

Εκτίμηση της ομοιότητας μεταξύ μουσικών ρυθμικών σχημάτων II. Ψυχοακουστικές και συνδεσιακές προσεγγίσεις

Αθανάσιος Φουλούλης
Μεταπτυχιακός φοιτητής
Τμήμα. Ηλ. Μηχ/κών &
Μηχ/κών Η/Υ
Α.Π.Θ. 54124
Θεσσαλονίκη
afululis@auth.gr

Γεώργιος Παπαδέλης
Επίκουρος Καθηγητής
Τμήμα Μουσικών
Σπουδών
Α.Π.Θ. 54124
Θεσσαλονίκη
papadeli@mus.auth.gr

Κων/νος Παστιάδης
Λέκτορας
Τμήμα Μουσικών
Σπουδών
Α.Π.Θ. 54124
Θεσσαλονίκη
pastiadi@auth.gr

ΠΕΡΙΛΗΨΗ

Στην εργασία αυτή παρουσιάζεται ένας τύπος αυτοοργανωμένου νευρωνικού δικτύου (SARDNET) που μπορεί να εφαρμοστεί στην αναπαράσταση και κατηγοριοποίηση ρυθμικών σχημάτων, και διερευνάται η αποτελεσματικότητα με την οποία αυτό μπορεί να χρησιμοποιηθεί στην ανάλυση ακολουθιών από χρονικά δεδομένα. Η πραγμάτευση αυτή αποτελεί προστάδιο επικείμενης μελέτης που αποσκοπεί στον εντοπισμό κατάλληλων «ποσοτικών δεικτών» για την εκτίμηση των αποστάσεων μεταξύ των «χαρτών» που παράγει το δίκτυο.

Η διερεύνηση της συμπεριφοράς του δικτύου έδειξε ότι αυτός ο τύπος νευρωνικού δικτύου μπορεί να αποδειχθεί αποτελεσματικός στην αναπαράσταση και αναγνώριση της ακρίβειας με την οποία εκτελούνται δοκιμασίες αναπαραγωγής μουσικών ρυθμικών σχημάτων.

Similarity estimation between musical rhythm patterns II. Psychoacoustic and Connectionist approaches

ABSTRACT

A self-organizing neural network for rhythm pattern mapping and classification called SARDNET, is described and its performance in the analysis of real musical timing data is evaluated as a preliminary stage of examining various indices of "distance measures" between the resulting "feature maps".

These preliminary results provide evidence that this type of network architecture may be proven successful on mapping and categorizing timing performance acuity in rhythm copying tasks.

Εισαγωγή

Η εφαρμογή μεθόδων ποσοτικής ανάλυσης για τον υπολογισμό «αποστάσεων» μεταξύ ρυθμικών σχημάτων αποτελεί ένα ισχυρό εργαλείο εκτίμησης της μεταξύ τους ομοιότητας ιδιαίτερα σε εφαρμογές ανάσυρσης πληροφοριών και μουσικού υλικού (τραγουδιών, οργανικής μουσικής κ.λ.π.) από βάσεις δεδομένων μουσικής. Αναπόσπαστο βέβαιο στοιχείο της θεώρησης αυτής είναι τα ερωτηματικά που γεννώνται σχετικά με τον βαθμό της «οικολογικής εγκυρότητας» των μεθόδων ποσοτικής ανάλυσης κατά την εφαρμογή τους στην περιγραφή γνωστικών διεργασιών, τον βαθμό δηλαδή στον οποίο αυτές μπορούν να περιγράψουν με ακρίβεια τον τρόπο με τον οποίο ο άνθρωπος ενεργεί κατά την εκτέλεση αντίστοιχων έργων σε πραγματικές συνθήκες.

Ένα από τα βασικά πλεονεκτήματα που παρουσιάζει η χρήση μοντέλων νευρωνικών δικτύων στην ανάλυση χαρακτηριστικών της ρυθμικής δομής μουσικών αποσπασμάτων σχετίζεται με την ιδιότητά τους να παρέχουν έναν αποδοτικό τρόπο συνοπτικής κωδίκευσης και αναπαράστασης της πληροφορίας που περιγράφει τη δομή αυτή, τόσο σε μακροχρονικό, όσο και σε μικροχρονικό επίπεδο. Από τους διάφορους τύπους αρχιτεκτονικής δικτύων που έχουν εφαρμοστεί, άλλοι είναι περισσότερο αποτελεσματικοί στο να προβλέπουν τη χρονική θέση κάθε επόμενου χτυπήματος σε μια ρυθμική ακολουθία εκμεταλλευόμενοι τη «γνώση» που αποκομίζουν από προηγούμενα χτυπήματα της ακολουθίας, ενώ άλλοι εξειδικεύονται στο να δημιουργούν αναπαραστάσεις ολόκληρης της ακολουθίας [1]. Στη δεύτερη κατηγορία ανήκουν και τα δίκτυα τύπου SARDNET (Sequential Activation Retention and Decay NETwork), τα οποία έχουν δοκιμαστεί με επιτυχία στην αναπαράσταση και κατηγοριοποίηση ακολουθιών από δυαδικούς ή πραγματικούς αριθμούς, καθώς επίσης και ακολουθιών από φωνήματα της αγγλικής [2].

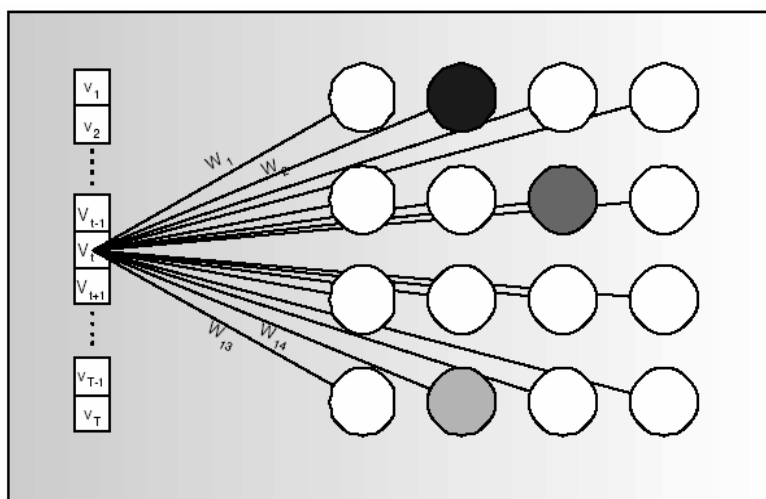
Έχοντας ως πρωταρχικό μας μέλημα την ανάδειξη ενός αποδοτικού μετασχηματισμού της πρωτογενούς περιγραφής σύντομων ρυθμικών σχημάτων ως ακολουθίας των χρονικών διαρκειών των επιμέρους χτυπημάτων τους, παρουσιάζουμε στη μελέτη αυτή τα πρώτα αποτελέσματα από την εφαρμογή δικτύου τύπου SARDNET στην δημιουργία «χαρτών αναπαράστασης» της ρυθμικής δομής. Το ενδιαφέρον μας επικεντρώνεται κυρίως στο ερώτημα κατά πόσο είναι δυνατό οι χάρτες αυτοί να αναπαραστήσουν με επαρκή λεπτομέρεια μικροχρονικές παραλλαγές του ίδιου ρυθμικού σχήματος, έτσι ώστε στη συνέχεια να είναι δυνατή η εφαρμογή ποσοτικών μεθόδων υπολογισμού των αποστάσεων ανάμεσα στις αναπαραστάσεις αυτές. Παράλληλα συζητείται συνοπτικά η προβληματική του συνδυασμού των αποτελεσμάτων εκτίμησης των «αποστάσεων» αυτών με αντίστοιχα εκτίμησης της ομοιότητας μεταξύ μικροπαραλλαγών του ίδιου ρυθμικού σχήματος από μουσικά έμπειρους ακροατές, τα οποία συλλέγονται με την εφαρμογή μεθόδων της Ψυχοακουστικής.

1. Η αρχιτεκτονική του SARDNET (Sequential Activation Retention and Decay NETwork)

Το δίκτυο αυτό ανήκει στην κατηγορία των δικτύων SOMs (Self-Organizing Maps) στα οποία εφαρμόζεται η μη ελεγχόμενη εκπαίδευση (unsupervised

learning). Τέτοια δίκτυα έχουν την ικανότητα να ανιχνεύουν τις σχέσεις μεταξύ των δεδομένων που δέχονται σαν είσοδο και να τις αφομοιώνουν ώστε να προσαρμόζουν την έξοδο τους με βάση αυτά. Τα δίκτυα αυτά μαθαίνουν να αναγνωρίζουν παρόμοια δεδομένα εισόδου με τρόπο ανάλογο με αυτόν που οι διάφορες ομάδες νευρωνικών κυκλωμάτων του ανθρώπινου εγκεφάλου αντιδρούν σε όμοια ερεθίσματα. Έτσι γειτονικοί νευρώνες σε ένα τέτοιο δίκτυο μαθαίνουν να αντιδρούν σε «κοντινά» δεδομένα εισόδου [9].

Είσοδο στο SARDNET αποτελούν ακολουθίες από n-διάστατα διανύσματα: $\mathbf{S} = \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \dots, \mathbf{V}_T$. Για παράδειγμα κάθε διάνυσμα εισόδου (\mathbf{V}_i) μπορεί να αντιστοιχεί σε ένα τμήμα ενός ρυθμικού σχήματος, όπως ένα μεμονωμένο χτύπημα, όπου οι διαστάσεις n του διανύσματος να αφορούν τη διάρκεια, την ένταση κ.λ.π. Έτσι το επίπεδο εισόδου του δικτύου αποτελείται από n κόμβους, έναν για κάθε διάσταση του διανύσματος εισόδου. Για παράδειγμα, το δίκτυο του σχήματος 1.1 αποτελείται από $m_R \times m_C$ κόμβους (νευρώνες) σε καθένα από τους οποίους αντιστοιχεί ένας συντελεστής βάρους \mathbf{W}_{jk} ο οποίος καθορίζει την δυνατότητα ενεργοποίησης του κόμβου σε κάθε διάνυσμα εισόδου. [2,6]



Σχήμα 1.1 Η αρχιτεκτονική του SARDNET. Στο σχήμα απεικονίζεται ένα δίκτυο με $m_R \times m_C = 4 \times 4$ κόμβους

Υπολογίζοντας την ευκλείδεια απόσταση του διανύσματος εισόδου \mathbf{V}_i με το συντελεστή βάρους του κάθε κόμβου \mathbf{W}_{jk} θεωρούμε ως «νικητή» κόμβο αυτόν για τον οποίο η τιμή της απόστασης είναι η ελάχιστη [7]:

$$d(v, w) = \sqrt{\sum_{i=1}^n |v_i - w_i|^2} \quad (1.1)$$

Ο νικητής κόμβος αποκτά τιμή ενεργοποίησης $\eta=1$, ενώ ο συντελεστής βάρους του κόμβου αυτού καθώς και οι συντελεστές μιας γειτονικής περιοχής κόμβων μεταβάλλονται με βάση τη σχέση [7]:

$$\Delta w_{jk} = \alpha (w_{jk,i} - v_i) \quad (1.2)$$

όπου το α καθορίζει τον ρυθμό εκμάθησης (learning rate) με βάση το οποίο θα γίνεται η μεταβολή των συντελεστών βάρους των κόμβων. Συνήθως η γειτονιά του «νικητή» κόμβου είναι εκτεταμένη στα πρώτα στάδια εκπαίδευσης του δικτύου και μικραίνει σταδιακά όσο αυτή εξελίσσεται [2, 8].

Σαν τελευταίο βήμα στην επεξεργασία του διανύσματος εισόδου είναι η εξασθένιση των τιμών ενεργοποίησης όλων των προηγούμενων ενεργοποιημένων κόμβων με βάση τη σχέση:

$$\eta_{jk}(t+1) = d\eta_{jk}, \quad 0 < d < 1 \quad (1.3)$$

Ένα σημαντικό χαρακτηριστικό του SARDNET είναι ότι από τη στιγμή που ένας νευρώνας ενεργοποιηθεί («κερδίσει») δεν έχει πλέον τη δυνατότητα να ενεργοποιηθεί εκ νέου από τα επόμενα διανύσματα της ακολουθίας. Με αυτό τον τρόπο κάθε διαφορετικό διάνυσμα αντιστοιχείται σε ένα διαφορετικό νευρώνα του δικτύου και έτσι κάθε ακολουθία T διανυσμάτων απεικονίζεται με T διαφορετικούς ενεργούς νευρώνες που έχουν διαδοχικές τιμές ενεργοποίησης [2,6].

Καθώς η εκπαίδευση του δικτύου προχωράει, τα διανύσματα εισόδου «κωδικοποιούνται» και αναπαρίστανται τελικά υπό τη μορφή ενός συνόλου συντελεστών βάρους του δικτύου. Όμως επειδή οι συντελεστές αυτοί μεταβάλλονται κατά γειτονικές περιοχές, γειτονικοί νευρώνες τείνουν εν τέλει να αποκτήσουν κοντινά βάρη [5].

2. Η παρούσα υλοποίηση

Στην παρούσα υλοποίηση κάθε ρυθμικό σχήμα αναπαραστάθηκε ως μια ακολουθία απόλυτων χρονικών διαρκειών (σε ms) που αντιστοιχούν στα διαστήματα μεταξύ των επιμέρους διαδοχικών χτυπημάτων του. Έτσι λοιπόν τα διανύσματα εισόδου ήταν μονοδιάστατα και το επίπεδο εισόδου του δικτύου αποτελούνταν μόνο από έναν κόμβο.

Αρχικά, το δίκτυο εκπαιδεύτηκε με τιμές χρονικών διαστημάτων που προέκυψαν από κανονικές κατανομές ενώ σε δεύτερη φάση με δεδομένα από εκτελέσεις σύντομων ρυθμικών σχημάτων από δύο χαρακτηριστικούς πληθυσμούς παιδιών ηλικίας μεταξύ 5 και 8 χρονών: μια ομάδα παιδιών με μουσική κατάρτιση (μαθητές ωδείου) και μια ομάδα παιδιών με μαθησιακές δυσκολίες. Ο αριθμός των νευρώνων του δικτύου ήταν 200 και 300 αντίστοιχα.

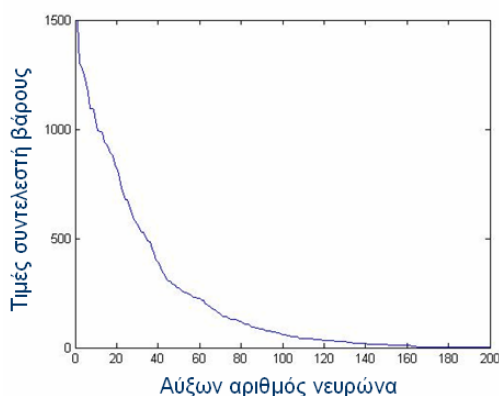
2.1 Εκπαίδευση με δεδομένα από κανονικές κατανομές

Το σετ εκπαίδευσης στην περίπτωση αυτή αποτέλεσαν χρονικά διαστήματα που προέκυψαν από κανονικές κατανομές. Στην περίπτωση που η τιμή της ρυθμικής αγωγής (tempo) στο επίπεδο των τετάρτων είναι 120 bpm, πράγμα που σημαίνει ότι η διάρκεια τους είναι 500ms, οι αντίστοιχες τιμές σε ms για τις υπόλοιπες κατηγορίες ρυθμικών αξιών έχουν ως εξής:

Πίνακας 2.1 Αντιστοιχία ρυθμικών αξιών και διαρκειών σε ms

1/2	1/4	1/8	1/16	1/32	1/64	1/128
1000ms	500ms	250ms	125ms	62,5ms	31,25ms	15,625ms

Παίρνοντας λοιπόν κανονικές κατανομές με μέση τιμή καθεμία από τις παραπάνω τιμές και κατάλληλες τυπικές αποκλίσεις πήραμε 20 τυχαίες τιμές για κάθε μια και χρησιμοποιήσαμε το παραπάνω σύνολο τιμών για την εκπαίδευση του δικτύου. Η γραφική παράσταση των συντελεστών βάρους σε συνάρτηση με τον αύξοντα αριθμό του κόμβου μετά το τέλος της εκπαίδευσης δίνεται στο σχήμα 2.1.



Σχήμα 2.1 Οι τιμές των συντελεστών βάρους κάθε κόμβου για την εκπαίδευση με δεδομένα από κανονικές κατανομές

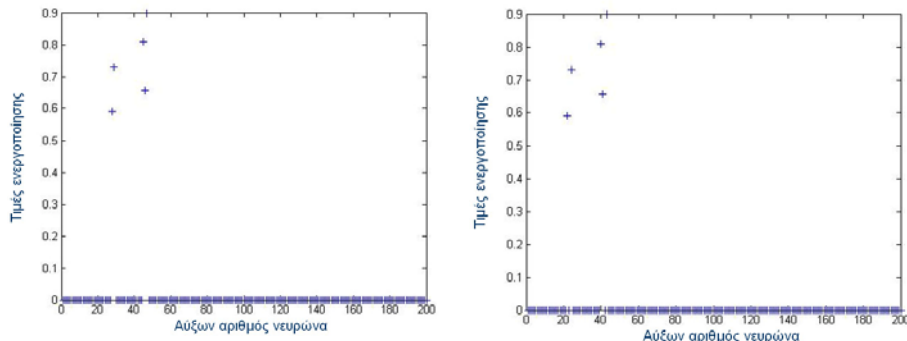
Στον πίνακα 2.2 απεικονίζονται τα δεδομένα εισόδου δηλαδή οι χρονικές αποστάσεις μεταξύ διαδοχικών χτυπημάτων για τρία ρυθμικά σχήματα. Οι αναπαραστάσεις που δίνει το SARDNET για τα δύο πρώτα ρυθμικά σχήματα δίνονται στο σχήμα 2.2 όπου είναι φανερή η ομοιότητα τους. Αν στη συνέχεια

Πίνακας 2.2 Αποστάσεις μεταξύ διαδοχικών χτυπημάτων για τρία ρυθμικά σχήματα

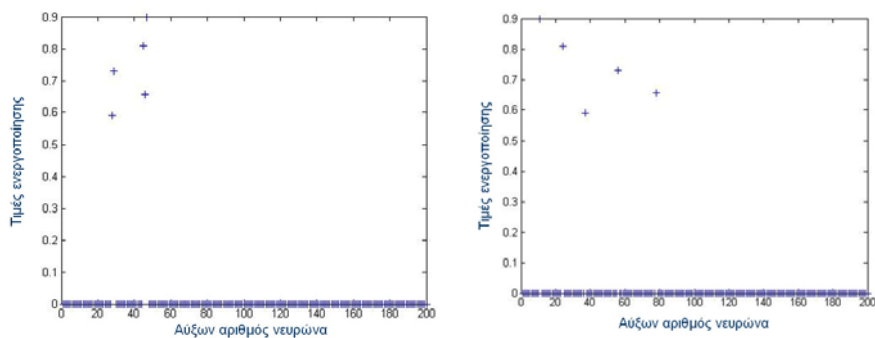
Ρυθμικό σχήμα 1	Ρυθμικό σχήμα 2	Ρυθμικό σχήμα 3
600, 300, 600, 300, 300	734, 367, 692, 387, 330	447,125,240,691,1000

συγκρίνουμε τις απεικονίσεις του SARDNET για τα ρυθμικά σχήματα 1 και 3 (σχήμα 2.3), παρατηρούμε εύκολα ότι πρόκειται για δύο εντελώς διαφορετικά παραδείγματα. Βλέπουμε συνεπώς ότι το SARDNET έχει την ικανότητα να αναγνωρίζει ακολυθίες δεδομένων που παρουσιάζουν χαρακτηριστική ομοιότητα.

Η ακρίβεια των αναπαραστάσεων του SARDNET καθορίζεται άμεσα από τον αριθμό των κόμβων (νευρώνων) του δικτύου. Αν για παράδειγμα ζητάμε ακρίβεια 1ms και έχουμε ένα εύρος τιμών από περίπου 300 έως 1000ms, δηλαδή περίπου 700ms θα χρειαστούμε έναν ανάλογο αριθμό κόμβων.



Σχήμα 2.2 Αναπαραστάσεις που δίνει το SARDNET για το 1^ο και 2^ο ρυθμικό σχήμα του πίνακα 2.2 αντίστοιχα.



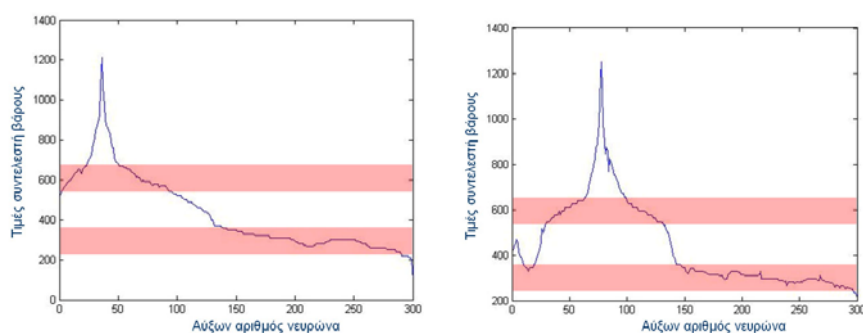
Σχήμα 2.3 Αναπαραστάσεις που δίνει το SARDNET για το 1^ο και 3^ο ρυθμικό σχήμα του πίνακα 2.2 αντίστοιχα.

2.2 Εκπαίδευση με δεδομένα από εκτελέσεις σύντομων μουσικών ρυθμικών σχημάτων από παιδιά ηλικίας 5-8 χρονών

Στην περίπτωση αυτή ως δεδομένα εισόδου χρησιμοποιήθηκαν χρονικά διαστήματα από μουσικές εκτελέσεις που πραγματοποιήθηκαν από παιδιά ηλικίας μεταξύ 5 και 8 χρονών, όταν τους ζητήθηκε να αναπαράγουν σύντομα ρυθμικά σχήματα με τέταρτα και όγδοα διάρκειας 600ms και 300ms αντίστοιχα. Το δίκτυο περιείχε 300 νευρώνες και εκπαιδεύτηκε με δύο διαφορετικά σύνολα δεδομένων: αρχικά με δεδομένα από εκτελέσεις ομάδας παιδιών με μαθησιακές δυσκολίες [3] και στη συνέχεια με δεδομένα από εκτελέσεις παιδιών αντίστοιχης ηλικίας με μουσική κατάρτιση (μαθητές ωδείου) [4].

Στο σχήμα 2.4 δίνονται οι τιμές των συντελεστών βάρους κάθε νευρώνα για τις δύο παραπάνω περιπτώσεις εκπαίδευσης. Όπως μπορεί κανείς να παρατηρήσει στην δεύτερη περίπτωση είναι ιδιαίτερα εμφανής η παρουσία χαρακτηριστικών απότομων μεταβολών της κλίσης της καμπύλης, οι οποίες «οριοθετούν» διακριτά

τμήματά της (τμήματα της καμπύλης εντός των σκιασμένων περιοχών) γύρω από τις τιμές 600 και 300ms που αντιστοιχούν στις διάρκειες των τετάρτων και ογδών αντίστοιχα μιας απόλυτα ισόρρυθμης εκτέλεσης. Από το γεγονός αυτό διαφαίνεται καταρχήν ότι το δίκτυο είναι ευαίσθητο στο να διακρίνει μεταξύ τους δεδομένα εκτέλεσης μουσικών ρυθμικών σχημάτων που αντιστοιχούν σε διαφορετικές κατηγορίες επίδοσης (π.χ. χαμηλή σε αντιδιαστολή με υψηλή ακρίβεια εκτέλεσης).



Σχήμα 2.4 Οι τιμές των συντελεστών βάρους κάθε νευρώνα για δύο διαφορετικές εκπαιδεύσεις του δικτύου

Συγκεκριμένα, λαμβάνοντας υπόψη τις δύο κατηγορίες δεδομένων εκπαίδευσης που χρησιμοποιήθηκαν, η διαμόρφωση των συντελεστών βάρους του δικτύου αναδεικνύει το βαθμό διακριτότητας με τον οποίο μπορεί να αναπαρασταθεί η ικανότητα μουσικά καταρτισμένων παιδιών να εκτελούν με μεγαλύτερη ακρίβεια τις επιμέρους αξίες σύντομων ρυθμικών σχημάτων, σε αντιδιαστολή με άλλους πληθυσμούς παιδιών χαμηλότερης δεξιότητας στο συγκεκριμένο έργο.

3. Συμπεράσματα

Στην παρούσα εργασία παρουσιάστηκε μια αρχική προσέγγιση στο πρόβλημα της εκτίμησης της ομοιότητας μεταξύ μουσικών ρυθμικών σχημάτων μέσα από συνδεδεσμένου τύπου θεωρήσεις, δηλαδή με την εφαρμογή νευρωνικών δικτύων τύπου SARDNET για την αναπαράσταση και κατηγοριοποίηση σύντομων ρυθμικών σχημάτων. Στο βαθμό που μας επιτρέπει η μέχρι τώρα ολοκλήρωση της έρευνάς μας, διαπιστώνουμε ότι το SARDNET έχει την ευαισθησία να αναγνωρίζει ακολουθίες δεδομένων που παρουσιάζουν ομοιότητα μεταξύ τους. Επίσης, φαίνεται ότι είναι σε θέση να κατηγοριοποιήσει ακολουθίες χρονικών διαρκειών που περιγράφουν τη δομή μουσικών ρυθμικών σχημάτων, αλλά και να αναπαραστήσει διαβαθμίσεις της ακρίβειας με την οποία αυτά εκτελούνται. Ως αποτέλεσμα, μελετώντας τα διαγράμματα των τιμών των συντελεστών βάρους των κόμβων μπορούμε να διακρίνουμε τα χαρακτηριστικά διαφορετικών ομάδων δεδομένων.

Από την άλλη πλευρά, οι αναπαραστάσεις των ρυθμικών σχημάτων που δίνει το SARDNET μπορεί καταρχήν να αποτελούν χρήσιμους μετασχηματισμούς των αρχικών δεδομένων, όμως δεν είναι δυνατόν μέσω αυτών να περιγραφεί ποσοτικά ο βαθμός ομοιότητας μεταξύ ρυθμικών σχημάτων. Για το λόγο αυτό ένας από τους άμεσους στόχους της μελέτης αυτής είναι η διερεύνηση δεικτών εκτίμησης της

«απόστασης» μεταξύ αναπαραστάσεων του τύπου αυτού. Κριτήριο για την αξιολόγηση της αποτελεσματικότητας των δεικτών αυτών θα αποτελέσουν αντίστοιχα δεδομένα εκτίμησης της ομοιότητας ρυθμικών σχημάτων από μουσικά έμπειρους ακροατές, τα οποία συλλέγονται από πειραματική έρευνά μας που βρίσκεται σε εξέλιξη κατά την τρέχουσα περίοδο. Επιπρόσθετα σχεδιάζεται επέκταση της έρευνας με την εισαγωγή και δεύτερης διάστασης στα δεδομένα εισόδου, όπως για παράδειγμα της σχετικής δυναμικής των επιμέρους χτυπημάτων.

4. Αναφορές

[1] E.W. Large & J.F. Kolen «Resonance and the Perception of Musical Meter» in *Musical Networks*, edited by N. Griffith & P. Todd, Bradford Books, pp 65-96 (1999).

[2] D. L. James & R. Miikkulainen «SARDNET: A Self-Organizing Feature Map for Sequences» Proceedings of the Advances in Neural Information Processing Systems, **volume 7**, Morgan Kaufmann, pp 577-584 (1995).

[3] K. Chatzikallia, M. Krommyda, P. Vakalos, G. Papadelis & P. Kardaras, «Investigating the relation of literacy skills to elementary musical ones in Greek children with various types of learning disabilities» Conference on Rhythm, Time and Temporal Organization. Institute for Music in Human and Social Development, Edinburgh, Scotland. (2006).

[4] G. Papadelis, I. Papadimitriou, H. Gatzoflias & K. Overy, «Developmental aspects of musical timing skills in Greek children aged 5-8. A pilot study» Conference on Rhythm, Time and Temporal Organization. Institute for Music in Human and Social Development, Edinburgh, Scotland. (2006).

[5] V. B. Rao & H. Rao *C++ Neural Networks and Fuzzy Logic* M&T Books, IDG Books Worldwide, (1995).

[6] J. M. Martins & E. R. Miranda «A Connectionist Architecture for the Evolution of Rhythms» Applications of Evolutionary Computing, EvoWorkshops 2006: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoINTERACTION, EvoMUSART, and EvoSTOC, Budapest, Hungary, pp 696-706 (2006).

[7] J. M. Martins, M. Gimenes, J. Manzolli, & Jr. A. Maia «Similarity Measures for Rhythmic Sequences», Proceedings of the 10th Brazilian Symposium of Musical Computation (SBCM), Belo Horizonte (Brazil), 2005.

[8] Self-Organizing Maps, Neural Network Toolbox, Matlab 7 - R14, (2004).

[9] J. A. Freeman & D. M. Skapura *Neural Networks: Algorithms, Applications, and Programming Techniques* (Computation and Neural Systems Series) Addison-Wesley Publishing Company, New York (1991).