

Βιοακουστική και “Shazam” : Μέθοδοι και υπολογιστικές εφαρμογές αναγνώρισης των ήχων της ορνιθοπανίδας στο φυσικό περιβάλλον

Κουτσογιάννης Γεώργιος
ΜΑ, τμήμα Μουσικών Σπουδών Ε.Κ.Π.Α.
geokouts@music.uoa.gr

Γεωργάκη Αναστασία
Αναπληρώτρια καθηγήτρια, τμήμα Μουσικών
Σπουδών Ε.Κ.Π.Α. - Διευθύντρια του Εργαστηρίου
Μουσικής Ακουστικής και Τεχνολογίας
georgaki@music.uoa.gr

Περίληψη

Τα τελευταία χρόνια, διάφορες εταιρείες σε συνεργασία με ερευνητικά εργαστήρια στο χώρο της Βιοακουστικής έχουν αναπτύξει λογισμικά ακουστικής ανάλυσης των καλεσμάτων των πουλιών, μέσω αυτόματης αναγνώρισης προτύπων, που ταυτοποιούν τη ρυθμική και μελωδική δομή του “τραγουδιού” τους σύμφωνα με τη βιολογική τους κατάταξη. Οι εφαρμογές αυτές ακολουθούν διαφορετική μεθοδολογία ως προς το κομμάτι της ανάλυσης και της αναγνώρισης, ωστόσο παρουσιάζουν κάποια κοινά μειονεκτήματα, όπως είναι η χαμηλή ακρίβεια, η έλλειψη μεθόδων επαγγελματικής ανάλυσης για μεγαλύτερο εύρος πουλιών, οι διάφορες τεχνικές δυσκολίες και η προϋπόθεση ύπαρξης εξειδικευμένου ανθρώπινου δυναμικού για την υλοποίησή τους. Στο πλαίσιο αυτής της ανακοίνωσης θα παρουσιάσουμε κάποιες από τις υπάρχουσες εφαρμογές και σημαντικές έρευνες που έχουν πραγματοποιηθεί στο πεδίο αυτό, ενώ παράλληλα θα σχολιάσουμε τις επικρατέστερες μεθόδους και τεχνικές που χρησιμοποιούν. Εν συνεχεία θα αναφερθούμε στο πρόβλημα της καταγραφής και της συγκριτικής μελέτης μεγαλύτερου εύρους πουλιών, καθώς και στη δημιουργία μίας νέας εφαρμογής αναγνώρισης που μελλοντικά θα περιλαμβάνει την ελληνική ορνιθοπανίδα. Τέλος, θα συζητήσουμε τα οφέλη της εφαρμογής αυτού του τύπου λογισμικών στη σχολική εκπαίδευση και τον ηχητικό και μουσικό εγγραμματισμό στο πλαίσιο της ακουστικής οικολογίας.

Λέξεις Κλειδιά

Ακουστική οικολογία, αναγνώριση πτηνών, ανάλυση ήχου, βιομουσικολογία, μηχανική μάθηση, μουσική τεχνολογία, φωνήματα πτηνών.

Εισαγωγή

Η πρώτη γνωστή καταγραφή του “τραγουδιού” των πουλιών (birdsong) αναφέρεται το 1889 από τον Ludwig Koch ενώ η σύγχρονη μελέτη αντίστοιχων θεμάτων στη σύγχρονη εποχή έχει σαν αφετηρία το έργο του ζωολόγου William Thorpe (Thorpe, 1958). Παράλληλα εξελίσσεται και η Βιοακουστική ως επιστήμη που έχει σαν αφετηρία τις μελέτες του Σλοβένου βιολόγου Ivan Regen, το 1925, σχετικά με τους ήχους των εντόμων [?]. Εκτοτε τίθενται οι βάσεις για την διερεύνηση της Βιοακουστικής σε άλλους ήχους της βιοπανίδας όπως η παραγωγή και η πρόσληψη της ορνιθοπανίδας. Η επιστήμη της Βιολογίας ερευνά και καταγράφει το τραγούδι των πουλιών σαν μέσο για την παρακολούθηση του οικοσυστήματος αναφορικά με τον πληθυσμό τους.

Στα χρόνια που μεσολάβησαν συντελέστηκαν πολλές εξελίξεις σε αυτό τον τομέα μέσω διεπιστημονικών προσεγγίσεων. Ο θεμελιωτής της Ακουστικής Οικολογίας M.Schafer και οι συνεργάτες ξεκίνησαν ένα πολλά υποσχόμενο και ελπιδοφόρο πρόγραμμα για το μέλλον της ανθρωπότητας σε σχέση με τη φύση στο πλαίσιο του “World Soundscape Project” (Schafer, 1977) μέσω της καταγραφής των ήχων που μας περιβάλλουν, συμπεριλαμβανομένου και των καλεσμάτων και κελαηδισμάτων των πουλιών. Μέσω των αναφορών στο χώρο της Βιοακουστικής, της Βιολογίας με έμφαση τη μελέτη των ήχων της πανίδας και της Ακουστικής Οικολογίας σχηματίζεται η νέα επιστήμη της Βιομουσικολογίας - όρος που πρωτοεμφανίζεται από τον Nils L. Wallin (Wallin, 1991). Η Βιομουσικολογία ως κλάδος ασχολείται με την μελέτη της μουσικής από βιολογικής άποψης, ενώ η Ζωομουσικολογία - όρος που χρησιμοποιείται αρχικά από τον Γάλλο συνθέτη Francois-Bernard

Mâche (Mâche, 1983) - είναι τομέας της μουσικολογίας και της ζωολογίας ο οποίος μελετά τις μουσικές πτυχές του παραγόμενου και προσλαμβανόμενου ήχου των ζώων (Georgaki, 2017). Τέλος, η Ορνιθομουσικολογία – η οποία αποτελεί κλάδο της Ζωομουσικολογίας - εστιάζει στα μελωδικο-ρυθμικά μοτίβα των πουλιών, μέσα από την παρατήρηση, την ανάλυση και την καταγραφή της μουσικής τους, με βασικό εργαλείο τη χρήση φασματογραφημάτων και μουσικής σημειογραφίας. Πρωτοπόρος του συγκεκριμένου κλάδου υπήρξε ο συνθέτης και ορνιθολόγος Olivier Messiaen, δάσκαλος του F.B. Mâche.

Τα τελευταία χρόνια τη σκυτάλη για τη μελέτη των καλεσμάτων των πουλιών έχει πάρει από την επιστήμη της Μουσικής Ακουστικής η επιστήμη της Εξελικτικής Νευροβιολογίας όπου ένα από τα κύρια ερευνητικά ενδιαφέροντα της είναι η συγκριτική προσέγγιση της μουσικής συμπεριφοράς που σχετίζεται με τους μηχανισμούς της μάθησης στον ανθρώπινο εγκέφαλο και στον εγκέφαλο των πουλιών. Παράλληλα, η Γνωστική Νευροεπιστήμη ασχολείται με την κατανόηση των νευρικών μηχανισμών που εμπλέκονται στην εκμάθηση του κελαηδίσματος στα πουλιά, με στόχο την κατανόηση της εκμάθησης της γλώσσας. Τέλος, στον τομέα της Πληροφορικής, νέα υπολογιστικά συστήματα στοχεύουν στην ανάλυση, αναγνώριση και επεξεργασία του ήχου των πουλιών ενώ οι τεχνικές που χρησιμοποιούνται είναι παρόμοιες με αυτές που χρησιμοποιούνται στην Αυτόματη Αναγνώριση Ομιλίας¹ στις Τηλεπικοινωνίες.

1. Ακουστικές προσεγγίσεις στην ανάλυση του καλέσματος των πουλιών

Ως αναγνώριση ενός ηχητικού σήματος, αναφέρεται ο προσδιορισμός του είδους της ηχητικής πηγής, της συχνότητας, καθώς και ο χρονικός προσδιορισμός του κάθε επιμέρους ήχου που το απαρτίζουν. Η ανάλυση ηχητικών σημάτων είναι μία διαδικασία με σκοπό την εξαγωγή γνώσης αναφορικά με το περιεχόμενο και την φύση των σημάτων αυτών. Οι τεχνικές ανάλυσης εφαρμόζονται στα καλέσματα των πουλιών προκειμένου να εξαχθούν τα ηχητικά χαρακτηριστικά που περιγράφουν το ηχητικό σήμα τους, με σκοπό την κατάτμηση (segmentation), την ταξινόμηση (classification), την ανάκτηση (retrieval) κ.ά.

1.1. Μουσικές δομές των καλεσμάτων των πουλιών - τρόποι επικοινωνίας

Η φωνητική επικοινωνία στα ωδικά πουλιά διακρίνεται σε δύο κατηγορίες: τα καλέσματα (calls) και τα τραγούδια (songs). Τα καλέσματα είναι απλές, σύντομες φωνητικές συνθέσεις και διακρίνονται με βάση τη λειτουργικότητά τους σε καλέσματα προειδοποίησης, καλέσματα τροφής κ.ά. Τα τραγούδια είναι πολύπλοκες φωνητικές συνθέσεις μεγαλύτερης διάρκειας, με σαφείς μουσικές δομές και μεγαλύτερο τονικό και ρυθμικό εύρος και παράγονται σχεδόν αποκλειστικά από αρσενικά πουλιά. Συνήθως ξεκινούν με εισαγωγικές νότες τις οποίες διαδέχονται ένα ή δύο μοτίβα², τα οποία είναι επαναλαμβανόμενες νότες ή συλλαβές³.

1.2. Αναγνώριση του καλέσματος των πουλιών - συνήθη προβλήματα

Στην αναγνώριση του birdsong, στον προσδιορισμό δηλαδή του είδους του πουλιού βάσει του ηχογραφημένου καλέσματος του, προκύπτουν διάφορες τεχνικές δυσκολίες. Η διαδικασία συλλογής δεδομένων από τα πουλιά είναι απαιτητική και χρονοβόρα, αφού απαραίτητη προϋπόθεση είναι η ύπαρξη ειδικού εξοπλισμού εντοπισμού και καταγραφής ήχου, καθώς και εξειδικευμένου ανθρώπινου δυναμικού. Αέρας, βροχή, ήχοι άλλων πουλιών και ζώων, θόρυβοι που προκαλούνται από την ανθρώπινη παρουσία, πουλιά που έχουν την δυνατότητα να μιμούνται ήχους άλλων πουλιών είναι συνήθη προβλήματα. Υπάρχουν επίσης διάφοροι παράγοντες που μπορούν να επηρεάσουν το κελάηδισμα, όπως η ηλικία (Williams, 2004), το φύλο του πουλιού, ή ο τόπος του πουλιού – αφού σε αστικά περιβάλλοντα τα πουλιά κελαηδούν σε υψηλότερες συχνότητες λόγω του θορύβου (Luther & Baptista, 2009). Σε παγκόσμια κλίμακα παρουσιάζονται διαφοροποιήσεις στο κελάηδισμα, ακόμα και σε ίδια είδη ή υποείδη. Πρόκειται για το ζήτημα των “τοπικών διαλέκτων”, όπου ίδια πουλιά παράγουν διαφορετικούς ήχους ανάλογα με τον τόπο τους.

2. Αναγνώριση Προτύπων

¹ Βλ. Automatic Speech Recognition (ASR).

² Οι εισαγωγικές αυτές νότες παραπέμπουν στο γνωστό απήχημα της Βυζαντινής Μουσικής όπου μια μονοσύλλαβη φράση “νε” ψάλλεται πριν την έναρξη του μέλους στον φθόγγο της βάσης του ήχου του εκάστοτε μέλους, τον οποίο και προετοιμάζει.

³ Ως “συλλαβή” ορίζεται ένας αδιάλειπτος ήχος, ο οποίος εμπεριέχει ένα ή περισσότερα ηχητικά ίχνη χρόνου-συχνότητας τα οποία αποκαλούνται “νότες” (Berwick et al., 2011).

Ένα μέρος αυτών των προβλημάτων προσπαθεί να επιλύσει η Αυτόματη Αναγνώριση Birdsong, η χρήση υπολογιστικών συστημάτων δηλαδή, τα οποία μπορούν να παρέχουν γρήγορα, ακριβή αποτελέσματα στην ταυτοποίηση πουλιών. Οι έρευνες σε αυτό το πεδίο σχετίζονται άμεσα με την Αυτόματη Αναγνώριση Ομιλίας αφού η ανθρώπινη ομιλία και το τραγούδι των πουλιών έχουν παρεμφερή, ιεραρχική δομή⁴. Η διαδικασία που ακολουθούν τα συστήματα αυτά ονομάζεται “Αναγνώριση Προτύπων” (Pattern Recognition) και περιλαμβάνει συγκεκριμένα στάδια. Αρχικά, τα ακατέργαστα δεδομένα ψηφιοποιούνται και προεπεξεργάζονται για να απομακρυνθεί ο ανεπιθύμητος θόρυβος. Ιδανική συχνότητα δειγματοληψίας είναι τα 44.100 Hz ή τα 48.000 Hz. Εάν τα αρχεία ήχου αποτελούνται από πολλά κανάλια, μετατρέπονται σε ένα μονό κανάλι. Στη συνέχεια, αφού τα δεδομένα χωριστούν σε πλαίσια (frames), γίνεται η εξαγωγή των ακουστικών χαρακτηριστικών τους. Τα δεδομένα χωρίζονται σε δεδομένα εκπαίδευσης (training data) και δεδομένα δοκιμής (testing data), και ο ταξινομητής εκπαιδεύεται στα δεδομένα εκπαίδευσης. Στο επόμενο στάδιο τα δεδομένα δοκιμής εισάγονται στον ταξινομητή για να προβεί σε αναγνώριση του είδους του πουλιού. Τέλος, όλα τα αποτελέσματα που θα προκύψουν από την αναγνώριση των δεδομένων δοκιμής συνυπολογίζονται για να εκτιμηθεί το ποσοστό ακρίβειας του ταξινομητή. Είναι επίσης εφικτό ο ταξινομητής να επανεκπαιδευτεί με νέα δεδομένα προκειμένου να βελτιωθεί αυτό το ποσοστό.

2.1. Προεπεξεργασία: αφαίρεση θορύβου και κατάτμηση

Ένα πολύ σύνθηρες πρόβλημα, που υποστηρίζεται από την πλειοψηφία των ερευνητών και επηρεάζει σημαντικά την ακρίβεια της αναγνώρισης, είναι η ύπαρξη θορύβου στις ηχογραφήσεις των πουλιών. Για να αντιμετωπιστεί, συνήθως χρησιμοποιούνται διαφόρων ειδών φίλτρα (high-pass, low-pass, band-pass κ.ά.), αφού πρώτα έχει ληφθεί το προφίλ θορύβου της καταγραφής. Αυτή η πρακτική έχει περιορισμούς και στην πράξη προκύπτουν δυσκολίες στην εφαρμογή της. Επιπρόσθετα, αλληλένδετα ζητήματα είναι αυτά του θορύβου και της κατάτμησης⁵: μια ηχογράφιση μπορεί να περιέχει εκτός από το κάλεσμα του πουλιού που μας ενδιαφέρει, και περιττή πληροφορία όπως μεγάλες παύσεις και κενά, με θόρυβο ή σιωπή. Σημαντικά εργαλεία, βοηθητικά ως προς τη διάκριση των συλλαβών, είναι η κυματομορφή και το φασματογράφημα. Μάλιστα οι de Oliveira et al. (2015) προτείνουν μία μέθοδο ανίχνευσης της ακουστικής δραστηριότητας των πουλιών, με βάση το μορφολογικό φιλτράρισμα του φασματογραφήματος, όταν αυτό επεξεργάζεται ως εικόνα. Ανάλογες τεχνικές έχουν χρησιμοποιήσει και άλλοι ερευνητές (π.χ. Potamitis 2014).

2.1. Τεχνικές Ανάλυσης – Εξαγωγή Χαρακτηριστικών (Feature Extraction)

Τα προς επεξεργασία και ανάλυση ηχητικά σήματα είναι στη συντριπτική τους πλειοψηφία σε ψηφιακή μορφή, συνεπώς, στο στάδιο αυτό χρησιμοποιούνται μέθοδοι από τη ψηφιακή επεξεργασία σήματος. Συγκεκριμένα χρησιμοποιείται η βραχυπρόθεσμη επεξεργασία (short-term processing), η οποία βασίζεται στην παραθύρωση (windowing), δηλαδή την τμηματική εφαρμογή παραθύρων τα οποία κάθε φορά μετακινούνται στον χρόνο και έτσι απομονώνουν ένα ξεχωριστό πλαίσιο (frame) του σήματος. Η επιτυχημένη εξαγωγή των κατάλληλων ακουστικών χαρακτηριστικών είναι ο σημαντικότερος παράγοντας στην διαδικασία αναγνώρισης μοτίβων και η ακρίβεια του ταξινομητή που θα εκπαιδευτεί, εν πολλοίς, εξαρτάται από αυτήν. Στην Αυτόματη Αναγνώριση του birdsong έχει διαχρονικά χρησιμοποιηθεί μία μεγάλη ποικιλία τέτοιων χαρακτηριστικών, απλών και σύνθετων, που αφορούν αφορούν είτε στην ένταση (ενέργεια) ολόκληρης της καταγραφής ή κάποιων παραθυροποιημένων τμημάτων της, είτε στη συχνότητα (spectrum, cepstrum).

Η Φασματική Ανάλυση είναι η βασικότερη μέθοδος ανάλυσης του ηχητικού σήματος. Το συχνοτικό φάσμα είναι η μέτρηση της ενέργειας του ηχητικού σήματος σαν μια συνάρτηση της συχνότητας. Πολύ διαδεδομένη είναι επίσης η φασματογραφική (Spectrogram) ανάλυση, η οποία απεικονίζει μία φασματική ανάλυση του αρχείου ήχου στον άξονα χρόνου-συχνότητας, όπου το πλάτος των αρμονικών καθορίζεται από την έντονη γραμμοσκίαση τους. Σε αυτό το σημείο θα πρέπει να αναφερθεί και η ανάλυση Fourier με τη χρήση του αλγορίθμου Fast Fourier Transform (FFT), η οποία μας παρέχει τη δυνατότητα να μεταβούμε από το δυναμικό πεδίο ενός ήχου της κυματομορφής του στο φασματικό πεδίο. Ο αλγόριθμος FFT παράγει σε σύντομο χρονικό διάστημα ανάλυση Fourier και χρησιμοποιείται σε ειδικά λογισμικά ακουστικής ανάλυσης και επεξεργασίας ακουστικών ψηφιακών σημάτων. Σημαντική μέθοδος ανάλυσης αποτελεί η Γραμμική Προγνωστική

⁴ Χρησιμοποιούν δηλαδή συλλαβές για να σχηματίσουν λέξεις, και λέξεις για να σχηματίσουν προτάσεις.

⁵ Με τον όρο “κατάτμηση” εννοούμε τη διαίρεση ή χωρισμό του ήχου σε τμήματα, επικεντρώνοντας στο είδος της πληροφορίας που μας ενδιαφέρει να ανακτήσουμε. Παραδείγματος χάρη, στο birdsong, μία πρακτική που ακολουθείται είναι ο χωρισμός σε “συλλαβές”. Είναι μία διαδικασία που σε μεγάλο βαθμό γίνεται χειροκίνητα, αφού συνήθως η αυτοματοποιημένη κατάτμηση αποφεύγεται.

Κωδικοποίηση (Linear Predictive Coding, LPC). Πρόκειται για ένα εργαλείο που χρησιμοποιείται κυρίως στην επεξεργασία ηχητικού σήματος και στην επεξεργασία ομιλίας, για την απεικόνιση της φασματικής περιβάλλουσας ενός ψηφιακού σήματος ομιλίας σε συμπιεσμένη μορφή, χρησιμοποιώντας τις πληροφορίες ενός γραμμικού προγνωστικού μοντέλου (Deng & O'Shaughnessy, 2003). Τα χαρακτηριστικά που εξάγονται μέσω αυτής της διαδικασίας, ονομάζονται Linear Prediction Cepstrum Coefficients (LPCC) και βρίσκουν ιδιαίτερη εφαρμογή τόσο στην Αυτόματη Αναγνώριση Ομιλίας όσο και στην Αυτόματη Αναγνώριση του Birdsong (π.χ. Lee et al., 2006). Τέλος, ευρύτερα διαδεδομένη είναι η ανάλυση cepstrum και η εξαγωγή των συντελεστών Mel Frequency Cepstral Coefficients, οι οποίοι προσφέρουν μια περιγραφή της φασματικής μορφής του ήχου. Τα MFCC είναι τα πλέον πολυχρησιμοποιημένα χαρακτηριστικά στην αναγνώριση ανθρώπινης ομιλίας (Makhoul & Schwartz, 1995)· τις τελευταίες δεκαετίες εφαρμόζονται κατά κόρον στην αναγνώριση των φωνημάτων των πουλιών (Stattner et al., 2013), αλλά και άλλων ζώων (Clemins & Johnson, 2003). Άλλοι τύποι ανάλυσης είναι ανάλυση wavelets⁶, Short-Time Fourier Transform⁷ (STFT) κ.ά.

2.2. Τεχνικές Ταξινόμησης

Απαραίτητη προϋπόθεση για να γίνει η αναγνώριση είναι η ύπαρξη ενός ταξινομητή (classifier), τον οποίο θα έχουμε εκπαιδεύσει στο να αναγνωρίζει αντίστοιχα πρότυπα. Η “μηχανική μάθηση”, υποπεδίο της επιστήμης των υπολογιστών, διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα (Kohavi & Provost, 1998) και να κάνουν προβλέψεις με βάση αυτά. Μία συγκεκριμένη προσέγγιση της μηχανικής μάθησης που βρίσκει εφαρμογή στην αναγνώριση ομιλίας είναι η Βαθιά Μάθηση (Deep Learning), η οποία δίνει τη δυνατότητα στους υπολογιστές να μαθαίνουν από σύνθετα δεδομένα. Παρότι το πεδίο αυτό παρέχει πολλές επιλογές και δυνατότητες, είναι ακόμη δύσκολο να γίνει μία ασφαλής, ποιοτική σύγκριση μεταξύ των αλγορίθμων ταξινόμησης, εξαιτίας των διαφορετικών μεθόδων που χρησιμοποιούν αλλά και της διαφορετικής ποιότητας δεδομένων που εισάγονται κάθε φορά.

Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks, ANN) είναι υπολογιστικά συστήματα που εκτελούν ορισμένες από τις χαρακτηριστικές λειτουργίες των βιολογικών νευρωνικών δικτύων. Χρησιμοποιούνται είτε για τη μοντελοποίηση σύνθετων σχέσεων μεταξύ δεδομένων εισόδου και εξόδου, είτε για την ανακάλυψη προτύπων στα δεδομένα, ή για τον εντοπισμό στατιστικής δομής σε μία άγνωστη κοινή κατανομή πιθανότητας μεταξύ των παρατηρούμενων μεταβλητών. Σε μία πρώιμη προσπάθεια χρησιμοποίησης ΤΝΔ, ο ταξινομητής αναγνώρισε σωστά περίπου το 80 - 85% των δειγμάτων (McIlraith & Card, 1995), ενώ στην πρώτη μεγάλη έρευνα σχετικά με την χρήση ΤΝΔ για μεγάλο σύνολο δεδομένων επιτεύχθηκε συνολική ακρίβεια 86,8% (Cai et al., 2007). Στα ΤΝΔ η μάθηση μπορεί να είναι είτε επιτηρούμενη είτε μη επιτηρούμενη⁸. Ένας Αυτορρυθμιζόμενος Χάρτης (Self-Organizing Map, SOM), είναι ένας τύπος ΤΝΔ που εκπαιδεύεται χρησιμοποιώντας μη επιτηρούμενη μάθηση. Σε μία έρευνα (Stowell & Plumbley, 2014), η μη επιτηρούμενη μάθηση προτείνεται ως επιτυχημένος τρόπος εκπαίδευσης ως προς την ακρίβεια του ταξινομητή, ενώ σε άλλες έρευνες που έχουν διεξαχθεί αναφορικά με την αξιοπιστία της μη επιτηρούμενης μάθησης (Selin et al., 2007) έχουν βρεθεί αντίθετα αποτελέσματα. Μία διαφορετική μέθοδος που ανήκει στην κατηγορία της μηχανικής μάθησης είναι οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines, SVM), οι οποίες αποτελούν ένα σύνολο μεθόδων επιτηρούμενης μάθησης που χρησιμοποιούνται για την ταξινόμηση και την παλινδρόμηση. Εδώ δίνεται ένα σύνολο παραδειγμάτων εκπαίδευσης και κάθε φορά δηλώνεται σε ποια από τις δύο κατηγορίες ανήκει το παράδειγμα. Η ΜΔΥ κατασκευάζει ένα μοντέλο που προβλέπει εάν το νέο παράδειγμα εμπίπτει στην μία κατηγορία ή την άλλη. Έχει αποδειχτεί ότι οι ΜΔΥ λειτουργούν πολύ καλά με τους συντελεστές MFCC (Fagerlund, 2007). Άλλη μέθοδος είναι τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models, HMM), ένα στατιστικό εργαλείο με ευρεία εφαρμογή στην επεξεργασία σήματος και ιδιαίτερα στην επεξεργασία ομιλίας. Στην αναγνώριση του birdsong παρέχουν στατιστικές προβλέψεις ως προς την διαδοχή των συλλαβών του καλέσματος (Kogan & Margoliash, 1998). Οι Jančonič και Kőküer (2016) δοκίμασαν ταξινομητές KMM για αναγνώριση πολλαπλών ειδών, κάτι που ακόμη θεωρείται πρόκληση για τους περισσότερους ερευνητές.

Πέραν αυτών των διαδεδομένων τεχνικών, σε μικρότερο βαθμό έχουν δοκιμαστεί και άλλες μέθοδοι. Επιγραμματικά, κάποιες από αυτές είναι: Τα Μεικτά Γκαουσιανά Μοντέλα (Gaussian Mixture Models, GMM),

⁶ Η ανάλυση με μετασχηματισμό κυματιδίων χρησιμοποιεί παράθυρα μεταβλητού μεγέθους και απεικόνιση χρόνου – κλίμακας.

⁷ Κατά το μετασχηματισμό Short-Time Fourier το σήμα αναλύεται σε ένα χρονικό παράθυρο κάθε φορά και η ακρίβεια της πληροφορίας εξαρτάται από το μέγεθος αυτού του παραθύρου. Χρησιμοποιεί απεικόνιση χρόνου – συχνότητας.

⁸ Στην επιτηρούμενη μάθηση το μοντέλο προσπαθεί να κατατάξει το νέο παράδειγμα που θα του εισαχθεί σε συγκεκριμένες, προκαθορισμένες κλάσεις, βάσει των δεδομένων στα οποία έχει εκπαιδευτεί. Στην μη επιτηρούμενη μάθηση το μοντέλο προσπαθεί να βρει κάποια πιθανή δομή που μπορεί να υπάρχει σε μη χαρακτηρισμένα δεδομένα, δηλαδή δεν υπάρχουν προκαθορισμένες κλάσεις.

η μέθοδος K-Κοντινότερων Γειτόνων (K-Nearest Neighbors, KNN), διάφοροι τύποι ΤΝΔ όπως Πολυεπίπεδοι Αισθητήρες (Multilayer Perceptron, MLP), Δένδρα Αποφάσεων (Decision Trees) και Τυχαία Δάση (Random Forests). Τέλος, υπάρχουν μέθοδοι αναγνώρισης που συνδυάζουν τεχνικές ταξινόμησης για να δημιουργήσουν υβριδικά συστήματα. Αυτές οι υβριδικές τεχνικές παρουσιάζουν καλύτερα αποτελέσματα και συνολικά υψηλότερη ακρίβεια σε σύγκριση με τις συμβατικές προσεγγίσεις.

3. Τα λογισμικά αναγνώρισης ηχητικών προτύπων της ορνιθοπανίδας

Ιδιαίτερο ενδιαφέρον παρουσιάζουν νέες τεχνικές ανάλυσης για την εξαγωγή χαρακτηριστικών, διαφορετικοί αλγόριθμοι για την ταξινόμηση, καινούργιοι συνδυασμοί τεχνικών για βελτιωμένα αποτελέσματα, που εξελίσσονται διαρκώς τα τελευταία χρόνια. Βάσει αυτών των ερευνών τα τελευταία χρόνια ορισμένες εταιρείες έχουν αναπτύξει λογισμικά ακουστικής ανάλυσης. Κάποια από αυτά είναι γενικότερης χρήσης, ενώ κάποια άλλα επικεντρώνονται αποκλειστικά στην ακουστική ανάλυση ζώων ή συγκεκριμένα στην αναγνώριση του birdsong. Παρακάτω παρουσιάζουμε συνοπτικά τα δημοφιλέστερα από αυτά τα λογισμικά:

- Το ARBIMON είναι ένα δίκτυο για την αποθήκευση, την κοινή χρήση και την ανάλυση ακουστικής πληροφορίας (Aide et al. 2013), ενώ παράλληλα για την αναγνώριση χρησιμοποιεί KMM.
- Το Avisoft-SASLab Pro δημιουργήθηκε από την Avisoft Bioacoustics το 1990, και είναι ένα εμπορικό λογισμικό γενικής χρήσης για την ανάλυση ήχου (Sprecht, 1993) το οποίο μπορεί να προβαίνει σε μέτρηση παραμέτρων του ήχου.
- Το Raven Pro είναι μία εφαρμογή η οποία αναπτύχθηκε από το Cornell Lab of Ornithology (Charif et al., 2010) και παρέχει εργαλεία για την ανάλυση και την απεικόνιση του ήχου.
- Το Song Scope της Wildlife Acoustics ενώ αρχικά υπήρξε εμπορικό προϊόν, πλέον διατίθεται δωρεάν στην ιστοσελίδα της εταιρείας. Χρησιμοποιεί φίλτρα για την μείωση του θορύβου, αλγόριθμο ανίχνευσης για τον διαχωρισμό των διάφορων καλεσμάτων σε κάθε ηχογράφιση, χαρακτηριστικά MFCC για την ανάλυση (Agranat, 2009) και KMM για την κατηγοριοποίηση.
- Το Sound Analysis Pro είναι ένα δωρεάν λογισμικό το οποίο δεν ενδείκνυται για καταγραφές πεδίου, αλλά προτείνεται ως εφαρμογή για να διερευνηθεί η μάθηση και η μίμηση του τραγουδιού στα πουλιά (Tchernichovski, 2012). Η αναγνώριση στηρίζεται σε εκτίμηση της Ευκλείδειας απόστασης.
- Το SoundID είναι ένα εμπορικό σύστημα αναγνώρισης ήχου το οποίο βρίσκει εφαρμογή στο πεδίο της βιοακουστικής, το οποίο εξάγει χαρακτηριστικά LPCC και χρησιμοποιεί σύγκριση γεωμετρικής απόστασης για την αναγνώριση (Jinpaï et al., 2012).

Η ραγδαία εξέλιξη και εξάπλωση των νέων τεχνολογιών, αλλά και η αύξηση του ενδιαφέροντος για ζητήματα ακουστικής οικολογίας, έχει οδηγήσει τα τελευταία χρόνια στην δημιουργία τέτοιου είδους λογισμικών για έξυπνα κινητά (smartphones). Αυτές οι εφαρμογές είναι πιο απλές στην χρήση, αλλά δεν παρέχουν εξειδικευμένες λειτουργίες ακουστικής ανάλυσης. Ωστόσο η μικρή ποικιλία ειδών που έχει αποθηκευτεί, οι δυσκολίες που προκύπτουν από την ύπαρξη ανεπιθύμητου θορύβου και τα χαμηλά ποσοστά ακρίβειας, είναι στοιχεία που δυσχεραίνουν την αναγνώριση προτύπων σε διαφορετικά μήκη και πλάτη της γής όπου τα ίδια είδη πουλιών αναπτύσσουν διαφορετικά γλωσσικά ιδιώματα.

Τέλος, τα πιο διαδεδομένα από αυτά τα λογισμικά είναι τρία: το Bird Song Id, το Song Sleuth και το ChirpOMatic.

4. Η εφαρμογή birdMelos

Στο πλαίσιο της διπλωματικής μου εργασίας για το μεταπτυχιακό πρόγραμμα Μουσικολογία με έμφαση της Μουσική Τεχνολογία, ΤΜΣ, ΕΚΠΑ (2017) σε συνεργασία με την Καθηγήτρια μου κα Αναστασία Γεωργάκη, σεδιάσαμε μία νέα εφαρμογή αναγνώρισης ρυθμικών και μελωδικών προτύπων των πουλιών (Κουτσογιάννης, 2019). Ο απώτερος στόχος αυτής της εφαρμογής είναι έχει εκπαιδευτικές εφαρμογές για την αναγνώριση και ταυτοποίηση των πουλιών της ελληνικής ορνιθοπανίδας. Για τις ανάγκες αυτής της εφαρμογής χρησιμοποιήθηκαν δεδομένα από το διαδίκτυο προκειμένου αφενός να δοκιμαστεί η απόδοση των διάφορων αλγορίθμων ταξινόμησης ως προς την ακρίβεια αναγνώρισης, αφετέρου να εξεταστούν τα διάφορα χαρακτηριστικά του ήχου ως προς την καταλληλότητά τους. Η χρησιμοποίηση του συνόλου αυτών των χαρακτηριστικών αποτελεί καινοτομία της εφαρμογής.

Επιλέχθηκαν τρία διαφορετικά πουλιά, τα οποία υπάρχουν και στον ελλαδικό χώρο, και των οποίων τα καλέσματα έχουν σαφείς και διακριτές δομές: ο γκιώνης (*otus scops*), ο κούκος (*cuculus canorus*) και η δεκαοχτούρα (*streptopelia decaocto*).

Ο κώδικας για την εξαγωγή χαρακτηριστικών και την εκπαίδευση του ταξινομητή γράφηκε στην πλατφόρμα του MATLAB, με χρήση της βιβλιοθήκης MIRtoolbox 1.7.1 (Lartillot & Toivianinen, 2007) και της εφαρμογής Classification Learner App. Τέλος, δημιουργήθηκε ένα εύχρηστο GUI interface για το χρήστη μέσω της εφαρμογής MATLAB App Designer. Το birdMelos είναι σε θέση να εξάγει χαρακτηριστικά από κάθε νέο δείγμα – stereo ή mono - και να προβαίνει σε αναγνώριση του είδους, ενώ παράλληλα διαθέτει δικά του φίλτρα αφαίρεσης θορύβου προς βελτίωση της ποιότητας της ακρίβειας του.

4.1.1. Προεπεξεργασία

Μέσω της διαδικτυακής πλατφόρμας www.xeno-canto.org ανακτήσαμε συνολικά 213 αρχεία ήχου με καλέσματα και κελαιδίσματα πουλιών. Το κριτήριο επιλογής αυτών των δεδομένων ήταν αφενός η καλύτερη δυνατή ποιότητα του δείγματος και αφετέρου – για πρακτικούς λόγους - η συνολική διάρκειά του να μην ξεπερνάει τα 40 sec. Από τα 71 δείγματα που αντιστοιχούσαν σε κάθε πουλί, τα 50 χρησιμοποιήθηκαν ως δεδομένα εκπαίδευσης του συστήματος και τα 21 χρησιμοποιήθηκαν ως δεδομένα δοκιμής. Η προεπεξεργασία των δεδομένων εκπαίδευσης έγινε σε διάφορα στάδια με χρήση του λογισμικού Audacity. Αρχικά τα αρχεία μετατράπηκαν από .mp3 σε .wav, και στη συνέχεια τα κανάλια από stereo μετατράπηκαν σε mono. Αφού το συχνοτικό εύρος κανενός από τα τρία πουλιά δεν ξεπερνούσε τα 2000 Hz, εφαρμόστηκε ένα low-pass φίλτρο προκειμένου να αφαιρεθούν συχνότητες άλλων πουλιών που ξεπερνούσαν αυτή τη συχνότητα. Χρησιμοποιήθηκε ένα ακόμα φίλτρο για τις χαμηλές συχνότητες, κάτω από τα 400 Hz, προκειμένου να αφαιρεθεί ο ανεπιθύμητος θόρυβος. Τελικώς αφαιρέθηκαν “κενά” σημεία στην αρχή και στο τέλος των δειγμάτων.

4.1.2. Εξαγωγή χαρακτηριστικών και εκπαίδευση ταξινομητή

Στην εφαρμογή birdMelos τα εξαγόμενα χαρακτηριστικά επιλέχθηκαν με το κριτήριο της μη εξάρτησης από χρονικές πληροφορίες καθώς κάτι τέτοιο μπορεί να επηρεάσει την ακρίβεια της αναγνώρισης, δεδομένου ότι στο σύστημα εισήχθησαν αρχεία διαφορετικής χρονικής διάρκειας. Δόθηκε έμφαση στις παρακάτω παραμέτρους: Ρίζα του Μέσου Τετραγώνου της Ενέργειας (Root Mean Square Energy), Ποσοστό Χαμηλής Ενέργειας (Low Energy Rate), Ρυθμός Διέλευσης από το Μηδέν (Zero-Crossing Rate), Φασματική Απόκλιση (Rolloff), Φασματική Κατανομή των Συχνοτήτων (Brightness), Φασματικό Κεντροειδές (Spectral Centroid), Τυπική Απόκλιση (Spectral Spread), Φασματική Ασυμμετρία (Spectral Skewness), Φασματική Επιπεδότητα (Spectral Flatness), Φασματική Εντροπία (Entropy of Spectrum), Φασματικοί Συντελεστές της Κλίμακας Mel (MFCC) και Φασματική Διακύμανση (Spectral Irregularity). Στο τελικό στάδιο της διαδικασίας εκπαιδεύτηκε ένας ταξινομητής, η επιλογή του οποίου έγινε με κριτήριο την μέγιστη δυνατή ακρίβεια. Συγκεκριμένα, με χρήση της εφαρμογής MATLAB Classification Learner, εισήχθη στο σύστημα η βάση δεδομένων με τα 150 δεδομένα εκπαίδευσης. Ορίστηκαν 24 predictors και 1 response (*otus scops* / *cuculus canorus* / *streptopelia decaocto*), έγινε διασταυρούμενη επικύρωση 5 τμημάτων (5 fold cross-validation)⁹ και εκπαιδεύτηκαν όλοι οι αλγόριθμοι ταξινόμησης. Καλύτερα ποσοστά πέτυχαν τα Δένδρα Αποφάσεων (93.3%) και συγκεκριμένα τα “Ensembles of Bagged Trees”¹⁰.

4.1.3. Αποτελέσματα

Τα δεδομένα δοκιμής αναγνωρίστηκαν από τον ταξινομητή με ποσοστό ακρίβειας 80% για το σύνολό τους. Συγκεκριμένα στα δεδομένα του γκιώνη πραγματοποιήθηκαν 13 σωστές αναγνωρίσεις (62%), στα δεδομένα του κούκου όλα τα δείγματα αναγνωρίστηκαν σωστά (100%), τέλος, στα δεδομένα της δεκαοχτούρας πραγματοποιήθηκαν 17 σωστές αναγνωρίσεις (81%). Από τις 8 εσφαλμένες αναγνωρίσεις στην περίπτωση του

⁹ Ο στόχος της διασταυρούμενης επικύρωσης (cross-validation) είναι να δοκιμάσει την ικανότητα του μοντέλου να προβλέπει νέα δεδομένα που δεν χρησιμοποιήθηκαν για την εκτίμησή του, προκειμένου να επισημάνει προβλήματα όπως πχ. η Υπερμοντελοποίηση (Overfitting) και να δώσει μια εικόνα για το πώς το μοντέλο θα γενικευθεί σε ένα άγνωστο σύνολο δεδομένων.

¹⁰ Η ονομασία τους προέρχεται από τη δένδροειδή δομή τους και αποτελούν ίσως τον δημοφιλέστερο αλγόριθμο μηχανικής μάθησης με επιτήρηση. Όσον αφορά συγκεκριμένα στα Ensembles of Bagged Trees, πρόκειται για εφαρμογή πολλαπλών Δένδρων Αποφάσεων με την τεχνική Bagging (Bootstrap Aggregation).

γκιώνη οι 6 έδειξαν “κούκο” και οι 2 έδειξαν “δεκαοχτούρα”, ενώ στην περίπτωση της δεξαοχτούρας και οι 4 εσφαλμένες αναγνωρίσεις έδειξαν “κούκο”. Βάσει των δεδομένων που χρησιμοποιήθηκαν η συνολική ακρίβεια του συστήματος κρίνεται ως αρκετά ικανοποιητική.

Μία προσπάθεια ερμηνείας των αποτελεσμάτων αφορά στην ποιότητα των δειγμάτων που χρησιμοποιήθηκαν για την εκπαίδευση του ταξινομητή και για τις δοκιμές που ακολούθησαν: τα αρχεία που αποκτήθηκαν από το διαδίκτυο δεν είχαν ηχογραφηθεί από τον ίδιο εξοπλισμό, ούτε υπό τις ίδιες συνθήκες. Θόρυβος και, κυρίως, η ύπαρξη άλλων πουλιών και ζώων αναπόφευκτα επηρέασαν την αναγνώριση. Στην περίπτωση του κούκου - όπου επιτεύχθηκε απόλυτη ακρίβεια – είναι ενδιαφέρον ότι ακόμα και οι λανθασμένες αναγνωρίσεις των άλλων πουλιών έδειξαν “κούκο” σε ποσοστό 83.3%. Μία πιθανή εξήγηση είναι η εξής: οιοδήποτε δείγμα εισάγουμε στο σύστημα για αναγνώριση, το σύστημα θα μας δώσει συνήθως το πλησιέστερο αποτέλεσμα στα δεδομένα εκπαίδευσής του. Ειδικά όσον αφορά στην σχέση κούκου - δεκαοχτούρας, τα δύο πουλιά μοιάζουν πολύ στο συχνοτικό πεδίο. Ο κούκος παράγει συχνότητες σε ένα μέσο εύρος από 700 έως 900 Hz, ενώ το συχνοτικό εύρος της δεκαοχτούρας είναι ελαφρώς χαμηλότερο. Αντίθετα ο γκιώνης παράγει πιο ψηλές συχνότητες, περίπου στα 1500 Hz. Λαμβάνοντας υπόψη τα παραπάνω, μπορεί να γίνει αντιληπτό γιατί και οι 4 λανθασμένες αναγνωρίσεις της δεκαοχτούρας έδειξαν “κούκο”. Τέλος, μία άλλη παράμετρος είναι η ενσωμάτωση κάποιων καλεσμάτων στα δεδομένα εκπαίδευσης που παρέκλιναν αρκετά από τα τυπικά καλέσματα του είδους. Σε κάποιες δοκιμές που έγιναν, επανεκπαιδεύτηκε ο ταξινομητής χωρίς να συμπεριληφθούν αυτά τα αρχεία και τα αποτελέσματα παρουσιάστηκαν σχετικά βελτιωμένα (η συνολική ακρίβεια της αναγνώρισης προσέγγισε το 82%).

5. Επίλογος

Κλείνοντας, όσον αφορά στον τομέα της εκπαίδευσης, τα τελευταία χρόνια στην χώρα μας έχουν ήδη εφαρμοστεί εκπαιδευτικές προτάσεις που αφορούν ηχητικούς περιπάτους και ανάλυση του ηχοτοπίου του σχολικού περιβάλλοντος καθώς και παρατήρηση του καλέσματος των πουλιών και ταυτοποίηση των ειδών οι οποίες στοχεύουν στην ανάπτυξη της δημιουργικότητας και της περιβαλλοντικής ευαισθητοποίησης των παιδιών καθώς και στον μουσικό και ηχητικό εγγραμματισμό [Γεωργάκη, 2016]. Είναι γεγονός ότι η μελέτη των ήχων των πουλιών μπορεί να λειτουργήσει ευεργετικά για τους μαθητές, σε επίπεδο περιβαλλοντικής ευαισθητοποίησης αλλά και εξοικείωσης με την τεχνολογία. Τα παιδιά μαθαίνουν να χρησιμοποιούν με διαφορετικό τρόπο ακόμα και τα καθημερινά τεχνολογικά μέσα που διαθέτουν - όπως το smartphone ή ο υπολογιστής, το μάθημα γίνεται πιο διαδραστικό και το εκπαιδευτικό αντικείμενο μεταφέρεται πλέον έξω από τον χώρο της σχολικής αίθουσας.

Είναι έκδηλο ότι υπάρχει ακόμα περιθώριο προς περαιτέρω έρευνα σχετικά με τους ήχους των πουλιών. Ζητούμενο και στόχος για τον μελλοντικό ερευνητή δεν μπορεί παρά να είναι η βελτίωση της υπάρχουσας τεχνολογίας ώστε να γίνει πλέον εφικτή η διεξοδική καταγραφή περισσότερων ειδών. Η καταγραφή της τοπικής ορνιθοπανίδας μπορεί, παράλληλα, να αποφέρει οφέλη τόσο στην πολιτιστική οικονομία όσο και στον πολιτιστικό τουρισμό. Θα πρέπει όμως μέσω μίας σύμπραξης της επιστημονικής κοινότητας με την κεντρική εξουσία, να οριστούν διεξοδικά οι μέθοδοι και οι πρακτικές ενός τέτοιου εγχειρήματος.

Βιβλιογραφικές Αναφορές

Agranat, I. (2009). *Automatically identifying animal species from their vocalizations*. – Proceedings of the 5th International Conference on Bio-Acoustics, Holywell Park.

Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G. & Alvarez, R. (2013). *Real-time bioacoustics monitoring and automated species identification*. – PeerJ 1: e103.

Berwick, R. C., Okanoya, K., Beckers, G. J., & Bolhuis, J. J. (2011). *Songs to syntax: the linguistics of birdsong*. Trends Cogn. Sci. 15, 113–121.

Cai, J., Ee, D., Pham, B., Roe, P. & Zhang, J. (2007). *Sensor Network for the monitoring of Ecosystem: Bird Species Recognition*. Retrieved from <http://eprints.qut.edu.au/11227/1/11227a.pdf>

Charif, R., Strickman, L. & Waack, A. (2010). *Raven Pro 1.4 user's manual. Revision 11*. – The Cornell Lab of Ornithology, Ithaca, NY. Retrieved from <https://www.birds.cornell.edu/brp/raven/Raven14UsersManual.pdf>

- Clemins, P. J. & Johnson, M. T. (2003). *Application of speech recognition to African elephant (Loxodonta africana) vocalizations*. – IEEE Int. Conf. Acoust. Speech Signal Process. 1: 484–487.
- de Oliveira, A. G., Ventura, T. M., Ganchev, T. D., de Figueiredo, J. M., Jahn, O., Marques, M. I. & Schuchmann, K.-L. (2015). *Bird acoustic activity detection based on morphological filtering of the spectrogram*. – Appl. Acoust. 98: 34–42.
- Deng, L. & O'Shaughnessy, D. (2003). *Speech processing: a dynamic and optimization-oriented approach*. Marcel Dekker. pp. 41–48.
- Fagerlund, S. (2007). *Bird species recognition using support vector machines*. EURASIP J. Appl. Signal Process., 2007(1):64-64.
- Γεωργάκη Αναστασία(2016), «Η δε γλαύξ ολολύζει» : η ακουστική επικοινωνία της ορνιθοπανίδας σαν μέσο για τον ηχητικό και μουσικό εγγραμματισμό στο ψηφιακό σχολείο, στα πρακτικά του 4ου συνεδρίου Ακουστικής Οικολογίας, Πανεπιστήμιο Αιγαίου.
- Georgaki, A. (2017). *Spectres d'une écho-mythologie dans l'oeuvre de François-Bernard Mâche dans le livre "F.B Mâche, le compositeur et le savant face à l'univers sonore;*, ed. Hermann, Paris (Broché).
- Jančovič, P. & Köküer, M. (2016). "Recognition of Multiple Bird Species Based on Penalised Maximum Likelihood and HMM-Based Modelling of Individual Vocalisation Elements." Interspeech 2016, 09, 2016.
- Jinnai, M., Boucher, N., Fukumi, M. & Taylor, H. (2012). *A new optimization method of the geometric distance in an automatic recognition system for bird vocalisations*. – Proceedings of the Acoustics 2012 Nantes Conference, pp. 2439–2445.
- Kogan, J. A. & Margoliash, D. (1998). *Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study*. – J. Acoust. Soc. Am. 103: 2185–2196.
- Kohavi, R. & Provost, F. (1998). «Glossary of terms». Machine Learning 30: 271–274.
- Lartillot, O. & Toivianen, P. (2007). "A Matlab Toolbox for Musical Feature Extraction From Audio", International Conference on Digital Audio Effects, Bordeaux.
- Lee, C. H., Lee, Y. K. & Huang, R. Z. (2006). *Automatic recognition of bird songs using cepstral coefficients*. – J. Inform. Technol. Appl. 1: 17–23.
- Luther, D. & Baptista, L. (2009). *Urban noise and the cultural evolution of bird songs*. 277. Proceedings of the Royal Society B: Biological Sciences.
- Mache F. B. (1983). *Musique, mythe, nature, ou les Dauphins d'Arion*.- Paris: Méridiens Klincksieck.
- Makhoul, J. & Schwartz, R. (1995). *State of the art in continuous speech recognition*. – Proc. Natl Acad. Sci. USA 92: 9956–9963.
- McIlraith, A. L. & Card, H. C. (1995). *Birdsong recognition with DSP and neural networks*. pp. 409–414 in IEEE WESCANEX'95 Proceedings, Winnipeg, Canada.
- Potamitis, I. (2014). *Automatic Classification of a Taxon-Rich Community Recorded in the Wild*. PLoS ONE 9(5): e96936.
- Schafer, R. M. (1977). *The Tuning of the World*.- Michigan: Knopf; original: University of Michigan.
- Selin, A., Turunen, J. & Tanntu, J. T. (2007). *Wavelets in recognition of bird sounds*. – EURASIP J. Appl. Signal Process. 2007: 141–141.

- Specht, R. (1993). AVISOFT – *sound analysis and synthesis laboratory Pro: a PC program for sonographic analysis*. – AVISOFT, Berlin, Germany.
- Stattner, E., Segretier, W., Collard, M., Hunel, P. & Vidot, N. (2013). *Song-based classification techniques for endangered bird conservation*. – ICML 2013 Workshop on Machine Learning for Bioacoustics, Atlanta, GA, USA.
- Stowell, D. & Plumbley, M. D. (2014). *Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning*. – PeerJ 2: e488.
- Tchernichovski, O. (2012). *Sound Analysis Pro 2011 user manual*. – Sound Analysis Pro. Retrieved from <http://www.soundanalysispro.com/manual-1>
- Thorpe, W. H. (1958). *The learning of song patterns by birds, with especial references to the song of the chaffinch, Fringilla coelebs*. Ibis 100:535-570.
- Wallin, N. L. (1991). *Biomusicology: Neurophysiological, Neuropsychological and Evolutionary Perspectives on the Origins and Purposes of Music*.- Stuyvesant, NY: Pendragon Press.
- Williams, H. (2004). *Birdsong and singing behavior*. – Ann. N. Y. Acad. Sci. 1016: 1–30.
- Κουτσογιάννης, Γ. (2019). *birdMelos: Δημιουργία εφαρμογής για την αυτόματη αναγνώριση των φωνημάτων των πτηνών*. Πανεπιστήμιο Αθηνών, Τμήμα Μουσικών Σπουδών.

Bioacoustics and “Shazam”: Methods and computational applications to identify bird-fauna sounds in the natural environment

Koutsogiannis George
MA, Department of Music Studies, U.O.A.
geokouts@music.uoa.gr

Georgaki Anastasia
Associate Professor of Music Technology at
University of Athens - Director of Laboratory
of Music Acoustics and Technology
georgaki@music.uoa.gr

Abstract

During the last few years, companies in association with bioacoustics laboratories have introduced particular acoustical analysis software for bird calls, via the so-called automatic pattern recognition method, which identifies with both the rhythmic and the melodic structure of birdsong, always in accordance with the species' biological classification. Such applications are based on an entirely different methodology, in the sense of analyzing and identification, however, they appear to have some common limitations such as low accuracy, lack of professional analysis methods for covering a broader range of species, various technical difficulties and last the manpower shortage to fulfill such attempts. In the context of these events, we are going to present some of the most contemporary applications and significant research findings in the field, while focusing on the most widely used methods and technics in practice. Next, we are going to elaborate on the issues with regard to the recording and the comparative study of a wider range of bird species, as well as the procedure of developing a new application to cover the Greek bird fauna in the future. Last, we will discuss the benefits of such applications in school education, through the lens of acoustic ecology.

Key Words

Acoustic ecology, biomusicology, bird vocalization, birdsong recognition, machine learning, music technology, sound analysis.