

ΥΠΟΛΟΓΙΣΤΙΚΑ ΜΟΝΤΕΛΑ ΤΗΣ ΜΟΥΣΙΚΗΣ ΝΟΗΣΗΣ

Γιώργος Βελισσαρίδης & Χριστίνα Αναγνωστοπούλου

Η υπολογιστική μοντελοποίηση της μουσικής νόησης είναι μία προσέγγιση στη μελέτη του νου με αντικείμενο την κατασκευή υπολογιστικών προγραμμάτων τα οποία προσομοιάζουν τον τρόπο που λειτουργεί ο νους όταν επεξεργάζεται μουσική. Σε αυτή την ανασκόπηση αρχικά συζητούνται γενικά ζητήματα σχετικά με την υπολογιστική μοντελοποίηση και στη συνέχεια παρουσιάζονται χαρακτηριστικά μοντέλα οργανωμένα με βάση τον τομέα της νόησης που μοντελοποιούν. Τα υπολογιστικά μοντέλα δέχονται δεδομένα σε συμβολική μορφή ή σε μορφή αρχείων ήχου ως είσοδο και χρησιμοποιούν αναπαραστάσεις γνώσης, προκαθορισμένες/χειροποίητες ή προϊόντα μηχανικής μάθησης. Δύο θεμελιώδη προβλήματα της μουσικής αντίληψης που επιδέχθηκαν υπολογιστική μοντελοποίηση νωρίς στην ιστορία του τομέα είναι η εύρεση τονικότητας και μέτρον. Άλλα ζητήματα μοντελοποίησης της μουσικής αντίληψης περιλαμβάνουν την ανίχνευση τονικού ύψους από ηχητικό σήμα, την ομαδοποίηση νοτών σε πρότυπα (patterns), σε φράσεις και σε μεγαλύτερες ενότητες, τον διαχωρισμό φωνών σε πολυφωνική μουσική, την αρμονική ανάλυση και την ονομασία τονικών υψών. Σχετικά με τη γνωσιακή επεξεργασία της μουσικής, έχουν κατασκευαστεί μοντέλα ανακάλυψης προτύπων και μουσικής ομοιότητας. Η προσδοκία είναι βασικό κομμάτι της μουσικής εμπειρίας, γι' αυτό έχουν προταθεί υπολογιστικά μοντέλα που υλοποιούν τη θεωρία IR, μοντέλα που βασίζονται σε αναλογίες με φυσικές δυνάμεις καθώς και το μοντέλο IDyOM, που στηρίζεται στην αναπαράσταση πολλαπλών οπτικών γωνιών. Άλλα τμήματα της μουσικής εμπειρίας που μοντελοποιούνται σε υπολογιστικά μοντέλα είναι η μουσική τάση, η αυξομείωση της ανάγκης για μουσική λύση, και τα μουσικά συναισθήματα που έχουν αποτελέσει κυρίως αντικείμενο εφαρμοσμένης έρευνας του τομέα ανάκτησης μουσικής πληροφορίας. Τα μοντέλα ακουστικής προσοχής αποσκοπούν στην εξήγηση των νοητικών μηχανισμών εστίασης γνωσιακών πόρων και τα μοντέλα μακροχρόνης μνήμης στην αποτύπωση των νοητικών μηχανισμών κατασκευής κατηγοριών και σχημάτων. Η εργασία της αλγοριθμικής σύνθεσης έχει στόχο την παραγωγή πρωτότυπης και αξιόλογης μουσικής και ορισμένα σχετικά μοντέλα είναι γνωσιακά. Το θέμα αυτό έχει προσεγγιστεί με συμβολική TN, όπως με κανόνες και γενετικές γραμματικές, αλλά και με στατιστικά μοντέλα μηχανικής μάθησης όπως μαρκοβιανά μοντέλα και νευρωνικά δίκτυα. Η γνωσιακή αρχιτεκτονική IDyOT επιχειρεί να ενοποιήσει πολλά κομμάτια της μουσικής νόησης σε ένα ενιαίο μοντέλο το οποίο εκτελεί πολλαπλές εργασίες επεξεργασίας ακολουθιακών δεδομένων.

Λέξεις-κλειδιά:

Μουσική, αντίληψη, νόηση, υπολογιστική, γνωσιακή μοντελοποίηση

ΕΙΣΑΓΩΓΗ

Η υπολογιστική μοντελοποίηση (computational modeling) αποτελεί από την απαρχή της γνωσιακής επιστήμης μία από τις βασικές προσεγγίσεις της, μαζί με την πειραματική ψυχολογία και τη νευροεπιστήμη, και είναι στενά συνδεδεμένη με τον τομέα της τεχνητής νοημοσύνης. Είναι μία προσέγγιση πειραματική και ποσοτική στη φύση της. Υπολογιστικό γνωσιακό μοντέλο ονομάζουμε ένα υπολογιστικό πρόγραμμα που προσομοιώνει διεργασίες και μηχανισμούς του νου, με σκοπό να δώσει εξηγήσεις για το πώς ο νους επεξεργάζεται πληροφορία. Τα μοντέλα αυτά έχουν βρει εφαρμογή σχεδόν σε όλους τους υποτομείς μελέτης του νου, μεταξύ των οποίων και στη μουσική νόηση. Κάποια από τα χαρακτηριστικότερα υπολογιστικά μοντέλα της μουσικής νόησης παρουσιάζονται παρακάτω, μαζί με γενικότερα ζητήματα που προκύπτουν κατά τη μοντελοποίηση της νόησης.

Πιο συγκεκριμένα, αρχικά ορίζονται κάποιες έννοιες και παρουσιάζονται γενικά θέματα της υπολογιστικής μοντελοποίησης. Ακολούθως, μελετώνται μοντέλα αντίληψης βασικών μουσικών χαρακτηριστικών, μουσικής νόησης, εμπειρίας, προσοχής και μνήμης. Τέλος, γίνεται λόγος για μοντέλα μουσικής σύνθεσης, και για ένα γενικότερο μοντέλο αντίληψης και νοητικής επεξεργασίας τόσο μουσικής όσο και γλώσσας.

Στο συγκεκριμένο κεφάλαιο έχουμε θεωρήσει ως γνωσιακό μοντέλο ένα υπολογιστικό μοντέλο που μοντελοποιεί ανθρώπινες μουσικές διαδικασίες, χωρίς απαραίτητα να επιχειρεί να εξομοιώσει τον τρόπο με τον οποίο γίνεται η διαδικασία γνωσιακά.

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΖΗΤΗΜΑΤΑ

Προτού παρουσιάσουμε υπολογιστικά μοντέλα, είναι χρήσιμο να αποσαφηνίσουμε τους όρους «μοντέλο» και «υπολογιστικό», καθώς και οι δύο κρύβουν μεγάλο βάθος.

Ένα μοντέλο στη γνωσιακή επιστήμη είναι μία περιγραφή ενός μέρους της (ανθρώπινης) νόησης, δηλαδή ορισμός των στοιχείων που το αποτελούν, των σχέσεων μεταξύ τους και διαδικασιών που περιλαμβάνουν τα στοιχεία αυτά. Μάλιστα, η εν λόγω περιγραφή δεν είναι μία απλή σχέση εισόδου-εξόδου (input-output), δηλαδή ερεθίσματος και απόκρισης σε αυτό, αλλά δίνει πληροφορίες για διαδικασίες και μηχανισμούς του νου, μέσω των οποίων επιτυγχάνεται τελικά η παραγωγή εξόδου για κάθε είσοδο (Temperley, 2012). Η λειτουργία ενός μοντέλου αποτελείται από τα εξής τρία στάδια: λήψη δεδομένων (μπορούμε να τα σκεφτούμε ως ερεθίσματα από το περιβάλλον), επεξεργασία αυτών των δεδομένων με βάση την τρέχουσα κατάσταση του μοντέλου και παραγωγή αποτελεσμάτων (μπορούμε να τα σκεφτούμε ως απόκριση του μοντέλου, συχνά μπορεί να αποτελεί πρόβλεψη). Αλγόριθμο ονομάζουμε τη σειρά βημάτων που εκτελεί ένα μοντέλο για να εκτελέσει κάποια διεργασία. Ένα μοντέλο μπορεί να χρησιμοποιηθεί τόσο για επεξήγηση του μέρους της νόησης στο οποίο εστιάζει, όσο και για παραγωγή προβλέψεων της συμπεριφοράς αυτού του μέρους της νόησης σε διάφορες καταστάσεις.

Τα περισσότερα υπολογιστικά μοντέλα της νόησης περιγράφουν κάποιον τομέα ή διεργασία του νου σε αφαιρετικό επίπεδο, δηλαδή με λιγότερη πληροφορία απ' ό τι το βιολογικό επίπεδο. Δεν επιχειρούν να συνδέσουν τους μηχανισμούς που προσδιορίζουν με τη φυσική υλοποίησή τους στο νευρολογικό επίπεδο του εγκεφάλου (με εξαίρεση την υπολογιστική νευροεπιστήμη). Ωστόσο, είναι κοινώς αποδεκτό στη γνωσιακή επιστήμη ότι το επίπεδο περιγραφής τους είναι εξίσου πραγματικό με το νευρολογικό επίπεδο, απλώς ψηλότερα στην ιεραρχία αφαίρεσης (από το αναλυτικότερο επίπεδο-χαμηλά, έως το πιο αφαιρετικό-ψηλά), που τους προσδίδει μοναδική χρησιμότητα για την κατανόηση του νου.

Σημαντικός είναι ο διαχωρισμός μεταξύ υπολογιστικών μοντέλων στη γνωσιακή επιστήμη και στην πληροφορική (ή υπολογιστική μηχανική /computational engineering). Στην πληροφορική, σκοπός είναι συνήθως η επίλυση κάποιου πρακτικού προβλήματος όσο πιο αποδοτικά και αποτελεσματικά γίνεται και όχι η εξήγηση του πώς ο νους λύνει το πρόβλημα αυτό. Ωστόσο, υπάρχει ανταλλαγή ιδεών μεταξύ της προσέγγισης της πληροφορικής και της γνωσιακής έρευνας.

Σχετικά με την έννοια του υπολογισμού στα μοντέλα που παρουσιάζονται παρακάτω, αυτή δεν έγκειται στο ότι υλοποιούνται σε ηλεκτρονικούς υπολογιστές (τα πρώτα από αυτά είναι τόσο απλά που μπορούν να υλοποιηθούν και με χαρτί και μολύβι!). Αντιθέτως, το ότι τα μοντέλα είναι υπολογιστικά σημαίνει ότι είναι πλήρως προσδιορισμένα ώστε να είναι εφικτή η υλοποίησή τους σε ηλεκτρονικούς υπολογιστές - σε αντίθεση με μοντέλα προσδιορισμένα σε φυσική γλώσσα που μπορεί να είναι ασαφή και ασυνεπή. Φυσικά, η χρήση υπολογιστών προσφέρει πολλά ακόμη οφέλη όπως πιο εύκολη και γρήγορη ανάπτυξη και έλεγχο μοντέλων. Επιπλέον, η τεράστια πολυπλοκότητα των μοντέλων (όπως τα νευρωνικά δίκτυα) και ο όγκος των δεδομένων που αυτά χρησιμοποιούν (τεράστια σώματα/corpora, δηλαδή συλλογές μουσικών έργων), κάνουν τη χρήση ηλεκτρονικών υπολογιστών απαραίτητη.

Ακόμη, για να επιβεβαιώσουμε την εγκυρότητα ενός γνωσιακού υπολογιστικού μοντέλου πρέπει να το επαληθεύσουμε συνδέοντάς το με ψυχολογικά και νευροβιολογικά πειράματα και γενικότερα με άλλους τομείς. Έτσι μπορούμε να διαπιστώσουμε εάν ένα μοντέλο μοντελοποιεί έναν υπαρκτό μηχανισμό του νου και όχι έναν μηχανισμό που παράγει καλά αποτελέσματα σε εργασίες του τμήματος της νόησης που προσομοιάζει, όμως το επιτυγχάνει με διαφορετικό τρόπο από τον ανθρώπινο νου. Αυτό το καταφέρνουμε παράγοντας προβλέψεις της συμπεριφοράς του τμήματος της νόησης που μοντελοποιείται, τόσο σε ήδη πειραματικά μελετημένες καταστάσεις, όσο και σε καταστάσεις που δεν έχουν μελετηθεί εμπειρικά ακόμα. Για να επαληθευτεί η εγκυρότητα ενός μοντέλου στην πρώτη περίπτωση, οι προβλέψεις του πρέπει να συμφωνούν με τα υπάρχοντα αποτελέσματα, ενώ στη δεύτερη πρέπει οι προβλέψεις να επιβεβαιωθούν από νέα πειράματα. Επιπλέον κριτήριο της ποιότητας ενός μοντέλου, ειδικά στην περίπτωση που έχουμε να επιλέξουμε μεταξύ πολλών μοντέλων, είναι το να παράγει σχετικά αναπάντεχες προβλέψεις για νέες καταστάσεις, επειδή τότε παρέχει περισσότερη γνώση από μοντέλα που παράγουν αναμενόμενες προβλέψεις για τις νέες καταστάσεις (Hoping, 2006). Επιπροσθέτως, ένα γνωσιακό μοντέλο πρέπει να έχει ικανότητα γενίκευσης (π.χ. να μπορεί να εξηγήσει πειραματικά δεδομένα που δεν λήφθηκαν υπόψη κατά την κατασκευή του), να είναι όσο πιο

απλό γίνεται (κριτήριο γνωστό ως το “ξυράφι του Όκαμ”), και, πιθανώς, να συνδέει υπάρχουσες θεωρίες (Purwins et al., 2008).

Ορισμένα από τα υπολογιστικά μοντέλα που παρουσιάζονται στο άρθρο δεν έχουν ακόμη υποβληθεί σε ρητή γνωσιακή επαλήθευση μέσω πειραματικών ή θεωρητικών προσεγγίσεων, όπως αυτές που αναφέρθηκαν προηγουμένως. Παρ' όλα αυτά, κρίνεται σκόπιμο να συμπεριληφθούν, καθώς αποτελούν σημαντικούς υποψήφιους μηχανισμούς που δύνανται να αποτυπώσουν γνωσιακές διεργασίες, έχουν δηλαδή γνωσιακή ευλογοφάνεια (cognitive plausibility).

Ένα σημαντικό γενικό πρόβλημα που ανακύπτει κατά την ανάπτυξη οποιουδήποτε υπολογιστικού μοντέλου της μουσικής νόησης είναι το εξής: τι είδους πληροφορία, ή είσοδο, θα δέχεται; Αρχεία ήχου, που διατηρούν σχεδόν αυτούσια την πληροφορία του ηχητικού σήματος, ή κάποιου είδους συμβολική μουσική αναπαράσταση, που εφαρμόζει απωλεστική συμπίεση στο ηχητικό σήμα, με αποτέλεσμα να χάνεται πληροφορία, αλλά η πληροφορία που προκύπτει να είναι διακριτή (discrete); Η πλειοψηφία των γνωσιακών μοντέλων της μουσικής νόησης μέχρι σήμερα χρησιμοποιεί συμβολικές αναπαραστάσεις. Αυτό οφείλεται στο ότι οι συμβολικές αναπαραστάσεις είναι πιο εύκολες στο χειρισμό (έχουν μικρότερο μέγεθος απ' ότι τα αρχεία ήχου και έτσι είναι πιο γρήγορη η επεξεργασία τους) και επίσης στο ότι πολλές εργασίες μπορούν να λυθούν πιο αποτελεσματικά μέσω συμβολικής πληροφορίας, απ' ότι μέσω αρχείων ήχου. Ίσως αυτό να φαίνεται περίεργο, καθώς ένα αρχείο ήχου ενός μουσικού κομματιού περιέχει περισσότερη πληροφορία απ' ότι μία συμβολική αναπαράσταση του ίδιου κομματιού. Μία πιθανή εξήγηση είναι ότι και ο νους δημιουργεί συμβολικές αναπαραστάσεις της μουσικής κατά την ακρόαση, οι οποίες αναπαραστάσεις διατηρούν μόνο τη σημαντική πληροφορία για τις εργασίες που ο νους ενδιαφέρεται να εκτελέσει (π.χ. εύρεση τονικότητας), αφήνοντας πίσω την περιττή. Θα εννοείται ότι όλα τα μοντέλα που παρουσιάζονται παρακάτω χρησιμοποιούν συμβολική αναπαράσταση ως είσοδο, εκτός αν αναφέρεται κάτι άλλο ρητά.

Τι είδους συμβολική αναπαράσταση όμως μπορεί να χρησιμοποιηθεί σε ένα υπολογιστικό μοντέλο της μουσικής νόησης; Το ερώτημα αυτό σχετίζεται με μία θεμελιώδη έννοια στη γνωσιακή επιστήμη, αυτή της νοητικής αναπαράστασης: ενός αντικειμένου του νου με σημασιολογικό περιεχόμενο που αναφέρεται στον εξωτερικό κόσμο ή σε κάποια αφαίρεσή του κόσμου. Ο τομέας που ασχολείται με τη δημιουργία αναπαραστάσεων για υπολογιστικά μοντέλα ονομάζεται Αναπαράσταση Γνώσης (knowledge representation). Είναι ένας σημαντικός τομέας στην Τεχνητή Νοημοσύνη, και στη μουσική περιλαμβάνει πολλά διαφορετικά σχήματα. Εκτός από το πρωτόκολλο MIDI, έχουν προταθεί επίσης το σύστημα CHARM (Wiggins & Smaill, 1998), το σύστημα πολλαπλών οπτικών γωνιών (multiple viewpoints - Conklin & Witten, 1995, Conklin & Anagnostopoulou, 2001), το General Chord Type (GCT - Cambouropoulos et al., 2014), και άλλα. Οι περισσότερες αναπαραστάσεις γνώσης στη μουσική συνήθως ξεχωρίζουν τις διαφορετικές μουσικές παραμέτρους και δημιουργούν ξεχωριστές αναπαραστάσεις για τις διαφορετικές μουσικές ιδιότητες. Μία εναλλακτική προσέγγιση από τη χειροποίητη κατασκευή αναπαραστάσεων από ανθρώπους είναι η χρήση μηχανικής μάθησης ώστε το υπολογιστικό μοντέλο να δημιουργεί το ίδιο τις αναπαραστάσεις γνώσης του με βάση τα δεδομένα που δέχε-

ται, κάτι που αποτελεί αντικείμενο του τομέα «μάθησης αναπαραστάσεων» /“representation learning” (Kim et al., 2020). Ανάλογα με το κομμάτι του νου που ενδιαφέρεται να μελετήσει κανείς με υπολογιστική μοντελοποίηση, είναι πιθανώς διαφορετική η αναπαραστάση γνώσης που δίνει τα καλύτερα αποτελέσματα για τις εργασίες που εκτελεί το κομμάτι αυτό.

Παρακάτω περιγράφουμε μερικές υπολογιστικές έρευνες που είτε έχουν παρουσιαστεί ως γνωσιακά μοντέλα, είτε ως υπολογιστικές τυποποιήσεις ανθρώπινων μουσικών διεργασιών που όμως πιστεύουμε ότι έχουν γνωσιακή ευλογοφάνεια. Ομαδοποιούμε τις έρευνες αυτές ανάλογα με το τι προσπαθούν να μοντελοποιήσουν.

Εύρεση τονικότητας

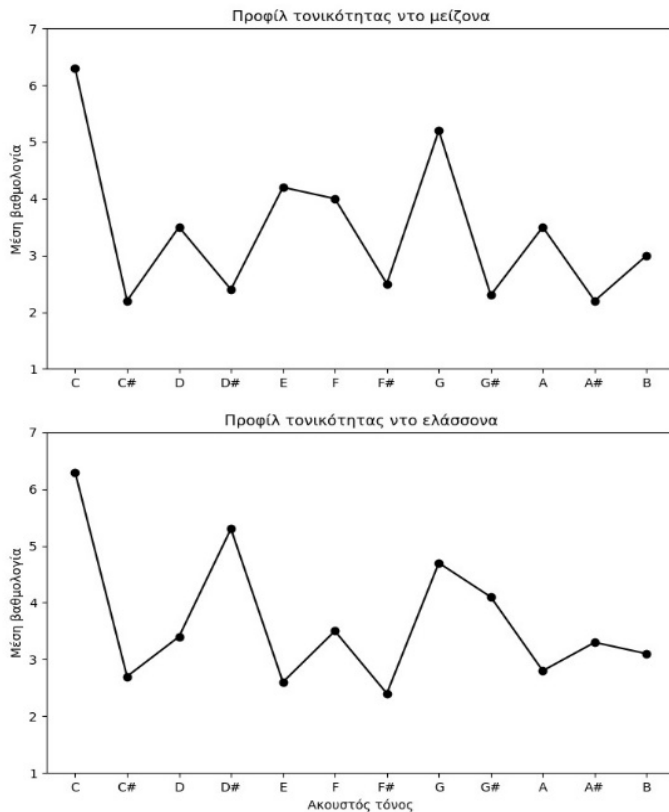
Η τονικότητα είναι πολύ βασικό χαρακτηριστικό σχεδόν κάθε είδους μουσικής, και ιδιαίτερα της δυτικής τονικής μουσικής. Διαφορετικοί πολιτισμοί χρησιμοποιούν διαφορετικούς μηχανισμούς τονικών κέντρων. Πειραματικές μελέτες έχουν επιδείξει τη σημασία της ανθρώπινης αντίληψης και νόησης στην εύρεση τονικότητας κατά την ακρόαση. Μάλιστα, έχει φανεί ότι ακόμη και μη μουσικά εκπαιδευμένοι ακροατές έχουν ευαισθησία σε τονικά και αρμονικά πρότυπα (Krumhansl, 2004). Είναι λοιπόν λογικό ότι το πρόβλημα του πώς ο νους εξάγει την τονικότητα ενός κομματιού όσο το ακούει ήταν από τα πρώτα που αντιμετωπίστηκε με τη δημιουργία υπολογιστικών μοντέλων.

Ένα από τα πρώτα μοντέλα που δημιουργήθηκαν και είχαν μεγάλη επιρροή είναι αυτό των Longuet-Higgins και Steedman (1971), το οποίο δημιούργησαν με σκοπό την υπολογιστική ανάλυση των φουγκών του καλώς-συγκερασμένου πληκτροφόρου του J.S. Bach. Το μοντέλο δέχεται ως είσοδο μονοφωνική μελωδία (το θέμα μίας φούγκας) την οποία επεξεργάζεται σειριακά, μία νότα τη φορά, ξεκινώντας από την αρχή του κομματιού. Σε κάθε νότα, όλες οι τονικότητες οι οποίες δεν περιλαμβάνουν τη νότα αυτή αποκλείονται από την αναζήτηση. Χρησιμοποιούνται επίσης πολλοί ευρετικοί κανόνες¹ για την αντιμετώπιση ειδικών περιπτώσεων. Παρά τον σωστό εντοπισμό των τονικοτήτων στις μελωδίες του καλώς συγκερασμένου πληκτροφόρου, το μοντέλο δεν τα πηγαίνει καλά σε πολλές άλλες μελωδίες - ιδιαίτερα όταν αυτές περιλαμβάνουν νότες εκτός της τονικότητας ή/και δεν χρησιμοποιούν όλες τις νότες μίας τονικότητας. Το ελάττωμα αυτό προκαλείται από την υπερεφαρμογή (overfitting) του μοντέλου στα δεδομένα για τα οποία αναπτύχθηκε, που μειώνει την ικανότητα γενίκευσής του σε άλλα δεδομένα (άλλες μελωδίες με διαφορετικά χαρακτηριστικά).

Μία διαφορετική προσέγγιση στο πρόβλημα της εύρεσης τονικότητας (key-finding) είναι αυτή των Krumhansl και Schmuckler (Krumhansl, 1990). Το μοντέλο

¹ Κανόνες που έχει παρατηρηθεί εμπειρικά ότι συχνά βρίσκουν γρήγορα λύσεις, που μπορεί όμως να μην είναι βέλτιστες ή και ορθές.

τους βασίζεται στην έννοια του «προφίλ τονικότητας» (key profile), ενός διάνυσματος² 12 τιμών. Κάθε τέτοιο προφίλ αντιπροσωπεύει τη σταθερότητα και καταλληλότητα κάθε τόνου του δωδεκάφθογου τονικού συστήματος σε σχέση με μία τονικότητα. Οι τιμές αυτές βρέθηκαν μέσω πειραμάτων στα οποία συμμετέχοντες-ακροατές κλήθηκαν να αξιολογήσουν κατά πόσο κάθε τονική τάξη³ που άκουγαν ταιρίαζε σε ένα τονικό περιεχόμενο, που είχε προηγουμένως εγκαθιδρυθεί μέσω ακρόασης της εκάστοτε τονικής συγχορδίας και τριών πτωτικών ακολουθιών (Krumhansl & Kessler, 1982) (Εικόνα 1).



Εικόνα 1. Τα προφίλ τονικότητας για τις τονικότητες ντο μείζονα (πάνω) και ντο ελάσσονα (κάτω), όπως αυτά προέκυψαν από τα πειράματα των Krumhansl και Kessler. Ο οριζόντιος άξονας αναπαριστά καθεμία από τις δώδεκα νότες του ισοσυγκερασμένου συστήματος και ο κατακόρυφος τη μέση βαθμολογία (σε κλίμακα από 1 έως 7) των συμμετεχόντων για το κατά πόσο καθεμία από τις νότες ταιριάζει με την αντίστοιχη τονικότητα. Οι νότες αυτές ακούγονταν αφού είχε ακουστεί η τονική συγχορδία και τρεις πτωτικές ακολουθίες της κάθε τονικότητας (προσαρμοσμένο σχήμα από τα σχετικά ευρήματα των Krumhansl & Kessler, 1982).

² διάνυσμα = διατεταγμένο σύνολο n τιμών, όπου το n είναι φυσικός αριθμός.

³ τονική τάξη/pitch class = νότα συγκεκριμένου ονόματος ανεξάρτητα της οκτάβας της.

Με δεδομένα τα 12 προφίλ για κάθε τονικότητα, το μοντέλο λαμβάνει ως είσοδο τη συμβολική αναπαράσταση ενός νέου μουσικού κομματιού και παράγει ένα νέο διάνυσμα για αυτό, κάθε τιμή του οποίου αντιστοιχεί στη συνολική διάρκεια της αντίστοιχης τονικής τάξης στο σύνολο του κομματιού. Στη συνέχεια, το νέο διάνυσμα συγκρίνεται με τα 12 διανύσματα των τονικοτήτων και τελικά επιλέγεται η τονικότητα με το διάνυσμα που παρουσιάζει τον καλύτερο συσχετισμό με το διάνυσμα του κομματιού. Συμπερασματικά, το μοντέλο αυτό βασίζεται στην κατανομή των τονικών τάξεων ενός κομματιού, χωρίς να λαμβάνει υπόψη τη χρονική διάταξη των νωτών ή το τονικό ύψος τους (οκτάβα όπου εμφανίζονται).

Η προσέγγιση αυτή ακολουθήθηκε και από μεταγενέστερα μοντέλα όπως αυτό του Leman (1995), όπου τα διανύσματα παράγονται από αρχεία ήχου, και της Chew (2002, 2014), όπου χρησιμοποιείται αναπαράσταση ελικοειδούς σπείρας για τους τόνους και σε κάθε τονικότητα αντιστοιχίζεται ένα σημείο του χώρου.

Παρά την επιτυχία των μοντέλων εύρεσης τονικότητας που κάνουν χρήση κατανωμών, υπάρχουν και διάφορες προκλήσεις. Μία από τις σημαντικότερες είναι ο χειρισμός μετατροπιών, αλλαγών τονικότητας εντός ενός κομματιού, που δυσκολεύει την εφαρμογή μοντέλων κατανομών καθώς, στην απλούστερη μορφή τους, η πληροφορία στην οποία βασίζονται αντλείται από ολόκληρο το κομμάτι, χωρίς να λαμβάνονται υπόψη αλλαγές τονικότητας. Επίσης, παρά την ισχύ της προσέγγισης κατανομών στο πρόβλημα εύρεσης τονικότητας, ως θεωρία του νου είναι υπερ-απλουστευτική, όπως φανερώνεται και από πειράματα που δείχνουν ότι οι ίδιοι τόνοι σε διαφορετική σειρά μπορεί να προκαλέσουν διαφορετική αντίληψη τονικότητας σε ακροατές (Matsunaga & Abe, 2005). Άλλου είδους πληροφορία που πειραματικά έχει βρεθεί ότι χρησιμοποιείται στην αντίληψη τονικότητας από τον νου είναι αρμονική πληροφορία καθώς και τυποποιημένα μελωδικά πρότυπα, πληροφορία που αν ενσωματωνόταν στα μοντέλα κατανομών ενδεχομένως θα βελτίωνε την απόδοσή τους. Ένας τρόπος να συμβεί αυτό είναι η χρήση νευρωνικών δικτύων, όπως στην εργασία των Korzeniowski και Widmer (2017) που πρότειναν ένα συνελκτικό νευρωνικό δίκτυο (convolutional neural network) το οποίο αποδίδει καλύτερα από τα άλλα συστήματα στη βιβλιογραφία και μάλιστα είναι ικανό να μάθει ένα ενοποιημένο μοντέλο για πολλά μουσικά είδη.

Μία εναλλακτική προσέγγιση στην εύρεση τονικότητας δίνεται στο γνωσιακό μοντέλο του Large και συνεργατών (2016), που συνδυάζουν την προσέγγιση κατανομών με την προσέγγιση δυναμικών συστημάτων. Το μοντέλο τους εφαρμόστηκε στην εργασία εύρεσης τονικότητας raga ινδικής κλασικής μουσικής και βρέθηκε σημαντική συσχέτιση μεταξύ των προβλέψεων του μοντέλου και των κρίσεων ανθρώπων με και χωρίς γνώση του μουσικού είδους. Τέλος, το γνωσιακό μοντέλο του Milne και συνεργατών (2015) βασίζεται στην υπόθεση ότι το φάσμα ενός ήχου είναι καθοριστικό για το ανθρώπινο ακουστικό σύστημα και εξηγεί τα ευρήματα των πειραμάτων των Krumbhansl και Kessler (1982). Επίσης, κάνει προβλέψεις σχετικά με την αντίληψη τονικότητας και τονικών τριάδων για άλλους τρόπους του ισοσυγκερασμένου συστήματος πέρα από την μείζονα και ελάσσονα κλίμακα, καθώς και μικροτονικών κλιμάκων.

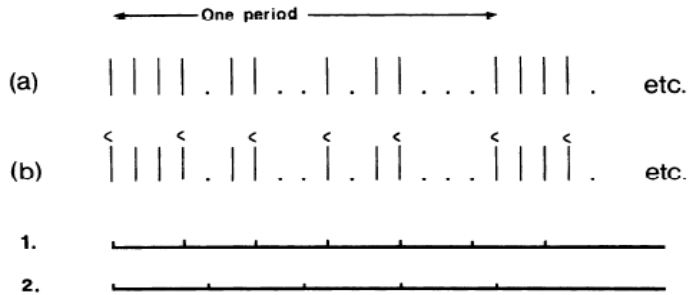
Εύρεση μέτρου

Το μέτρο αποτελεί και αυτό θεμελιώδες συστατικό της μουσικής κατανόησης, λόγω και της επίδρασης που έχει σε άλλα μέρη της μουσικής εμπειρίας (π.χ. χρονική προσδοκία), και μελετήθηκε κι αυτό νωρίς στην ιστορία της υπολογιστικής και γνωσιακής προσέγγισης στη μουσικολογία. Το μέτρο αποτελείται από χτύπους, σημεία στον χρόνο, καθένας από τους οποίους λαμβάνει υποκειμενικό τονισμό στο μυαλό του ακροατή (που συχνά δεν συνάδει με συμβάντα στην ίδια τη μουσική). Οι διαφορετικοί χτύποι συσχετίζονται ιεραρχικά και ένα ενδιάμεσο επίπεδο χτύπου αναπαριστά τον βασικό παλμό (tactus) ενός κομματιού. Τα μοντέλα εύρεσης μέτρου (meter-finding) παράγουν είτε μόνο τον παλμό είτε και τον παλμό και επιπρόσθετα χαμηλότερα (με μικρότερες διάρκειες) ή υψηλότερα (με μεγαλύτερες διάρκειες) επίπεδα.

Η μελέτη των Longuet-Higgins και Steedman (1971) περιέχει, εκτός από το μοντέλο εύρεσης τονικότητας που παρουσιάστηκε παραπάνω, και ένα μοντέλο εύρεσης μέτρου. Όπως το μοντέλο εύρεσης τονικότητας, έτσι και αυτό το μοντέλο δέχεται μονοφωνική είσοδο (φούγκα του καλώς συγκερασμένου πληκτροφόρου) την οποία επεξεργάζεται σειριακά, ξεκινώντας από την αρχή. Γίνεται η υπόθεση πως η πρώτη νότα του κομματιού ορίζει ένα μετρικό επίπεδο και κάθε επόμενη μεγαλύτερης διάρκειας νότα ορίζει ένα υψηλότερο επίπεδο. Γίνεται επίσης η υπόθεση της «συμφωνίας» (congruence), ομοίως με το μοντέλο εύρεσης τονικότητας, που αξιώνει ότι δεν θα υπάρξει γεγονός που συγκρούεται με τη σωστή δομή, προτού η δομή αυτή έχει εδραιωθεί. Παρά την απλότητά του, το μοντέλο πάσχει τόσο από υπερ-εφαρμογή, όσο και από το πρόβλημα ότι δεν μπορεί να παραγάγει μετρικά επίπεδα χαμηλότερα από το επίπεδο της πρώτης νότας.

Το μοντέλο εύρεσης μέτρου των Povel και Essens (1985) ακολουθεί μία πολύ διαφορετική προσέγγιση στο πρόβλημα: λαμβάνει υπόψη όλα τα δυνατά «ρολόγια» (όπου ρολόι είναι ένα επίπεδο χτύπων που υπερτίθεται στη μελωδία υπό εξέταση) και επιλέγει ένα μόνο από αυτά – εκείνο που κρίνεται καλύτερο. Πιο συγκεκριμένα, ένα ρολόι αξιολογείται με βάση την ευθυγράμμιση των χτύπων του με γεγονότα της μελωδίας εισόδου. Το μέγεθος που υπολογίζεται είναι πόσοι χτύποι ενός ρολογιού αποτυγχάνουν να ευθυγραμμιστούν με γεγονότα της εισόδου, δηλαδή περιλαμβάνουν χτύπους που συμπίπτουν με ασθενή γεγονότα της μελωδίας ή με απουσία γεγονότων, και για κάθε αστοχία του επιβάλλεται η αντίστοιχη ποινή. Αφού υπολογιστεί η συνολική ποινή για κάθε δυνατό ρολόι επιλέγεται εκείνο με τη χαμηλότερη συνολική ποινή (Εικόνα 2).

Και τα δύο μοντέλα εύρεσης μέτρου έχουν τον εξής περιορισμό: υποθέτουν ότι η είσοδός τους αποτελείται από νότες με διάρκειες που είναι ακριβή πολλαπλάσια μίας μικρής μετρικής μονάδας - δηλαδή ότι είναι κβαντισμένη (quantized input). Στην πραγματικότητα όμως η ανθρώπινη εκτέλεση της μουσικής δεν είναι απολύτως ακριβής ρυθμικά, συχνά εκούσια, για ερμηνευτικούς σκοπούς. Ένα μοντέλο εύρεσης μέτρου που ανέχεται ρυθμικές αποκλίσεις είναι αυτό του Rosenthal (1992). Το μοντέλο εντοπίζει ζεύγη έναρξης νοτών, ρυθμικά γειτονικών ή μη, σε μία μελωδία και χρησιμοποιεί το χρονικό διάστημα μεταξύ καθενός από τα ζεύγη αυτά ως ένα μετρικό επίπεδο. Στη συνέχεια ελέγχει τι συμβαίνει στη μελωδία στον επόμενο χτύπο



Εικόνα 2. Ένα μετρικό πρότυπο (*pattern*), όπως παρουσιάζεται (a) και όπως πιθανώς γίνεται αντιληπτό. (b) Δύο ρολόγια, το (1) που υπονοείται έντονα από την αρχή του προτύπου, αλλά το οποίο στη συνέχεια αποδεικνύεται ανεπαρκές και το (2), που υπονοείται λίγο στην αρχή του προτύπου το οποίο όμως είναι πιο επαρκές (Povel & Essens, 1985, σελ. 417. Αναδημοσιεύεται με την άδεια του εκδότη).

του μετρικού επιπέδου που παρήχθη, αλλά χρησιμοποιώντας ένα χρονικό «παράθυρο» γύρω από την ακριβή τιμή του χτύπου, ώστε αν υπάρχει επόμενη νότα που ακολουθεί αυτό το μετρικό επίπεδο που δεν είναι τοποθετημένη ακριβώς στο χτύπο αλλά βρίσκεται εντός του παραθύρου να γίνει αποδεκτή ως συνέχεια του μετρικού επιπέδου. Έτσι παράγονται πολλά μετρικά επίπεδα που ομαδοποιούνται σε «οικογένειες» επιπέδων που συσχετίζονται με απλούς λόγους.

Μία ακόμα προσέγγιση στο πρόβλημα εύρεσης μέτρου αποτελεί η χρήση πιθανοτήτων και του κανόνα του Bayes. Υπό αυτό το πρίσμα η εύρεση μέτρου μετατρέπεται στο πρόβλημα εύρεσης της μετρικής δομής που μεγιστοποιεί την πιθανότητα της δομής αυτής δεδομένης της μουσικής επιφάνειας (όπου μουσική επιφάνεια είναι για παράδειγμα οι διάρκειες νοτών μίας μελωδίας). Πολλές έρευνες για εύρεση μέτρου έχουν εφαρμόσει την Bayesian προσέγγιση, όπως αυτές του Cemgil και συνεργατών (2000) που προτείνουν δύο μοντέλα: το ένα δέχεται κβαντισμένη είσοδο και παράγει πολλαπλά μετρικά επίπεδα, ενώ το άλλο δέχεται μη κβαντισμένη είσοδο και εξάγει ένα μετρικό επίπεδο. Ακόμη, το μοντέλο του Temperley (2007) παράγει τρία μετρικά επίπεδα από μη κβαντισμένη είσοδο.

Επιπλέον, η ερευνητική οδός των δυναμικών συστημάτων στην εύρεση μέτρου έχει επηρεάσει βαθιά τον κλάδο, κυρίως μέσω της εργασίας των Large, Jones και συνεργατών (1989, 1999, 2019). Σύμφωνα με την οπτική αυτή, η αντίληψη του μέτρου υλοποιείται μέσω ταλαντωτών (*oscillators*) – μηχανισμών που παράγουν ο καθένας ένα περιοδικό κύμα. Όταν ένας ταλαντωτής εκτίθεται σε κάποια ρυθμική είσοδο (μουσικό έργο), προσαρμόζει την περίοδο και τη φάση του ώστε οι κορυφές του κύματος του ταλαντωτή να συμπίπτουν με γεγονότα της εισόδου. Αφού ένας ταλαντωτής τεθεί σε κίνηση συνεχίζει εκ φύσεως να κινείται και οι κορυφές του κύματος που παράγει αναπαριστούν την προσδοκία μελλοντικών γεγονότων, ενώ το ύψος των κορυφών (η απόστασή τους από τη θέση ισορροπίας) αντιστοιχεί στην βεβαιότητα του συστήματος για το προσδοκώμενο γεγονός - υψηλή βεβαιότητα σημαίνει μεγάλο ύψος και αντιστρόφως. Μπορούν να χρησιμοποιηθούν πολλοί ταλαντωτές ταυτόχρονα για την αναπαράσταση πολλών μετρικών επιπέδων.

Παρά τις ποικίλες προσεγγίσεις στο πρόβλημα της εύρεσης μέτρου υπάρχουν και πολλοί άλλοι παράγοντες που επηρεάζουν την αντίληψη του μέτρου, οι οποίοι δεν λαμβάνονται υπόψη στα μοντέλα που παρουσιάστηκαν παραπάνω και έχουν να κάνουν κατά κύριο λόγο με το τονικό ύψος (Prince et al., 2009 · Prince & Schmuckler, 2012). Παράδειγμα αποτελεί η χρήση επαναλαμβανόμενων μελωδικών προτύπων, τα οποία υπονοούν ότι το μέτρο του κομματιού είναι τέτοιο ώστε να τονίζεται πάντα η ίδια νότα του προτύπου καθώς και ο αρμονικός ρυθμός, δηλαδή ο ρυθμός εναλλαγής συγχορδιών (Hannon et al., 2004).

Ανίχνευση τονικού ύψους

Το ερώτημα του πώς το ανθρώπινο ακουστικό σύστημα μετατρέπει το συνεχές ηχητικό σήμα σε πρότυπο διακριτών νοτών είναι ένα από τα πιο θεμελιώδη στη μελέτη της μουσικής νόησης καθώς αποτελεί απαραίτητο βήμα για την εκτέλεση πολλών άλλων νοητικών μουσικών διεργασιών. Παρά το βασικό της ρόλο, η ανίχνευση τονικού ύψους (pitch identification/detection) αποτελεί δύσκολη εργασία που δυσχεραίνεται από το γεγονός ότι είναι δυνατόν ένας ακροατής να αναγνωρίσει τόνους των οποίων η συχνότητα δεν παρουσιάζεται καθόλου στο σήμα, φαινόμενο γνωστό ως απύσφα θεμέλιος/missing fundamental (Fletcher, 1924). Μία κατεύθυνση έρευνας ανίχνευσης τονικού ύψους έχει τις ρίζες της στην ακουστική ψυχολογία, την ψυχοακουστική και τη νευροεπιστήμη (de Cheveigné, 2005). Η κατεύθυνση της πληροφορικής από την άλλη ονομάζει το πρόβλημα «αυτόματη μεταγραφή» (automatic transcription) και τα σχετικά μοντέλα προορίζονται κυρίως για πρακτικούς σκοπούς που ανήκουν στον κλάδο ανάκτησης μουσικής πληροφορίας (MIR - music information retrieval, Klapuri, 2004).

Μία βασική προσέγγιση των μοντέλων ανίχνευσης τονικού ύψους είναι η επεξεργασία του φάσματος της κυματομορφής του ηχητικού σήματος, που αναπαριστά το επίπεδο ενέργειας των συχνοτήτων που συνθέτουν την κυματομορφή. Μάλιστα, ένα από τα πρώτα μοντέλα αυτής της οικογένειας κατάφερε να αποτυπώσει το φαινόμενο της απύσφας θεμελίου, μελετώντας το πρότυπο των κορυφών του φάσματος και εντοπίζοντας τη συχνότητα που οι αρμονικοί της συμπίπτουν με περισσότερες από τις κορυφές αυτές σε σχέση με κάθε άλλη συχνότητα (Terhardt, 1974).

Μία εναλλακτική προσέγγιση είναι η επεξεργασία της ίδιας της κυματομορφής του ηχητικού σήματος. Σκοπός είναι να εντοπιστεί η συχνότητα που παρουσιάζει την ισχυρότερη περιοδικότητα σε κάθε (μικρό) χρονικό διάστημα, που θεωρείται η θεμέλιος (fundamental). Ο εντοπισμός της θεμελίου μπορεί να γίνει χρησιμοποιώντας αυτό-συσχέτιση, ένα μαθηματικό μέγεθος που εκφράζει την ομοιότητα ενός σήματος με τον εαυτό του σε διαφορετικά χρονικά διαστήματα, όπως στον αλγόριθμο ανίχνευσης τονικού ύψους YIN (de Cheveigné & Kawahara, 2002). Μία άλλη μέθοδος αποτελεί η επεξεργασία της περιβάλλουσας πλάτους του σήματος (Tolonen & Karjalainen, 2000). Σχετικά με τη γνωσιακή ευλογοφάνεια των παραπάνω μεθόδων δεν υπάρχει ακόμη γενική συναίνεση, καθώς η λειτουργία του βιολογικού υποστρώματος είναι ακόμα υπό μελέτη. Συγκεκριμένα, το ερώτημα του αν ο εγκέφαλος χρησιμοποιεί εξειδικευμένες δομές για την ακουστική επεξεργα-

σία, όπως νευρωνικούς ταλαντωτές (Doelling et al., 2019), ή αν μετατρέπει το ακουστικό ερέθισμα σε μία ακολουθία αποκρίσεων σε αυτό (Shamma, 2001) είναι ανοιχτό.

Η εργασία δυσκολεύει περισσότερο όταν επεξεργαζόμαστε πολυφωνική μουσική, και ουσιαστικά ανάγεται στο πρόβλημα ομαδοποίησης αρμονικών της κάθε θεμελιώδους συχνότητας με τον σωστό τρόπο. Σε αυτό βοηθάει το γεγονός ότι οι αρμονικοί της ίδιας νότας γενικά αλλάζουν με παρόμοιους τρόπους (π.χ. ίδια στιγμή έναρξης και τέλους, ίδιες αλλαγές δυναμικής, ίδιο βιμπράτο κλπ.) και οδηγεί σε συστήματα που χρησιμοποιούν σχετικές ευρετικές (Rosenthal & Okuno, 1998). Ο Benetos και συνεργάτες (2018) δίνουν μία επισκόπηση πιο πρόσφατων μοντέλων αυτόματης μεταγραφής για πολυφωνική μουσική.

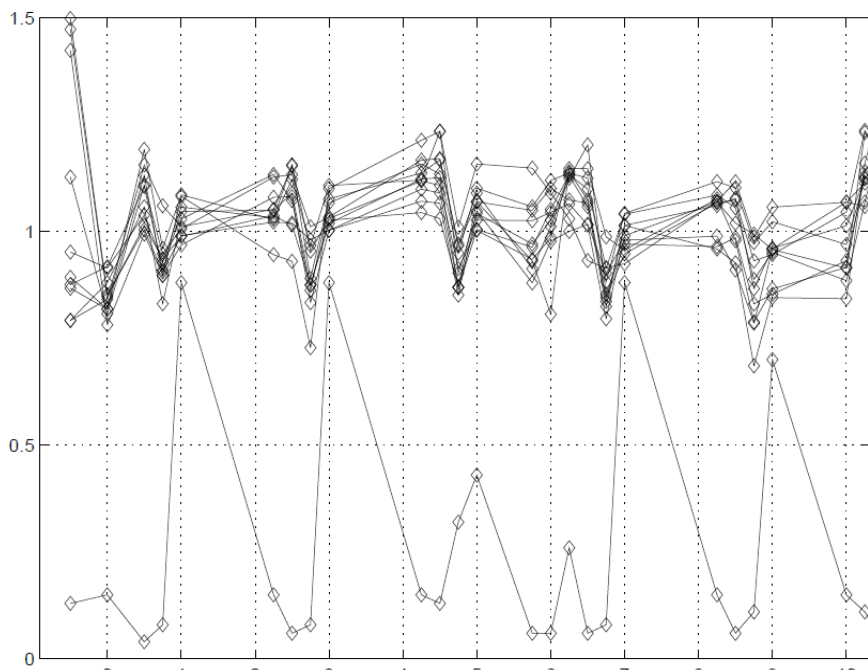
Ομαδοποίηση

Κατά την ακρόαση μίας μελωδίας, οι άνθρωποι συνήθως ομαδοποιούν τις νότες σε σύντομα χρονικά τεμάχια (chunks) που αντιστοιχούν σε πρότυπα και φράσεις, καθώς και σε μεγαλύτερες ενότητες που αντιστοιχούν σε περιόδους και τμήματα. Ένα απλό αλλά αποδοτικό μοντέλο ομαδοποίησης (grouping) μελωδίας είναι αυτών των Tenney και Polansky (1980), το οποίο βασίζεται στην ιδέα ότι τα όρια φράσεων τείνουν να συμπίπτουν με σημαντικές αλλαγές, όχι μόνο στα μελωδικά διαστήματα αλλά και σε άλλες μουσικές διαστάσεις - μεγάλα χρονικά διαστήματα και μεγάλες διαφορές δυναμικής ή/και χροιάς. Αυτός ο τρόπος ομαδοποίησης σχετίζεται και με τη Gestalt ψυχολογία, οπότε το μοντέλο έχει βάση σε γνωσιακές αρχές. Ο τρόπος λειτουργίας του μοντέλου είναι ο εξής: σε κάθε ζεύγος γειτονικών νοτών μίας μελωδίας ανατίθεται μία τιμή που αντιστοιχεί στην «απόσταση» των νοτών και που υπολογίζεται ως σταθμισμένο άθροισμα των παραπάνω τεσσάρων διαστάσεων. Στη συνέχεια, τα ζεύγη που αποτελούν τοπικά μέγιστα, δηλαδή έχουν τιμή «απόστασης» μεγαλύτερη από τα ζεύγη που έρχονται πριν και μετά από αυτό, θεωρούνται όρια τεμαχίων και χρησιμοποιούνται κατόπιν στο σχηματισμό τμημάτων μεγαλύτερου επιπέδου, με βάση τον ίδιο μηχανισμό.

Το μοντέλο Local Boundary Detection του Καμπουρόπουλου (2001) ακολουθεί παρόμοια προσέγγιση χρησιμοποιώντας ως δεδομένα τα διαστήματα μεταξύ των νοτών μίας μελωδίας, τη χρονική τους απόσταση και τις παύσεις. Επίσης, το μοντέλο έχει επεκταθεί ώστε να λαμβάνει υπόψη του μελωδικές επαναλήψεις, που συχνά τις αντιλαμβάνομαστε ως όρια φράσεων (Cambouropoulos, 2006) (Εικόνα 3).

Η *Γενετική Θεωρία της Τονικής Μουσικής / Generative Theory of Tonal Music* ή πιο απλά GTTM (βλ. Τσούγκρα και Δρακουλάκη στον ίδιο τόμο) των Lerdahl και Jackendoff (1983) παρουσιάζει διάφορα κριτήρια για ομαδοποίηση τα οποία έχουν υλοποιηθεί, μεταξύ άλλων μοντέλων, και στο υπολογιστικό μοντέλο των Frankland & Cohen (2004). Οι συγγραφείς συγκρίνουν τις προβλέψεις του μοντέλου τους με αποκρίσεις ανθρώπων σε σειρά ψυχολογικών πειραμάτων, οπότε θεμελιώνουν και τη θεωρία GTTM ως μια γνωσιακή θεωρία, που πιθανώς αποτυπώνει διεργασίες του νου. Ακόμα, το μοντέλο του Bod (2002) κάνει χρήση στατιστικής πληροφορίας για

τον αριθμό φράσεων ανά μελωδία, νοτών ανά φράση, κ.ά., ενός μουσικού σώματος, ώστε να παραγάγει αναπαράσταση μιας μελωδίας σε μορφή δέντρου το οποίο χωρίζεται σε φράσεις και ακολούθως σε νότες.



Εικόνα 3. Τα πρώτα μέτρα της σπουδής νούμερο 3, έργο 10 του F. Chopin. Η χαμηλότερη καμπύλη αναπαριστά το προφίλ ισχύος ορίου που υπολογίζει το μοντέλο Local Boundary Detection του Καμπουρόπουλου, όπου οι κορυφές της υποδηλώνουν πιθανά τοπικά όρια (τέλη προτύπων ή φράσεων). Οι υπόλοιπες καμπύλες αναπαριστούν το χρονισμό διαφορετικών εκτελέσεων του κομματιού (όσο χαμηλότερα τόσο πιο αργή εκτέλεση) (Cambourououlos, 2001, σελ. 7).

Όλα τα παραπάνω μοντέλα δέχονται μόνο μελωδίες ως είσοδο, όμως η περισσότερη μουσική που ακούμε σήμερα είναι πολυφωνική. Πειραματικές μελέτες δείχνουν ωστόσο ότι η ομαδοποίηση της πολυφωνικής μουσικής είναι στενά συσχετισμένη με αυτή της μονοφωνικής μουσικής, πιθανώς επειδή κατά την ακρόαση πολυφωνίας, εξάγουμε πρώτα την κύρια μελωδία, εκτελούμε ομαδοποίηση σε αυτή και μετά εφαρμόζουμε την ίδια ομαδοποίηση σε όλο το κομμάτι (Bruderer et al., 2010). Μία άλλη πρόκληση για τα υπολογιστικά μοντέλα ομαδοποίησης είναι η αντιμετώπιση των υψηλότερων επιπέδων οργάνωσης: τα περισσότερα από αυτά ασχολούνται με το επίπεδο νοτών και φράσεων, αγνοώντας το επίπεδο των μεγαλύτερων ενοτήτων, παρόλο που νευροεπιστημονικές έρευνες έχουν επιδείξει ότι κατά την ακρόαση

οι άνθρωποι οργανώνουν τη μουσική ιεραρχικά (Koelsch et al., 2013). Πιθανή λύση σε αυτό το πρόβλημα δίνουν οι μέθοδοι ιεραρχικής συσταδοποίησης (hierarchical clustering), όπως αυτές στα μοντέλα των McFee και Ellis (2014a, 2014b) και του de Berardinis και συνεργατών (2020).

Διαχωρισμός φωνών

Ένα συγγενές με την ομαδοποίηση πρόβλημα που προκύπτει σε πολυφωνική μουσική είναι αυτό του διαχωρισμού φωνών (voice separation/stream segregation), δηλαδή του χωρισμού των νοτών ενός πολυφωνικού κομματιού σε διακριτές ροές (streams) ή φωνές. Μία ιδιαίτερη προσέγγιση είναι αυτή του Gjerdingen (1994) που στηρίζεται στο φαινόμενο εμφανούς κίνησης (apparent motion effect). Αυτό έγκειται στην αντίληψη κίνησης που δεν υπάρχει στην πραγματικότητα, και έχει μελετηθεί εκτενώς στην Gestalt ψυχολογία (Wertheimer, 1912/2012). Στη μουσική, το φαινόμενο παρουσιάζεται για παράδειγμα στην αντίληψη της κίνησης μίας μελωδίας από μία νότα σε μία άλλη ως συνεχόμενη ακόμα και όταν παίζεται από όργανα που δεν μπορούν να παραγάγουν συχνότητες ανάμεσα στις συχνότητες των δύο νοτών, όπως το πιάνο. Το μοντέλο διαχωρισμού φωνών του Gjerdingen, που έχει ως βάση το μοντέλο αντίληψης κίνησης στην όραση των Grossberg και Rudd (1989), τοποθετεί τις νότες ενός κομματιού σε ένα δισδιάστατο διάγραμμα συχνότητας-χρόνου και δημιουργεί “ίχνη κίνησης” που αντιστοιχούν σε μουσικές ροές ή φωνές, και μπορεί να μοντελοποιήσει μουσικά φαινόμενα όπως τρίλλιες και παράλληλες φωνές.

Ένα άλλο μοντέλο διαχωρισμού φωνών είναι αυτό των Kirilin και Utgoff (2005), το οποίο λαμβάνει αποφάσεις για το αν δύο νότες ανήκουν στην ίδια φωνή βασισμένο στην τονική τους εγγύτητα, τη χρονική τους απόσταση κ.ά. Ο Temperley (2001) προτείνει ένα μοντέλο στο οποίο λαμβάνονται υπόψη πολλές αναλύσεις και επιλέγεται αυτή που έχει σχετικά λίγες φωνές, λίγα μεγάλα πηδήματα και λίγες μεγάλες παύσεις εντός των φωνών. Ακόμη, το μοντέλο διαχωρισμού φωνών του Καμπουρόπουλου (2008) βασίζεται σε κάποιες από τις αντιληπτικές αρχές που ο μουσικός ψυχολόγος Huron (2001) εντόπισε ως πηγή των κανόνων καλής κίνησης μεμονωμένων φωνών (voice leading).

Αρμονική ανάλυση

Υπολογιστικά μοντέλα έχουν αναπτυχθεί και για τις διεργασίες σχετικές με την αντίληψη αρμονίας σε πολυφωνική μουσική, που στη βιβλιογραφία ονομάζονται μοντέλα αρμονικής ανάλυσης (harmonic analysis). Οι βασικές πληροφορίες που πρέπει να εξάγει ένα τέτοιο μοντέλο είναι η βάση της κάθε συγχορδίας, η ποιότητα της (αν είναι μείζονα ή ελάσσονα, μεθ’ εβδόμης κ.λπ.) και η βαθμίδα της εντός της τονικότητας. Οι δυσκολίες στην εργασία αυτή είναι ο εντοπισμός σημείων μετατροπιών, η αντιμετώπιση τόνων εκτός συγχορδίας, ημιτελών και εννοούμενων συγχορδιών, και η διασύνδεσή τους με το μέτρο και άλλες μουσικές διαστάσεις. Ένα μοντέλο αρμο-

νικής ανάλυσης που δημιουργήθηκε νωρίς στην ιστορία του κλάδου είναι αυτό του Winograd (1968), το οποίο χρησιμοποιεί περίπλοκους κανόνες και πολύ μεγάλη ποσότητα ενσωματωμένης γνώσης σχετικά με την αρμονία.

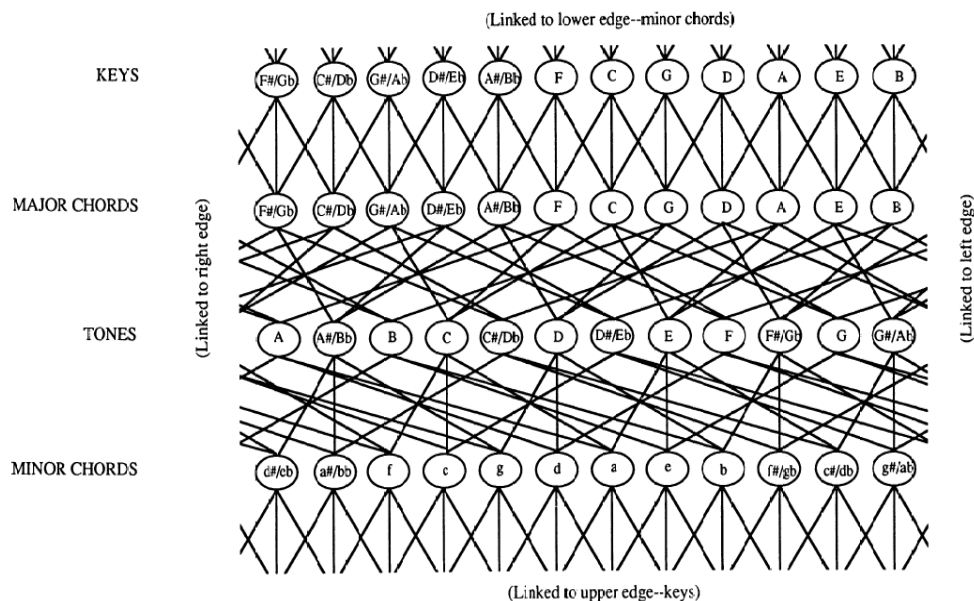
Το μοντέλο των Temperley και Sleator (1999) αξιολογεί πιθανές αναλύσεις σύμφωνα με τη συμβατότητα κάθε τόνου με την τρέχουσα βάση (λαμβάνοντας υπόψη περαστικούς τόνους), την ευθυγράμμιση των αρμονικών αλλαγών με ισχυρούς χτύπους και την απόσταση γειτονικών συγχορδιών στον κύκλο των 5ών. Ένα άλλο μοντέλο το οποίο κάνει χρήση Bayesian πιθανοτήτων (βλ. ενότητα “Εύρεση μέτρου”) είναι αυτό των Raphael και Stoddard (2004), που εντοπίζει συγχρόνως και την τονικότητα ενός κομματιού, επιδεικνύοντας έτσι το συσχετισμό μεταξύ της εύρεσης τονικότητας και της αρμονικής ανάλυσης.

Το πρόβλημα της αρμονικής ανάλυσης έχει προσεγγιστεί και με συνδεδετιστικές⁴ (connectionist) μεθόδους. Το ιστορικό συνδεδετιστικό μοντέλο αρμονικής ανάλυσης MUSACT του Bharucha (1987) αποτελείται από τρία επίπεδα κόμβων που αναπαριστούν τόνους, συγχορδίες και τονικότητες (Εικόνα 4). Ο τρόπος λειτουργίας του είναι ο εξής: κατά την σειριακή επεξεργασία ενός κομματιού εισόδου, οι νότες του κομματιού ανά πάσα στιγμή ενεργοποιούν τους αντίστοιχους κόμβους τόνων, που ενεργοποιούν κόμβους συγχορδιών που περιέχουν αυτούς τους τόνους, και αυτοί οι κόμβοι συγχορδιών με τη σειρά τους ενεργοποιούν κόμβους τονικότητων στις οποίες ανήκουν οι συγχορδίες. Αρχικά δηλαδή η ενεργοποίηση εξαπλώνεται ανωφερώς (bottom-up), στη συνέχεια όμως εξαπλώνεται και κατωφερώς (top-down), καθώς οι ενεργοί κόμβοι τονικότητων ενεργοποιούν σε μικρότερο βαθμό και κόμβους συγχορδιών που περιλαμβάνονται στην τονικότητα αλλά δεν έχουν ακουστεί, και ομοίως οι κόμβοι συγχορδιών για κόμβους τόνων που ανήκουν στη συγχορδία αλλά δεν έχουν ακουστεί. Κατ’ αυτό τον τρόπο το μοντέλο εξηγεί το φαινόμενο της αρμονικής προέγερσης (harmonic priming), δηλαδή της αυξημένης ευαισθησίας ακροατών για συγχορδίες που συνδέονται με το αρμονικό περιεχόμενο (προηγούμενες συγχορδίες).

Η ισχύς ενεργοποίησης του κάθε κόμβου, στον βαθμό που οφείλεται σε ενεργοποίηση γειτονικού του κόμβου, καθορίζεται από βάρη (αριθμητικές τιμές) στην ακμή που ενώνει τους δύο κόμβους. Στην αρχική μελέτη δοκιμάστηκαν πολλές διαφορετικές τιμές για τα βάρη αυτά, ενώ σε μετέπειτα μελέτη χρησιμοποιήθηκε η τεχνική αυτό-οργανωμένου χάρτη (self-organizing map⁵) ώστε το ίδιο το μοντέλο να μάθει τα κατάλληλα βάρη μεταξύ νοτών/συγχορδιών/τονικότητων από παραδείγματα εκπαίδευσης (Tillmann et al., 2000).

⁴ Ο συνδεδετισμός (connectionism) είναι προσέγγιση της γνωσιακής επιστήμης και μέθοδος της τεχνητής νοημοσύνης που χρησιμοποιεί τεχνητά νευρωνικά δίκτυα, εμπνευσμένα από τα βιολογικά, τα οποία αποτελούνται από κόμβους-μονάδες (νευρώνες) που συνδέονται μεταξύ τους με ακμές, είναι οργανωμένοι σε επίπεδα (layers) και ανταλλάσσουν σήματα.

⁵ Υπολογιστική μέθοδος για την ανάλυση δεδομένων, που παρουσιάστηκε από τον Kohonen (1982). Σχετίζεται με το συνδεδετισμό και τη γνωσιακή επιστήμη και αποσκοπεί στη μοντελοποίηση της χωρικής οργάνωσης των λειτουργιών του φλοιού του εγκεφάλου (Kohonen, 2013).



Εικόνα 4. Το δίκτυο MUSACT αναπαριστά σχέσεις μεταξύ τόνων, συγχορδιών και τονικοτήτων. Συνδέσεις μεταξύ κόμβων δηλώνουν ότι ένας τόνος ανήκει σε μία συγχορδία ή ότι μία συγχορδία ανήκει σε μία τονικότητα. Κάποιο μουσικό περιεχόμενο ενεργοποιεί κόμβους τόνων και η ενεργοποίηση εξαπλώνεται στο δίκτυο μέχρι να επιτευχθεί κατάσταση ισορροπίας. Το πρότυπο ενεργοποίησης στην κατάσταση ισορροπίας αναπαριστά τις προσδοκίες συγχορδιών και τις υπονοούμενες τονικότητες και επηρεάζει το βαθμό συμφωνίας/διαφωνίας και αναγνώρισης των μουσικών γεγονότων που ακολουθούν (Bharucha, 1987, σελ. 16. Αναδημοσιεύεται με την άδεια του εκδότη).

Ονομασία τόνων / τονικών υψών

Το ζήτημα της ονομασίας τόνων (pitch spelling) έγκειται στην ανάθεση μίας από τις επτά συλλαβές νοτών της δυτικής μουσικής κλίμακας ακολουθούμενης από καμία, μία ή δύο υφέσεις ή διέσεις (π.χ. «μι ύφεση» ή «σι διπλή διέση») σε κάθε τόνο που εντοπίζεται σε ένα κομμάτι. Λόγω της εναρμόνιας ισοδυναμίας του δωδεκάφθογγου συστήματος χορδίσματος, είναι δυνατό να ονομάσουμε ένα τόνο με πολλά διαφορετικά τέτοια ονόματα, ωστόσο στη μουσική πράξη φαίνεται πως οι έμπειροι μουσικοί κάνουν συγκεκριμένες επιλογές ονομασίας. Πιθανοί νοητικοί μηχανισμοί με τους οποίους επιτυγχάνεται η ονομασία τόνων έχουν υλοποιηθεί ήδη στο μοντέλο εύρεσης τονικότητας των Longuet-Higgins και Steedman, αλλά και σε υπολογιστικά μοντέλα που προτάθηκαν μετέπειτα (Temperley, 2001 · Cambouropoulos, 2003 · Chew & Chen, 2005 · Meredith, 2006 · Honingh, 2009). Μια βασική αρχή αυτών των συστημάτων είναι η προτίμηση ονομάτων που εντοπίζονται πιο συμπαγώς σε κάποιο χώρο τόνων, συνήθως τη γραμμή των πεμπτών (ξεδιπλωμένος κύκλος των πεμπτών). Άλλο κριτήριο ονομασίας τόνων αποτελεί η κίνηση μεμονωμένων φωνών (voice leading),

δηλαδή η ονομασία τόνων με βάση το που οδηγούν στη φωνή όπου εμφανίζονται (π.χ. ρε δίεση αν οδηγεί σε μι και μι ύφεση αν οδηγεί σε ρε). Ακόμη, ένα ερώτημα που προκύπτει είναι το εξής: εντοπίζονται πρώτα τα ονόματα νοτών, και στη συνέχεια χρησιμοποιούνται για τον εντοπισμό της αρμονίας και της τονικότητας ή προκύπτουν αφού η τονικότητα και η αρμονία έχουν αποσαφηνιστεί; (Temperley, 2012). Τα παραπάνω μοντέλα δίνουν το καθένα διαφορετική απάντηση στο ερώτημα αυτό.

Τέλος, οι Micchi και συν. (2020) παρουσίασαν πρόσφατα ένα περίπλοκο νευρωνικό δίκτυο το οποίο εκπαίδευσε σε σώμα μουσικής που περιλαμβάνει καταγραφή της βαθμίδας κάθε συγχορδίας για να επιτύχει ταυτόχρονα αρμονική ανάλυση και ονομασία τόνων. Ωστόσο, τόσο αυτό το μοντέλο όσο και γενικότερα τα μεγάλα νευρωνικά δίκτυα της τελευταίας δεκαετίας, πάσχουν από το εξής πρόβλημα: είναι δύσκολο να αναπαραστήσουμε, με την τρέχουσα κατανόησή μας, τη γνώση που αποκτούν κατά την εκπαίδευσή τους σε υψηλό επίπεδο, δηλαδή με όρους αντιληπτούς σε ανθρώπους (με άλλα λόγια αποτελούν συχνά μαύρα κουτιά/black boxes). Επομένως, παρά τη νευρολογική έμπνευση των νευρωνικών δικτύων, είναι κάπως επίφοβο να τα χαρακτηρίσουμε νοητικά μοντέλα εφόσον δεν καταλαβαίνουμε τους εσωτερικούς υπολογισμούς τους – παρόλο που συχνά είναι τα πρώτα που κατηγοριοποιούνται ως γνωστικά μοντέλα.

Ανακάλυψη προτύπων

Ανακάλυψη προτύπων (pattern discovery) είναι η εργασία εντοπισμού επαναλαμβανόμενων μοτίβων (ή προτύπων) σε ένα ή περισσότερα μουσικά έργα. Το ποια ακριβώς μπορούν να είναι τα μοτίβα αυτά εξαρτάται από υποκειμενικούς παράγοντες σε μεγαλύτερο βαθμό από άλλα κομμάτια της μουσικής αντίληψης, όπως η ανίχνευση τονικού ύψους ή η εύρεση μέτρου. Πολλές προσεγγίσεις στη μουσική ανάλυση (μοτιβική, παραδειγματική, και άλλες) σχετίζονται με την ανακάλυψη προτύπων. Στην υπολογιστική έρευνα έχουν προταθεί μοντέλα ανακάλυψης προτύπων για μελωδίες, όπως αυτά των Rolland (1999), των Lartillot & Saint-James (2004), Conklin and Anagnostopoulou (2001), ενώ το μοντέλο των Meredith και συνεργατών (2002) εντοπίζει πρότυπα σε πολυφωνική μουσική.

Ένα ζήτημα που ανακύπτει στην ανακάλυψη προτύπων είναι πως συχνά η επανάληψη των προτύπων δεν είναι τέλεια, επομένως χρειάζεται να γίνονται αποδεκτές και κατά προσέγγιση αντιστοιχίσεις. Γι' αυτό είναι σημαντικό να χρησιμοποιηθεί η κατάλληλη αναπαράσταση γνώσης ώστε οι διαφορές και ομοιότητες μεταξύ προτύπων να μπορούν να τυποποιηθούν και να ποσοτικοποιηθούν. Επιπλέον, ο αριθμός προτύπων που εντοπίζονται με όλους τους παραπάνω αλγορίθμους σε ένα κομμάτι είναι μεγάλος, όμως μόνο λίγα από αυτά τα πρότυπα είναι όντως σημαντικά για την αντίληψη ή τη μουσική ανάλυση, άρα είναι απαραίτητη και η προσθήκη κάποιας μεθόδου επιλογής των σημαντικών προτύπων. Μία πιθανή λύση σε αυτό το ζήτημα δίνεται από την προσέγγιση πολλαπλών οπτικών γωνιών (multiple viewpoints), στην οποία χρησιμοποιείται πολυδιάστατη αναπαράσταση της μουσικής επιφάνειας η οποία περιλαμβάνει, εκτός από τις βασικές διαστάσεις τόνου και διάρκειας της κάθε

νότας, και διαστάσεις για τη βαθμίδα εντός της κλίμακας, μετρική τοποθέτηση, μελωδικό περίγραμμα (contour) καθώς και συνδυασμούς των παραπάνω (Conklin & Witten, 1995). Με βάση τις διαστάσεις αυτές γίνεται επιλογή των προτύπων που εμφανίζονται συχνότερα (με χρήση πιθανοτήτων και συγκεκριμένα μαρκοβιανών αλυσίδων). Μία επέκταση της αναπαράστασης πολλαπλών οπτικών γωνιών παρουσιάστηκε από τους Conklin και Anagnostopoulou (2006), με την προσθήκη των τμηματικών οπτικών γωνιών (segmental viewpoints). Οι τμηματικές οπτικές γωνίες, σε αντίθεση με τις μελωδικές οπτικές γωνίες που παρουσιάστηκαν παραπάνω, δεν αναφέρονται σε πρότυπα νοτών αλλά πρότυπα από μικρότερα πρότυπα, που με τη σειρά τους αποτελούνται από πρότυπα νοτών. Οι συγγραφείς προτείνουν επίσης έναν αλγόριθμο ανακάλυψης τμηματικών προτύπων σε μεγάλα μουσικά σώματα που βασίζεται σε δέντρα επιθεμάτων⁶ (suffix trees).

τ	$[\cdot]_{\tau}$	$[\tau]$	Derived from
st	start-time of event	{0, 1, 2, ...}	st
pitch	pitch, in {C \sharp 4, ..., G \sharp 5}	{60, ..., 79}	pitch
duration	quarter note, eighth note, etc.	{1, 2, 3, 4, 6, 8, 12, 16}	duration
keysig	1 sharp, 1 flat, etc.	{-4, ..., 4}	keysig
timesig	3/4 time, 4/4 time	{12, 16}	timesig
fermata	event under / not under fermata?	{T, F}	fermata
deltast	rest, no rest	{0, 4}	st
gis221	difference in start-time	{1, ..., 20}	st
posinbar	position of event in bar	{0, ..., 15}	st
fib	first / not first in bar	{T, F}	st
seqint	sequential melodic interval	Z	pitch
contour	rising, falling, static	{-1, 0, 1}	pitch
referent	referent of piece	{0, ..., 11}	keysig
intfref	vertical interval from referent	{0, ..., 11}	pitch
inscale	in / not in scale	{T, F}	pitch
intfib	interval from first event in bar	[seqint]	pitch
intfip	interval from first event in piece	[seqint]	pitch
intphbeg	interval from phrase beginning	[seqint]	pitch
thrbar	seqint at bars	[seqint] \times Z ⁺	pitch, st
lphrase	length of phrase	Z ⁺	fermata, st
thrph	seqint at phrases	[seqint] \times Z ⁺	pitch, st
thrqu	seqint at quarters	[seqint] \times Z ⁺	pitch, st

Εικόνα 5. Οι βασικές οπτικές γωνίες και κάποιες παράγωγες οπτικές γωνίες για την ανάλυση χορικών του J.S. Bach. Η πρώτη στήλη αναπαριστά το όνομα της οπτικής γωνίας, η δεύτερη σύντομη περιγραφή της, η τρίτη αναφέρει το σύνολο από το οποίο λαμβάνει τιμές και η τέταρτη το όνομα της οπτικής γωνίας από την οποία παράχθηκε (ή το ίδιο της όνομα, αν πρόκειται για βασικά οπτική γωνία) (Conklin & Witten, 1995, σελ. 20. Αναδημοσιεύεται με την άδεια του εκδότη).

⁶ Δενδρική δομή δεδομένων στην πληροφορική. Περιλαμβάνει όλα τα επιθέματα ενός κειμένου και επιτρέπει την ταχεία εκτέλεση πράξεων, όπως αναζήτηση υποκειμένων ή εύρεση του μεγαλύτερου κοινού υποκειμένου μεταξύ δύο κειμένων, με κόστος όμως την αυξημένη ανάγκη αποθηκευτικού χώρου (Weiner, 1973 · Ukkonen, 1995).



Viewpoint	Viewpoint sequence
Pitch	[72,71,69,67,69,71,72,74,72,71,69,71,72,96,97]
Duration	[24,24,24,24,24,24,24,24,24,24,24,12,12,48,24]
Contour	[-,-,=,+,+,+,+,-,-,+,+,-,-]
Pcint	[11,10,10,0,2,2,1,2,10,11,2,1,9,10]
Intref	[0,11,9,7,7,9,11,0,2,0,11,9,11,0,9,7]
Select(contour, new(contour))	[-,=,+,+,-,-]
Ratio(duration)	[1,1,1,1,1,1,1,1,1,1,1,1/2,1,4,1/2]
New(countour)	[1,0,0,1,1,0,0,0,1,0,0,1,0,1,0]
Beats	[4,4,8]
Shape	[descending, ascending, descending]
Lift(contour)	[[[-,-,-], [+,,+,,+], [-,-,-,+,,+,-,-]]]
Lift(select(contour, new(contour)))	[[[-], [+], [-,+,-]]]
Pair(pcint)	[[{0,1,2,3,4,5,6,7,8,9,10,11}, {0,2,7,9,11}]]
Set(intref)	[[{0,7,9,11}, {0,7,9,11}, {0,2,7,9,11}]]
Thread(pcint, last)	[5,7]
Thread(contour, highest)	[=,+]
Ratio(beats)	[1,2]
Link(beats, thread(pcint, last))	[<4,5>, <8,7>]

Εικόνα 6. Ακολουθίες οπτικών γωνιών για μελωδία χορικού του J.S.Bach (BWV 255). Το πάνω μέρος του πίνακα (μέχρι και το shape) παρουσιάζει μελωδικές οπτικές γωνίες ενώ το κάτω μέρος του πίνακα παρουσιάζει τμηματικές (segmental) οπτικές γωνίες (προσαρμοσμένη εκδοχή από Conklin & Anagnostopoulou, 2006).

Ομοιότητα

Συγγενές με την ανακάλυψη προτύπων είναι το θέμα της μουσικής ομοιότητας (music similarity). Η αντίληψη της μουσικής ομοιότητας είναι ένα καιρικό θέμα της γνωσιακής μουσικολογίας, μια που η μουσική, στην πλειοψηφία της κάνει χρήση της επανάληψης, της ομοιότητας, της παραλλαγής. Η Margulis στο βιβλίο της “On Repeat” (2014, σελ. 176) αναφέρει σχετικά ότι “... η επανάληψη ωθεί την προσοχή σε διαφορετικά επίπεδα και εκφάνσεις του ερεθίσματος: κάτω, προς πιο λεπτά, ανεπαίσθητα χαρακτηριστικά ή πάνω, προς μεγάλης κλίμακας, δομικά χαρακτηριστικά. Η παραλλαγή, κατά κάποιον τρόπο, μιμείται αυτή την υποκειμενική αντιληπτική

διεργασία. Παίρνει τις μεταβαλλόμενες ιδιότητες που συνήθως επιβάλλει ο ακροατής στη μουσική και τους αλλάζει τοποθέτηση μέσα στην ίδια τη μουσική.”

Μπορούμε να διακρίνουμε ομοιότητα σε όλες τις μουσικές διαστάσεις: μελωδική, αρμονική, ρυθμική, υφής, ηχοχρώματος, διάθεσης, φόρμας, κλπ. Εξ αυτών, κυρίως δύο σημασίες του όρου έχουν υλοποιηθεί σε υπολογιστικά μοντέλα, που αντικατοπτρίζουν το διαχωρισμό μεταξύ ομοιότητας μεταξύ έργων, *inter-opus*, και ομοιότητας εντός ενός μουσικού έργου, *intra-opus* (Delière, 2007).

Η πρώτη σημασία αντιστοιχεί στην υψηλού επιπέδου ομοιότητα και εντοπίζεται σε γενικά μουσικά χαρακτηριστικά όπως στίλ, διάθεση ή ρυθμός. Παράδειγμα υπολογιστικού μοντέλου που μελετά την υψηλού επιπέδου ομοιότητα είναι αυτό του Eerola και συνεργατών (2001), το οποίο βασίζεται σε κρίσεις ανθρώπων για την ομοιότητα μεταξύ παραδοσιακών μελωδιών. Το μοντέλο χρησιμοποιεί τη στατιστική τεχνική πολλαπλής παλινδρόμησης (*multiple regression*) ώστε να αξιολογήσει το ρόλο της κατανομής τόνων των παραδοσιακών μελωδιών στις ανθρώπινες κρίσεις ομοιότητας. Δύο από τους δημιουργούς του προηγούμενου μοντέλου δημιούργησαν αργότερα ένα άλλο μοντέλο με δυνατότητα μάθησης συμπιεσμένων αναπαραστάσεων μελωδιών, χρησιμοποιώντας διάφορα μουσικά χαρακτηριστικά, το οποίο υπολογίζει τη μουσική ομοιότητα μεταξύ μελωδιών συγκρίνοντας τις συμπιεσμένες αναπαραστάσεις τους (Toiviniainen & Eerola, 2002). Επίσης, το μοντέλο των Ens & Pasquier (2020) υπολογίζει την ομοιότητα μεταξύ ενός μουσικού κομματιού και του στίλ που καθορίζεται από ένα μουσικό σώμα κομματιών (σε αναπαράσταση MIDI). Έχει βρεθεί συσχέτιση μεταξύ των αποτελεσμάτων του μοντέλου και της κρίσης ακροατών για την ομοιότητα κομματιών σε στίλ χορικών Bach.

Η δεύτερη σημασία της μουσικής ομοιότητας αντιστοιχεί στο χαμηλό επίπεδο, δηλαδή σε ομοιότητα μεταξύ συγκεκριμένων μελωδικών και ρυθμικών προτύπων, συχνά εντός του ίδιου μουσικού έργου. Η σημασία αυτή προσεγγίζεται από το μοντέλο του Schmuckler (1999), που εφαρμόζει ανάλυση Fourier σε τονικά πρότυπα, εντοπίζοντας έτσι ομοιότητα μεταξύ μελωδικών περιγραμμάτων (*melodic contour*) σε διαφορετικές χρονικές κλίμακες. Ακόμη, οι Αναγνωστοπούλου & Westermann (1997) ερευνούν τη μελωδική ομοιότητα στα πλαίσια της παραδειγματικής μουσικής ανάλυσης (Nattiez & Barry, 1982). Το μοντέλο τους απαιτεί από τον χρήστη/μουσικό αναλυτή να προσδιορίσει ορισμένα χαρακτηριστικά, βάσει των οποίων θα γίνει κατηγοριοποίηση των μουσικών τμημάτων σε κατηγορίες από ένα αυτό-οργανωμένο νευρωνικό δίκτυο (*self-organizing neural network*). Σε παρόμοιο μήκος κύματος, οι Αναγνωστοπούλου και Καμπουρόπουλος (2017) παρουσιάζουν δύο υπολογιστικά μοντέλα που υπολογίζουν με διαφορετικούς τρόπους μελωδική ομοιότητα και προορίζονται επίσης για παραδειγματική ανάλυση. Το πρώτο εξ αυτών σχετίζεται με την ανακάλυψη προτύπων και το δεύτερο με την κατηγοριοποίηση τμημάτων της μουσικής επιφάνειας σε τάξεις.

Πέρα από τη νοητική μελέτη της μουσικής ομοιότητας, αυτή έχει μελετηθεί εκτενώς και στον τομέα ανάκτησης μουσικής πληροφορίας, όπου χρησιμεύει στην ανάπτυξη συστημάτων συστάσεων (*recommender systems*) καθώς και στην ανάπτυξη βάσεων δεδομένων που μπορεί κανείς να αναζητήσει μουσική τραγουδώντας ένα τμήμα μιας μελωδίας (*query-by-humming systems*). Παρά τη μη-γνωσιακή στο-

χευση των μοντέλων αυτών, οι αλγόριθμοι που χρησιμοποιούν μπορούν να μελετηθούν και σε γνωσιακό πλαίσιο (Müllensiefen & Frieler, 2004). Το μοντέλο του Tyrke (2003) και συνεργατών, για παράδειγμα, υπολογίζει την ομοιότητα μεταξύ δύο μελωδιών εντοπίζοντας τον αποδοτικότερο τρόπο να τροποποιηθεί η μία ώστε να προκύψει η άλλη, επιβάλλοντας ποινές για κάθε διαγραφή/προσθήκη, ενώ το μέγεθος των ποινών αυτών εξαρτάται και από τη σημασία της κάθε νότας. Ακόμη, το διαδραστικό σύστημα συστάσεων “Moodplay” βασίζεται στην ομοιότητα υψηλού επιπέδου διαθέσεων καλλιτεχνών (Andjelković et al., 2019). Τέλος, ο Bogdanov και συνεργάτες (2011) προτείνουν ένα μέτρο σύγκρισης μουσικής ομοιότητας που συνδυάζει τις υψηλού και χαμηλού επιπέδου σημασίες/εκφάνσεις της.

Προσδοκία

Η ιδέα ότι η προσδοκία παίζει σημαντικό ρόλο στη μουσική εμπειρία και τα συναισθήματα υπάρχει τουλάχιστον από τότε που ο Meyer εξέδωσε το βιβλίο του *Emotion and Meaning in Music* (1956). Πιο πρόσφατα, το βιβλίο *Sweet Anticipation* του Huron (2006) εξερευνά τη μουσική προσδοκία και τη σχέση της με την αισθητική από ψυχολογική σκοπιά και έχει επηρεάσει βαθιά τον τομέα της γνωσιακής μουσικολογίας.

Μία μουσική θεωρία προσδοκίας που αφορά συγκεκριμένα τη μελωδική προσδοκία είναι η θεωρία *υπαινιγμού-πραγματοποίησης* (Implication-Realization) του Narmour (1990). Ονομάζεται έτσι καθώς μοντελοποιεί τις περισσότερο ή λιγότερο προσδοκώμενες συνέχειες που υπαινίσσεται ένα μη καταληκτικό μελωδικό διάστημα, μία από τις οποίες πραγματοποιείται στο διάστημα που ακολουθεί. Ο Narmour κάνει διαχωρισμό μεταξύ έμφυτων και μαθημένων μελωδικών προσδοκιών. Οι έμφυτες προσδοκίες προκύπτουν ανωφερώς (bottom-up) από βιολογικούς περιορισμούς του ακουστικού μονοπατιού του εγκεφάλου, και αντιπροσωπεύονται από πέντε αρχές συνέχισης, βασισμένες στη Gestalt ψυχολογία. Αντίθετα, οι μαθημένες, κατωφερείς (top-down) προσδοκίες προκύπτουν από έκθεση σε μουσική και στατιστική επεξεργασία της μουσικής αυτής. Κατά την ακρόαση μπορεί να υπερισχύσουν οι μαθημένες προσδοκίες έναντι των έμφυτων ή αντίστροφα. Οι βασικές μελωδικές δομές της θεωρίας του Narmour αλληλεπιδρούν και συνδυάζονται με περίπλοκους τρόπους και σε πολλαπλά επίπεδα ιεραρχίας. Ένα τμήμα της θεωρίας ελέγχθηκε πειραματικά από τις Cuddy και Lunney (1995) και βρήκε σημαντική υποστήριξη, ανεξάρτητα από το επίπεδο μουσικής εκπαίδευσης των συμμετεχόντων. Το υπολογιστικό μοντέλο του Schellenberg (1997) υλοποιεί μία απλουστευμένη εκδοχή της θεωρίας, χωρίς να χάνει προβλεπτική ικανότητα, επιδεικνύοντας τον πιθανό πλεονασμό στη θεωρία του Narmour.

Η θεωρία υπαινιγμού-πραγματοποίησης αντιμετωπίζει τη μελωδική προσδοκία χωρίς να λαμβάνει υπόψη την επίδραση της αρμονίας, ωστόσο μεταγενέστερες προσεγγίσεις τη συσχετίζουν με το αρμονικό και τονικό πλαίσιο. Εμπνευσμένος από την σκέψη του γνωσιακού επιστήμονα Hofstadter για την αναλογία (2001), ο Larson δημιούργησε μία θεωρία που μοντελοποιεί τη μελωδική προσδοκία σε αναλογία με την

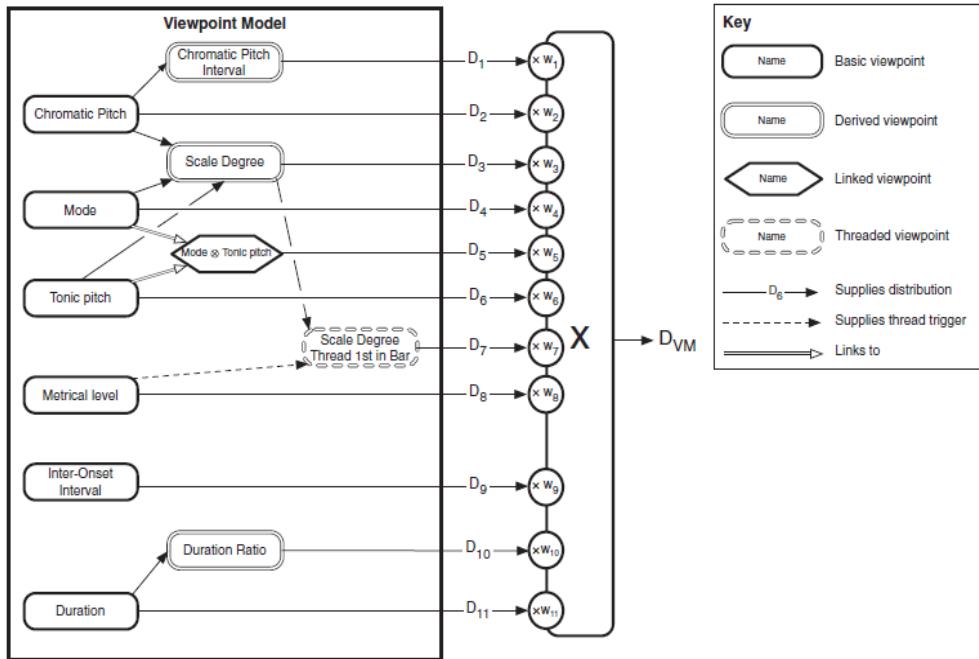
ενσώματη (embodied) αντίληψη μας για τις φυσικές δυνάμεις, την οποία υλοποίησε και σε υπολογιστικό μοντέλο (2004). Πιο συγκεκριμένα, η «βαρύτητα» ευνοεί μελωδική πτώση, η «αδράνεια» ευνοεί συνέχιση της τρέχουσας κατεύθυνσης (ανοδική ή καθοδική) και ο «μαγνητισμός» ευνοεί κίνηση προς τόνους που είναι πιο σταθεροί στην τρέχουσα συγχορδία (και μάλιστα παράγει προβλέψεις για πολλές επόμενες νότες μίας μελωδίας, καθεμία εκ των οποίων είναι όλο και πιο σταθερή). Το θεωρητικό μοντέλο προσδοκίας του Lerdahl, που στηρίζεται στη θεωρία τονικού χώρου⁷ (tonal pitch space, 1988), είναι παρόμοιο με αυτό του Larson και βασίζεται στην υπόθεση ότι η προσδοκία μίας μελωδικής συνέχειας, εντός ενός μελωδικού πλαισίου, είναι ανάλογη της αρμονικής σταθερότητάς της και αντιστρόφως ανάλογη του τετραγώνου της απόστασής της από την τρέχουσα νότα, σε αναλογία με τη δύναμη της βαρύτητας. Ακόμη, το θεωρητικό γνωσιακό μοντέλο της Margulis (2005) συνδυάζει κριτήρια μελωδικής κίνησης της θεωρίας του Narmour και τονικά κριτήρια της θεωρίας του Larson.

Ένα υπολογιστικό μοντέλο μελωδικής προσδοκίας που κάνει χρήση της αναπαράστασης γνώσης πολλαπλών οπτικών γωνιών (βλ. ενότητα “Ανακάλυψη προτύπων”) είναι το IDyOM (Information Dynamics of Music) των Pearce και Wiggins (2006, 2012), και βασίζεται σε μεγάλο βαθμό στην εργασία των Conklin και Witten (1995) (Εικόνα 6). Σε αντίθεση με τη θεωρία υπαινιγμού-πραγματοποίησης, το IDyOM μαθαίνει να παράγει μελωδικές προσδοκίες με στατιστικό τρόπο, μέσω έκθεσης σε μουσικά ερεθίσματα, και συγκεκριμένα χρησιμοποιώντας μαρκοβιανές αλυσίδες (ή αλλιώς n-grams). Επομένως δεν διαχωρίζει μεταξύ έμφυτων και μαθημένων προσδοκιών, αλλά όλες οι προσδοκίες του συστήματος προκύπτουν ανωφερώς από τη στατιστική επεξεργασία των δεδομένων (μουσικά σώματα). Κατά την επεξεργασία κάποιας συγκεκριμένης μελωδίας, μπορεί να χρησιμοποιηθεί το ήδη εκπαιδευμένο μοντέλο (που μοντελοποιεί τη μακρόχρονη μνήμη) ή να συνδυαστεί με ένα μοντέλο που εκτίθεται μόνο στην τρέχουσα μελωδία (που μοντελοποιεί τη βραχύχρονη μνήμη). Επίσης, υπάρχει η δυνατότητα επιλογής για το αν τα χρησιμοποιούμενα μοντέλα θα μαθαίνουν και από την τρέχουσα μελωδία ή θα είναι στατικά. Ένα ακόμη σημαντικό στοιχείο του IDyOM είναι ότι το ίδιο επιλέγει με μεθόδους βελτιστοποίησης ποια αναπαράσταση, δηλαδή ποιες οπτικές γωνίες θα χρησιμοποιήσει για να επιτύχει τη μεγαλύτερη συμπίεση των δεδομένων του, χαρακτηριστικό που βασίζεται στην υπόθεση ότι και ο νους συμπιέζει την πληροφορία που περιέχει με αποδοτικό τρόπο.

Ψυχολογικά πειράματα επέδειξαν ότι το IDyOM μοντελοποιεί με μεγαλύτερη ακρίβεια την ανθρώπινη μελωδική προσδοκία απ’ ότι η υλοποίηση της θεωρίας υπαινιγμού-πραγματοποίησης του Schellenberg (Pearce et al., 2010) και από άλλα μοντέλα μελωδικής προσδοκίας, αλλά και ότι αποτελεί ένα αρκετά καλό μοντέλο ομαδοποίησης. Το γεγονός αυτό είναι μη αναμενόμενο καθώς το IDyOM δεν αναπτύχθηκε ώστε να είναι μοντέλο ομαδοποίησης, και, όπως είδαμε στην ενότητα “Βασικές έννοιες και ζητήματα”, αποτελεί ένδειξη της ποιότητας του μοντέλου. Επιπλέον,

⁷ Τονικός χώρος = χωρική αναπαράσταση των νοτών που παρέχει ένα εύληπτο μέτρο της απόστασης μεταξύ συγχορδιών και τονικοτήτων.

το IDyOM έχει χρησιμοποιηθεί σε νευροβιολογικά πειράματα που έχουν δείξει ότι η προσδοκία του μοντέλου συσχετίζεται με τη μουσική μάθηση και δημιουργικότητα (Zioga et al., 2020) αλλά και ότι η μη πραγματοποίηση προσδοκιών συσχετίζεται με την ανθρώπινη συναισθηματική αντίδραση σε μουσική (Egermann et al., 2013).



Εικόνα 6. Διάγραμμα που απεικονίζει την εσωτερική δομή του μοντέλου IDyOM. Κάθε οπτική γωνία παρέχει μία κατανομή πιθανότητας για την πρόβλεψη του επόμενου μουσικού γεγονότος (τα D_i , για $i = 1-11$) και ένα βάρος που καθορίζει τη σημαντικότητά της οπτικής γωνίας σε σχέση με τις άλλες οπτικές γωνίες (τα w_i , για $i = 1-11$). Ο συνδυασμός των επιμέρους κατανομών παρέχει την τελική κατανομή του μοντέλου, βάσει της οποίας γίνονται οι προβλέψεις του. Η παραπάνω περιγραφή ισχύει τόσο για το εκπαιδευμένο μοντέλο (μακρόχρονη μνήμη) όσο και για το μοντέλο που εκτίθεται μόνο στην τρέχουσα μελωδία (βραχύχρονη μνήμη) (Pearce & Wiggins, 2012, σελ. 633. Αναδημοσιεύεται με την άδεια του εκδότη).

Τάση

Συσχετισμένη με την προσδοκία είναι και η τονική τάση⁸ (tension), η αυξομείωση της ανάγκης για λύση κάποιου μουσικού γεγονότος σε κάποιο άλλο (π.χ. μίας απομακρυσμένης από το τονικό κέντρο, διάφωνης συγχορδίας στην τονική συγχορδία),

⁸ Επιλέξαμε να μεταφράσουμε τη λέξη “tension” με τη λέξη “τάση” αντί για την “ένταση”, ώστε να αποφευχθούν παρεξηγήσεις με την ένταση ως δυναμική.

που παίζει σημαντικό ρόλο στη συναισθηματική πλευρά της μουσικής εμπειρίας. Το μοντέλο τονικής τάσης των Lerdahl και Krumhansl (2007) βασίζεται στη θεωρία τονικής αφαίρεσης⁹ (pitch reduction), που παρουσιάστηκε στη *GTTM* (βλ. ενότητα “Ομαδοποίηση” · Lerdahl & Jackendoff, 1983), καθώς και στη θεωρία τονικού χώρου του Lerdahl (1988). Ένα τμήμα του υπολογισμού της τάσης ενός μουσικού γεγονότος, η ιεραρχική τάση (hierarchical tension) προκύπτει από δύο πηγές: από το βάθος του γεγονότος αυτού στο αφαιρετικό δέντρο (reductional tree) που παράγεται (όσο μεγαλύτερο βάθος, τόσο μεγαλύτερη τάση), αλλά και από την απόσταση στον τονικό χώρο μεταξύ του γεγονότος και του προγόνου του στο δέντρο (όσο μεγαλύτερη απόσταση, τόσο μεγαλύτερη τάση). Στον υπολογισμό τάσης συνυπολογίζονται ακόμη η ακολουθιακή τάση (sequential tension), δηλαδή η απόσταση στον τονικό χώρο μεταξύ γειτονικών γεγονότων, και ο βαθμός διαφωνίας της κάθε συγχορδίας της μουσικής επιφάνειας.

Ένα άλλο μοντέλο τονικής τάσης αποτελεί τμήμα του συστήματος αυτόματης παραγωγής μουσικής MorphueS (Herremans & Chew, 2016), και βασίζεται στο μοντέλο spiral array που αναφέρθηκε στην ενότητα εύρεσης μέτρου (βλ. ενότητα “Εύρεση τονικότητας” · Chew, 2002). Η αυξομείωση της τάσης εντός ενός κομματιού γίνεται αρχικά χωρίζοντας το κομμάτι σε ίσα μέρη που αντιστοιχούν σε «σύννεφα» σημείων στην τονική σπείρα και στη συνέχεια υπολογίζοντας τη διασπορά των νοτών του κάθε σύννεφου νοτών (μετρά τη διαφωνία), την ορμή μεταξύ σύννεφων (μετρά την τονική αλλαγή) και την απόστασή του τρέχοντος σύννεφου από το κέντρο της καθολικής τονικότητας.

Συναισθήματα

Η αισθητική εμπειρία και τα συναισθήματα που προκαλούνται από τη μουσική είναι για πολλούς ανθρώπους ο βασικός λόγος που ακούν μουσική (Schäfer et al., 2013). Στη γνωσιακή μουσικολογία, το συναίσθημα έχει μελετηθεί από πολλές διαφορετικές σκοπιές (Juslin & Sloboda, 2010), ωστόσο υπολογιστικά μοντέλα με στόχευση την προσέγγιση της λειτουργίας των ανθρώπινων συναισθημάτων λείπουν από τη βιβλιογραφία. Ταυτόχρονα, ο εφαρμοσμένος κλάδος της μουσικής αναγνώρισης συναισθημάτων (MER - Music Emotion Recognition), που αποτελεί τμήμα του τομέα ανάκτησης μουσικής πληροφορίας, έχει παραγάγει πάμπολλα μοντέλα αναγνώρισης μουσικών συναισθημάτων από χαρακτηριστικά του ηχητικού σήματος (Panda et al., 2020). Σκοπός των μοντέλων αυτών είναι η ταξινόμηση μουσικών κομματιών σε συναισθηματικές κατηγορίες από τις οποίες μπορεί να γίνει επιλογή για ακρόαση, με βάση τις συναισθηματικές ανάγκες χρηστών (Yang et al., 2018). Τα μοντέλα αυτά όμως δεν έχουν δημιουργηθεί με βάση τη γνωσιακή ευλογοφάνεια.

Μία διάκριση που γίνεται στη μελέτη των μουσικών συναισθημάτων είναι αυτή μεταξύ των εκπεφρασμένων συναισθημάτων (expressed/perceived emotions) ενός

⁹ Τονική αφαίρεση = εντοπισμός ιεραρχικών σχέσεων μεταξύ των νοτών.

μουσικού κομματιού και των αισθανόμενων συναισθημάτων (felt/induced emotions) που νιώθει ένας ακροατής ακούγοντάς το, τα οποία συνήθως παρουσιάζουν συσχέτιση αλλά ενίοτε αποκλίνουν (Evans & Schubert, 2008). Ακόμη, τα περισσότερα μοντέλα μουσικών συναισθημάτων χρησιμοποιούν είτε διακριτή αναπαράσταση (discrete) είτε αναπαράσταση διαστάσεων (dimensional) για τη μοντελοποίηση των συναισθημάτων (Eerola & Vuoskoski, 2012). Τα διακριτά μοντέλα βασίζονται στην υπόθεση ότι υπάρχουν κάποιες βασικές κατηγορίες συναισθημάτων (χαρά, λύπη, θυμός, αποστροφή, κ.ά. · Juslin & Laukka, 2004). Τα μοντέλα διαστάσεων βασίζονται στην υπόθεση ότι τα συναισθήματα είναι συνδυασμοί των τιμών άλλων μεγεθών και μπορούμε να τα δούμε ως σημεία στο πολυδιάστατο χώρο που ορίζουν τα μεγέθη αυτά. Κατά κανόνα χρησιμοποιείται ο διδιάστατος χώρος του Russell (1980), όπου οι άξονες αναπαριστούν το σθένος (valence) και τη διέγερση (arousal), και που έχει σε μεγάλο βαθμό επιβεβαιωθεί από νευροεπιστημονικές και ψυχολογικές έρευνες (Barrett & Russell, 1999 · Posner et al., 2005). Το ζήτημα το μουσικού συναισθήματος αποτελεί επίσης αντικείμενο του μέρους Δ του παρόντος τόμου.

Μοντελοποίηση προσοχής και μνήμης

Μέχρι τώρα εξετάσαμε υπολογιστικά μοντέλα εξειδικευμένα σε κάποιο τομέα, τα οποία εκτελούν μία ή λίγες μουσικές εργασίες. Ωστόσο, στη βιβλιογραφία έχουν προταθεί και γενικότερα υπολογιστικά μοντέλα της προσοχής και μνήμης, τα οποία σχετίζονται με πολλούς τομείς και εργασίες της μουσικής νόησης.

Στην καθημερινή ζωή, η συντριπτική πλειοψηφία των ήχων με τους οποίους ερχόμαστε σε επαφή είναι συνδυασμοί πολλαπλών ηχητικών πηγών. Οι ηχητικές ροές από κάθε πηγή πρέπει να διαχωριστούν για να γίνουν κατανοητές, εργασία που είναι γνωστή ως ακουστική ανάλυση σκηνής (auditory scene analysis · Bregman, 1990). Ο ρόλος της προσοχής είναι να εστιάσει τους αισθητηριακούς και γνωσιακούς πόρους μόνο στην πληροφορία του περιβάλλοντος που είναι σχετική ανά πάσα στιγμή, στην κατάλληλη ροή. Το αντικείμενο εστίασης της προσοχής ρυθμίζεται τόσο από αισθητηριακά ερεθίσματα του περιβάλλοντος (ανωφερώς) όσο και από γνωσιακούς παράγοντες, όπως προσδοκίες ή μαθημένα σχήματα (κατωφερώς). Οι Kaya και Elhilali (2017) δίνουν επισκόπηση για υπολογιστικά μοντέλα ακουστικής προσοχής, που ακολουθούν ποικίλες προσεγγίσεις και μεθόδους.

Οι μουσικές έννοιες (νότες, συγχορδίες, ρυθμούς, κ.λπ.) που χρησιμοποιούνται στις διάφορες μουσικές διεργασίες είναι αποθηκευμένες στη μακρόχρονη μνήμη ως κατηγορίες. Οι κατηγορίες είναι κατασκευασμένες ή/και μαθημένες νοητικές δομές που επιτρέπουν σε ένα εύρος μουσικών ερεθισμάτων να γίνουν αντιληπτά ως μέλη της ίδιας κατηγορίας, όπως για παράδειγμα δύο τονικά ύψη με συχνότητες που διαφέρουν μόνο κατά λίγα cent να γίνουν αντιληπτά ως η ίδια νότα (Purwins et al., 2008). Αυτή η κατηγοριοποίηση των συνεχών μεγεθών της αντίληψης σε λίγες διακριτές κατηγορίες έχει σκοπό την μείωση του αποθηκευτικού φόρτου στη μνήμη (Snyder, 2000). Οι κατηγορίες είναι στατικές κατασκευές και δεν μπορούν να αντικατοπτρίσουν ακολουθίες ή συνδυασμούς μουσικών γεγονότων. Για αυτό στη γνω-

σιακή επιστήμη υπάρχει η έννοια του σχήματος (Bartlett, 1932/1995), που σχετίζεται και με προσδοκίες για ανάπτυξη μουσικών γεγονότων.

Οι αυτο-οργανωμένοι χάρτες που συναντήσαμε σε προηγούμενες ενότητες έχουν χρησιμοποιηθεί για την κατασκευή μουσικών σχημάτων (Leman, 1995) και κατηγοριών (Purwins et al., 2007). Η γνωσιακή θεωρία προσαρμοστικού συγχρονισμού (adaptive resonance theory/ART · Grossberg, 1976 · Grossberg, 2013) αντιμετωπίζει το δίλημμα μεταξύ σταθερότητας-πλαστικότητας στην κατασκευή κατηγοριών βρίσκοντας ισορροπία μεταξύ την αντίληψης του περιβάλλοντος με ήδη υπάρχουσες κατηγορίες (σταθερότητα) και της δημιουργίας νέων κατηγοριών (πλαστικότητα). Στη μουσική έχει χρησιμοποιηθεί στα πλαίσια του μοντέλου αναγνώρισης τονικού ύψους του Taylor και Greenhough (1994). Τέλος, το μοντέλο δημιουργίας ιεραρχίας εννοιών COBWEB (Fisher, 1987) εφαρμόζει σταδιακή συσταδοποίηση στα αντιληπτικά δεδομένα με σκοπό να μεγιστοποιήσει την ικανότητα συμπερασματολογίας του. Έχει χρησιμοποιηθεί στα πλαίσια του μοντέλου του Marger και συνεργατών (2007), που επιχειρεί να εξηγήσει την ανάδυση των μουσικών κατηγοριών των κλιμάκων, προτύπων και αρμονίας και έχει βρεθεί συνεπές με τη μουσική θεωρία.

ΜΟΥΣΙΚΗ ΣΥΝΘΕΣΗ

Ήδη από τις πρώτες δεκαετίες μετά την εφεύρεση των ηλεκτρονικών υπολογιστών, οι άνθρωποι έσπευσαν να τους χρησιμοποιήσουν και για μουσική σύνθεση (Hiller & Isaacson, 1959), και έκτοτε έχει δημιουργηθεί μεγάλο πλήθος υπολογιστικών συστημάτων «αυτόματης/αλγοριθμικής μουσικής σύνθεσης» (Nierhaus, 2009). Για την ανάπτυξή τους χρησιμοποιούνται ποικίλες τεχνικές της τεχνητής νοημοσύνης, από τις πιο κλασικές μέχρι και τις πιο πρόσφατες. Τα περισσότερα από αυτά τα συστήματα είναι στοχευμένα στην παραγωγή αξιολογής και πρωτότυπης μουσικής, ενώ ορισμένα από αυτά είναι επιπλέον γνωσιακά – δηλαδή επιχειρούν να μοντελοποιήσουν τις νοητικές διαδικασίες των συνθετών. Για αυτό το λόγο, η πλειοψηφία των μοντέλων που αναφέρονται παρακάτω δεν έχουν γνωσιακή στόχευση ούτε έχουν δοκιμαστεί ως γνωσιακά μοντέλα, ωστόσο εκτιμούμε ότι είναι σε κάποιο βαθμό γνωσιακά ευλογοφανή. Θα παρουσιάσουμε δύο από τις σημαντικότερες γενικές κατευθύνσεις μοντελοποίησης στον τομέα της αυτόματης σύνθεσης: τη συμβολική TN και τη μηχανική μάθηση. Αξιοσημείωτο είναι ότι πολλά από τα μοντέλα που έχουν προταθεί στη βιβλιογραφία ακολουθούν υβριδική προσέγγιση, συνδυάζοντας τις παραπάνω κατευθύνσεις. Μία επισκόπηση του τομέα της τεχνητής νοημοσύνης (TN) στην εργασία της αυτόματης σύνθεσης δίνεται από τους Fernandez και Vico (2013).

Συμβολικά μοντέλα κανόνων και γραμματικών

Μία μέθοδος δημιουργίας μοντέλων σύνθεσης είναι η συλλογή μουσικής γνώσης (συνήθως για κάποιο συγκεκριμένο στυλ) σε μορφή κανόνων και η ενσωμάτωσή της σε υπολογιστικά συστήματα που χρησιμοποιούνται για την παραγωγή νέων κομμα-

τών μουσικής. Τέτοια συστήματα ονομάζονται συστήματα βασισμένα σε κανόνες (rule-based systems) και ανήκουν στον υπο-τομέα της ΤΝ που ονομάζεται συμβολική ΤΝ (symbolic AI). Χαρακτηριστικό ιστορικό παράδειγμα τέτοιου μοντέλου είναι αυτό του Ebeioğlu (1990) που χρησιμοποιεί πάνω από 350 κανόνες για εναρμόνιση¹⁰ χορικών σε στίλ Bach. Οι Anders και Miranda (2011) δίνουν μία επισκόπηση της προσέγγισης περιορισμών, που στηρίζεται επίσης σε κανόνες, στην εργασία της εναρμόνισης μελωδίας. Ο Widmer (1992) υλοποίησε ένα ιδιαίτερα γνωστικά ευλογοφανές μοντέλο για εναρμόνιση μελωδιών το οποίο μαθαίνει τους κανόνες του αυτόματα από ένα σώμα μελωδιών και εναρμονίσεών τους. Ο Rowe (1992), βασισμένος στην αρχιτεκτονική “κοινωνίας του νου” (society of mind) του Minsky (1988), παρουσίασε το διαδραστικό σύστημα cypher, αποτελούμενο από δύο τμήματα: έναν “ακροατή”, που αναλύει MIDI δεδομένα και έναν “εκτελεστή”, που παράγει νέα μουσική σε πραγματικό χρόνο. Ακόμα, ο Pereira και συνεργάτες (1997) υλοποίησαν ένα γνωστικό μοντέλο που χρησιμοποιεί συλλογιστική βάσει περιπτώσεων (case-based reasoning) και υλοποιεί τη θεωρία δημιουργικότητας του Wallas (1926). Συστήματα βασισμένα σε κανόνες μπορούν να χρησιμοποιηθούν και στη μελέτη των δομικών λίθων ενός συγκεκριμένου στίλ, αλλά προϋποθέτουν τη συλλογή τεράστιας ποσότητας γνώσης για το στίλ αυτό (που γίνεται εξαιρετικά δύσκολο για πιο σύνθετα στίλ από τα χορικά του Bach).

Παραλλαγή των κανόνων αποτελεί η κατασκευή και χρήση κάποιας γενετικής γραμματικής (generative grammar). Με την προσέγγιση αυτή, η μουσική επιφάνεια του παραγόμενου μουσικού κομματιού προκύπτει ως το τελικό προϊόν μετασχηματισμών μη-τερματικών συμβόλων υψηλού επιπέδου σε τερματικά σύμβολα, που αναπαριστούν μουσικές έννοιες πολλαπλών επιπέδων (Baroni, 1999).

Ένα από τα πρώτα συστήματα παραγωγής μουσικής μέσω γενετικών γραμματικών είναι αυτό των Baroni και Jacobini (1978) που παράγει μελωδίες χορικών σε στίλ Bach. Μετέπειτα έρευνες όρισαν γενετικές γραμματικές που επιχειρούν να αποτυπώσουν ευρύτερα μουσικά είδη, όπως τζαζ αρμονικές ακολουθίες (Steedman, 1984 · Granroth-Wilding & Steedman, 2014), ρυθμούς tabla στην ινδική κλασική μουσική (Kirpen & Bel, 1989) ή και ολόκληρη την τονική αρμονία (Rohrmeier, 2011). Οι γενετικές γραμματικές έχουν συνδυαστεί επίσης με εξελικτικούς αλγόριθμους¹¹, όπως στην εργασία του Reddin και συνεργατών (2009).

Η λειτουργία των συστημάτων κανόνων ή γενετικής γραμματικής είναι σχετικά εύκολα κατανοητή από ανθρώπους, το οποίο σημαίνει ότι μπορούμε να ακολουθήσουμε την πορεία κατασκευής όλων των χαρακτηριστικών της παραγόμενης μουσικής σύνθεσης. Ωστόσο, τέτοια συστήματα είναι περιορισμένα λόγω της αδυναμίας προσαρμογής τους, που προκύπτει από την έλλειψη μηχανισμών αυτόματης αλλαγής, προσθήκης ή/και διαγραφής κανόνων (π.χ. αδυναμία παραγωγής μουσικής σε

¹⁰ Θεωρούμε την εναρμόνιση (harmonisation) ως μία μορφή σύνθεσης.

¹¹ Εξελικτικός αλγόριθμος: μέθοδος βελτιστοποίησης η οποία ξεκινά με ένα πληθυσμό, εφαρμόζει επαναλαμβανόμενα τη συνάρτηση καταλληλότητας (fitness function) σε κάθε μέλος του πληθυσμού, δημιουργώντας έτσι την επόμενη γενιά που περιλαμβάνει τα πιο κατάλληλα μέλη στα οποία εφαρμόζει διασταυρώσεις ή/και μεταλλάξεις, μέχρι ο πληθυσμός να ικανοποιεί κάποιο κριτήριο τερματισμού (Eiben et al., 2015).

διαφορετικά στίλ απ' αυτό για το οποίο αρχικά δημιουργήθηκαν). Επίσης, παρά δεκαετίες φιλοσοφικών μελετών, ψυχολογικών και νευροβιολογικών πειραμάτων, δεν υπάρχουν ακόμη θεωρητικά ή εμπειρικά δεδομένα τα οποία να αποδεικνύουν σε σημαντικό βαθμό τη γνωσιακή ευλογοφάνεια των κανόνων ή των γραμματικών (Pereplyotchik, 2017).

Στατιστικά μοντέλα μηχανικής μάθησης

Τα στατιστικά μοντέλα εκπαιδεύονται σε δεδομένα για να μάθουν να επιτελούν τη λειτουργία τους, και γι' αυτό κατατάσσονται στον υπο-τομέα της τεχνητής νοημοσύνης που ονομάζεται μηχανική μάθηση (machine learning). Πολλά συστήματα μηχανικής μάθησης δεν χρησιμοποιούν συμβολική ΤΝ, ωστόσο υπάρχει σημαντική επικάλυψη μεταξύ των τομέων. Λόγω της αδυναμίας των πρώιμων αυτών μοντέλων να παραγάγουν αξιόλογα μουσικά αποτελέσματα (Ames, 1989), τις δεκαετίες του '70 και '80 η προσοχή στράφηκε σε χειροποίητα (hand-crafted) μοντέλα αμιγούς συμβολικής ΤΝ και ήταν μόλις τη δεκαετία του '90 που επέστρεψαν στο προσκήνιο (Conklin, 2003), όπου έκτοτε έχουν παραμείνει.

Η στατιστική προσέγγιση στην εργασία της μουσικής σύνθεσης βασίζεται στην εξής αρχή: «ένα μουσικό στίλ συνίσταται από πολλών ειδών γεγονότα που ακολουθούν κάποια στατιστική κατανομή και το στίλ αυτό μπορεί να προσομοιωθεί –τουλάχιστον με προσεγγιστικό τρόπο– παράγοντας νέα μουσική με την ίδια κατανομή» (Temperley, 2012). Τα γεγονότα που αναφέρονται δεν είναι μεμονωμένες νότες αλλά λαμβάνεται υπόψη και το περιεχόμενό τους, που μπορεί να επιτευχθεί χρησιμοποιώντας μαρκοβιανά μοντέλα (Markov models) και συγκεκριμένα μαρκοβιανές αλυσίδες (Markov chains · Norris, 1998). Πιο αναλυτικά, μία μαρκοβιανή αλυσίδα αναπαριστά την πιθανότητα ενός γεγονότος (π.χ. αλλαγή κατεύθυνσης μιας μελωδίας) δεδομένου ενός συγκεκριμένου αριθμού προηγούμενων γεγονότων. Ο αριθμός αυτός καθορίζει την τάξη του μοντέλου: αν είναι μηδέν τότε αναπαρίσταται η πιθανότητα ενός γεγονότος εκτός περιεχόμενου, αν είναι ένα τότε αναπαρίσταται η πιθανότητα ενός γεγονότος δεδομένου του προηγούμενου γεγονότος (π.χ. η πιθανότητα κατάβασης μιας μελωδίας, δεδομένου ότι αμέσως πριν ανέβηκε), αν είναι δύο τότε αναπαρίσταται η πιθανότητα ενός γεγονότος δεδομένων των δύο προηγούμενων γεγονότων, και ούτω καθεξής. Οι πιθανότητες αυτές μπορούν να εξαχθούν από ένα μουσικό σώμα και να χρησιμοποιηθούν ώστε να παραχθεί νέα μουσική στοχαστικά, δηλαδή με χρήση τυχαιότητας που όμως προσεγγίζει την ίδια στατιστική κατανομή με αυτή του μουσικού σώματος.

Παράδειγμα ενός τέτοιου μοντέλου αποτελεί αυτό των Ponsford και συν. (1999), που χρησιμοποιεί μαρκοβιανές αλυσίδες τρίτης και τέταρτης τάξης για την παραγωγή Sarabande σε μπαρόκ στίλ. Η εργασία των Conklin και Witten που εισήγαγε την αναπαράσταση πολλαπλών οπτικών γωνιών (βλ. ενότητες “Ανακάλυψη προτύπων” και “Ομοιότητα” · Conklin & Witten, 1995), εκτός από προβλέψεις για τη συνέχεια μίας μελωδίας, χρησιμοποιήθηκε και για την παραγωγή μελωδιών σε στίλ χορικού Bach. Το σύστημα IDyOT (βλ. ενότητα “ομοιότητα” · Pearce & Wiggins,

2006), που βασίζεται στην αναπαράσταση πολλαπλών γωνιών, έχει επεκταθεί ώστε να εφαρμοστεί και στην εργασία σύνθεσης μουσικής, όπως στο σύστημα παραγωγής τετράφωνων χορικών του Whorley και συνεργατών (2013). Η προσέγγιση πολλαπλών οπτικών γωνιών έχει χρησιμοποιηθεί και στο σύστημα “Continuator” του Pachet (2003) που συνθέτει μουσική σε πραγματικό χρόνο στα πλαίσια αυτοσχεδιαστικής διάδρασης με ανθρώπους μουσικούς. Επίσης, ο Pachet και συνεργάτες (2011) παρουσίασαν μία προσέγγιση που συνδυάζει την ικανοποίηση περιορισμών (βασισμένη σε κανόνες) με τις μαρκοβιανές αλυσίδες και μπορεί να χρησιμοποιηθεί για την ανάπτυξη μοντέλων παραγωγής μελωδιών σε οποιοδήποτε μουσικό ιδίωμα. Παραλλαγή των μαρκοβιανών μοντέλων αποτελούν τα κρυφά μαρκοβιανά μοντέλα (hidden Markov models), στα οποία ένα σύνολο μη παρατηρήσιμων (“κρυφών”) καταστάσεων επιδρούν σε ένα σύνολο παρατηρήσιμων καταστάσεων και στόχος είναι να μοντελοποιηθούν οι κρυφές καταστάσεις βάσει των παρατηρήσεων. Ένα σύστημα που χρησιμοποιεί την παραλλαγή αυτή για σύνθεση μουσικής είναι εκείνο του Μαυρομάτη (2009), το οποίο συνθέτει ψαλμούς Βυζαντινής μουσικής. Τέλος, ο Καλιακάτσος-Παπακώστας και συνεργάτες παρουσιάζουν το σύστημα υποβόηθησης εναρμόνισης μελωδίας CHAMELEON (2017), που χρησιμοποιεί τόσο συμβολική TN όσο και κρυφά μαρκοβιανά μοντέλα και το οποίο στηρίζεται στη θεωρία εννοιολογικής ανάμιξης (conceptual blending · Fauconnier & Turner, 2003) για να συνδυάσει αρμονικά ιδιώματα.

Τα νευρωνικά δίκτυα αποτελούν επίσης μοντέλο στατιστικής μάθησης, άνθισαν αρχικά τη δεκαετία του '80. Το πλεονέκτημα των νευρωνικών δικτύων έναντι των μαρκοβιανών μοντέλων είναι ότι μπορούν να λάβουν υπόψη τους γεγονότα που συνέβησαν οποτεδήποτε στο παρελθόν, και όχι μέχρι κάποιο συγκεκριμένο όριο όπως τα μαρκοβιανά μοντέλα. Ωστόσο, όπως αναφέρθηκε και στην ενότητα “ονομασία τόνων”, πάσχουν από το μειονέκτημα ότι οι εσωτερικοί υπολογισμοί τους δεν είναι γενικά κατανοητοί σε υψηλό επίπεδο από ανθρώπους (αν και οι χαμηλού επιπέδου υπολογισμοί είναι). Παρόλο που γίνονται προσπάθειες οι υπολογισμοί αυτοί να γίνουν ξεκάθαροι με ανθρώπινους όρους, φαίνεται ότι ο τομέας της «εξηγήσιμης TN» (explainable AI) είναι ακόμα μακριά απ' το να δώσει ασφαλείς απαντήσεις.

Ένα από τα πρώτα νευρωνικά δίκτυα αυτόματης σύνθεσης της πρώιμης περιόδου ανάπτυξης τους είναι αυτό του Todd (1989), το οποίο συνθέτει μελωδίες. Το μοντέλο του Mozer (1994) ενσωματώνει ψυχολογικά θεμελιωμένες αναπαραστάσεις για το τονικό ύψος, τη χρονική διάρκεια και την αρμονική δομή, έχει εκπαιδευτεί σε μελωδίες του J.S.Bach και σε παραδοσιακές μελωδίες και παράγει νέες μελωδίες σε αυτά τα στυλ, που ωστόσο πάσχουν όσον αφορά την καθολική συνοχή τους. Ο Toivainen (1995) παρουσιάζει ένα μοντέλο για τζαζ αυτοσχεδιασμό στο bebop ιδίωμα, που πάσχει επίσης από το πρόβλημα της συνοχής στο επίπεδο ολοκλήρης της μελωδίας. Για την αντιμετώπιση του ζητήματος αυτού, οι Eck και Schmidhuber (2002) υπήρξαν από τους πρώτους που χρησιμοποίησαν ανατροφοδοτούμενα (recurrent) δίκτυα με μακρόχρονη-βραχύχρονη μνήμη (Long Short-Term Memory, LSTM) που εφάρμοσαν στην παραγωγή μπλουζ μουσικής. Το διαδραστικό μοντέλο τζαζ αυτοσχεδιασμού της Franklin (2001) αποτελείται από δύο φάσεις: στην πρώτη ένα ανατροφοδοτούμενο δίκτυο εκπαιδεύεται σε τζαζ μελωδίες ενώ στη δεύτερη το δίκτυο

εκπαιδευεται περαιτέρω με ενισχυτική μάθηση (reinforcement learning) βάσει ευρετικών κανόνων.

Τα νευρωνικά δίκτυα επέδειξαν σημαντική πρόοδο στην εργασία της αλγοριθμικής σύνθεσης στα τέλη της δεκαετίας του 2010, κάνοντας χρήση βαθιάς μάθησης (νευρωνικών δικτύων με πολλά στοιβαγμένα επίπεδα). Σημαντικοί λόγοι της επιτυχίας τους είναι η συνεχώς μεγαλύτερη διαθεσιμότητα δεδομένων και η αύξηση της ταχύτητας επεξεργασίας των υπολογιστών. Ο Hadjeres και συνεργάτες παρουσιάζουν το μοντέλο DeepBach (2017) το οποίο παράγει χορικά σε στίλ Bach, εξαρχής ή συμπληρώνοντας φωνές που λείπουν, χρησιμοποιώντας βαθιά ανατροφοδοτούμενα νευρωνικά δίκτυα που λαμβάνουν υπόψη τόσο τα μουσικά γεγονότα που προηγούνται σε σχέση με το τρέχον σημείο παραγωγής όσο και αυτά που το ακολουθούν. Το μοντέλο Music Transformer της Huang και συνεργατών (2018) χρησιμοποιεί την αρχιτεκτονική νευρωνικών δικτύων transformer (Vaswani, 2017) και την προσαρμόζει ώστε να παράγει μουσική με μεγαλύτερη συνοχή απ' ό τι ήταν προηγουμένως εφικτό. Οι Briot και Pachet (2020) και οι Hernandez-Olivan και Beltran (2021) δίνουν ανασκοπήσεις για μεθόδους αλγοριθμικής σύνθεσης με βαθιά μάθηση. Τα τελευταία χρόνια η βελτίωση της απόδοσης συστημάτων βαθιάς μάθησης έχει συνεχιστεί, ωστόσο προς κατευθύνσεις που είναι αμφίβολο ότι αντικατοπτρίζουν γνωσιακές διεργασίες.

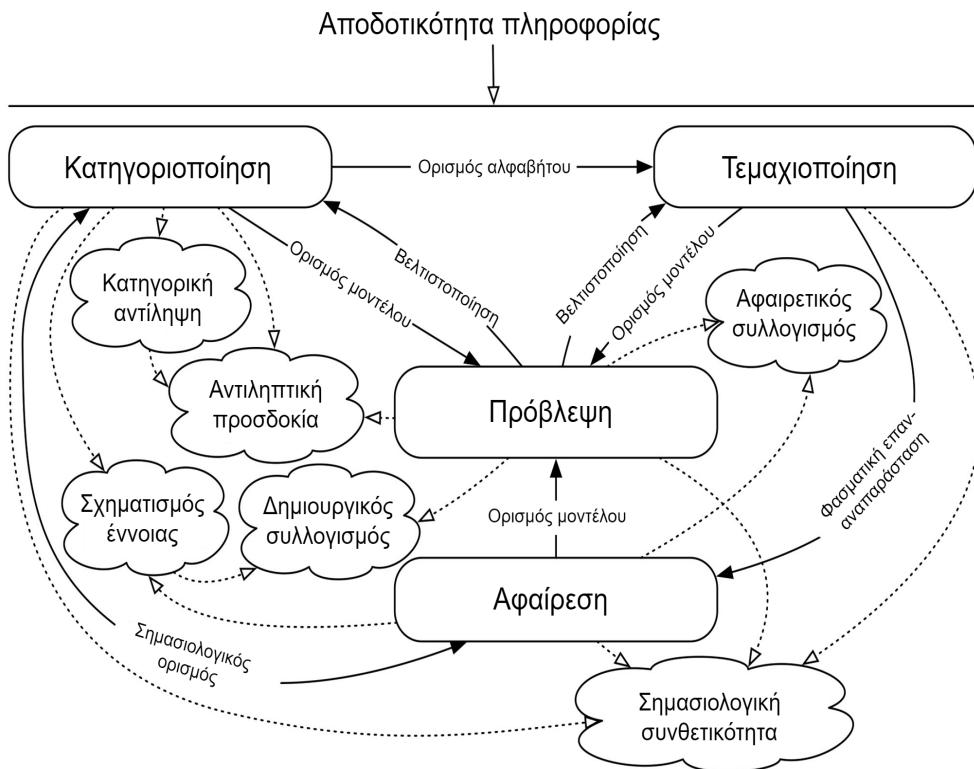
ΠΕΡΙΠΤΩΣΗ ΜΕΛΕΤΗΣ: ΤΟ ΜΟΝΤΕΛΟ IDyOT

Το μοντέλο IDyOT, «Δυναμική πληροφορίας της σκέψης» (Information Dynamics of Thinking), και η θεωρία στην οποία στηρίζεται βρίσκονται σε ενεργή ανάπτυξη την τελευταία δεκαετία από τον Wiggins και συνεργάτες του (2020). Το IDyOT βασίζεται στο μοντέλο IDyOM (βλ. ενότητα “Ομοιότητα”) και σχετίζεται με πολλές από τις προσεγγίσεις και τα μοντέλα που παρουσιάστηκαν παραπάνω, αλλά αποτελεί μία προσπάθεια ενοποίησης τους σε ένα κοινό γενικό πλαίσιο και επέκτασής τους. Αυτό φανερώνεται, μεταξύ άλλων, στο ότι δεν περιορίζεται στη μοντελοποίηση της μουσικής νόησης, αλλά στη νόηση κάθε είδους σειριακού ερεθίσματος, όπως της φυσικής γλώσσας. Το IDyOT μοντελοποιεί τη λειτουργία του νου σε αφαιρετικό, όχι βιολογικό επίπεδο, αν και είναι συμβατό με τη γνώση μας για τη νευροβιολογία.

Πιο συγκεκριμένα, το μοντέλο αποτελεί γνωσιακή αρχιτεκτονική, δηλαδή ενοποιημένη θεωρία της νόησης, και χρησιμοποιεί απλές αλλά ισχυρές μαθηματικές δομές, βάσει των οποίων αναδύεται πολύπλοκη συμπεριφορά. Η βασική αρχή που καθορίζει τη λειτουργία του μοντέλου είναι η αποδοτικότητα πληροφορίας (information efficiency) η οποία εκδηλώνεται:

- στις προβλέψεις του συστήματος, που κατευθύνουν τον τρόπο με τον οποίο αντιλαμβάνεται το περιβάλλον του,
- στην επιδίωξή του να κατασκευάσει το πιο αποδοτικό μοντέλο του περιβάλλοντός του,
- στην εφαρμογή της ενεργούς προσοχής μόνο στις δομές που παρέχουν την περισσότερο πληροφορία σε σχέση με άλλες.

Δεν υπάρχουν ρητοί στόχοι ή μοντελοποίηση συναισθημάτων στην αρχιτεκτονική, ενώ το μοναδικό κίνητρο είναι η περιέργεια: το IDyOT συνεχώς προβλέπει τι πρόκειται να συμβεί στο περιβάλλον του και παράγει νέες πιθανές συνέχειες στις ακολουθίες που μαθαίνει. Αυτή η ιδιότητα παραγωγής νέων ακολουθιών μάλιστα είναι που δίνει στο IDyOT τις δημιουργικές ικανότητές του, που παρουσιάζονται παρακάτω (Εικόνα 7).



Εικόνα 7. Επισκόπηση των τμημάτων και του κύκλου λειτουργίας του IDyOT. Τα στρογγυλεμένα κουτιά αναπαριστούν τις βασικές λειτουργίες του μοντέλου. Τα κουτιά σε σχήμα σύννεφου αναπαριστούν τα φαινόμενα που το μοντέλο εξηγεί. Οι ετικέτες των βελών αντιστοιχούν σε λειτουργίες που συνδέουν τις βασικές λειτουργίες. Η θεμελιώδης αρχή είναι η αποδοτικότητα πληροφορίας (απόδοση στα Ελληνικά του κύκλου λειτουργίας του IDyOT, Wiggins & Sanjekdar, 2019).

Η πλέον θεμελιώδης λειτουργία του IDyOT είναι η τεμαχιοποίηση (chunking) του ερεθίσματος. Προκειμένου να εντοπιστούν όρια σε μία ακολουθία, το σύστημα κάνει χρήση της ιδέας ότι οι προβλέψεις για το τι έπεται σε μία ακολουθία γίνονται ολοένα και πιο σίγουρες όσο πλησιάζουμε το τέλος ενός τμήματος/τεμαχίου (segment/chunk).

Η μνήμη του IDyOT διακρίνεται σε επεισοδιακή και σημασιολογική. Η επεισοδιακή μνήμη αποτελεί καταγραφή ακολουθιών που γίνονται αντιληπτές από το περιβάλλον, σε κάποιο χαμηλό επίπεδο (το χαμηλότερο απ' όλα απορρίπτεται καθώς περιέχει υπερβολικά πολλή πληροφορία) αλλά και σε υψηλότερα από αυτό επίπεδα, τα οποία προκύπτουν μέσω τεμαχιοποίησης. Η σημασιολογική μνήμη περιέχει τα σύμβολα από τα οποία αποτελούνται οι ακολουθίες της επεισοδιακής μνήμης και τις σχέσεις μεταξύ συμβόλων του ίδιου επιπέδου και διαφορετικών επιπέδων. Πιο αναλυτικά, συνίσταται από μία ιεραρχία εννοιολογικών χώρων (Gärdenfors, 2000), γεωμετρικών δομών που γεφυρώνουν τον διακριτό κόσμο των συμβόλων με την πολύπλοκη, πολυδιάστατη δομή των αντιληπτικών δεδομένων μέσω των «διαστάσεων ποιότητας» τους (quality dimensions). Οι διαστάσεις αυτές αναπαριστούν έννοιες που είναι αντιληπτές σε ανθρώπους (π.χ. ο εννοιολογικός χώρος του χρώματος αποτελείται από διαστάσεις έντασης, φωτεινότητας και απόχρωσης). Κάθε εννοιολογικός χώρος στη σημασιολογική μνήμη αποτελείται από σημεία τα οποία είναι φασματικές αναπαραστάσεις του συμβόλου/τεμαχίου χαμηλότερου επιπέδου με το οποίο ταυτίζονται. Χρησιμοποιώντας φασματικές αναπαραστάσεις επιτυγχάνεται αφαίρεση της πληροφορίας χρονισμού του συμβόλου/τεμαχίου χαμηλότερου επιπέδου (δηλαδή η ταχύτητα εκτέλεσής του). Επίσης, αναπαραστάσεις στους χώρους αυτούς των οποίων οι δομές είναι αρκετά παρόμοιες κατηγοριοποιούνται στην ίδια κατηγορία με αποτέλεσμα να αφαιρούνται λεπτομέρειες του χαμηλότερου επιπέδου. Ακόμα, η λειτουργία της *μνημονικής παγίωσης* (memory consolidation) λαμβάνει χώρα όταν το σύστημα δεν δέχεται νέα δεδομένα (σε αντιστοιχία με τον ύπνο στους ανθρώπους) και βελτιστοποιεί τις αναπαραστάσεις που έμαθε το σύστημα όταν επεξεργάστηκε νέα δεδομένα. Η παγίωση συνιστά ουσιαστικά επαν-αναπαράσταση (re-representation) των μαθημένων δεδομένων ώστε να είναι πιο αποδοτική η αποθήκευση και ανάκλησή τους.

Με βάσει όλα τα παραπάνω, μπορούμε τώρα να ερευνήσουμε που εντοπίζεται δημιουργικότητα στη συμπεριφορά ενός πράκτορα IDyOT. Μία απάντηση είναι πως όταν το σύστημα δεν δέχεται νέα αντιληπτικά δεδομένα μπορεί να προβλέψει ακολουθίες τις οποίες δεν έχει ξανασυναντήσει στο περιβάλλον του, και τις οποίες συνθέτει δημιουργικά με βάση τις δομές της μνήμης του. Αυτή η διαδικασία μάλιστα μπορεί να εξηγήσει το φαινόμενο δημιουργικής έμπνευσης (ή *aha moment*), όπου μία δημιουργική ιδέα εμφανίζεται ξαφνικά στη συνείδηση κάποιου, συχνά λύνοντας ένα πρόβλημα το οποίο είχε τεθεί ωρίτερα, και το οποίο ο ασυνείδητος νους επεξεργάζοταν μέχρι να βρει τη λύση που προκάλεσε την έμπνευση. Μπορούμε όμως να εντοπίσουμε δημιουργικότητα και στην επαν-αναπαράσταση των συμβόλων του συστήματος που συμβαίνει κατά τη μνημονική παγίωση, καθώς και αυτή αποτελεί μία μορφή εξερεύνησης των χώρων της σημασιολογικής μνήμης.

Το IDyOT είναι ένα σχετικά νέο μοντέλο και κατά συνέπεια έχουν υλοποιηθεί μόνο απλοποιημένες εκδοχές του, οι οποίες έχουν δοκιμαστεί σε γλωσσικές και όχι μουσικές εργασίες. Ωστόσο, τα αποτελέσματα των πειραμάτων αυτών έχουν βρεθεί υποσχόμενα (Wiggins & Sanjekdar, 2019). Πρόσφατα, η τρέχουσα ερευνητική ομάδα του IDyOT πρότεινε μία νέα φασματική αναπαράσταση γνώσης για το IDyOT (Homer et al., 2024a), και ένα νέο μέτρο πληροφορίας που θα χρησιμοποιηθεί στην αρχιτεκτονική (Homer et al., 2024b). Ένα σημαντικό επόμενο βήμα για το IDyOT

είναι η πλήρης υλοποίηση και επικύρωση της θεωρίας. Η μουσική θα παίξει σημαντικό ρόλο στην επαλήθευση του μοντέλου ως γνωσιακό, επειδή δεν επιβαρύνεται από την αναφορά στον κόσμο που αποτελεί βασικό χαρακτηριστικό της γλώσσας - αντιθέτως λειτουργεί στο δικό της, κλειστό κόσμο. Έτσι, η μουσική αποτελεί ιδανικό τομέα για έλεγχο των εσωτερικών μηχανισμών του IDyOT καθώς και για σύγκριση της επίδοσής του σε σχέση με τον προκάτοχό του, IDyOM.

ΕΠΙΛΟΓΟΣ

Στο παρόν κεφάλαιο πραγματοποιήσαμε μία σύντομη ανασκόπηση των σημαντικότερων υπολογιστικών προσεγγίσεων στη μουσική νόηση, χωρισμένων ανά μουσική διεργασία. Τα μοντέλα που περιγράφηκαν είναι είτε ξεκάθαρα γνωσιακά (π.χ. IDyOM, Pearce & Wiggins, 2012), είτε υπολογιστικά μοντέλα που όμως μοντελοποιούν ανθρώπινες νοητικές διαδικασίες και που πιθανώς έχουν κάποια γνωσιακή ευλογοφάνεια. Αυτό μένει να καθοριστεί από προσεχείς ψυχολογικές μελέτες που χρειάζεται να ελέγξουν μοντέλα ως προς τη γνωσιακή τους αξία. Λόγω του περιορισμένου μεγέθους του άρθρου, δεν μπόρεσε να γίνει αναφορά σε πολλά μοντέλα, κυρίως νεότερα, τα οποία συχνά βασίζονται στα παρουσιαζόμενα, τα τροποποιούν και επεκτείνουν. Στην ενότητα της μουσικής σύνθεσης, που είναι ένα τεράστιο κεφάλαιο για τον τομέα της μουσικής και τεχνητής νοημοσύνης, και για την πληρότητα ενός κεφαλαίου που ασχολείται με την υπολογιστική μοντελοποίηση, παρουσιάσαμε επίσης τις βασικότερες προσεγγίσεις της TN, όχι μόνο στη μουσική σύνθεση αλλά και γενικότερα στη μουσική αντίληψη και ανάλυση.

Ο τομέας της υπολογιστικής μοντελοποίησης της μουσικής νόησης είναι ένας από τους πλέον υποσχόμενους για να δώσει έγκυρες απαντήσεις στα θεμελιώδη ερωτήματα του πώς ο νους αντιλαμβάνεται, επεξεργάζεται και παράγει μουσική. Παρά την σημαντική πρόοδο των τελευταίων πενήντα ετών, υπάρχουν ακόμα πολλά ανοιχτά προβλήματα και προκλήσεις. Ένα από αυτά είναι η ενσωμάτωση γνώσης πολλαπλών επιπέδων στις αποκρίσεις των μοντέλων, σε αντίθεση με την εστίαση μόνο σε τοπικά χαρακτηριστικά ή χαρακτηριστικά μίας μόνο μουσικής διάστασης. Μία άλλη πρόκληση είναι η κατασκευή γενικών μοντέλων που εξηγούν συγχρόνως πολλαπλά φαινόμενα, και όχι μόνο μουσικά, όπως επιχειρεί να κάνει το IDyOT. Ακόμη, τίθεται το ζήτημα της αξιολόγησης των μοντέλων, που για να εκτελεστεί ορθά απαιτεί τον έλεγχό τους σε μεγάλο αριθμό μουσικών εργασιών που περιλαμβάνουν όσα περισσότερα μουσικά είδη γίνεται, καθώς και χρήση κοινών ελέγχων μεταξύ μοντέλων, ώστε να είναι εφικτή η απευθείας σύγκριση τους. Επιπλέον, η επιτυχία των νευρωνικών δικτύων την τελευταία δεκαετία εγείρει το ζήτημα της αξιολόγησής τους ως νοητικά μοντέλα και της κατανόησης των εσωτερικών μηχανισμών τους.

Η δημιουργία ικανότερων μοντέλων της μουσικής νόησης από τα υπάρχοντα, πέρα από την επιστημονική αξία τους, θα οδηγήσει και σε πρακτικά εργαλεία με εφαρμογές στον τομέα ανάκτησης μουσικής πληροφορίας, στη μουσική εκπαίδευση, τη μουσικοθεραπεία, σε εργασίες σχετικές με τα μουσικά συναισθήματα καθώς και στην υποστήριξη της μουσικής σύνθεσης και του αυτοσχεδιασμού.

Computational models of Music Cognition

Yorgos Velissaridis & Christina Anagnostopoulou

COMPUTATIONAL MODELLING OF MUSIC COGNITION is an approach to understanding the musical mind by building computer programs that simulate the way it operates. In this review we first discuss general issues regarding computational modelling and then present characteristic models organised by task. Computational modelling of music cognition focuses on cognitive plausibility rather than efficiency and might be evaluated using experimental behavioural and neurobiological data (Purwins et al., 2008). Computational models make use of specific knowledge representation schemes, built-in (Conklin & Anagnostopoulou, 2001) or learned from their input, which can be symbolic or raw audio. Two of the historically first music perception problems to be tackled using computational modelling were key and metre-finding, to which there exist multiple different approaches (Longuet-Higgins & Steedman, 1971 · Krumhansl & Kessler, 1982; Povel & Essens, 1985 · Large & Jones, 1999). Other issues regarding perception include pitch identification, the extraction of notes given an audio signal (de Cheveigné, 2005), the grouping of notes into motives, phrases and larger sections (Cambouropoulos, 2006) and voice segregation in polyphonic music (Gjerdingen, 1994). Higher up the perception-to-cognition scale we find computational models of harmonic analysis, employing techniques like Bayesian modelling (Raphael & Stoddard, 2004) and neural networks (Bharucha, 1987; Tillmann et al., 2000). Closely related are models of pitch spelling (Chew & Chen, 2005), including models which do both harmonic analysis and pitch spelling concurrently (Micchi et al., 2020). Moving further into the realm of cognitive processing, computational models have been proposed for pattern discovery (Conklin & Anagnostopoulou, 2006), and for music similarity. Global similarity is related to style or mood (Eerola et al., 2001) and pattern similarity with melodic likeness (Rolland, 1999). Anticipation is another crucial part of musical experience (Huron, 2006) that has been tackled by computational modelling, specifically through implementations of Narmour's Implication-Realization theory (Schellenberg, 1997). Another model of melodic expectation based on the multiple-viewpoint representation scheme is IDyOM (Pearce & Wiggins, 2006), which outperforms other models and has been found to be cognitively plausible (Egermann et al., 2013; Pearce et al., 2010). Related to anticipation is tonal tension, the fluctuation of the need for resolution which has been modelled using the pitch reduction theory of the Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1983) by Lerdahl and Krumhansl (2007) as well as by Herremans and Chew (2016). Regarding musical emotions, there is a lack of cognitive computational models but a plethora of applied models that perform classification and recommendation tasks, which utilise discrete and/or dimensional representations (Eerola & Vuoskoski, 2012). Algorithmic composition models aim to produce novel and valuable music while some are also cognitively plausible. Symbolic AI systems use rules

(Ebcioğlu, 1990) or generative grammars (Steedman, 1984) that guide the creation of new compositions, but despite the interpretability of these approaches they lack flexibility. Statistical machine learning models are based on the distribution assumption and include Markov chains (Ponsford et al., 1999) and other models that extend Markov models (Pachet, 2003) or combine statistical learning with symbolic AI (Kalikatsos-Papakostas et al., 2017). Neural networks have been used for music composition since the 80's (Todd, 1989), but their recent success has increased interest in them (Briot & Pachet, 2020), although the lack of interpretability renders their status as a cognitive model uncertain. Models of auditory attention attempt to capture the mechanisms of focusing resources only on relevant environmental stimuli (Kaya & Elhilali, 2017) and models of musical long-term memory deal with schema (Leman, 1995) and category formation (Marxer et al., 2007). Finally, the IDyOT cognitive architecture aims to integrate many areas of music cognition into one unified model that performs multiple tasks related to sequence processing (Wiggins, 2020).

Keywords:

Music, perception, cognition, computational, modelling

Βιβλιογραφικές αναφορές / References

- Ames, C. (1989). The Markov process as a compositional model: A survey and tutorial. *Leonardo*, 22(2), 175-187.
- Anagnostopoulou, C., & Cambouropoulos, E. (2017). Semiotic analysis and computational modelling: Two case studies on works by Debussy and Xenakis. In *Music semiotics: A network of significations* (pp. 129-146). Routledge.
- Anagnostopoulou, C., & Westermann, G. (1997, September). Classification in music: A computational model for paradigmatic analysis. In *ICMC*.
- Anders, T., & Miranda, E. R. (2011). Constraint programming systems for modeling music theories and composition. *ACM Computing Surveys (CSUR)*, 43(4), 1-38.
- Andjelkovic, I., Parra, D., & O'Donovan, J. (2019). Moodplay: interactive music recommendation based on artists' mood similarity. *International Journal of Human-Computer Studies*, 121, 142-159.
- Baroni, M. (1999). Musical grammar and the study of cognitive processes of composition. *Musicae Scientiae*, 3(1), 3-21.
- Baroni, M., & Jacoboni, C. (1978). *Proposal for a grammar of melody: The Bach chorales* (Vol. 4). Gaetan Morin Editeur Ltee.
- Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, 8(1), 10-14.
- Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Kippen, J., & Bel, B. (1989). The identification and modelling of a percussion 'language,' and the Emergence of Musical Concepts in a machine-learning experimental set-up. *Computers and the Humanities*, 23, 199-214.
- Benetos, E., Dixon, S., Duan, Z., & Ewert, S. (2018). Automatic music transcription: An overview. *IEEE Signal Processing Magazine*, 36(1), 20-30.
- Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5(1), 1-30.

- Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research*, 17, 289–308.
- Bogdanov, D., Serra, J., Wack, N., Herrera, P., & Serra, X. (2011). Unifying low-level and high-level music similarity measures. *IEEE Transactions on Multimedia*, 13(4), 687–701.
- Bregman, A. S. (1984, July). Auditory scene analysis. In *Proceedings of the 7th International Conference on Pattern Recognition* (pp. 168-175).
- Briot, J.-P., & Pachet, F. (2020). Deep learning for music generation: Challenges and directions. *Neural Computing and Applications*, 32(4), 981–993.
- Bruderer, M. J., Mckinney, M. F., & Kohlrausch, A. (2010). The perception of structural boundaries in polyphonic representations of Western popular music. *Musicae Scientiae*, 14(1_suppl), 115–155.
- Cambouropoulos, E. (2001, September). The local boundary detection model (LBDM) and its application in the study of expressive timing. In *ICMC* (p. 8).
- Cambouropoulos, E. (2003). Pitch spelling: A computational model. *Music Perception*, 20(4), 411–429.
- Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation. *Music Perception*, 23(3), 249–268.
- Cambouropoulos, E. (2008). Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception*, 26(1), 75–94.
- Cambouropoulos, E., Kaliakatsos-Papakostas, M. A., & Tsougras, C. (2014, September). An idiom-independent representation of chords for computational music analysis and generation. In *ICMC*.
- Cemgil, A. T., Desain, P., & Kappen, B. (2000). Rhythm quantization for transcription. *Computer Music Journal*, 24(2), 60-76.
- Cemgil, A. T., Kappen, B., Desain, P., & Honing, H. (2000). On tempo tracking: Tempogram Representation and Kalman filtering. *Journal of New Music Research*, 29(4), 259–273.
- Chew, E. (2002, August). The spiral array: An algorithm for determining key boundaries. In *International Conference on Music and Artificial Intelligence* (pp. 18-31). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Chew, E. (2014). Mathematical and computational modeling of tonality. *AMC*, 10(12), 141.
- Chew, E., & Chen, Y.-C. (2005). Real-time pitch spelling using the spiral array. *Computer Music Journal*, 29(2), 61–76.
- Conklin, D. (2003, April). Music generation from statistical models. In *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences* (pp. 30-35).
- Conklin, D., & Anagnostopoulou, C. (2001). Representation and discovery of multiple viewpoint patterns. In *ICMC* (pp. 479-485).
- Conklin, D., & Anagnostopoulou, C. (2006). Segmental pattern discovery in music. *INFORMS Journal on Computing*, 18(3), 285–293.
- Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1), 51–73.
- Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics*, 57, 451-462.
- de Berardinis, J., Vamvakaris, M., Cangelosi, A., & Coutinho, E. (2020). Unveiling the hierarchical structure of music by multi-resolution community detection. *Transactions of the International Society for Music Information Retrieval*, 3(1), 82-97.
- De Cheveigne, A. (2005). Pitch perception models. In Ch. J. Plack, R.R. Fay, A.J. Oxenham & A.N. Popper (Eds), *Pitch: Neural coding and perception* (pp. 169-233). Springer.
- De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4), 1917-1930.

- Deliège, I. (2007). Similarity relations in listening to music: How do they come into play?. *Musicae Scientiae*, 11(1_suppl), 9-37.
- Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, 116(20), 10113-10121.
- Ebcioğlu, K. (1990). An expert system for harmonizing chorales in the style of JS Bach. *The Journal of Logic Programming*, 8(1-2), 145-185.
- Eck, D., & Schmidhuber, J. (2002). *A first look at music composition using LSTM recurrent neural networks*.
- Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception*, 18(3), 275-296.
- Eerola, T., & Vuoskoski, J. K. (2012). A review of music and emotion studies: approaches, emotion models, and stimuli. *Music Perception*, 30(3), 307-340.
- Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3), 533-553.
- Eiben, A. E., Smith, J. E., Eiben, A. E., & Smith, J. E. (2015). What is an evolutionary algorithm?. *Introduction to Evolutionary Computing*, 25-48.
- Ens, J., & Pasquier, P. (2020). Quantifying musical style: Ranking symbolic music based on similarity to a style. *arXiv preprint arXiv:2003.06226*.
- Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12(1), 75-99.
- Fauconnier, G., & Turner, M. (2003). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- Fernandez, J. D., & Vico, F. (2013). AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research*, 48, 513-582.
- Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2, 139-172.
- Fletcher, H. (1924). The physical criterion for determining the pitch of a musical tone. *Physical Review*, 23(3), 427.
- Frankland, B. W., & Cohen, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. *Music Perception*, 21(4), 499-543.
- Franklin, J. A. (2001, March). Multi-phase learning for jazz improvisation and interaction. In *Proceedings of the Eight Biennial Symposium on Art and Technology* (pp. 51-60).
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. MIT Press.
- Gjerdingen, R. O. (1994). Apparent motion in music? *Music Perception*, 11(4), 335-370.
- Granroth-Wilding, M., & Steedman, M. (2014). A robust Parser-interpreter for jazz chord sequences. *Journal of New Music Research*, 43(4), 355-374.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23(3), 121-134.
- Grossberg, S. (2013). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 37, 1-47.
- Grossberg, S., & Rudd, M. E. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks*, 2(6), 421-450.
- Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5), 956.
- Hadjeres, G., Pachet, F., & Nielsen, F. (2017, July). Deepbach: a steerable model for bach chorales generation. In *International conference on machine learning* (pp. 1362-1371). PMLR.

- Hernandez-Olivan, C., & Beltran, J. R. (2021). *Music composition with Deep Learning: A review* (arXiv:2108.12290). arXiv.
- Herremans, D., & Chew, E. (2016). MorpheuS: Automatic music generation with recurrent pattern constraints and tension profiles. *2016 IEEE Region 10 Conference (TENCON)*, 282–285.
- Hiller, L., & Isaacson, L. (1959). *Experimental music. Composition with an electronic computer*. Greenwood Publishing Group Inc.
- Hofstadter, D. R. (2001). Epilogue: Analogy as the core of cognition. In D. Gentner, K.J. Holyoak, & B.N. Kokinov (Eds). *The analogical mind: perspectives from Cognitive Science*. MIT Press
- Homer, S. T., Harley, N., & Wiggins, G. A. (2024a). Modelling of musical perception using spectral knowledge representation. *Journal of Cognition*, 7(1).
- Homer, S. T., Harley, N., & Wiggins, G. A. (2024b). Contrast information dynamics: A novel information measure for cognitive modelling. *Entropy*, 26, 638.
- Honingh, A. K. (2009). Compactness in the Euler-Lattice: A parsimonious pitch spelling model. *Musicae Scientiae*, 13(1), 117–138.
- Honing, H. (2006). Computational modeling of music cognition: A case study on model selection. *Music Perception*, 23(5), 365–376.
- Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., ... & Eck, D. (2018). Music transformer. *arXiv preprint arXiv:1809.04281*.
- Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, 19(1), 1–64.
- Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.
- Jones, M. (2019). *Time will tell: A theory of dynamic attending*. Oxford University Press.
- Jones, M., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459–491.
- Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3), 217–238.
- Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of music and emotion: Theory, research, and applications*. Oxford University Press.
- Kaliakatsos-Papakostas, M., Queiroz, M., Tsougras, C., & Cambouropoulos, E. (2017). Conceptual blending of harmonic spaces for creative melodic harmonisation. *Journal of New Music Research*, 46(4), 305–328.
- Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714), 20160101.
- Kim, J., Urbano, J., Liem, C. C., & Hanjalic, A. (2020). One deep music representation to rule them all? A comparative analysis of different representation learning strategies. *Neural Computing and Applications*, 32(4), 1067–1093.
- Kirlin, P. B., & Utgoff, P. E. (2005). VOISE: Learning to segregate voices in explicit and implicit polyphony. In *ISMIR* (pp. 552–557).
- Klapuri, A. P. (2004). Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3), 269–282.
- Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*, 110(38), 15443–15448.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59–69.
- Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37, 52–65.

- Korzeniowski, F., & Widmer, G. (2017, August). End-to-end musical key estimation using a convolutional neural network. In *2017 25th European Signal Processing Conference (EU-SIPCO)* (pp. 966-970). IEEE.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.
- Krumhansl, C. L. (2004). The cognition of tonality – as we know it today. *Journal of New Music Research*, 33(3), 253–268.
- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89(4), 334.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119–159.
- Large, E. W., Kim, J. C., Flaig, N. K., Bharucha, J. J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Perception*, 33(3), 319–331.
- Larson, S. (2004). Musical forces and melodic expectations: Comparing computer models and experimental results. *Music Perception*, 21(4), 457–498.
- Lartillot, O., & Saint-James, E. (2004). Automating motivic analysis through the application of perceptual rules. *Music Query: Methods, Models, and User Studies*, 13, 73–92.
- Leman, M. (1995). A model of retroactive tone-center perception. *Music Perception*, 12(4), 439–471.
- Leman, M. (1995). *Music and schema theory: cognitive foundations of systematic musicology*. Springer Science & Business Media.
- Lerdahl, F. (1988). Tonal pitch space. *Music Perception*, 5, 315–350.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press.
- Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception*, 24(4), 329–366.
- Steedman, M., & Longuet-Higgins, H. C. (1971). On interpreting Bach. *Machine Intelligence*, 6(221), 221–241.
- Margulis, E. H. (2005). A model of melodic expectation. *Music Perception*, 22(4), 663–714.
- Margulis, E. H. (2014). *On repeat: How music plays the mind*. Oxford University Press, USA.
- Marxer, R., Holonowicz, P., Purwins, H., & Hazan, A. (2007). Dynamical hierarchical self-organization of harmonic, motivic, and pitch categories. *Music, Brain and Cognition. Part, 2*.
- Matsunaga, R., & Abe, J.-I. (2005). Cues for key perception of a melody. *Music Perception*, 23(2), 153–164.
- Mavromatis, P. (2009). Minimum description length modelling of musical structure. *Journal of Mathematics and Music*, 3(3), 117–136.
- McFee, B., & Ellis, D. P. (2014, May). Learning to segment songs with ordinal linear discriminant analysis. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5197-5201). IEEE.
- McFee, B., & Ellis, D. (2014, October). Analyzing song structure with spectral clustering. In *ISMIR* (pp. 405-410).
- Meredith, D. (2006). The *ps13* pitch spelling algorithm. *Journal of New Music Research*, 35(2), 121–159.
- Meredith, D., Lemström, K., & Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, 31(4), 321–345.
- Meyer, L. B. (1956). *Emotion and meaning in music*. Univ. of Chicago Press.
- Micchi, G., Gotham, M., & Giraud, M. (2020). Not all roads lead to Rome: Pitch representation and model architecture for automatic harmonic analysis. *Transactions of the International Society for Music Information Retrieval*, 3(1), 42–54.
- Milne, A. J., Laney, R., & Sharp, D. B. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception*, 32(4), 364–393.

- Minsky, M. (1988). *Society of mind*. Simon and Schuster.
- Mozer, M. C. (1994). Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing. *Connection Science*, 6(2-3), 247-280.
- Müllensiefen, D., & Frieler, K. (2004). Measuring melodic similarity: Human vs. Algorithmic Judgments. *Computing in Musicology*, 13, 147-176.
- Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. (pp. xiv, 485). University of Chicago Press.
- Nattiez, J. J., & Barry, A. (1982). Varèse's 'Density 21.5': A study in semiological analysis. *Music Analysis*, 1(3), 243-340.
- Nierhaus, G. (2009). *Algorithmic composition: paradigms of automated music generation*. Springer Science & Business Media.
- Norris, J. R. (1998). *Markov chains* (No. 2). Cambridge University Press.
- Pachet, F. (2003). The Continuator: Musical interaction with style. *Journal of New Music Research*, 32(3), 333-341.
- Pachet, F., Roy, P., & Barbieri, G. (2011). Finite-Length Markov processes with constraints. In *22nd International Joint Conference on Artificial Intelligence*.
- Panda, R., Malheiro, R., & Paiva, R. P. (2020). Audio features for music emotion recognition: A survey. *IEEE Transactions on Affective Computing*, 14(1), 68-88.
- Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, 39(10), 1367-1391.
- Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: the influence of context and learning. *Music Perception*, 23(5), 377-405.
- Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4), 625-652.
- Pereira, F. C., Grilo, C. F. A., Macedo, L., & Cardoso, F. A. B. (1997). Composing music with case-based reasoning. In *International Conference on Computational Models of Creative Cognition*.
- Pereplyotchik, D. (2017). *Psychosyntax: the nature of grammar and its place in the mind* (Vol. 129). Springer.
- Ponsford, D., Wiggins, G., & Mellish, C. (1999). Statistical learning of harmonic movement. *Journal of New Music Research*, 28(2), 150-177.
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715-734.
- Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4), 411-440.
- Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine?. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1598.
- Prince, J. B., & Schmuckler, M. A. (2012). The tonal-metric hierarchy: A corpus analysis. *Music Perception*, 31(3), 254-270.
- Purwins, H., Blankertz, B., & Obermayer, K. (2007). Toroidal models in tonal theory and pitch-class analysis. *Computing in Musicology*, 15.
- Purwins, H., Herrera, P., Grachten, M., Hazan, A., Marxer, R., & Serra, X. (2008). Computational models of music perception and cognition I: The perceptual and cognitive processing chain. *Physics of Life Reviews*, 5(3), 151-168.
- Raphael, C., & Stoddard, J. (2004). Functional harmonic analysis using probabilistic models. *Computer Music Journal*, 28(3), 45-52.

- Reddin, J., McDermott, J., & O'Neill, M. (2009). Elevated pitch: automated grammatical evolution of short compositions. In *applications of evolutionary computing: EvoWorkshops 2009: EvoCOMNET, EvoENVIRONMENT, EvoFIN, EvoGAMES, EvoHOT, EvoIASP, EvoINTERACTION, EvoMUSART, EvoNUM, EvoSTOC, EvoTRANSLOG*, Tübingen, Germany, April 15-17, 2009. *Proceedings* (pp. 579-584). Springer Berlin Heidelberg.
- Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1), 35-53.
- Rolland, P.-Y. (1999). Discovering patterns in musical sequences. *Journal of New Music Research*, 28(4), 334-350.
- Rosenthal, D. (1992). Emulation of human rhythm perception. *Computer Music Journal*, 16(1), 64.
- Rosenthal, D. F., & Okuno, H. G. (1998). Computational Auditory Scene Analysis. In *Proceedings of the Ijcai-95 Workshop*. CRC press.
- Rowe, R. (1992). Machine listening and composing with cypher. *Computer Music Journal*, 16(1), 43-63.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161.
- Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, 4.
- Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception*, 14(3), 295-318.
- Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception*, 16(3), 295-326.
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5(8), 340-348.
- Snyder, B. (2000). *Music and memory: An introduction*. MIT press.
- Steedman, M. J. (1984). A Generative Grammar for Jazz Chord Sequences. *Music Perception*, 2(1), 52-77.
- Taylor, I., & Greenhough, M. (1994). Modelling pitch perception with adaptive resonance theory artificial neural networks. *Connection Science*, 6(2-3), 135-154.
- Temperley, D. (2001). *The Cognition of Basic Musical Structures*. (pp. xvi, 404). The MIT Press.
- Temperley, D. (2007). *Music and probability*. MIT Press.
- Temperley, D. (2012). Computational models of music cognition. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 327-368). Elsevier.
- Temperley, D., & Sleator, D. (1999). Modeling meter and harmony: A preference-rule approach. *Computer Music Journal*, 23(1), 10-27.
- Tenney, J., & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory*, 24(2), 205.
- Terhardt, E. (1974). Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, 55(5), 1061-1069.
- Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, 107(4), 885-913.
- Todd, P. (1989). A Connectionist approach to algorithmic composition. *Computer Music Journal*, 13(4), 27-43.
- Toiviainen, P. (1995). Modeling the target-note technique of bebop-style jazz improvisation: An artificial neural network approach. *Music Perception*, 12(4), 399-413.
- Toiviainen, P., & Eerola, T. (2002). A computational model of melodic similarity based on multiple representations and self-organizing maps. In *Proceedings of the seventh international conference on music perception and cognition*, Sydney. Causal Productions, Adelaide (pp. 236-239).

- Tolonen, T., & Karjalainen, M. (2000). A computationally efficient multipitch analysis model. *IEEE Transactions on Speech and Audio Processing*, 8(6), 708-716.
- Typke, R., Giannopoulos, P., Veltkamp, R. C., Wiering, F., & van Oostrum, R. (2003). Using transportation distances for measuring melodic similarity. *ISMIR*
- Ukkonen, E. (1995). On-line construction of suffix trees. *Algorithmica*, 14(3), 249-260.
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Wallas, G. (1926). The art of thought. *Franklin Watts*.
- Weiner, P. (1973, October). Linear pattern matching algorithms. In *14th Annual Symposium on Switching and Automata Theory (swat 1973)* (pp. 1-11). IEEE.
- Wertheimer, M. (2012). Experimental studies on seeing motion. *Zeitschrift für Psychologie*, 61(1), 161-265. In L. Spillmann (Ed), *On perceived motion and figural organization*
- Whorley, R. P., Wiggins, G. A., Rhodes, C., & Pearce, M. T. (2013). Multiple viewpoint systems: Time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. *Journal of New Music Research*, 42(3), 237-266.
- Widmer, G. (1992). Qualitative perception modeling and intelligent musical learning. *Computer Music Journal*, 16(2), 51-68.
- Wiggins, G. A. (2020). Creativity, information, and consciousness: The information dynamics of thinking. *Physics of Life Reviews*, 34-35, 1-39.
- Wiggins, G. A., & Sanjekdar, A. (2019). Learning and consolidation as re-representation: Revisiting the meaning of memory. *Frontiers in Psychology*, 10, 802.
- Wiggins, G., & Smaill, A. (2013). Musical knowledge: what can Artificial Intelligence bring to the musician?. In *Readings in music and artificial intelligence* (pp. 29-46). Routledge.
- Winograd, T. (1968). Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory*, 12(1), 2.
- Yang, X., Dong, Y., & Li, J. (2018). Review of data features-based music emotion recognition methods. *Multimedia Systems*, 24(4), 365-389.
- Zioga, I., Harrison, P. M., Pearce, M. T., Bhattacharya, J., & Luft, C. D. B. (2020). From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity. *NeuroImage*, 206, 116311.

Ο Γιώργος Βελισσαρίδης έλαβε το πτυχίο του από το Τμήμα Μουσικών Σπουδών του ΕΚΠΑ, όπου ακολούθησε την κατεύθυνση μουσικής τεχνολογίας. Η πτυχιακή του εργασία, υπό την επίβλεψη της Χριστίνας Αναγνωστοπούλου, αφορούσε την υπολογιστική μουσική νόηση και την ενσωμάτωση του μοντέλου αντίληψης και νόησης ακολουθιών IDyOT με το μοντέλο κινητοποίησης και συναισθήματος MicroPsi. Επί του παρόντος σπουδάζει πληροφορική στο ΟΠΑ και έχει γίνει δεκτός σε διδακτορικό πρόγραμμα μουσικής και τεχνητής νοημοσύνης στο Queen Mary University of London. Έχει βαθύ ενδιαφέρον για την υπολογιστική γνωσιακή επιστήμη ως τρόπο να κατανοήσουμε τους εαυτούς μας και να δημιουργήσουμε τεχνητούς νοήμονες πράκτορες.

Yorgos Velissaridis got his MA from NKUA's Music Studies Department, where he studied music technology. His thesis, under the supervision of Christina Anagnostopoulou, was on computational music cognition and the integration of the IDyOT model of sequence perception and cognition with the MicroPsi model of motivation and emotion. He is currently studying for a BSc in computer science at AUEB and has been admitted to a doctoral programme in music and AI at Queen Mary University of London. He has a deep interest in computational cognitive science as a means for understanding ourselves and building artificial intelligent agents.

Η **Χριστίνα Αναγνωστοπούλου** είναι αναπληρώτρια καθηγήτρια στο Τμήμα Μουσικών Σπουδών του ΕΚΠΑ, και διευθύντρια του Εργαστηρίου Μουσικής, Γνωσιακών Επιστημών και Κοινότητας. Έχει σπουδάσει Μουσικολογία και Τεχνητή Νοημοσύνη, και έχει διδάξει στα Πανεπιστήμια του Εδιμβούργου, της Γλασκώβης και του Queen's Belfast. Στο Μπέλφαστ έλαβε το βραβείο καλύτερης διδασκαλίας, τελείωσε το μεταπτυχιακό στα παιδαγωγικά της Τριτοβάθμιας Εκπαίδευσης, και μονιμοποιήθηκε. Στην Αθήνα ήρθε το 2006, και από τότε εργάζεται ως μέλος ΔΕΠ στο Πανεπιστήμιο Αθηνών. Τα ερευνητικά της ενδιαφέροντα σχετίζονται με την υπολογιστική μουσικολογία, μουσική στην κοινότητα, γνωσιακές επιστήμες στη μουσική και γενικότερα τη συστηματική μουσικολογία.

Christina Anagnostopoulou is an associate professor at the Department of Music Studies, National and Kapodistrian University of Athens, and the Director of the Music Cognition, Computation and Community Lab (MCCC Lab). She studied Music and Artificial Intelligence, and she has taught at the Universities of Edinburgh, Glasgow, and Queen's Belfast on various topics related to music and AI. In 2005 she received the Teaching Award at Queen's Belfast, she graduated from her postgraduate degree in Higher Education teaching (PGCHET) and got tenured. She came to Athens in 2006, and has been working at the University of Athens since. Her research interests are related to computational musicology, music in the community, cognitive sciences of music and in general to systematic musicology.