

Σύγκριση και Ομαδοποίηση Αμφιωτικών Χωρικών Φίλτρων (HRTF) με Χρήση Μεθόδων Μεταεπεξεργασίας και Μείωσης Διαστάσεων

Κωνσταντίνος Μπακογιάννης*, Αρετή Ανδρεοπούλου
Εργαστήριο Μουσικής Ακουστικής Τεχνολογίας (LabMAT)
Τμήμα Μουσικών Σπουδών, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
*k.bakogiannis@music.uoa.gr

ΠΕΡΙΛΗΨΗ

Η παρούσα μελέτη επικεντρώνεται στη σύγκριση και ομαδοποίηση δεδομένων HRTF με χρήση τεχνικών μεταεπεξεργασίας και μείωσης διαστάσεων. Για τη διασφάλιση συνέπειας και συγκρισιμότητας, εφαρμόσαμε βήματα μεταεπεξεργασίας όπως επαναδειγματοληψία, ζωνοπερατό φίλτράρισμα και κανονικοποίηση στάθμης. Υπολογίστηκαν δύο βασικές μετρικές, η αντιληπτική ομοιότητα και η ακρίβεια εντοπισμού, είτε σε κοινές θέσεις μέτρησης μεταξύ ζευγών datasets είτε συνολικά για κάθε dataset. Χρησιμοποιήθηκε η Ανάλυση Κύριων Συνιστωσών (PCA) για τη μείωση των διαστάσεων των δεδομένων, επιτρέποντας την ομαδοποίηση των datasets βάσει βασικών χαρακτηριστικών. Επιπλέον, υλοποιήθηκε φασματική ομαδοποίηση για την ανίχνευση ομάδων HRTFs και τον προσδιορισμό αντιπροσωπευτικών σημείων (centroids). Η μεθοδολογία δοκιμάστηκε σε 72 HRTF datasets από έξι βάσεις δεδομένων. Τα αποτελέσματα έδειξαν ότι επιτυγχάνεται σημαντική μείωση των διαστάσεων διατηρώντας πάνω από το 90% της διακύμανσης των δεδομένων.

Comparison and Clustering of HRTF Datasets through Post-Processing and Dimensionality Reduction Methods

ABSTRACT

This paper presents the database matching technique for optimizing personalized Head-Related Transfer Function (HRTF) selection from publicly available datasets. The goal is to identify the most suitable HRTF dataset for each user, enhancing the immersive audio experience in virtual and augmented reality applications. Two key metrics, perceptual similarity and localization accuracy, were calculated for all datasets, either at common spatial positions between dataset pairs or across each dataset independently. Principal Component Analysis (PCA) was used to reduce data dimensions, enabling clustering of datasets based on key characteristics. Additionally, a spectral clustering approach was employed to group HRTFs and identify representative ones (centroids). The methodology was tested on 72 HRTF datasets from six databases. Results demonstrated that significant dimensionality reduction is possible while preserving over 90% of the data variance.

Εισαγωγή

Η εμπυθιστική εμπειρία και η ακριβής απόδοση της χωρικής πληροφορίας του ηχητικού περιβάλλοντος αποτελούν βασικές προϋποθέσεις για μια επιτυχημένη, ολοκληρωμένη, ρεαλιστική και ευχάριστη εμπειρία αμφιωτικής ακρόασης, ειδικά σε εφαρμογές εκτεταμένης πραγματικότητας (extended reality). Στόχος είναι μια προσωποποιημένη εμπειρία, η οποία να προσεγγίζει όσο το δυνατόν περισσότερο τον πραγματικό τρόπο με τον οποίο το κάθε άτομο αντιλαμβάνεται τη χωρική ηχητική πληροφορία στο φυσικό περιβάλλον. Αυτό επιτυγχάνεται μέσω της χρήσης ψηφιακών αμφιωτικών χωρικών φίλτρων, τα οποία ενσωματώνουν την κρουστική απόκριση των κυριότερων δομών του ανθρώπινου σώματος, όπως τα αυτιά, το κεφάλι και ο κορμός, τα οποία είναι υπεύθυνα για τη χωρική ακουστική αντίληψη. Τα φίλτρα αυτά, που είναι μοναδικά για κάθε άτομο, μπορούν να χρησιμοποιηθούν είτε στο πεδίο του χρόνου είτε στο πεδίο των συχνοτήτων, ως Head Related Impulse Responses (HRIR) και Head Related Transfer Functions (HRTF) αντίστοιχα.

Μελέτες έχουν δείξει ότι τα HRTF που προέρχονται από την ίδια βάση δεδομένων εμφανίζουν συχνά κοινά χαρακτηριστικά, τα οποία οφείλονται στην επίδραση της διαδικασίας μέτρησης [1], [2]. Η εργασία αυτή επικεντρώνεται στις τεχνικές μετα-επεξεργασίας των δεδομένων με στόχο τον περιορισμό της επίδρασης των μεθόδων ηχογράφησης, ώστε να διευκολυνθεί η ακριβέστερη σύγκριση των HRTF μεταξύ διαφορετικών βάσεων δεδομένων. Ζητούμενο πολλών εργασιών είναι η ομαδοποίηση και η σύγκριση των διαφορετικών HRTF datasets με βάση κάποια κοινά χαρακτηριστικά και ιδιότητές τους. Το παραπάνω απαιτεί την ποσοτικοποίηση της επίδρασής τους σε αντιληπτές παραμέτρους, η οποία επιτυγχάνεται μέσω της χρήσης κατάλληλων μετρικών (π.χ., [3], [4], [5]). Δεδομένης της πληθώρας διαθέσιμων HRTF, τα οποία περιγράφονται από διαφορετικές μετρικές και παρουσιάζουν διαφορές ανά συχνοτική περιοχή ή/και χωρικό σημείο, οι τεχνικές σύγκρισης και ομαδοποίησης καλούνται να χειριστούν δεδομένα πολλών διαστάσεων. Για την καλύτερη επεξεργασία αυτών των δεδομένων, εφαρμόζονται τεχνικές μείωσης διαστάσεων, οι οποίες περιγράφονται στην εργασία.

1. Μεθοδολογία

Στην παρακάτω ενότητα θα παρουσιαστούν οι γενικές αρχές των τεχνικών σύγκρισης και ομαδοποίησης των HRTFs καθώς και οι λεπτομέρειες της υλοποίησής τους.

1.1 Συλλογή και μετα-επεξεργασία δεδομένων

Για τη συλλογή των HRTFs, αξιοποιήθηκε το αποθετήριο της SOFA, το οποίο περιλαμβάνει ανοιχτές βιβλιοθήκες HRTF σε ενιαίο τύπο αρχείου, με συγκεκριμένη παραμετροποίηση. [6] Αυτή η κοινή μορφή διευκολύνει την επεξεργασία και την αξιοποίηση των δεδομένων. Επιλέχθηκαν βιβλιοθήκες στις οποίες τα HRTFs είχαν προκύψει από ηχογραφήσεις πραγματικών υποκειμένων, διασφαλίζοντας έτσι ρεαλιστικές κρουστικές αποκρίσεις. Για την ανάλυση χρησιμοποιήθηκαν εκδοχές των HRTFs που είχαν υποστεί ισοστάθμιση ελεύθερου πεδίου (free-field equalization),

ώστε να ελαχιστοποιηθεί η επίδραση των μικροφώνων και των άλλων μετρητικών διατάξεων.

Για να εξασφαλιστεί η συγκρισιμότητα των HRTFs, υιοθετήσαμε μια σειρά από βήματα μετα-επεξεργασίας. Μέσω αυτής της διαδικασίας διασφαλίστηκε ότι τα HRTFs θα διαθέτουν τα αναγκαία κοινά χαρακτηριστικά, τα οποία με ενιαίο τρόπο θα επιτρέψουν την εφαρμογή τους στα επόμενα στάδια του υπολογισμού των μετρικών. Παράλληλα, οι τεχνικές αυτές συνέβαλαν στον περιορισμό της επίδρασης της μετρικής διάταξης και των διαφορετικών μεθόδων ηχογράφησης, βελτιώνοντας έτσι την ακρίβεια και την αξιοπιστία των συγκρίσεων.

Τα βήματα της μετα-επεξεργασίας που υιοθετήσαμε σύμφωνα και με τη διαθέσιμη βιβλιογραφία ([1], [2]) είναι τα εξής:

- **Μετατροπή των HRTFs σε DTFs:** Τα DTFs (Directional Transfer Functions) επικεντρώνονται στα κατευθυντικά στοιχεία του ήχου, αφαιρώντας μη κατευθυντικές επιδράσεις που εξαρτώνται από τη συχνότητα.
- **Μετασηματισμός σε κοινή συχνότητα δειγματοληψίας (Uniform Frequency Resampling):** Επιλέξαμε τη χαμηλότερη συχνότητα δειγματοληψίας (44.100 Hz) μεταξύ των διαθέσιμων HRTF.
- **Περιορισμός συχνοτικού εύρους με ζωνο-περατό (band-pass) φίλτρο:** Επιλέξαμε ένα βαθυπερατό φίλτρο με συχνότητα αποκοπής τα 8 kHz, η οποία διατηρεί τα βασικά χωρικά στοιχεία, συμπεριλαμβανομένης της φασματικής πληροφορίας που σχετίζεται με το πτερύγιο του αυτιού, ενώ εξαλείφει τον υψηλής συχνότητας θόρυβο που εισάγεται από τον εξοπλισμό ηχογράφησης. Επιπλέον, εφαρμόστηκε ένα υψυπερατό φίλτρο με αποκοπή στα 200 Hz, για να αφαιρέσει τα χαμηλής συχνότητας artifacts (π.χ. θόρυβος δωματίου, βουητό εξοπλισμού). Οι συχνότητες κάτω των 200 Hz μεταφέρουν λιγότερη λεπτομερή πληροφορία για τη διαφορά χρόνου άφιξης μεταξύ των αυτιών (Interaural Time Difference - ITD) και συμβάλλουν ελάχιστα στη λεπτομερή χωρική εντόπιση. Αυτό το ζωνοπερατό φίλτρο (200 Hz έως 8 kHz) συλλαμβάνει το κρίσιμο φάσμα για την ακριβή ανάλυση του χωρικού ήχου, διατηρώντας τόσο τα στοιχεία ITD και διαφοράς στάθμης μεταξύ των αυτιών (Interaural Level Difference - ILD), όσο και τη φασματική απόκριση από το πτερύγιο του αυτιού για οριζόντιο και κατακόρυφο εντοπισμό.
- **Αφαίρεση DC Offset:** Εξισορρόπηση του σήματος γύρω από το μηδέν για αποφυγή παραμορφώσεων στις χαμηλές συχνότητες.
- **Περικοπή Κρουστικής Απόκρισης (Impulse Response Truncation):** Εφαρμόστηκε ορθογώνιο παράθυρο για την περικοπή της κρουστικής απόκρισης (IR) σε ένα ομοιόμορφο μήκος. Το μήκος των δειγμάτων βασίστηκε στη μικρότερη αναφερόμενη διάρκεια κρουστικής απόκρισης (5.3 ms, που αντιστοιχεί σε 233.73 δείγματα). Για λόγους υπολογιστικής αποδοτικότητας, το ομοιόμορφο μήκος ορίστηκε σε 256 δείγματα (η επόμενη δύναμη του δύο). Το σημείο έναρξης ορίστηκε 40 δείγματα πριν από την πρώτη ανίχνευση της έναρξης, η οποία ορίστηκε ως το σημείο όπου το σήμα ξεπέρασε κατά 10 dB την κορυφαία τιμή. Αυτό εξάλειψε τις πρώιμες ανακλάσεις και εξασφάλισε ότι μόνο η άμεση διαδρομή του ήχου καταγράφηκε.
- **Κανονικοποίηση Επίπεδου (Level Gains Normalization):** Λόγω διαφορών στο πλάτος των σημάτων τα οποία ηχογραφήθηκαν και στην απόσταση των ηχείων

μεταξύ των διατάξεων, Εφαρμόστηκε κανονικοποίηση RMS σε όλα τα HRIR, ώστε να διασφαλιστεί ομοιόμορφη ένταση ήχου μεταξύ των datasets.

- **Μηδενισμός πρώτου και τελευταίου δείγματος (Zeroing Boundary Samples):** Αυτό εξασφαλίζει καθαρά όρια στο σήμα και προσφέρει ομαλές μεταβάσεις κατά τη διάρκεια της συνέλιξης, αποτρέποντας "κλικ" ή ασυνέχειες στο τελικό ηχητικό αποτέλεσμα.

1.2 Επιλογή μετρικών

Μετά τη δημιουργία ενός συνόλου μετα-επεξεργασμένων HRTF datasets, με στόχο τη διασφάλιση της συνέπειας και της συγκρισιμότητας, το επόμενο βήμα είναι ο υπολογισμός των βασικών μετρικών. Αυτές οι μετρικές είτε εφαρμόζονται στις κοινές θέσεις μέτρησης μεταξύ δύο datasets, είτε σε ολόκληρο το dataset. Χρησιμοποιήσαμε δύο κύριες μετρικές από το Auditory Modelling Toolbox: μία μετρική αντιληπτικής ομοιότητας (perceptual similarity) και μία μετρική ακρίβειας εντοπισμού ήχου (localization accuracy).

- **Μετρική Αντιληπτικής Ομοιότητας:** Βασισμένη στη Binaural Perceptual Similarity των McKenzie κ.α. [7] (διαθέσιμη στο AMT McKenzie2021), η μετρική αυτή υπολογίζεται για κάθε κοινή θέση μεταξύ δύο datasets. Παρέχει ένα μέτρο του πόσο παρόμοια αντιλαμβάνεται ο χρήστης τις φασματικές διαφορές των δύο HRTF, λαμβάνοντας υπόψη και τα δύο αυτιά και παρέχοντας αποτελέσματα για συγκεκριμένες θέσεις ακρόασης.
- **Μετρική Ακρίβειας Εντοπισμού Ήχου:** Αυτή η μετρική, βασισμένη στο Bayesian Spherical Sound Localization των Barumerli κ.α. [8](διαθέσιμη στο AMT Barumerli2023), εφαρμόζεται στο σύνολο του dataset. Αξιολογεί το πόσο ένα HRTF dataset οδηγεί στον ακριβή εντοπισμό ηχητικών πηγών στον σφαιρικό ακουστικό χώρο, χρησιμοποιώντας προχωρημένα Bayesian μοντέλα για την πρόβλεψη λαθών εντοπισμού.

Αυτές οι μετρικές παρέχουν πολύτιμες ακουστικές πληροφορίες και μπορούν να αξιοποιηθούν σε ένα ευρύ φάσμα εφαρμογών, που διακρίνονται σε δύο κατηγορίες. Η πρώτη μετρική μπορεί να χρησιμοποιηθεί ως εργαλείο σε εφαρμογές όπου η διατήρηση της φυσικότητας των ηχοχρωμάτων είναι καθοριστική, αποφεύγοντας αλλοιώσεις που ενδέχεται να προκαλέσουν δυσαρέσκεια ή έλλειψη φυσικότητας κατά την ακρόαση (π.χ., εφαρμογές μουσικής ακρόασης). Η δεύτερη μετρική είναι κατάλληλη για εφαρμογές στις οποίες ο ακριβής εντοπισμός της προέλευσης του ήχου είναι κρίσιμος (π.χ., χειρισμός μηχανημάτων, περιηγήσεις σε χώρους, video games κ.λπ.).

Παρακάτω θα εξειδικεύσουμε τις διαφορές στην υλοποίηση που εντοπίζονται μεταξύ των μετρικών που εφαρμόζονται στα κοινά σημεία δύο datasets και των μετρικών που εφαρμόζονται σε ολόκληρο το dataset.

1.3 Σύγκριση datasets με μετρικές που εφαρμόζονται στα μεταξύ τους κοινά σημεία

Οι μετρικές αυτές συγκρίνουν την επίδραση δύο διαφορετικών HRTF datasets και εξάγουν διαφορετικές τιμές για τα κοινά σημεία των dataset αυτών. Το πρώτο βήμα επομένως σε κάθε δυαδική σύγκριση είναι ο εντοπισμός των κοινών σημείων. Για να

ληφθεί υπόψη η αδυναμία του ανθρώπου να διακρίνει ότι δύο πηγές εκπέμπουν από διαφορετικά σημεία του χώρου αν η απόστασή τους είναι μέχρι μια συγκεκριμένη τιμή (φαινόμενο γνωστό ως localization blur), εφαρμόζουμε μια ανεκτική μέθοδο για τον καθορισμό κοινών θέσεων μεταξύ των HRTF datasets. Με βάση αυτό θεωρούνται ως κοινά σημεία όσων η απόσταση είναι χαμηλότερη κάποιων τιμών οι οποίες καθορίζονται ανάλογα με την αζιμούθια θέση τους και το επίπεδο ανύψωσης. Για τον καθορισμό των ορίων αυτών χρησιμοποιούνται οι τιμές στη σχετική βιβλιογραφία για το μέγεθος Minimum Audible Angle. [9], [10] Όταν οι θέσεις από το dataset A αντιστοιχούν σε πολλαπλές θέσεις από το dataset B (εντός των ορίων), υπολογίζουμε τον μέσο όρο των μετρικών για να αποφύγουμε την υπερ-αντιπροσώπευση. Αν υπάρχει ακριβής αντιστοιχία, προτιμούμε αυτή έναντι των λιγότερο ακριβών αντιστοιχιών.

Όλες οι μετρικές κανονικοποιούνται με βάση τις ελάχιστες και μέγιστες τιμές που βρέθηκαν μεταξύ όλων των συγκρίσεων κάθε ζεύγους datasets, για να εξασφαλιστεί συνέπεια. Οι κανονικοποιημένες τιμές κυμαίνονται από 0 έως 1, επιτρέποντας την άμεση σύγκριση μεταξύ διαφορετικών μετρικών. Τέλος, εφαρμόζουμε ένα βάρος στις αντίστοιχες μετρικές, διασφαλίζοντας ότι οι θέσεις που καλύπτουν μεγαλύτερη περιοχή επηρεάζουν περισσότερο τα τελικά αποτελέσματα. Κάθε κοινή θέση αποκτά βάρος αναλογικά με την περιοχή (στερεά γωνία) που αντιπροσωπεύει. Για να προσδιορίσουμε το βάρος αυτό χρησιμοποιούμε τον αλγόριθμο 1-NN για να χωρίσουμε τη σφαίρα σε συστάδες και να προσδιορίσουμε τον αριθμό των σημείων της σφαίρας τα οποία είναι εγγύτερα σε κάθε κοινή θέση.

1.4 Σύγκριση datasets με μετρικές που εφαρμόζονται σε ολόκληρα το dataset

Σε αυτή την προσέγγιση, οι μετρικές εφαρμόζονται όχι μέσω της σύγκρισης δύο HRTF datasets, αλλά υπολογίζονται απευθείας για κάθε dataset ανεξάρτητα. Κάθε dataset περιγράφεται από ένα σύνολο μετρικών που αντιπροσωπεύουν τις συνολικές του ιδιότητες, χωρίς να γίνεται αναφορά σε κοινές θέσεις ή σε συγκρίσεις με άλλα datasets.

Αντί να επικεντρωνόμαστε σε κοινές θέσεις μεταξύ δύο datasets, κάθε dataset αναλύεται συνολικά, και για κάθε dataset εξάγεται μία μοναδική τιμή για κάθε μετρική, η οποία συνοψίζει τα ακουστικά χαρακτηριστικά του, δίνοντας μία συνολική εκτίμηση της ακουστικής απόδοσης του HRTF.

Ο πίνακας που προκύπτει έχει διαστάσεις $d \times m$, όπου d είναι ο αριθμός των datasets και m είναι ο αριθμός των μετρικών που χρησιμοποιούνται. Αυτός ο πίνακας επιτρέπει την άμεση σύγκριση μεταξύ των datasets, χωρίς την ανάγκη ανεύρεσης κοινών θέσεων, καθώς οι μετρικές αναφέρονται σε ολόκληρο το dataset.

1.5 Ομαδοποίηση datasets με μετρικές που εφαρμόζονται στα μεταξύ τους κοινά σημεία

Τα μοναδικά ζεύγη HRTFs που σχηματίζουν τις δυαδικές συγκρίσεις είναι $d*(d-1)/2$, όπου d είναι ο αριθμός των datasets. Αν p είναι το σύνολο των κοινών θέσεων που εντοπίζονται μεταξύ όλων των δυαδικών συγκρίσεων και m ο αριθμός των μετρικών, τότε ο πίνακας που περιλαμβάνει όλες τις συγκρίσεις έχει διαστάσεις $d*(d-1)/2 \times p \times m$. Δεδομένου ότι δεν έχουν όλες οι συγκρίσεις τα ίδια κοινά σημεία, τα κενά στον πίνακα συμπληρώνονται με τεχνικές αναπλήρωσης, όπως η χρήση

μηδενικών ή μέσων όρων, για να διατηρηθεί η συνοχή των δεδομένων και να μπορεί να εφαρμοστεί η τεχνική μείωσης δεδομένων PCA (Principal Component Analysis), η οποία έχει εφαρμοστεί ξανά σε έρευνες σχετικές με HRTF (π.χ., [11], [12], [13], [14]).

Η PCA εφαρμόζεται στα δεδομένα που προκύπτουν από τις δυαδικές συγκρίσεις μεταξύ των HRTF datasets, με σκοπό να μειώσει τις διαστάσεις του πίνακα δεδομένων, διατηρώντας τις κύριες συνιστώσες (principal components) που συγκρατούν το μεγαλύτερο μέρος της πληροφορίας και της διακύμανσης των δεδομένων. Στην πράξη, ο αριθμός των κύριων συνιστωσών k που διατηρούνται εξαρτάται από το πόσο επιθυμούμε να μειώσουμε τις διαστάσεις και πόση διακύμανση θέλουμε να διατηρήσουμε (π.χ., 90% της συνολικής διακύμανσης). Μετά την PCA, κάθε ζεύγος σύγκρισης αντιπροσωπεύεται από k κύριες συνιστώσες που συνοψίζουν την πληροφορία από τις διάφορες μετρικές.

Αυτή η διαδικασία μπορεί να επεκταθεί ώστε να προκύψουν συμπεράσματα για τα ίδια τα datasets. Αρχικά, τα αποτελέσματα της PCA για τις δυαδικές συγκρίσεις ομαδοποιούνται με χρήση τεχνικών clustering (όπως το k -means) για να δημιουργηθεί ένας πίνακας ομοιοτήτων (similarity matrix) μεταξύ των datasets. Κάθε ζεύγος datasets αποκτά ένα βαθμό ομοιότητας, ανάλογα με το πόσο συχνά ταξινομείται στην ίδια ομάδα μέσω του clustering.

Στη συνέχεια, χρησιμοποιείται φασματική ομαδοποίηση (spectral clustering) πάνω στον πίνακα ομοιοτήτων. Δημιουργείται ένας γράφος (graph) όπου οι κόμβοι αντιστοιχούν στα datasets και οι ακμές αποτυπώνουν τον βαθμό ομοιότητας μεταξύ τους. Η φασματική ανάλυση του γράφου (μέσω του Laplacian matrix) επιτρέπει τον εντοπισμό φυσικών ομάδων μεταξύ των datasets, που παρουσιάζουν κοινά χαρακτηριστικά. Έτσι, καταλήγουμε σε clusters των datasets, και κάθε ομάδα (cluster) αντιπροσωπεύεται από έναν κεντρικό εκπρόσωπο (centroid), ο οποίος είναι το dataset που βρίσκεται πιο κοντά στο κέντρο της ομάδας και θεωρείται το πιο αντιπροσωπευτικό για την ομάδα του.

Με αυτόν τον τρόπο, μέσω της συνδυασμένης εφαρμογής PCA και φασματικής ομαδοποίησης, μπορούμε να προσδιορίσουμε τα αντιπροσωπευτικά HRTF datasets, τα οποία μπορούν στη συνέχεια να χρησιμοποιηθούν για περαιτέρω ανάλυση ή δοκιμή.

1.6 Ομαδοποίηση datasets με μετρικές που εφαρμόζονται σε ολόκληρα το dataset

Αντίθετα με την περίπτωση των μετρικών που εφαρμόζονται στα μεταξύ των datasets κοινά σημεία, στην περίπτωση των μετρικών που εφαρμόζονται σε ολόκληρο το dataset δε χρειάζεται κάποια προεργασία για την ομαδοποίηση των datasets. Η τεχνική PCA (Principal Component Analysis) μπορεί άμεσα να εφαρμοστεί για τη μείωση των διαστάσεων των δεδομένων, όπως ακριβώς και στην περίπτωση των συγκρίσεων σε κοινές θέσεις. Η PCA θα εξάγει ένα σύνολο κύριων συνιστωσών (principal components) που αντιπροσωπεύουν τη μεγαλύτερη διακύμανση στα δεδομένα, επιτρέποντας την ομαδοποίηση (clustering) των datasets σε ομάδες βάσει των ακουστικών χαρακτηριστικών τους. Έτσι, ο αρχικός πίνακας που με διαστάσεις $d \times m$, όπου d είναι ο αριθμός των datasets και m είναι ο αριθμός των μετρικών που χρησιμοποιούνται, εκπίπτει σε πίνακα διαστάσεων $d \times n$, όπου n ο αριθμός των principal components και $n < m$.

Οι αλγόριθμοι ομαδοποίησης, όπως το k-means ή το hierarchical clustering, μπορούν να χρησιμοποιηθούν για να αναζητηθούν φυσικές ομάδες των HRTF datasets, με βάση τις κύριες συνιστώσες που προέκυψαν από την PCA. Οι ομάδες αυτές αντιπροσωπεύουν HRTF με κοινά χαρακτηριστικά και μπορούν να συνοψιστούν μέσω των centroids τους, τα οποία αποτελούν τα πιο αντιπροσωπευτικά HRTF για κάθε ομάδα.

2. Αποτελέσματα μελέτης περίπτωσης

Ως μελέτη περίπτωσης, εφαρμόσαμε την παραπάνω μεθοδολογία σε 72 διαφορετικά HRTF datasets, 12 από κάθε μία από 6 διαφορετικές βάσεις δεδομένων. Για τη μετρική της αντιληπτικής ομοιότητας, επιλέξαμε δύο τιμές ($m = 2$): τον μέσο όρο αντιληπτικής ομοιότητας μεταξύ των δύο αυτιών και την τιμή για το αυτί που βρίσκεται πιο κοντά στην πηγή (ipsilateral). Η μετρική αυτή υπολογίζεται στα κοινά σημεία των ζευγών σύγκρισης $(72 \cdot (72-1)/2 = 2556$ δυαδικές συγκρίσεις). Ο συνολικός αριθμός κοινών θέσεων που εντοπίστηκαν (p) ήταν 428, οδηγώντας σε έναν πίνακα διαστάσεων $2556 \times 428 \times 2$.

Η τεχνική PCA (Principal Component Analysis) εφαρμόστηκε για μείωση των διαστάσεων, αναλύοντας τη διακύμανση σε 856 κύριες συνιστώσες. Από αυτές, οι πρώτες 10 εξηγούν το 75% της συνολικής διακύμανσης, ενώ το 90% εξηγείται από τις 24 πρώτες συνιστώσες. Αυτό σημαίνει ότι καταφέραμε να μειώσουμε τις διαστάσεις από 2556×856 σε 2556×24 , διατηρώντας το 90% της διακύμανσης των δεδομένων.

Αντίστοιχα, για τη μετρική της ακρίβειας εντοπισμού ήχου, η οποία υπολογίζεται για κάθε ένα HRTF dataset, προκύπτουν 5 τιμές, δημιουργώντας έναν πίνακα διαστάσεων 72×5 . Η τεχνική PCA (Principal Component Analysis) ανέλυσε τη διακύμανση σε 5 κύριες συνιστώσες. Από αυτές, η πρώτη εξηγεί το 58% της συνολικής διακύμανσης, ενώ οι πρώτες τρεις εξηγούν το 97%. Αυτό σημαίνει ότι μειώσαμε τις διαστάσεις από 72×5 σε 72×3 , διατηρώντας το 97% της διακύμανσης των δεδομένων και χάνοντας μόνο το 3%.

3. Συμπεράσματα

Στην παρούσα εργασία, παρουσιάσαμε τις γενικές αρχές των τεχνικών σύγκρισης και ομαδοποίησης των HRTFs καθώς, επικεντρώνοντας στη μετα-επεξεργασία των datasets και στις μεθόδους μείωσης διαστάσεων. Χρησιμοποιώντας μετρικές που ποσοτικοποιούν την αντιληπτική ομοιότητα και την ακρίβεια εντοπισμού ήχου, συγκρίναμε 72 datasets σε κοινές χωρικές θέσεις και αναλύσαμε τα αποτελέσματα με τη βοήθεια της τεχνικής PCA για μείωση των διαστάσεων. Η εφαρμογή της PCA μάς επέτρεψε να αποτυπώσουμε το μεγαλύτερο μέρος της διακύμανσης των δεδομένων με λιγότερες συνιστώσες, οδηγώντας σε αποδοτική ομαδοποίηση των HRTF με βάση τα κύρια ακουστικά χαρακτηριστικά. Στη συνέχεια, με φασματική ομαδοποίηση (spectral clustering) εντοπίστηκαν φυσικές ομάδες δεδομένων, με κάθε ομάδα να εκπροσωπείται από το κεντρικό HRTF (centroid), το οποίο θεωρείται το πιο αντιπροσωπευτικό.

Η μελλοντική εργασία επικεντρώνεται κυρίως στην εφαρμογή των παραπάνω στην τεχνική του database matching [15] αξιοποιώντας ένα ευρύτερο σύνολο

διαθέσιμων HRTF, το οποίο ανέρχεται σε περίπου 1000 datasets. Η υλοποίηση και αξιολόγηση αυτής της μεθόδου έχει τη δυνατότητα να προσφέρει βελτιωμένες και προσωποποιημένες ακουστικές εμπειρίες στους χρήστες, ενισχύοντας τη φυσικότητα και την αίσθηση ρεαλισμού στις αντίστοιχες εφαρμογές, καθώς θα συμβάλει στην παροχή εξατομικευμένων HRTF.

4. Ευχαριστίες

Το ερευνητικό έργο υποστηρίχτηκε από το ερευνητικό πρόγραμμα SONICOM (www.sonicom.eu) το οποίο χρηματοδοτείται από το πρόγραμμα “Ορίζοντας Ευρώπη 2020” της Ευρωπαϊκής Ένωσης (European Union’s Horizon 2020) με αριθμό έργου No. 101017743.

5. Αναφορές

- [1] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, “Inter-laboratory round robin HRTF measurement comparison,” *IEEE J Sel Top Signal Process*, vol. 9, no. 5, pp. 895–906, 2015.
- [2] J. Pauwels and L. Picinali, “On the relevance of the differences between HRTF measurement setups for machine learning,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, pp. 1–5.
- [3] P. Bremen, M. M. van Wanrooij, and A. J. van Opstal, “Pinna cues determine orienting response modes to synchronous sounds in elevation,” *Journal of Neuroscience*, vol. 30, no. 1, pp. 194–204, 2010.
- [4] A. Roginska, T. S. Santoro, and G. H. Wakefield, “Stimulus-dependent HRTF preference,” in *Audio Engineering Society Convention 129*, Audio Engineering Society, 2010.
- [5] B. F. G. Katz and G. Parsehian, “Perceptually based head-related transfer function database optimization,” *J Acoust Soc Am*, vol. 131, no. 2, pp. EL99–EL105, 2012.
- [6] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, “Spatially oriented format for acoustics 2.1: Introduction and recent advances,” *Journal of the Audio Engineering Society*, vol. 70, no. 7/8, pp. 565–584, 2022.
- [7] T. McKenzie, C. Armstrong, L. Ward, D. T. Murphy, and G. Kearney, “Predicting the colouration between binaural signals,” *Applied Sciences*, vol. 12, no. 5, p. 2441, 2022.

- [8] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, “A Bayesian model for human directional localization of broadband static sound sources,” *Acta Acustica*, vol. 7, p. 12, 2023.
- [9] J. C. Middlebrooks and D. M. Green, “Sound and localization by human listeners,” *Annu. Rev. Psychol.*, vol. 42, pp. 135–159, 1991, doi: 10.1146/annurev.ps.42.020191.001031.
- [10] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.
- [11] G. Marentakis and J. Hözl, “Compression efficiency and Signal Distortion of common PCA bases for HRTF Modelling,” in *Proceedings of the 18th Sound and Music Computing Conference*, Axa sas/SMC Network, 2021.
- [12] J. Hözl, “An initial Investigation into HRTF Adaptation using PCA,” *IEM Project Thesis, Institut für elektronische musik und akustik. Graz, Austria*, 2012.
- [13] F. Grijalva, B. Escobar, B. A. A. Acurio, and R. Álvarez, “Analysis and synthesis of hrtfs using principal component analysis,” in *2019 IEEE Fourth Ecuador Technical Chapters Meeting (ETCM)*, IEEE, 2019, pp. 1–6.
- [14] M. Zhang, Z. Ge, T. Liu, X. Wu, and T. Qu, “Modeling of individual HRTFs based on spatial principal component analysis,” *IEEE/ACM Trans Audio Speech Lang Process*, vol. 28, pp. 785–797, 2020.
- [15] A. Andreopoulou and A. Roginska, “Database matching of sparsely measured head-related transfer functions,” *Journal of the Audio Engineering Society*, vol. 65, no. 7/8, pp. 552–561, 2017.