αναδρομικό επίπεδο 512 κελιών LSTM με παράμετρο dropout 0.25. Το επίπεδο εξόδου αποτελείται από έναν κόμβο για κάθε ένα από τα εννιά επιμέρους κρουστά όργανα που έχουν επιλεγεί για να συμμετέχουν στα παραγόμενα μοτίβα τυμπάνων. Για συνάρτηση ενεργοποίησης στο επίπεδο εξόδου χρησιμοποιήθηκε η σιγμοειδής (sigmoid), ενώ για συνάρτηση κόστους χρησιμοποιήθηκε η Δυναμική Διασταυρωμένη Εντροπία (Binary Cross-Entropy).



Σχήμα 2.2 (α) Φασματογράφημα σήματος ήχου (β) Σήμα Ισχύος Onset

2.4 Μετεπεξεργασία

Το τελικό στάδιο της μετεπεξεργασίας ξεκινάει με τη στρογγυλοποίηση των πιθανοτήτων που προκύπτουν από την έξοδο του νευρωνικού δικτύου στην κοντινότερη δυαδική τιμή η οποία αναπαριστά την ύπαρξη ή όχι νότας MIDI σε κάποια χρονική στιγμή. Στη συνέχεια, οι δυαδικές αναπαραστάσεις που προκύπτουν φιλτράρονται εισάγοντας λογικούς και φυσικούς περιορισμούς οι οποίοι ενεργοποιούνται στις χρονικές στιγμές όπου υπάρχει παίξιμο πολλών νοτών MIDI και εξασφαλίζουν πως τα παραγόμενα μοτίβα τυμπάνων θα είναι όσο το δυνατόν πιο ανθρώπινα. Η διαδικασία ολοκληρώνεται αξιοποιώντας την πληροφορία για τη μετρική δομή του μουσικού κομματιού που έχει ανακτηθεί προηγουμένως ώστε η δυαδική αναπαράσταση των παραγόμενων μοτίβων τυμπάνων να μετατραπεί στο τελικό αρχείο MIDI που περιέχει όλη τη συμβολική πληροφορία.

3. Αξιολόγηση

Η αξιολόγηση του συστήματος πραγματοποιήθηκε αφού το νευρωνικό δίκτυο είχε εκπαιδευτεί για 225 εποχές. Λόγω της φύσης του προβλήματος, η προσέγγιση έγινε με υποκειμενικά κριτήρια και έτσι διεξήχθη μία διαδικτυακή έρευνα στην οποία 45 συμμετέχοντες κλήθηκαν να ακούσουν και να βαθμολογήσουν ηχητικά δείγματα τα οποία προήλθαν από μουσικά κομμάτια που δεν είχαν χρησιμοποιηθεί για την εκπαίδευση του συστήματος.

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.

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 [16] για επίπεδο σημαντικότητας 5%. Τα ποιοτικά χαρακτηριστικά των δειγμάτων που κλήθηκαν να βαθμολογήσουν οι συμμετέχοντες σε μια γραμμική κλίμακα από το 0 έως το 10 ήταν τα εξής:

- I. *Αισθητική*: πόσο ικανοποιητικό είναι ηχητικά το παίξιμο των τυμπάνων.
- II. *Προσαρμοστικότητα*: πόσο ταιριάζει με τον ρυθμό της μουσικής το παίξιμο των τυμπάνων.
- III. *Δημιουργικότητα*: πόσο ενδιαφέρον, εναλλακτικό ή απρόβλεπτο είναι το παίξιμο των τυμπάνων.
- IV. *Φυσικότητα*: πόσο ανθρώπινο ακούγεται το παίξιμο των τυμπάνων.
- V. *Ομοιότητα*: πόσο μοιάζουν ηχητικά τα δύο δείγματα ενός ζεύγους

Βάσει των δημογραφικών στοιχείων τους 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 (45/45)
- 2) Όσοι είναι εξοικειωμένοι με τη μουσική των Iron Maiden (29/45)
- 3) Όσοι δεν είναι εξοικειωμένοι με τη μουσική των Iron Maiden (16/45)
- 4) Όσοι ξέρουν να παίζουν τύμπανα (12/45)
- 5) Όσοι ξέρουν να παίζουν κάποιο άλλο όργανο ή έχουν κάποια άλλη μουσική ενασχόληση (23/45)
- 6) Όσοι δεν έχουν καμία μουσική εμπειρία (10/45)

Τα συγκεντρωτικά αποτελέσματα των μέσων βαθμολογιών για τα χαρακτηριστικά των ανθρώπινων και των τεχνητών δειγμάτων παρουσιάζονται στον Πίνακα 3.1. Επιπλέον, ζητήθηκε από τους συμμετέχοντες να προσπαθήσουν να αναγνωρίσουν ποιο από τα δύο δείγματα κάθε ζεύγους ήταν αυτό που παρήχθη τεχνητά από το σύστημα. Τα αποτελέσματα αυτής της διαδικασίας παρουσιάζονται στον Πίνακα 3.2.

Πίνακας 3.1 Αποτελέσματα μέσων βαθμολογιών δειγμάτων

	I		II		III		IV		V
	A	T	A	T	A	T	A	T	
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
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
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
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
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
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)
(Η σκίαση υποδεικνύει στατιστικά σημαντική διαφορά)

Πίνακας 3.2 Αποτελέσματα αναγνώρισης τεχνητών δειγμάτων

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

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. Επιπλέον, κατά μέσο όρο τα ποσοστά επιτυχίας και αποτυχίας αναγνώρισης των τεχνητών δειγμάτων βρίσκονται στο ίδιο επίπεδο. Συνεπώς, μπορεί να δικαιολογηθεί η θεώρηση πως τα παραγόμενα μοτίβα τυμπάνων του συστήματος βρίσκονται στο ίδιο επίπεδο με τα αντίστοιχα μοτίβα τυμπάνων που θα έπαιζε ένας άνθρωπος, επικυρώνοντας έτσι την αποτελεσματικότητα του συστήματος.

4. Συμπεράσματα

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.

Η χρήση του OSS για την παραγωγή μοτίβων τυμπάνων αποτελεί καινοτόμο στοιχείο και μπορεί να αποτελέσει σημαντικό θεμέλιο σε παρόμοιες μελλοντικές προσεγγίσεις, ενώ 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.

Κάποιες από τις μελλοντικές επεκτάσεις που επιδέχεται το σύστημα που παρουσιάστηκε είναι ο περαιτέρω πειραματισμός με τις διάφορες παραμέτρους και την αρχιτεκτονική του νευρωνικού δικτύου, η εκπαίδευση αυτού με περισσότερα δεδομένα και ιδιαίτερα με δεδομένα τραγουδιών διαφορετικού μουσικού είδους και τέλος, η υλοποίηση μιας ολοκληρωμένης εφαρμογής παραγωγής μοτίβων τυμπάνων με γραφικό περιβάλλον φιλικό προς τον χρήστη.

5. Βιβλιογραφία

- [1] N. Tokui «Can GAN Originate New Electronic Dance Music Genres? - Generating Novel Rhythm Patterns Using GAN with Genre Ambiguity Loss» (2020).
- [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*, pp 91-92 (2019).
- [3] N. Tokui «Towards Democratizing Music Production with AI - Design of Variational Autoencoder-Based Rhythm Generator as a DAW Plugin » (2020).
- [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*, pp. 212-220 (2020).
- [5] K. Lee, W. Kim, and J. Nam «PocketVAE: A Two-step Model for Groove Generation and Control» (2021).
- [6] P. Hutchings «Talking Drums: Generating Drum Grooves with Neural Networks» in *1st International Workshop on Deep Learning and Music joint with CNN*, pp. 43–47 (2017).
- [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*, pp. 2269–2279 (2019).
- [8] D. Makris, M. Kaliakatsos-Papakostas, I. Karydis, and K. L. Kermanidis «Combining LSTM and Feed Forward Neural Networks for Conditional Rhythm Composition» *Engineering Applications of Neural Networks*, **744**, pp. 570–582 (2017).
- [9] D. Makris, M. Kaliakatsos-Papakostas, and K. L. Kermanidis «DeepDrum: An Adaptive Conditional Neural Network for Generating Drum Rhythms» (2018).
- [10] D. Makris, M. Kaliakatsos-Papakostas, I. Karydis, and K. L. Kermanidis «Conditional Neural Sequence Learners for Generating Drums’ Rhythms» *Neural Computing and Applications*, **31**, pp. 1793–1804 (2019).
- [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*, pp. 179–194 (2022).
- [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*, pp. 847–854 (2019).
- [13] I. Sutskever, O. Vinyals, and Q. V. Le «Sequence to Sequence Learning with Neural Networks» *Advances in Neural Information Processing Systems*, **27** (2014).
- [14] B. Haki and S. Jordà «A Bassline Generation System Based on Sequence-to-Sequence Learning» in *International Conference on New Interfaces for Musical Expression*, pp. 204–209 (2019).
- [15] 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*, **22**, pp. 1765–1776 (2014).
- [16] B. L. Welch «The Significance of the Difference Between Two Means when the Population Variances are Unequal» *Biometrika*, **29**, pp. 350–362 (1938).