ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΕΠΙΣΤΗΜΩΝ

Τμήμα Μηχανικών Μουσικής Τεχνολογίας και Ακουστικής

Πτυχιακή εργασία

ΑΥΤΟΜΑΤΗ ΜΙΞΗ ΗΧΟΓΡΑΦΗΣΕΩΝ ΠΑΡΑΓΟΜΕΝΕΣ ΑΠΟ ΧΡΗΣΤΕΣ ΠΟΥ ΠΑΡΑΚΟΛΟΥΘΟΥΝ ΤΟ ΙΔΙΟ ΔΗΜΟΣΙΟ ΓΕΓΟΝΟΣ

Χρήστος Βαλσάμης

Επιβλέπων: ***Νικόλαος Στεφανάκης***

Επίκουρος Καθηγητής

Ρέθυμνο 2018

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ

Πτυχιακή εργασία

ΑΥΤΟΜΑΤΗ ΜΙΞΗ ΗΧΟΓΡΑΦΗΣΕΩΝ ΠΑΡΑΓΟΜΕΝΕΣ ΑΠΟ ΧΡΗΣΤΕΣ ΠΟΥ ΠΑΡΑΚΟΛΟΥΘΟΥΝ ΤΟ ΙΔΙΟ ΔΗΜΟΣΙΟ ΓΕΓΟΝΟΣ

του

Χρήστου Βαλσάμη

Επιβλέπων: *Νικόλαος Στεφανάκης*

Επίκουρος Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την Ιουνίου 2011.

.....	.....	.....
Νικόλαος Στεφανάκης	Νεκτάριος Παπαδογιάννης	Παναγιώτης Ζέρβας
Επίκουρος Καθηγητής	Καθηγητής	Επίκουρος Καθηγητής

Ρέθυμνο, Ιούνιος 2018

Πνευματικά δικαιώματα

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Μηχανικών Μουσικής Τεχνολογίας και Ακουστικής του Τεχνολογικού Εκπαιδευτικού Ιδρύματος Κρήτης δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

ΠΕΡΙΛΗΨΗ

Η εργασία αυτή προτείνει τρόπους για τη συνδυαστική αξιοποίηση της ακουστικής πληροφορίας από καταγραφές παραγόμενες από χρήστες φορητών συσκευών (User Generated Recordings – UGRs) που παρακολουθούν το ίδιο δημόσιο γεγονός. Μέχρι σήμερα, έχει προταθεί ένα πλήθος από τεχνικές για την ομαδοποίηση αυτών των καταγραφών και για την τοποθέτησή τους σε ένα κοινό χρονικό άξονα. Η παρούσα εργασία εστιάζει στο μετέπειτα κομμάτι επεξεργασίας, αυτό της μίξης των ηχογραφήσεων, με σκοπό την παραγωγή ενός νέου ηχητικού αρχείου που συνδυάζει την ακουστική πληροφορία από όλες τις διαθέσιμες ηχητικές καταγραφές. Το πρώτο βήμα πριν τη μίξη είναι η κανονικοποίηση των ηχογραφήσεων, η οποία αποσκοπεί στο να φέρει το πρωτογενές υλικό σε μία κοινή στάθμη σήματος. Αφού τα αρχεία κανονικοποιηθούν, ακολουθεί η διαδικασία της μίξης. Για να να αντιμετωπιστεί το γεγονός ότι το πλήθος των διαθέσιμων ηχογραφήσεων μεταβάλλεται με το χρόνο, είναι απαραίτητος ο προσδιορισμός χρονικά μεταβαλλόμενων βαρών μίξης, τα οποία προσαρμόζονται ανάλογα με το πλήθος των ηχογραφήσεων που συμμετέχουν στη μίξη ανά πάσα χρονική στιγμή. Προτείνουμε μια μεθοδολογία για τον αυτόματο προσδιορισμό αυτών των βαρών η οποία βασίζεται στην υπόθεση της ανεξαρτησίας των ηχητικών καταγραφών από διαφορετικούς χρήστες. Η αξιολόγηση της όλης υλοποίησης έγινε μέσα από ακουστικό τέστ το οποίο σχεδιάστηκε έχοντας ως βάση πραγματικές καταγραφές χρηστών από διάφορα δημόσια γεγονότα. Τα αποτελέσματα του τέστ αναδεικνύουν την αποτελεσματικότητα της προτεινόμενης τεχνικής.

ABSTRACT

This thesis presents a technique for the synergistic exploitation of the audio recordings that are produced by users of mobile devices (User Generated Recordings – UGRs) attending the same public event. Until today, several techniques have been presented on how to group UGRs from the same event and how to align then along the same temporal axis. Assuming a collection of correctly synchronized UGRs, the focus of this thesis is on how to mix the available recordings with the scope to produce a new audio stream of increased duration and improved quality. A first step before mixing is the normalization of the recordings, which aims to bring the audio signals at a common level. After normalization, the mixing process follows. As UGRs start and stop at arbitrary time instants, a mixing technique based on time-varying gains is proposed, derived as a function of the number of UGRs that participate in the mixing process at each point in time. We propose a methodology for the automatic calculation of these gains based on the assumption of independence between the UGRs. A specially designed listening test was designed based on real UGRs from different public events. The results of the test verify the suitability of the presented approach for automatic mixing of UGRs.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΑΠΟΔΟΣΗ ΟΡΩΝ

UGC	User-Generated Content
UGR	User-Generated audio Recording
PCM	Pulse Code Modulation
w-1	Τεχνική μίξης με σταθερά βάρη ίσα με 1
$w\text{-}N^{-1}$	Τεχνική μίξης με μεταβλητά βάρη ίσα με $1/N^{-1}$
$w\text{-}N^{-1/2}$	Τεχνική μίξης με μεταβλητά βάρη ίσα με $1/N^{-1/2}$
w-adapt	Τεχνική μίξης με μεταβλητά βάρη που προέρχονται από προσαρμοστική τεχνική

1 Εισαγωγή

1.1 Ηχητικό υλικό προερχόμενο από χρήστη (User Generated audio Recording– UGR)

Ζούμε στην εποχή των φορητών συσκευών, των «έξυπνων» φορητών τηλεφώνων (smartphones), των υπολογιστών τύπου tablet, των συσκευών που δύνανται να φορεθούν (wearables), γενικότερα των συσκευών που είναι ικανές να καταγράφουν κάθε στιγμή της ζωής μας και των εκδηλώσεων που παρακολουθούμε. Οι οπτικοακουστικές εγγραφές από αυτές τις συσκευές διατίθενται συνήθως από τους χρήστες μέσω των κοινωνικών δικτύων (social networks) και του μεγάλου αριθμού ιστότοπων που παρέχουν περιεχόμενο βίντεο (π.χ., YouTube). Συνήθως, κάθε μεμονωμένη εγγραφή παρέχει μια περιορισμένη οπτικοακουστική εμπειρία του γεγονότος, κυρίως λόγω της περιορισμένης άποψης του χώρου, της κακής - συνήθως - οπτικής και ακουστικής ποιότητας του σήματος και της ελλιπούς διάρκειας κάλυψης (συνήθως λίγα λεπτά του γεγονότος). Σε δημόσιες εκδηλώσεις, πολλαπλές καταγραφές βίντεο μπορούν να ληφθούν από διάφορους χρήστες, πιθανώς από διαφορετικά χρονικά διαστήματα, θέσεις και γωνίες λήψης. Σε αντίθεση με το τυπικό σενάριο, όπου κάθε εγγραφή παρέχεται για κατανάλωση ως ανεξάρτητο αρχείο πολυμέσων, στην παρούσα πτυχιακή εργασία μελετάται ο συνδυασμός και η από κοινού επεξεργασία των διαφορετικών καταγραφών ώστε να παραχθεί μια βελτιωμένη και πληρέστερη οπτικοακουστική αναπαράσταση του καταγεγραμμένου γεγονότος.

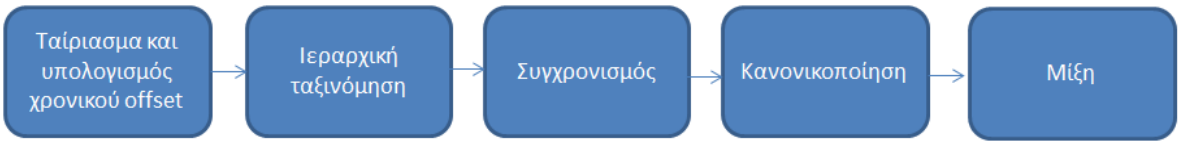
Η αξιοποίηση αυτού του πλούτου του περιεχομένου παραγόμενο από χρήστη (User-Generated Content, UGC) προσφέρει μια αξιοσημείωτη ευκαιρία για τον εμπλουτισμό της τηλεοπτικής κάλυψης ενός γεγονότος. Για παράδειγμα, η τηλεοπτική κάλυψη ενός ποδοσφαιρικού αγώνα προσφέρει μια στατική εμπειρία στον τηλεθεατή, ο οποίος δεν μπορεί να αλληλεπιδράσει με το περιεχόμενο, π.χ. να δει από άλλη γωνία λήψης για κάποιο χρονικό διάστημα τον αγώνα, ή να «βιώσει» τον παλμό της εξέδρας. Ταυτόχρονα, πλέον είναι συνηθισμένο για έναν τηλεθεατή, ενώ παρακολουθεί ένα αγώνα στην τηλεόραση να χρησιμοποιεί παράλληλα μια φορητή συσκευή για να «συνομιλεί» με άλλους τηλεθεατές σε κοινωνικά δίκτυα ή να καταναλώνει επιπλέον υλικό σχετικό με τον αγώνα που δεν προσφέρει η τηλεοπτική κάλυψη (π.χ. ιστορικά στοιχεία, στατιστικά στοιχεία, UGC, κλπ.)

Αυτή η συνεχώς αυξανόμενη τάση για χρήση συνοδευτικής συσκευής (companion device) κατά την τηλεθέαση ενός γεγονότος σε συνδυασμό με το UGC αποτελεί ένα σημαντικό κίνητρο για τη δημιουργία νέων προϊόντων και υπηρεσιών, με σημαντική δυνατότητα για εμπλουτισμό της εμπειρίας του τηλεθεατή. Σε αυτό θα μπορούσε να συμβάλλει η δημιουργία μιας διαδικτυακής κοινότητας (online community) που αφορά το συγκεκριμένο κάθε φορά γεγονός, με την πρόσβαση σε αυτό το επιπλέον περιεχόμενο για τους τηλεθεατές να γίνεται με κάποια χρέωση από τον πάροχο, όπως μια μικρή μηνιαία συνδρομή. Από την άλλη πλευρά, ο πάροχος θα μπορούσε να προσφέρει κάποια κίνητρα στους θεατές του γεγονότος για να αποθέσουν τις εγγραφές τους στην διαδικτυακή κοινότητα που αφορά το γεγονός κατά την ώρα που διεξάγεται, π.χ. με κάποιου είδους «αναγνώριση» σε σχέση με τα άλλα μέλη της κοινότητας. Η φιλοσοφία της επιβράβευσης της υποβολής περιεχομένου μπορεί να προκύψει με βάση τις έως τώρα τεχνικές επιβράβευσης στα πλαίσια του πληθοπορισμού (crowdsourcing).

Γενικότερα, η συνεργατική συνεισφορά και η μετέπειτα επεξεργασία του περιεχομένου παραγόμενου από χρήστη, μπορεί όχι μόνο να εμπλουτίσει την επαγγελματική παραγωγή, αλλά και να αποτελέσει μέσο για την παροχή οπτικοακουστικής κάλυψης σε πολλές δημόσιες εκδηλώσεις όπου η επαγγελματική κάλυψη απουσιάζει. Παρά την τεράστια αυτή δυναμική, οι τεχνικές και τα εργαλεία που σχετίζονται με το UGC δεν έχουν αναπτυχθεί και αξιοποιηθεί ιδιαίτερα έως τώρα. Συμπληρώνοντας αυτό το κενό, η παρούσα μελέτη επικεντρώνεται στο κομμάτι της ηχητικής πληροφορίας, με σκοπό το να προτείνει αυτόματα εργαλεία για τη συνδυαστική αξιοποίηση πολλαπλών ηχητικών καταγραφών από το ίδιο γεγονός, προκειμένου να παραχθεί μια βελτιωμένη -σε ποιότητα και περιεχόμενο- αναπαράσταση του ακουστικού γεγονότος.

1.2 Στάδια επεξεργασίας του ηχητικού υλικού

Για να μπορέσουνε να ενωθούνε οι πολλαπλές ηχητικές καταγραφές από τους διαφορετικούς χρήστες, το ηχητικό υλικό πρέπει να περάσει από διάφορα στάδια επεξεργασίας, όπως φαίνεται και στο Σχήμα 1.



<u>Σχήμα 1</u>: Τα διάφορα στάδια επεξεργασίας των UGR, από τη χρονική οργάνωση του περιεχομένου μέχρι τη διαδικασία της μίξης.

1.2.1 Ταίριασμα

Ο όρος *ταίριασμα* (audio matching) αναφέρεται στην λήψη μιας απόφασης σχετικά με το αν δύο ηχογραφήσεις περιέχουν κοινό περιεχόμενο, δηλαδή αν οι καταγραφές έγιναν κατά τον ίδιο χώρο και χρόνο [1-4]. Δεδομένη μιας βάσης με πολλά UGR, η δράση αυτή λαμβάνει υπόψιν όλους του συνδυασμούς των αρχείων ανά δύο. Για να παρθεί η απόφαση αν δύο αρχεία ταιριάζουν και επομένως επικαλύπτονται χρονικά, γίνεται εκτίμηση μιας μετρικής ομοιότητας μέσα από μία διαδικασία η οποία παρουσιάζει ομοιότητες με την πράξη της ετεροσυσχέτισης. Αρχεία ήχου για τα οποία η μετρική ομοιότητας είναι πάνω από ένα προκαθορισμένο κατώφλι θεωρούνται ότι ταιριάζουνε (positive match), αλλιώς θεωρούνται ως αταίριαστα (negative match). Η ετεροσυσχέτιση υπολογίζεται για όλες τις σχετικές χρονικές μετατοπίσεις μεταξύ των δύο αρχείων, και η τιμή της ετεροσυσχέτισης με τη μεγαλύτερη απόλυτη τιμή λαμβάνεται υπόψιν ως η τιμή της ομοιότητας. Ωστόσο, η πράξη της ετεροσυσχέτισης δε γίνεται με βάση το αρχικό PCM σήμα αυτό καθεαυτό, αλλά με βάση τα λεγόμενα ηχητικά αποτυπώματα του σήματος (audio fingerprints).

1.2.2 Ηχητικά αποτυπώματα

Αν και η εξαγωγή ηχητικών αποτυπωμάτων (audio fingerptints) δεν αποτελεί αντικείμενο της μελέτης αυτής, αξίζει να γίνει μια αναφορά στο τι είναι τα ηχητικά αποτυπώματα και πως γιατί αυτά εξελίχθηκαν μέσα στα τελευταία χρόνια. Τα ηχητικά αποτυπώματα δημιουργήθηκαν και εξελίχθηκαν λόγω της ανάγκης να πραγματοποιούμε γρήγορη και αξιόπιστη ταυτοποίηση μεταξύ ενός μικρής διάρκειας ηχητικού δείγματος και μιας μεγάλης βάσης δεδομένων. Η πιο συνηθισμένη εφαρμογή είναι στην αναγνώριση μουσικού υλικού (π.χ. εφαρμογή Shazam), όπου κάποιος χρήστης παρέχει ένα μικρής διάρκειας ηχητικό δείγμα το οποίο κατέγραψε με κάποια φορητή συσκευή με την απαίτηση να μάθει το όνομα

του καλλιτέχνη και τον τίτλο του έργου από το οποίο προέρχεται. Έχοντας μία βάση δεδομένων με γνωστά αρχεία (πχ., τραγούδια) τα ηχητικά αποτυπώματα του δείγματος συγκρίνονται με αυτά της βάσης. Το όνομα του κομματιού και του καλλιτέχνη προκύπτει από το αρχείο βάσης που παρουσιάζει τη μέγιστη ομοιότητα με το ηχητικό δείγμα.

Για τα ηχητικά αποτυπώματα έχουν προταθεί πολλές διαφορετικές τεχνικές, όπου κάθε μία επικεντρώνεται σε διαφορετικά χαρακτηριστικά του σήματος [4-7]. Η χρήση τους στην οργάνωση περιεχομένου παραγόμενου από χρήστη είναι σχετικά πρόσφατη, αλλά ουσιαστικά δε διαφοροποιείται σημαντικά σε σχέση με την ταυτοποίηση μουσικού υλικού. Σκοπός σε αυτήν την περίπτωση είναι η ομαδοποίηση και συγχρονισμός των ηχογραφήσεων που προέρχονται από το ίδιο ηχητικό γεγονός με απότερο σκοπό, την αυτόματη τοποθέτησή του πάνω στο χρονικό άξονα του γεγονότος. Η χρήση των fingerprints αντι του PCM σήματος κατά την πράξη της ετεροσυσχέτισης επιταχύνει σημαντικά την μέτρηση της ομοιότητας και το συγχρονισμό των αρχείων και αυτό έχει να κάνει με το ότι όγκος δεδομένων που καταλαμβάνει το ηχητικό αποτύπωμα είναι πολύ μικρότερος από αυτόν του αρχικού σήματος (της τάξης του 1/1000 και λιγότερο).

1.2.3 Συγχρονισμός των αρχείων

Ο σωστός συγχρονισμός των αρχείων ήχου που επικαλύπτονται χρονικά είναι μια σημαντική απαίτηση για το όλο σύστημα. Εφόσον κριθεί ότι τα δύο αρχεία πληρούν το κριτήριο ομοιότητας, τότε λαμβάνεται υπόψιν και η χρονική τους συσχέτιση (time offset), δηλαδή η πληροφορία σχετικά με το πότε ξεκίνησε η ηχογράφηση στη μία συσκευή σε σχέση με την άλλη. Έστω δύο αρχεία *i* και *j* που επικαλύπτονται χρονικά και έστω τ_{ij} η χρονική τους συσχέτιση η οποία υποδηλώνει την ποσότητα χρόνου που η συσκευή *j* ξεκίνησε να ηχογραφεί μετά την συσκευή *i*. Προφανώς, αν η συσκευή *j* ξεκίνησε να ηχογραφεί πριν την συσκευή *i*, τότε το τ_{ij} θα πάρει αρνητική τιμή. Επιπλέον θα ισχύει $\tau_{ji} = -\tau_{ij}$. Η μετρική ομοιότητας και η χρονική συσχέτιση υπολογίζονται και καταχωρούνται για όλους τους συνδυασμούς ανά ζεύγη τα οποία πληρούν μια μετρική ομοιότητας που είναι μεγαλύτερη από ένα προκαθορισμένο κατώφλι. Ωστόσο, για το σωστό συγχρονισμό όλων των ηχογραφήσεων πάνω σε ένα κοινό άξονα, ένα μόνον μέρος από τις χρονικές συσχετίσεις είναι απαραίτητο. Για παράδειγμα, έστω δύο αρχεία *i* και *j* τα οποία επικαλύπτονται χρονικά με το αρχείο *k* αλλά όχι μεταξύ τους, παρόλο που και τα τρία αυτά αρχεία προέρχονται από

το ίδιο ηχητικό γεγονός (πχ, το ίδιο τραγούδι σε μια συναυλία). Θα θέλαμε ιδανικά να ξέρουμε ότι και τα τρία αυτά αρχεία περιέχουν πληροφορία αναφορικά με το ίδιο γεγονός και άρα να τα εντάξουμε στην ίδια ομάδα. Αυτή ακριβώς η δουλειά επιτελείται στη διεργασία με το όνομα *ιεραρχική ταξινόμηση*. Η διεργασία αυτή θα κρίνει ποια αρχεία προέρχονται από το ίδιο γεγονός, ανεξάρτητα από το αν κάποια ζεύγη αρχείων δεν πληρούν το κριτήριο ομοιότητας. Αντίστοιχα, η διεργασία του συγχρονισμού θα διατάξει όλα τα αρχεία που τοποθετήθηκαν στην ίδια ομάδα πάνω σε ένα κοινό άξονα, επιλέγοντας την ανά-ζεύγη πληροφορία με τη μεγαλύτερη αξιοπιστία. Για παράδειγμα, αν η ομοιότητα μεταξύ των αρχείων i και j είναι πολύ ασθενής, τότε αυτά τα δύο αρχεία μπορούν να συγχρονιστούν έμμεσα μεταξύ τους, μέσω του αρχείου k. Είναι εύκολο να παρατηρήσει κάποιος ότι ενώ η πληροφορία τ_{ij} μπορεί να μην υπάρχει ή να είναι λανθασμένη, τότε τα δύο αρχεία μπορούν να συγχρονιστούν μεταξύ τους μέσω του k αξιοποιώντας τη σχέση $\tau_{ij}=\tau_{ik}+\tau_{kj}$.

1.2.4 Κανονικοποίηση

Η διεργασία της *κανονικοποίησης* έχει ως στόχο να φέρει τα σήματα που επικαλύπτονται χρονικά σε μια κοινή στάθμη, έτσι ώστε κάθε ηχογράφηση να έχει την ίδια βαρύτητα κατά τη διαδικασία της μίξης. Δεδομένου ότι οι ηχογραφήσεις προέρχονται από διαφορετικές συσκευές και θέσεις στο χώρο, η κανονικοποίηση αποτρέπει μεταξύ άλλων φαινόμενα του τύπου ένα UGR με υψηλή στάθμη σήματος να επισκιάζει ένα UGR που είναι διαθέσιμο σε πιο ασθενή στάθμη σήματος.

1.2.5 Μίξη

Τέλος η διεργασία της μίξης αποσκοπεί στο να υπερθέσει την ηχητική πληροφορία από πολλές χρονικά επικαλυπτόμενες καταγραφές, οδηγώντας έτσι στην δημιουργία μιας νέας ηχητικής αναπαράστασης του εκάστοτε γεγονότος, η οποία έχει μεγαλύτερη διάρκεια και καλύτερη ποιότητα συγκριτικά με κάθε μεμονωμένη καταγραφή. Αν και υπάρχουν πολλές διαφορετικές προσεγγίσεις που θα μπορούσε να ακολουθήσει κανείς για τη συνεργατική αξιοποίηση των συγχρονισμένων αρχείων, σε αυτή τη διατριβή η μελέτη εστιάζεται στην αυτοματοποίηση της μίξης μερικά επικαλυπτόμενων ηχογραφήσεων. Όπως θα γίνει κατανοητό σε μετέπειτα ενότητα, η μίξη μερικά επικαλυπτόμενων ηχογραφήσεων παρουσιάζει ιδιαίτερες δυσκολίες που σχετίζονται με τις μεταβολές στην ένταση, που

δημιουργούνται στα χρονικά σημεία όπου ένα ή περισσότερα UGRs αρχίζουν ή σταματούν να συνεισφέρουν στη μίξη.

2 Μεθοδολογία

2.1 Εισαγωγή

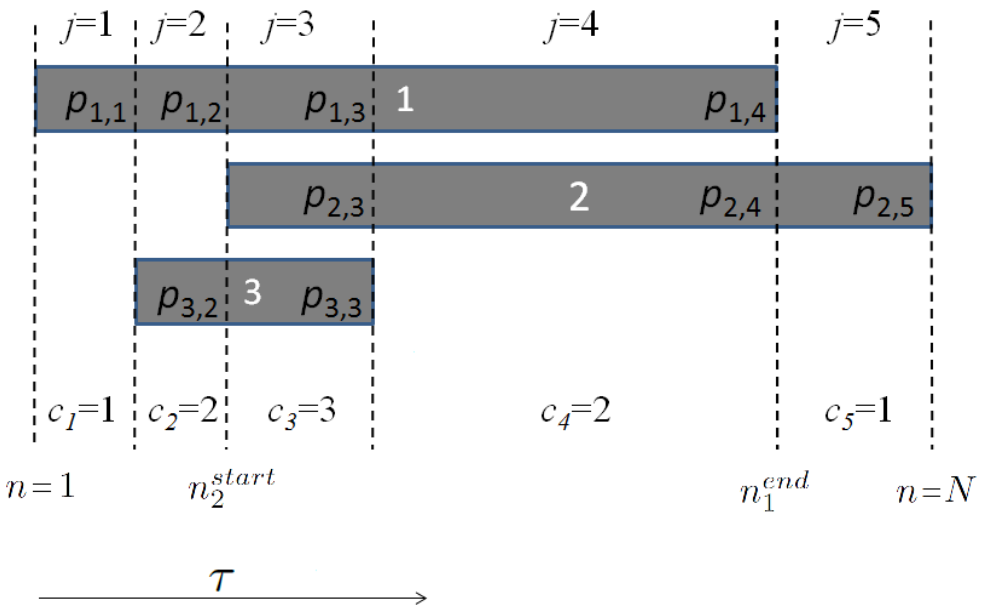
Θεωρούμε ένα σύνολο από *M* χρονικά επικαλυπτόμενες ηχογραφήσεις παραγόμενες από χρήστη (UGRs) οι οποίες έχουν τον ίδιο ρυθμό δειγματοληψίας F_s. Υποθέτουμε ότι όλα τα *M* ηχητικά αρχεία είναι σωστά ταξινομημένα σε ένα κοινό άξονα του χρόνου όπως παρουσιάζεται στο Σχήμα 2. Καθώς οι χρήστες θα ξεκινήσουν και θα σταματήσουν να ηχογραφούν σε τυχαίες χρονικές στιγμές, κάθε ηχητικό αρχείο θα έχει διαφορετική διάρκεια και θέση στον άξονα χρόνου. Ορίζουμε μια κατάτμηση των ηχογραφήσεων σε χρονικά πλαίσια συγκεκριμένου πλήθους δειγμάτων και έστω ότι τ είναι ο δείκτης χρονοπλαισίου. Μπορούμε να αναφερθούμε σε οποιοδήποτε χρονική στιγμή της διαδικασίας με βάση το δείκτη χρονοπλαισίου *τ,* ή με βάση το δείκτη δείγματος *n*. Όπως φαίνεται και στο Σχήμα 2, ορίζουμε το συμβολισμό n_m^{start} και n_m^{end} για να αναφερθούμε πρώτο και τελευταίο δείγμα του m-ιοστού ηχητικού αρχείου. Τα σημεία χρόνου n_m^{start} και $n_m^{end}+1,\ m=1,\dots,M$ αντιπροσωπεύουν τα σημεία μετάβασης του διαγράμματος, τα σημεία στο χρόνο δηλαδή, όπου αλλάζει το πλήθος των ηχητικών αρχείων που συμμετέχουνε στη μίξη. Χρησιμοποιούμε επίσης το συμβολισμό

$$\mathbb{U}_m = \{n_m^{(start)}, n_m^{(start)}+1, \dots, n_m^{(end)}\} \quad (1)$$

για να ορίσουμε το σετ με τους δείκτες δείγματος για τα οποία το ηχητικό αρχείο M περιέχει ηχητική πληροφορία. Σε ότι ακολουθεί θεωρείται πως για κάθε αρχείο i στη συλλογή υπάρχει τουλάχιστον ένα άλλο αρχείο j τέτοιο ώστε $\mathbb{U}_i \cap \mathbb{U}_j \neq \emptyset$. Με άλλα λόγια, κάθε ηχητικό αρχείο επικαλύπτεται (πλήρως ή μερικώς) με τουλάχιστον άλλο ένα αρχείο στην ομάδα. Θα αναφερθούμε στο $\mathbb{U}_m$ ως τη χρονική περιοχή όπου το αρχείο *M* είναι ενεργό, υπονοώντας πως η m-ιοστή ηχογράφηση δεν περιέχει καμία ηχητική πληροφορία σε χρονικά σημεία με δείκτη δείγματος μικρότερο από το ελάχιστο και μεγαλύτερο από το μέγιστο δείκτη εντός του σετ $\mathbb{U}_m$.

Όπως φαίνεται και στο Σχήμα 3, επεξεργαζόμενοι τις διαθέσιμες ηχογραφήσεις συνδυαστικά, μπορούμε να δημιουργήσουμε ένα ηχητικό αρχείο το οποίο έχει μεγαλύτερη χρονική διάρκεια από κάθε άλλο ηχητικό αρχείο που είναι υπό επεξεργασία. Χωρίς απώλεια της

γενικότητας, υποθέτουμε ότι η ηχητική ακολουθία παράγεται ως αποτέλεσμα του συνδυασμού όλων των διαθέσιμων ηχητικών σημάτων η οποία εκτείνεται από $n = 1$ ως $n = N$, τα χρονικά όρια του γεγονότος που καλύπτεται από τις M ηχογραφήσεις. Επίσης ορίζουμε το $\boldsymbol{c}$ ως ένα $N \times 1$ θετικό ακέραιο διάνυσμα δείχνοντας τον αριθμό των κομματιών που είναι ενεργά σε κάθε χρονική στιγμή. Προφανώς $1 \leq \boldsymbol{c}[n] \leq M, \forall n$.



<u>Σχήμα 2</u>: Τρεις αλληλεπικαλυπτόμενες ηχητικές ηχογραφήσεις καθορίζουν έξι σημεία μετάβασης και πέντε συνεχή χρονικά τμήματα. Τα χρονικά σημεία συμβολίζονται με *n* και τα χρονικά τμήματα με *l*.

Ορίζουμε τώρα τον μετασχηματισμό $x_m[n] \rightarrow \hat{x}_m[n]$ για το m-ιοστό ηχητικό σήμα ως εξής

$$\hat{x}_m[n] = \begin{cases} x_m[n - n_m^{start} + 1] & \varepsilon\alpha\nu\ n \in \mathbb{U}_m \\ 0, & \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases}, \qquad (2)$$

ο οποίος απλοποιεί το συμβολισμό που απαιτείται για να αναφερόμαστε σε συγκεκριμένες χρονικές στιγμές του γεγονότος όπως αυτές έχουν καταγραφεί από διαφορετικά UGRs.

2.2 Στατιστικές παραδοχές

Όλες η ηχογραφήσεις υποτίθεται ότι έχουν συγχρονιστεί εκ των προτέρων και έχουν ληφθεί με κοινή συχνότητα δειγματοληψίας F_s, ώστε να υπάρχει αντιστοιχία ένα προς ένα μεταξύ του δείκτη σήματος n και του χρόνου $t = \frac{n}{F_s}$. Οι τεχνικές μίξης που παρουσιάζονται

παρακάτω βασίζονται στη στατιστική υπόθεση ότι δύο διαφορετικές ηχογραφήσεις είναι ασυσχέτιστες μεταξύ τους, δηλαδή

$$E\{\hat{x}_i[n]\hat{x}_j[n]\} = 0 \text{ όταν } i \neq j. \quad (3)$$

Γενικά, δύο συγχρονισμένες ηχογραφήσεις από το ίδιο γεγονός όντως έχουν κάποιου είδους συσχέτιση (δεν θα ήταν εφικτό να τις συγχρονίσουμε άμα δεν είχαν). Όμως, για λόγους όπως

1) η διαφορετική τοποθεσία ηχογράφησης,
2) η διαφορετική συχνοτική απόκριση της κάθε συσκευής,
3) ο διαφορετικός θόρυβος που υπερτίθεται σε κάθε ηχογράφηση,

η συσχέτιση με την έννοια της Εξ. (3), παρότι δεν είναι ακριβώς αληθής, αναμένεται να είναι σχετικά ασθενής. Μία πιο ακριβής μαθηματική διατύπωση αυτής της παραδοχής μπορεί να γίνει μέσω της σχέσης

$$-\epsilon \leq \frac{E\{\hat{x}_i[n]\hat{x}_j[n]\}}{\sqrt{E\{\hat{x}_i^2[n]\}E\{\hat{x}_j^2[n]\}}} \leq \epsilon \text{ όταν } i \neq j, \quad (4)$$

όπου $\epsilon << 1$, υπονοώντας ότι δύο διαφορετικές ηχογραφήσεις είναι ασθενώς συσχετισμένες.

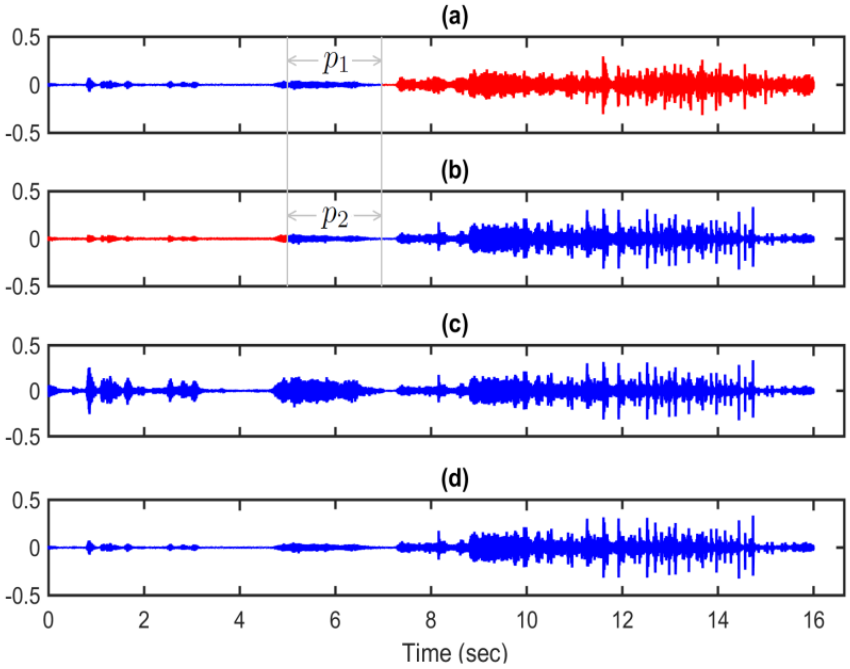
2.3 Κανονικοποίηση των UGRs

Ακόμα και όταν οι ηχογραφήσεις είναι σωστά συγχρονισμένες, δεν είναι ακόμα έτοιμες για τη διαδικασία της μίξης γιατί οι εντάσεις από το ένα αρχείο ήχου στο άλλο μπορεί να διαφέρουνε σημαντικά. Αυτή η διαφοροποίηση μπορεί να οφείλεται σε διάφορους λόγους όπως για παράδειγμα οι διαφορές στη θέση των συσκευών κατά τη φάση της ηχογράφησης, τα διαφορετικά χαρακτηριστικά από συσκευή σε συσκευή κλπ. Η κανονικοποίηση των ηχογραφήσεων ως προς τη στάθμη του σήματος παρέχει διάφορα οφέλη, όπως το να εξασφαλίσει ότι όλα τα ηχητικά αρχεία έχουν την ίδια βαρύτητα στην διαδικασία της μίξης, αποφεύγοντας για παράδειγμα ηχογραφήσεις που έχουν πραγματοποιηθεί σε κοντινή απόσταση από την ηχητική πηγή, να καλύπτουν αυτές που έγιναν σε μεγαλύτερη απόσταση. Επίσης, η ακριβής κανονικοποίηση είναι σημαντική για την δημιουργία μιας μίξης χωρίς διακοπές και ακουστές μεταβολές στην έντασης, οι οποίες αναμένεται να συμβούν στο σημείο μετάβασης όπου ένα ηχητικό αρχείο αρχίζει ή σταματάει να συμμετέχει στη διαδικασία της μίξης.

Η διαδικασία της κανονικοποίησης μπορεί να αναπαρασταθεί μαθηματικά μέσω της σχέσης

$$\tilde{x}_{\mathrm{m}}[\mathrm{n}] = \mu_{\mathrm{m}}\hat{\mathrm{x}}_{\mathrm{m}}[\mathrm{n}]\,, \tag{5}$$

όπου μ_m είναι ένα θετικό και πραγματικό βάρος. Αυτή η διαδικασία σκοπεύει να αντισταθμίσει τις διαφορές στη στάθμη του σήματος ανάμεσα στις διαφορετικές ηχογραφήσεις, δίνοντας ένα μοναδικό βάρος σε κάθε ηχογράφηση. Ως παράδειγμα για να γίνει καταναοητή η σοβαρότητα του ζητήματος, θεωρείστε την περίπτωση δύο πλήρως επικαλυπτώμενων ηχητικών ηχογραφήσεων, οι οποίες λήφθηκαν από το ίδιο πραγματικό δημόσιο γεγονός και παρουσιάζονται στο Σχήμα 3 (a) και (b). Μπορεί εύκολα να φανεί από τις αντίστοιχες κυματομορφές ότι οι δύο ηχογραφήσεις είναι σωστά τοποθετημένες κατά μήκος του χρονικού άξονα και έχουν συλλάβει την ίδια χρονική στιγμή από το δημόσιο γεγονός. Υποθέστε τώρα, ότι μόνο το μπλέ τμήμα από κάθε ηχογράφηση είναι διαθέσιμο, π.χ., τμήμα $t \in [0 \quad 7]$ s σχετικά με την ηχογράφηση 1 που φαίνεται στο (a) και το τμήμα $t \in [5 \quad 16]$ s σχετικά με την ηχογράφηση 2 που φαίνεται στο (b). Εφόσον κάθε ηχητικό αρχείο, φέρει συγκεκριμένο κομμάτι του γεγονότος, θα ήταν ωφέλιμο να συγχωνεύσουμε τις δύο ηχογραφήσεις με σκοπό να δημιουργήσουμε μία πιο ολοκληρωμένη αναπαράσταση του λαμβανόμενου γεγονότος. Στο συγκεκριμένο παράδειγμα, αυτό θα επιτευχθεί συνδιάζοντας το τμήμα [0 7] s από την ηχογράφηση 1 με το τμήμα (7 16] s από την ηχογράφηση 2. Το πρόβλημα είναι ότι, εάν οι δύο αρχικές ηχογραφήσεις έχουν πολύ διαφορετικές στάθμες σήματος, τότε το προκύπτον ηχητικό σήμα θα χαρακτηρίζεται από ξαφνική μετάβαση στάθμης στο $t = 7$ s. Όμως, οι απότομες αλλαγές στάθμης είναι γνωστές ως πηγές σύγχυσης για τους ακροατές [4]. Επιπλέον, για μερικές συνεργατικές μεθόδους παραγωγής ήχου οι οποίες έχουν προταθεί σχετικά πρόσφατα [8,10], είναι σημαντικό οι στιγμιαίες διαφορές στις ισχείς σήματος των διαφορετικών συνιστόσων μίξης να είναι μικρές. Αυτό ωθεί στη χρήση κάποιας μορφής κλιμάκωσης για την ελαχιστοποίηση των διαφορών της στάθμης σήματος μεταξύ των διαφορετικών ηχογραφήσεων. Παρακάτω, παρουσιάζονται δύο διαφορετικές προσεγγίσεις.



<u>Σχήμα 3</u>: Δύο συχγρονισμένα UGRs από το ίδιο γεγονός στο (a) και (b). Το αποτέλσμα από την ένωση των δύο αρχείων όταν η κανονικοποίηση βασίζεται στη μέση ισχύ του κάθε σήματος ξεχωριστά φαίνεται στο (c) και όταν η κανονικοποίηση βασίζεται σε σχέσεις ισχύος στο (d).

2.3.1 APN: Κανονικοποίηση βασισμένη στη μέση ισχύ

Υπό την προϋπόθεση ότι το ακουστικό συμβάν είναι μια εργοδική διαδικασία, μπορούμε να υποθέσουμε οτι τα αποκτώμενα σήματα έχουν συνεχείς ισχείς στον χρόνο και μπορούμε να υπολογίσουμε τη μέση στάθμη σήματος με οποιοδήποτε μέγεθος δείγματος. Η κανονικοποίηση σε αυτή την περίπτωση δεν έχει τόση σημασία, επομένως μπορεί να επιτευχθεί πραγματοποιόντας έναν υπολογισμό της μέσης ισχύς του σήματος, υπολογιζόμενο καθ'όλη τη διάρκεια της κάθε ηχογράφησης. Συγκεκριμένα, εάν N_m είναι η διάρκεια της m-ιοστής ηχογράφησης σε δείγματα, μια πρώτη προσέγγιση για την κανονικοποίηση είναι μέσω της διαδικασίας $\tilde{x}_m[n] = g_m \hat{x}[n]$ με το g_m να ορίζεται ως

$$g_m = \frac{g_0}{\sqrt{\frac{1}{N_m}\sum_{i=1}^{N_m} x_m^2[n]}}, \tag{6}$$

και g_0 αναφέρεται στη μέση ισχύ σήματος αναφοράς. Σε ότι ακολουθεί, ορίζουμε το g_0 να είναι ίσο με το αντίστροφο της τετραγωνικής ρίζας της μέσης ισχύος της πρώτης ηχογράφησης. Ως συνέπεια, το πρώτο στοιχείο στο προκύπτον διάνυσμα κέρδους θα είναι

πάντα ίσο με 1. Θα αναφερθούμε σε αυτή τη μέθοδο ως κανονικοποίηση με βάση τη μέση ισχύ του σήματος (Average Power Normalization -APN) [8]. Όσον αφορά το παράδειγμα του Σχήματος 2, το αποτέλεσμα της συγχώνευσης των δύο ηχογραφήσεων χρησιμοποιώντας τη μέθοδο APN παρουσιάζεται στο σχήμα (c). Μπορούμε να δούμε ότι η προσπάθεια να ισοσταθμίσουμε τις μέσες ισχείς από τα δύο ηχητικά αρχεία έχει ως αποτέλεσμα την υπερενίσχυση της πρώτης ηχογράφησης. Αυτό οφείλεται στο γεγονός ότι τα πραγματικά ακουστικά γεγονότα δεν είναι εργοδικά και μπορεί να εμφανίζουν σημαντικές διακυμάνσεις της ενέργειας στο χρόνο. Ως συνέπεια, η APN προσέγγιση δεν μπορεί να εγγυηθεί ένα ικονοποιητικό αποτέλεσμα όταν το ηχητικό γεγονός χαρακτηρίζεται από σημαντικές δυναμικές μεταβολές.

2.3.2 RPN: Κανονικοποίηση με βάση τη σχετική ισχύ μεταξύ σημάτων

Διαισθητικά, μια καλύτερη προσέγγιση για την κανονικοποίηση των δύο ηχογραφήσεων είναι μέσω σχέσεων ισχύος του σήματος, χρησιμοποιώντας ένα μέτρο της ενέργειας του σήματος κατά το χρονικό διάστημα όπου οι δύο ηχογραφήσεις επικαλύπτονται. Το p_1 και p_2 δηλώνουν την ενέργεια σήματος της πρώτης και της δεύτερης ηχογράφησης, αντιστοίχως, μετρώνται κατά μήκος του τμήματος $t \in [5 \quad 7]$ s. Αμα χρησιμοποιήσουμε την ηχογράφηση 2 ως αναφορά ($g_2 = 1$), μπορούμε να κλιμακώσουμε το πρώτο ηχητικό αρχείο με $g_1 = \sqrt{p_2/p_1}$ και το αποτέλεσμα της συγχόνευσης των δύο ηχογραφήσεων παρουσιάζεται στο Σχήμα 3(d). Προφανώς, αυτή η προσέγγιση σέβεται καλύτερα τις διακυμάνσεις στη δυναμική του πραγματικού γεγονότος. Παρ'όλα αυτά, η γενίκευση αυτής της προσέγγισης στην περίπτωση περισσότερων από δύο ηχογραφήσεων δεν είναι τόσο ασήμαντη, όπως φαίνεται στο εξής.

Η διαδικασία η οποία προτείνεται για την κανονικοποίηση των UGRs αποτελεί μια παραλλαγή μεθόδου που έχει δημοσιευτεί [8]. Ορίζουμε την ακριβή ενέργεια του ηχητικού αρχείου ως $\boldsymbol{p}_m^{(0)}$ που είναι ένα μη αρνητικό $N \times 1$ διάνυσμα το οποίο εμπεριέχει την ενέργεια του m-ιοστού ηχητικού αρχείου σε κάθε διακριτή τιμή του χρόνου, π.χ., $\boldsymbol{p}_m^{(0)} = [p_m[1], \dots, p_m[n], \dots, p_m[N]]$. Για τις χρονικές τιμές n που δεν ανήκουν στο χρονικό όριο του αρχείου m, καταχωρείται μια μηδενική τιμή, π.χ.,

$$\boldsymbol{p}_m^{(0)}[n] = \begin{cases} \hat{x}_m^2[n] \; \varepsilon\acute{\alpha}\nu \; n \in \mathbb{U}_m \\ \\ 0, \qquad \alpha\lambda\lambda\iota\acute{\omega}\varsigma. \end{cases} \tag{7}$$

Σύμφωνα με το παράδειγμα στο Σχήμα 2, η ενέργεια της ηχογράφησης $m = 2$ θα έχει μηδενικές τιμές για $n = 1$ μέχρι το σημείο $n = n_2^{start} - 1$ καθώς επίσης και για $n = n_2^{end} + 1$ μέχρι $n = N$. Παρατηρήστε τώρα ότι η συνολική ενέργεια του ηχητικού σήματος στο αρχείο m μπορεί να εξαχθεί από την σχέση $E_{\boldsymbol{m}} = \sum_{\boldsymbol{n}\in\mathbb{U}_{\boldsymbol{m}}} \boldsymbol{p}_m^{(0)}[n] = \sum_{n=1}^{N} \boldsymbol{p}_m^{(0)}[n]$. Η προτεινόμενη προσέγγιση για γενικευμένο RPN παρουσιάζεται στον Αλγόριθμο 1. Για την υλοποίηση αυτού του αλγορίθμου ορίζουμε το διάνυσμα αντίστροφης πολλαπλότητας ως

$$\boldsymbol{\theta} = \left[\frac{1}{\mathrm{c}[1]}, \ldots, \frac{1}{\mathrm{c}[N]}\right]^{\mathrm{T}}, \tag{8}$$

όπου $c[n]$ είναι το πλήθος των ενεργών ηχογραφήσεων κατά τη χρονική στιγμή n σύμφωνα και με την ανάλυση στο Κεφ. 3.1.

Αλγόριθμος 1: Επαναληπτική κανονικοποίηση

Είσοδος: Αρχική ενέργεια $\boldsymbol{p}_m^{(0)}$, $\forall m$,

Είσοδος: Αντίστροφο διάνυσμα πολλαπλότητας $\boldsymbol{\theta}$

Είσοδος: Αριθμός επαναλήψεων I

Έξοδος: βάρη κανονικοποίησης μ_m, $\forall m$

for $i = 1$ to I do

$\boldsymbol{q}^{(i)} = \sum_{m=1}^{M} \boldsymbol{p}_m^{(i-1)}$

$\boldsymbol{P}^{(i)} = \boldsymbol{\theta} \cdot \boldsymbol{q}^i (\cdot \; \sigma\upsilon\mu\beta o\lambda\acute{\iota}\zeta\varepsilon\iota \; \tau o\nu \; \pi o\lambda\lambda\alpha\pi\lambda\alpha\sigma\iota\alpha\sigma\mu\acute{o} \; \sigma\tau o\iota\chi\varepsilon\acute{\iota}o\upsilon \; \varepsilon\pi\acute{\iota} \; \sigma\tau o\iota\chi\varepsilon\acute{\iota}o\upsilon)$

for $m = 1$ to M do

$\boldsymbol{p}_m^{(i)} \leftarrow \lambda_m^{(i)} \boldsymbol{p}_m^{(i-1)}$

where $\lambda_m^{(i)} = \frac{\sum_{n\in\mathbb{U}_m} \boldsymbol{P}^{(i)}[n]}{\sum_{n\in\mathbb{U}_m} \boldsymbol{p}_m^{(i-1)}[n]}$

end for

end for

$\mu_m = \sqrt{\prod_{i=1}^{I} \lambda_m^{(i)}}\ ,\ m = 1, \ldots, M.$

Ο αλγόριθμος αυτός είναι σχεδιασμένος ώστε σε κάθε επανάληψη να βελτιώνεται η σύγκλιση ως προς τη σχέση

$$\sum_{n \in \mathbb{U}_m} \lambda_m \boldsymbol{p}_m^{(i-1)}[n] = \sum_{n \in \mathbb{U}_m} \boldsymbol{P}^{(i)}[n]. \quad (9)$$

Δηλαδή, η συνολική ενέργεια στο *m*-ιοστό UGR να προσεγγίζει τον αριθμητικό μέσο της ενέργειας από όλα τα UGRs υπολογισμένη κατά μήκος του χρονικού άξονα που το *m*-ιοστό UGR είναι ενεργό.

Διαπιστώσαμε από διάφορα πειράματα ότι ένα μικρό πλήθος από *I*=5 έως 10 επαναλήψεις αρκεί για να επιτευχθεί μια σύγκλιση με σφάλμα μικρότερο από 0.1% ως προς την Εξ. (9). Αξίζει ωστόσο να αναφερθούμε και σε μια εναλλακτική εκδοχή, όπου ο αλγόριθμος δεν σταματάει μετά από ένα συγκεκριμένο αριθμό επαναλήψεων, αλλά όταν η παρακάτω συνθήκη βρεθεί αληθής:

$$\left|\lambda_m^{(i)} - 1\right| \leq \epsilon, \forall m, \quad (10)$$

όπου $\epsilon \ll 1$ είναι μια προκαθορισμένη θετική τιμή . Σύμφωνα με την τελευταία συνθήκη, ο αλγόριθμος θα σταματά όταν οι αλλαγές στα βάρη κανονικοποίησης από τη μια επανάληψη στην επόμενη είναι αμελητέες.

2.4 Προβλήματα κατά την αυτοματοποίηση της μίξης

Στο σχήμα που ακολουθεί παρακάτω παρουσιάζονται τρείς διαφορετικοί μέθοδοι μίξης. Η προτεινόμενη μέθοδος με τα χρονικά μεταβαλλόμενα βάρη της μορφής

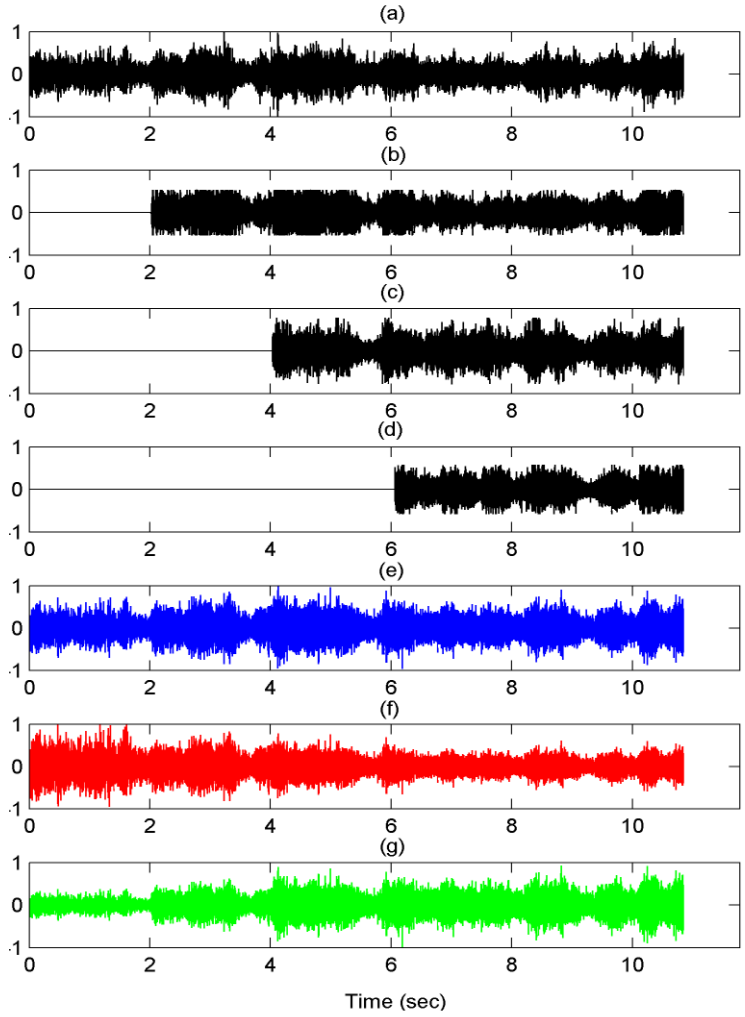
$$w_m[n] = \frac{1}{\sqrt{C[n]}}, \forall m \quad (11)$$

η μέθοδος με τα χρονικά μεταβαλλόμενα βάρη της μορφής

$$w_m[n] = \frac{1}{C[n]}, \forall m \quad (12)$$

και η μέθοδος με τα σταθερά βάρη της μορφής $w_m[n] = 1, \forall m$. Η πρώτη μέθοδος δηλώνει ότι τα βάρη της συνολικής μίξης διαιρούνται με την τετραγωνική ρίζα του αριθμού των ηχητικών αρχείων που λαμβάνουν μέρος στη μίξη, η δεύτερη μέθοδος δηλώνει την διαίρεση

με τον αριθμό των ηχητικών αρχείων που είναι στη μίξη και η τρίτη μέθοδος δηλώνει ότι τα βάρη δεν μεταβάλλονται όταν περισσότερες ηχογραφήσεις αρχίζουν να συμμετέχουν στη μίξη. Οι κυματομορφές που ανταποκρίνονται σε κάθε μία από αυτές τις μεθόδους παρουσιάζονται στο Σχήμα 4 ως (e), (f) και (g) αντίστοιχα.



<u>Σχήμα 4</u>: Κυματομορφές που παράγονται από τη μίξη τεσσάρων μερικά επικαλυπτόμενων ηχογραφήσεων. Οι αρχικές ηχογραφήσεις εμφανίζονται με το μαύρο χρώμα, το αποτέλεσμα της μίξης με τη μέθοδο w-$N^{-1/2}$ εμφανίζεται με μπλέ χρώμα στο (e) ενώ με το κόκκινο και πράσινο χρώμα εμφανίζονται οι μέθοδοι w-N^{-1} και w-1 στο (f) και (g) αντίστοιχα.

Επικεντρώνοντας την προσοχή στην περίπτωση της πράσινης κυματομορφής στο διάγραμμα (g) παρατηρούμε ότι το πλάτος του σήματος τη στιγμή που αρχίζουν να συμμετέχουν περισσότερα αρχεία στη μίξη αυξάνεται σταδιακά. Αρχικά αναπαράγεται μόνο ένα αρχείο που διαμορφώνει το αποτέλεσμα ενώ στα επόμενα δευτερόλεπτα προστίθενται και άλλα αρχεία στη μίξη. Όταν όλα τα αρχεία θα έχουν υπερτεθεί στο χρόνο t=10s το αποτέλεσμα

είναι το σήμα να έχει μεγαλύτερη ενέργεια απ' ότι στην αρχή που υπήρχε μόνο ένα αρχείο. Το αντίθετο ακριβώς συμβαίνει στο διάγραμμα (f) με την κόκκινη κυματομορφή, όπου φαίνεται ότι όσο αυξάνεται ο αριθμός των ηχογραφήσεων που εισέρχονται στη μίξη, η στάθμη το σήματος μειώνεται. Δεομένου ότι αυτές οι μεταβολές της έντασης δεν προέρχονται από τις διακυμάνσεις της ακουστικής σκηνής αλλά από τον τρόπο που γίνεται η μίξη, είναι αναμενόμενο ότι υπονομεύουν την εμπειρία ακρόασης [9].

2.5 Απλή μίξη

Όπως φαίνεται στο Σχήμα 2, κάθε ηχογράφηση ξεκινάει και σταματάει σε αυθαίρετες χρονικές στιγμές, και ο αριθμός των διαθέσιμων ηχητικών κομματιών διαφέρει κάθε χρονική στιγμή. Οι χρονικές στιγμές στις οποίες ο αριθμός των ηχητικών αρχείων που συμμετέχουν στη μίξη αλλάζει ονομάζονται σημεία μετάβασης. Σύμφωνα με το παράδειγμα στο Σχήμα 2, παρατηρήστε ότι οι τοποθεσίες των σημείων μετάβασης είναι $k_1 = 1,\ k_2 = n_3^{start},\ k_3 = n_2^{start},\ k_4 = n_3^{end} + 1,\ k_5 = n_1^{end} + 1$ και $k_6 = n_2^{end} + 1 = N + 1$. Τα χρονικά τμήματα ανάμεσα σε δύο διαδοχικά σημεία μετάβασης χαρακτηρίζονται από την ιδιότητα ότι το πλήθος των UGRs που συμμετέχουν στη διαδικασία της μίξης είναι σταθερό. Σύμφωνα με το παράδειγμα του Σχήματος 2, τα τρία ηχητικά αρχεία καθορίζουν 5 σταθερά τμήματα χρόνου, τα οποία απαριθμούνται με $j = 1, \dots, 5$. Για να εξασφαλίσουμε ότι το αποτέλεσμα της μίξης που προκύπτει από τον συνδυασμό των M ηχητικών αρχείων δεν θα παρουσιάσει απότομες αλλαγές έντασης όταν τα ηχητικά αρχεία αρχίζουν και σταματούν να συμμετέχουν στη μίξη (π.χ. όταν περνάμε τα σημεία μετάβασης), οι στάθμες με τις οποίες τα διαφορετικά σήματα ήχου θα αθροίζονται κατά τη διαδικασία της μίξης θα πρέπει να διαφέρουν σε κάθε χρονικό διάστημα j. Αυτό είναι το βασικό κίνητρο για ας επιλογές ηχητικής μίξης που αναφέρονται στη συνέχεια.

Οι τεχνικές που παρουσιάζονται σε ό,τι ακολουθεί προϋποθέτουν ότι τα ηχητικά αρχεία έχουν κανονικοποιηθεί, χρησιμοποιώντας για παράδειγμα την διαδικασία της επαναληπτικής κανονικοποίησης που περιγράφεται στην προηγούμενη ενότητα. Ας ορίσουμε το σετ $\mathbb{D}_j = \{k_j, k_j + 1, \dots, k_{j+1} - 1\}$ το οποίο περιέχει όλους τους δείκτες των δειγμάτων που ανήκουνε στο j-ιοστό χρονικό τμήμα. Ας, χρησιμοποιήσουμε το συμβολισμό $\boldsymbol{y}_{m,j}$ για

να απευθυνθούμε στο τμήμα του διανύσματος του m-ιοστού σήματος που ανήκει στο j-ιοστό χρονικό τμήμα. Αυτό το διάνυσμα μπορεί να κατασκευαστεί ως

$$\boldsymbol{y}_{m,j} = \tilde{x}_m[n]|_{n \in \mathbb{D}_j} = [\tilde{x}_m[k_j], \dots, \tilde{x}_m[k_{j+1} - 1]]^T, \tag{13}$$

όταν το m-ιοστό UGR είναι ενεργό κατά μήκος του j-ιοστού χρονικού τμήματος και

$$\boldsymbol{y}_{m,j} = \emptyset \tag{14}$$

διαφορετικά, όπου με $\emptyset$. συμβολίζεται το κενό σετ. Αναφορικά με το χρονικό τμήμα με δείκτη j μπορεί τώρα να κατασκευαστεί ο πίνακας

$$\boldsymbol{y}_j = [\boldsymbol{y}_{1,j}, \dots, \boldsymbol{y}_{M,j}], \tag{15}$$

ο οποίος απαρτίζεται από c_j στήλες, όπου $\boldsymbol{c_j}$ είναι το πλήθος των ενεργών ηχητικών αρχείων στο χρονικό τμήμα με δείκτη j. Αξιοποιώντας όλα τα διαθέσιμα σήματα για το συγκεκριμένο χρονικό τμήμα, η διαδικασία της μίξης μπορεί να υλοποιηθεί ως

$$\boldsymbol{s}_j = \boldsymbol{Y}_j \boldsymbol{w}_j = \frac{1}{\sqrt{c_j}} \boldsymbol{Y}_j \mathbb{1}_{c_j \times 1}, \tag{16}$$

όπου $\mathbb{1}_{c_j \times 1}$ είναι ένα $c_j \times 1$ διάνυσμα στήλης γεμάτο με μονάδες και

$$\boldsymbol{w}_j = \frac{1}{\sqrt{c_j}} \mathbb{1}_{c_j \times 1} \tag{17}$$

είναι το διάνυσμα με τα βάρη της μίξης.

Αυτή η διαδικασία μίξης επαναλαμβάνεται για όλα τα χρονικά τμήματα $j = 1, \dots, J$ και το συνολικό σήμα εξόδου προέρχεται από τη συνένωση όλων των επιμέρους χρονικών τμημάτων ως

$$\boldsymbol{s} = [\boldsymbol{s}_1^T, \dots, \boldsymbol{s}_J^T]^T. \tag{18}$$

Το γεγονός ότι η Εξ. (17) είναι σταθμισμένη από την τετραγωνική ρίζα του αριθμού των ενεργών ηχητικών αρχείων υποδηλώνει ότι αυτά τα αρχεία που συμμετέχουν στη διαδικασία της μίξης θεωρούνται στατιστικά ανεξάρτητα. Πράγματι, αυτή η υπόθεση δεν ισχύει πλήρως, αφού εάν διαφορετικά ηχητικά κανάλια ήταν εντελώς διαφορετικά μεταξύ τους, δεν θα ήταν εξαρχής δυνατόν να συγχρονιστούν. Παρ ’όλα αυτά, είναι συνετό να υποθέσουμε ότι ακόμα και όταν αυτά τα κανάλια είναι συγχρονισμένα, ο βαθμός συσχέτισης μεταξύ τους είναι μάλλον μικρός, έτσι ώστε θα έπρεπε να αναμένεται μια αύξηση των 3 dB στην ισχύ σήματος μίξης κάθε φορά που ο αριθμός των συμμετεχόντων αρχείων διπλασιάζεται. Στην περίπτωση

όμως που τα διαφορετικά ηχητικά αρχεία παρουσιάζουν ισχυρότερους συσχετισμούς, η κατ' αυτόν τον τρόπο συνένωση των σημάτων μπορεί να εμφανίζει φαινόμενα δημιουργικής ή καταστρεπτικής συμβολής. Καθότι αυτό μπορεί να οδηγήσει σε ανεπιθύμητες αλλαγές της στάθμης στα σημεία μετάβασης, ορίζουμε σε αυτά που ακολουθούν μια προσέγγιση για την περαιτέρω στάθμιση των βαρών μίξης έτσι ώστε να αποφύγουμε αυτό το πρόβλημα.

2.6 Προσαρμοστική μίξη

Για να εξαλείψουμε ακόμα περισσότερο τις ακουστές αυξομειώσεις έντασης μεταξύ δύο συνεχών σταθερών χρονικών τμημάτων στην λαμβανόμενη μίξη, μπορούμε να χρησιμοποιήσουμε μία εκτίμηση της ενέργειας του παραγόμενου σήματος σε κάθε χρονικό τμήμα j και να τη χρησιμοποιήσουμε με σκοπό να κλιμακώσουμε τα βάρη της μίξης που αναφέρονται στην Εξ. (17) ώστε να επιτευχθεί μια στοχευμένη στάθμη ενέργειας $\tilde{q}_j$. Πράγματι, χρησιμοποιώντας τα βάρη $\boldsymbol{w}_j$ από την Εξ. (15), η ισχύς του σήματος της μίξης στο j χρονικό τμήμα είναι

$$q_j = \boldsymbol{w}_j^T \boldsymbol{Y}_j^T \boldsymbol{Y}_j \boldsymbol{w}_j. \tag{19}$$

Τώρα μπορεί να δημιουργηθεί ένα καινούργιο διάνυσμα στάθμισης βασιζόμενο στην αναλογία μεταξύ της στοχευμένης και της πραγματικής ισχύος σήματος ως εξής

$$\tilde{\boldsymbol{w}}_j = \frac{\sqrt{\tilde{q}_j}}{\sqrt{q_j}} \boldsymbol{w}_j \tag{20}$$

Μία λογική επιλογή για τη στοχευμένη ισχύ στο j χρονικό τμήμα είναι

$$\tilde{q}_j = \frac{1}{\sqrt{c_j}} tr\{\boldsymbol{Y}_j^T \boldsymbol{Y}_j\} \tag{21}$$

Όπου $tr\{\cdot\}$ είναι το ίχνος του πίνακα. Σε αυτή την περίπτωση, τα σταθμίσματα μίξης γίνονται

$$\tilde{\boldsymbol{w}}_j = \frac{1}{\sqrt{c_j}} \sqrt{\frac{tr\{\boldsymbol{Y}_j^T \boldsymbol{Y}_j\}}{\mathbb{1}^T \boldsymbol{Y}_j^T \boldsymbol{Y}_j \mathbb{1}}} \mathbb{1}, \tag{22}$$

με $\mathbb{1}$ να συμβολίζει το $\boldsymbol{c_j} \times \mathbf{1}$ μοναδιαίο διάνυσμα. Σύμφωνα με την Εξ. (16) πάλι, η μίξη στο j χρονικό τμήμα μπορεί να υλοποιηθεί ως $\boldsymbol{s_l} = \boldsymbol{Y_l}\tilde{\boldsymbol{w}}_{\boldsymbol{l}}$ και η τελική μίξη θα είναι το αποτέλεσμα από τη συνένωση όλων των χρονικών τμημάτων, σύμφωνα με την Εξ. (18).

3 Αξιολόγηση

Σκοπός των προτεινόμενων τεχνικών είναι να παραχθεί μια ηχητική μίξη η οποία δε θα παρουσιάζει αισθητές μεταβολές κατά τα σημεία μετάβασης. Για να ελεγχτεί πόσο καλά επιτυγχάνεται αυτό, διαπιστώσαμε ότι η αξιολόγηση πρέπει να γίνει με βάση δύο διαφορετικού τύπου υποθέσεις. Ο πρώτος τύπος επικεντρώνεται στις μεταβολές της έντασης, που είναι και το βασικό χαρακτηριστικό που λαμβάνεται υπόψιν κατά το σχεδιασμό των τεχνικών μίξης. Οι δύο υποθέσεις πάνω στις οποίες θα γίνει η αξιολόγηση είναι οι εξής:

Ε0: Μεταβολές στην ένταση δεν γίνονται αντιληπτές.

Ε1: Μεταβολές στην ένταση γίνονται αντιληπτές.

Είναι σημαντικό για τη σωστή διεκπεραίωση της αξιολόγησης ο ακροατής να μην παρασυρθεί από άλλους παράγοντες που μπορούν να μεταβάλλονται κατά τη διαδικασία της ακρόασης και να δώσει απαντήσεις αποκλειστικά ως προς το εξεταζόμενο χαρακτηριστικό. Ως ένα μεγάλο βαθμό, είδαμε ότι ενώ μεταβολές στην ένταση μπορεί να μην είναι ακουστές, είναι ωστόσο αισθητές κάποιες μεταβολές οι οποίες σχετίζονται περισσότερο με το χαρακτηριστικό της ποιότητας του ηχητικού υλικού. Για να μπορέσουμε λοιπόν να διαχωρίσουμε αισθητές μεταβολές που αφορούν την ένταση από αυτές που αφορούν την ποιότητα, διατυπώνουμε άλλον ένα τύπο υπόθεσης που αφορά αποκλειστικά την ποιότητα ως εξής:

Π0: Μεταβολές στην ποιότητα δεν γίνονται αντιληπτές.

Π1: Μεταβολές στην ποιότητα γίνονται αντιληπτές.

3.1 Συλλογή ηχητικών δεδομένων

Για την δημιουργία του ερωτηματολογίου χρειάστηκε να τοποθετηθούν ηχογραφήσεις από μια πληθώρα ηχητικών γεγονότων. Παρακάτω ακολουθεί ο πίνακας με τις πληροφορίες για τις τοποθεσίες που έλαβαν μέρος οι ηχογραφήσεις των ηχητικών γεγονότων.

Δημόσιο γεγονός	Χώρος	Αριθμός Εμφανίσεων στο ερωτηματολόγιο	Τοποθεσία
Θέατρο	Κλειστός χώρος	9	Αμφιθέατρο πανεπιστημίου Κρήτης
Ποδόσφ. αγώνας	Ανοιχτός χώρος	5	Γήπεδο ΟΦΗ
Μουσική Παράσταση	Ανοιχτοί και κλειστοί χώροι	24	Πλατεία Ελευθερίας, Πανεπιστήμιο Κρήτης, cine-Studio

Πίνακας 1: Πληροφορίες για την τοποθεσία που πήρε μέρος το κάθε γεγονός καθώς και τον αριθμό εμφανίσεων στο ερωτηματολόγιο

Μικρότερη διάρκεια ηχητικού αρχείου : 00:09 λεπτά

Μεγαλύτερη διάρκεια ηχητικού αρχείου : 00:17 λεπτά

Συνολική διάρκεια ηχητικών αρχείων : 07: 28 λεπτά

3.1.1 Μέθοδοι που χρησιμοποιήθηκαν

Εκτός από τις δυο προτεινόμενες τεχνικές μίξης με μεταβλητά βάρη που προτείνονται στο προηγούμενο κεφάλαιο, παρουσιάζονται αποτελέσματα για δύο επιπλέον προσεγγίσεις. Η πρώτη προσέγγιση αφορά τη χρήση σταθερών βαρών που δε μεταβάλλονται καθόλου κατά τη διάρκεια της μίξης. Η χρήση αυτή της αφελούς προσέγγισης αποσκοπεί στο να μας δώσει μία ένδειξη του τι θα συνέβαινε αν δε λαμβάναμε κανένα μέτρο για να αντισταθμίσουμε τις μεταβολές στο πλήθος των ηχογραφήσεων. Μια επιπλέον τεχνική που εξετάζεται είναι αυτή της χρήσης μεταβλητών βαρών που είναι αντιστρόφως ανάλογα του πλήθους των ηχογραφήσεων, δηλαδή ίσα με $1/c_j$ όπου c_j είναι το πλήθος των ηχογραφήσεων που συμμετέχουνε στο j-ιοστό χρονικό τμήμα. Αν και η απλή αυτή προσέγγιση φαίνεται να παρέχει κάποια λύση για την εξάλειψη του προβλήματος, δεν είναι σε συμφωνία με την υπόθεση της ανεξαρτησίας που διατυπώνεται στην ενότητα 2.2.

Για συντομία στην παρουσίαση και ανάλυση των αποτελεσμάτων, ακολουθείται ο παρακάτω συμβολισμός για αναφορά στις διαφορετικές μεθόδους μίξης:

w-1: Συμβολίζουμε κατ' αυτόν τον τρόπο τη μέθοδο μίξης που χρησιμοποιεί σταθερά βάρη ίσα με 1. Τα βάρη δηλαδή παραμένουν σταθερά καθ' όλη τη διάρκεια της μίξης ηχητικών αρχείων.

w-N^{-1}: Συμβολίζουμε κατ' αυτόν τον τρόπο τη μέθοδο μίξης που χρησιμοποιεί μεταβλητά ίσα με $1/c_j$ όπου c_j είναι το πλήθος των ηχογραφήσεων που συμμετέχουνε στο j-ιοστό χρονικό τμήμα.

w-$N^{-1/2}$: Συμβολίζουμε κατ' αυτόν τον τρόπο τη μέθοδο μίξης που αναπτύχθηκε στο Κεφάλαιο 2.5 και που χρησιμοποιεί μεταβλητά βάρη ίσα με $\frac{1}{\sqrt{c_j}}$. όπου c_j είναι το πλήθος των ηχογραφήσεων που συμμετέχουνε στο j-ιοστό χρονικό τμήμα.

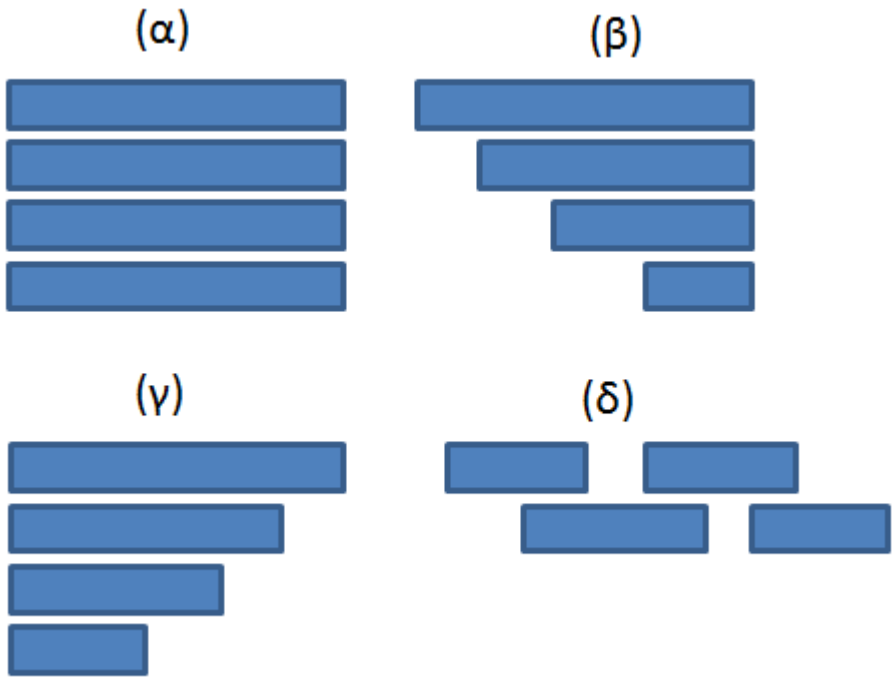
w-adapt: Συμβολίζουμε κατ' αυτόν τον τρόπο την προσαρμοστική μέθοδο μίξης που χρησιμοποιεί μεταβλητά βάρη και αναπτύχθηκε στο Κεφάλαιο 2.6.

3.1.2 Προ-επεξεργασία των ηχητικών αρχείων

Για τις ανάγκες της αξιολόγησης θεσπίστηκαν κάποια κριτήρια που να κάνουν το τεστ όσο γίνεται πιο «φιλικό» για τον ακροατή και ταυτόχρονα να βοηθούν στην αξιοπιστία των αποτελεσμάτων. Καταρχάς, θεωρήθηκε θεμιτό να έχουμε αποσπάσματα από διαφορετικούς τύπους δημοσίων γεγονότων, ώστε τα συμπεράσματα που εξάγονται να είναι αντιπροσωπευτικά για μεγάλο εύρος ηχητικού περιεχομένου. Επιπλέον, αποσπάσματα από το ίδιο δημόσιο γεγονός θα πρέπει να είναι όσο γίνεται ανακατεμένα και να μην εμφανίζονται συνεχόμενα στο τεστ, ώστε ο ακροατής να μην κουράζεται από την επανάληψη του ίδιου ηχητικού περιεχομένου κάθε φορά. Επίσης, η συνολική διάρκεια των ηχητικών αρχείων που παρουσιάζονται στο τεστ δεν πρέπει να υπερβαίνει κάποιο συγκεκριμένο όριο, ώστε το τεστ να μη γίνεται κουραστικό για τον ακροατή. Για να επιτευχθεί αυτό, αποφασίστηκε το κάθε απόσπασμα να είναι διάρκειας μικρότερης των 15 δευτερολέπτων.

Επίσης για τις ανάγκες της αξιολόγησης θεωρήθηκε σημαντικό το κάθε αρχείο ήχου να περιέχει αρκετά σημεία μετάβασης ώστε δίνεται η ευκαιρία στον ακροατή να εντοπίσει το πρόβλημα. Για να γίνει αυτό, η διαδικασία που ακολουθήθηκε είναι η εξής. Αποσπάσματα από κάθε συλλογή επιλέχθηκαν από περιοχές του χρόνου που υπήρχε ολική χρονική επικάλυψη από τουλάχιστον τέσσερα αρχεία ήχου, σχηματίζοντας δηλαδή δομές σαν αυτή

του Σχήματος 5(α). Στη συνέχεια, η διάρκεια τριών από των τεσσάρων αυτών ηχογραφήσεων τροποποιήθηκε με σκοπό να δημιουργηθούν τρία σημεία μετάβασης, όπως φαίνεται στο Σχήμα 5(β) και (γ). Κατ' αυτόν τον τρόπο δημιουργήθηκαν μεταβάσεις τύπου σκάλας, όπου το πλήθος των ηχογραφήσεων που συμμετέχουν στη μίξη αυξάνεται διαδοχικά από 1 σε 4 ή μειώνεται διαδοχικά από 4 σε 1. Εξαίρεση αποτελούν κάποια αποσπάσματα από μια μουσική εκδήλωση που έλαβε χώρα στο cine-Studio στο Ηράκλειο, όπου το πλήθος των χρονικά επικαλυπτόμενων ηχογραφήσεων ήταν μόνο δύο. Σε αυτήν την περίπτωση, σημεία μετάβασης δημιουργήθηκαν με βάση την κατάτμηση των ηχογραφήσεων σύμφωνα με το Σχήμα 5(δ). Η συγκεκριμένη προσέγγιση αποδείχτηκε ιδιαίτερα αποτελεσματική για δημιουργία των ηχητικών δεδομένων που είναι πλούσια σε σημεία μετάβασης, χωρίς να αυξάνεται ιδιαίτερα ο όγκος της πληροφορίας που παρουσιάζεται στον ακροατή και επομένως, η διάρκεια του τεστ.



<u>Σχήμα 5</u>: Προ-επεξεργασία των ηχογραφήσεων σε περιοχές ολικής επικάλυψης για δημιουργία των δεδομένων αξιολόγησης.

3.2 Ερωτηματολόγιο

Το ερωτηματολόγιο δημιουργήθηκε στο PowerPoint και εμπεριέχει από ένα αρχείο ήχου σε κάθε ερώτηση όπου ο χρήστης θα χρειαστεί να το ακούσει, και στη συνέχεια θα πρέπει να απαντήσει σε μία από τις 4 επιλογές που του παρατίθενται. Για τη δημιουργία του ερωτηματολογίου επιλέξαμε το κάθε αρχείο ήχου (ή απόσπασμα) να εμφανίζεται σε διαφορετικές παραλλαγές έτσι ώστε κάθε παραλλαγή να προκύπτει από επεξεργασία με διαφορετική μέθοδο μίξης κάθε φορά. Βασική προϋπόθεση είναι οι διαφορετικές παραλλαγές από το ίδιο ηχητικό απόσπασμα να μην εμφανίζονται σε διαδοχικές ερωτήσεις, ώστε η απόκριση του ακροατή να επηρεάζεται όσο το δυνατόν λιγότερο από τις προηγούμενες απαντήσεις του. Έγινε ξεκάθαρο επίσης ότι ο ακροατής πρέπει να φοράει ακουστικά για τη συμπλήρωση του τεστ, ώστε να επηρεάζεται όσο το δυνατόν λιγότερο από θόρυβο περιβάλλοντος. Στο Σχήμα 6 παρουσιάζεται ένα παράδειγμα όπου φαίνεται πως οι διαφορετικές επιλογές διατυπώθηκαν στο ερωτηματολόγιο.

Αντιλαμβάνεστε κάποια μεταβολή στην ένταση ή στην ποιότητα του ήχου;
Συμπληρώστε τον παρακάτω πίνακα βάζοντας «Χ» στο αντίστοιχο πλαίσιο.

Δεν αντιλαμβάνομαι κάποια μεταβολή	Αντιλαμβάνομαι μεταβολή στην ένταση	Αντιλαμβάνομαι μεταβολή στην ποιότητα	Αντιλαμβάνομαι μεταβολή και στην ένταση και στην ποιότητα
Ε0 και Π0	Ε1 και Π0	Ε0 και Π1	Ε1 και Π1

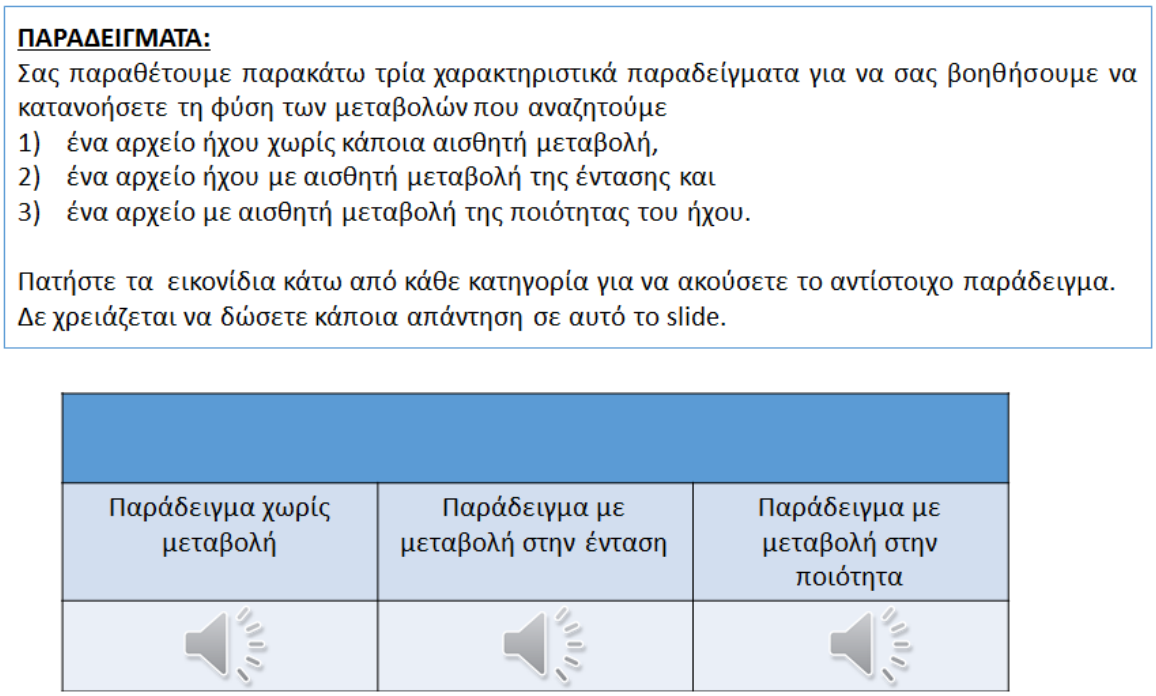
Σχήμα 6: Δείγμα δομής ερωτηματολογίου.

Βλέπουμε ότι κάθε απάντηση δίνει ένα πόντο σε δύο υποθέσεις κάθε φορά. Για παράδειγμα, όταν ένας ακροατής επιλέγει το πρώτο κουτί «Δεν αντιλαμβάνομαι κάποια μεταβολή» τότε αυτόματα επαληθεύονται τόσο η υπόθεση Ε0 όσο και η Π0, αφού αυτό σημαίνει ότι ούτε μεταβολές στην ένταση, ούτε στην ποιότητα έγιναν αισθητές. Από την άλλη, όταν για παράδειγμα για κάποιο ερώτημα διατυπώνεται ότι «αντιλαμβάνομαι μεταβολή μόνο στην ποιότητα» τότε μπορούμε να θεωρήσουμε ότι δεν επιβεβαιώνεται μόνο η Π1 αλλά και η Ε0.

Με βάση αυτή τη λογική επομένως, έχουμε προσθέσει στην τελευταία γραμμή του Σχήματος 6 ποια κουτιά (απαντήσεις) σχετίζονται με ποιες υποθέσεις. Παρατηρείστε ότι από μία μόνο απάντηση εξάγονται συμπεράσματα και για τα δυο χαρακτηριστικά κάθε φορά.

Για τον έλεγχο της αξιοπιστίας της διεργασίας, εκτός από τις 4 μεθόδους μίξης, προσθέσαμε και ένα αρχείο αναφοράς όπου το ηχητικό απόσπασμα δεν έχει υποστεί επεξεργασία με κάποια μέθοδο, αλλά εξάγεται απευθείας από κάποια ηχογράφηση. Προφανώς σε αυτήν την περίπτωση δεν υπάρχουν σημεία μετάβασης και επομένως ο ακροατής που το συμπληρώνει δεν θα πρέπει να αντιληφθεί κάποια μεταβολή, ούτε όσον αφορά την ένταση ούτε όσον αφορά την ποιότητα. Αυτό έγινε με σκοπό την επαλήθευση των αποτελεσμάτων του ερωτηματολογίου, ότι δηλαδή είναι σωστά δομημένο, αλλά και ο ακροατής έχει τηρήσει τις προϋποθέσεις για να το συμπληρώσει.

Το ερωτηματολόγιο αποτελείται από 37 ερωτήσεις, έτσι ώστε τα ηχητικά αποσπάσματα που αφορούν κάθε μέθοδο να έχουν λίγο πολύ τον ίδιο αριθμό. Στις πρώτες σελίδες του ερωτηματολογίου αναγράφονται οδηγίες για την συμπλήρωσή του, καθώς και ένα παράδειγμα με σκοπό την εξοικείωση του ακροατή με του διαφορετικού τύπου μεταβολές (έντασης και ποιότητας) που θα πρέπει να παρατηρήσει (βλ. Σχήμα 7).



Σχήμα 7: Ηχητικό παράδειγμα για την εξοικείωση του ακροατή με τους διαφορετικούς τύπους μεταβολών.

3.3 Αποτελέσματα

Το ερωτηματολόγιο απαντήθηκε από 32 άτομα και παρακάτω ακολουθούν τα αποτελέσματα των απαντήσεων ως προς την κάθε μέθοδο.

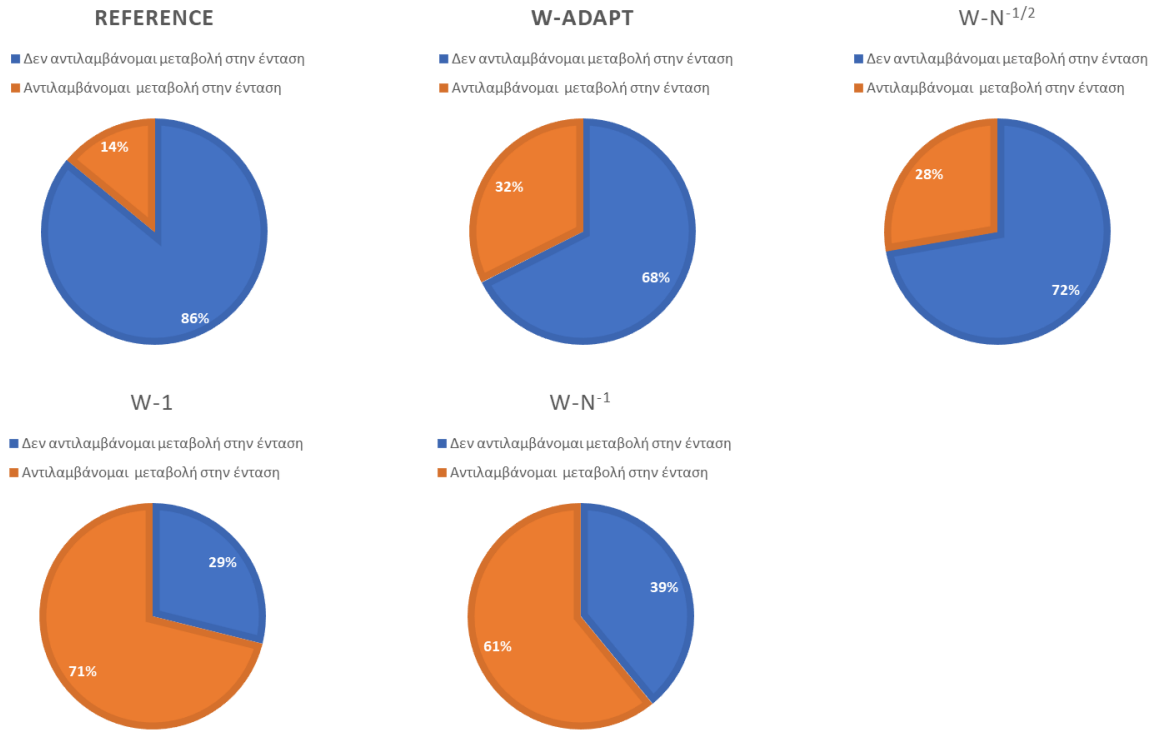
3.3.1 Αποτελέσματα απαντήσεων ως προς τη μέθοδο

Η ανάλυση των αποτελεσμάτων έγινε με βάση τέσσερις διαφορετικές υποθέσεις Ε0, Ε1, Π0 και Π1 που αναφέρονται στην ενότητα 3.2. Οι αριθμοί προκύπτουν ως εξής, όσοι ακροατές δήλωσαν για παράδειγμα ότι αντιλαμβάνονται μεταβολή μόνο στην ένταση, αυτομάτως δεν αντιλαμβάνονται μεταβολή στην ποιότητα. Αντιθέτως αυτοί που αντιλαμβάνονται μεταβολή μόνο στην ποιότητα δεν αντιλαμβάνονται στην ένταση. Με βάση αυτή τη λογική επομένως, προστέθηκαν οι αριθμοί των απαντήσεων των ηχητικών αρχείων που ανήκουν στην ίδια μέθοδο.

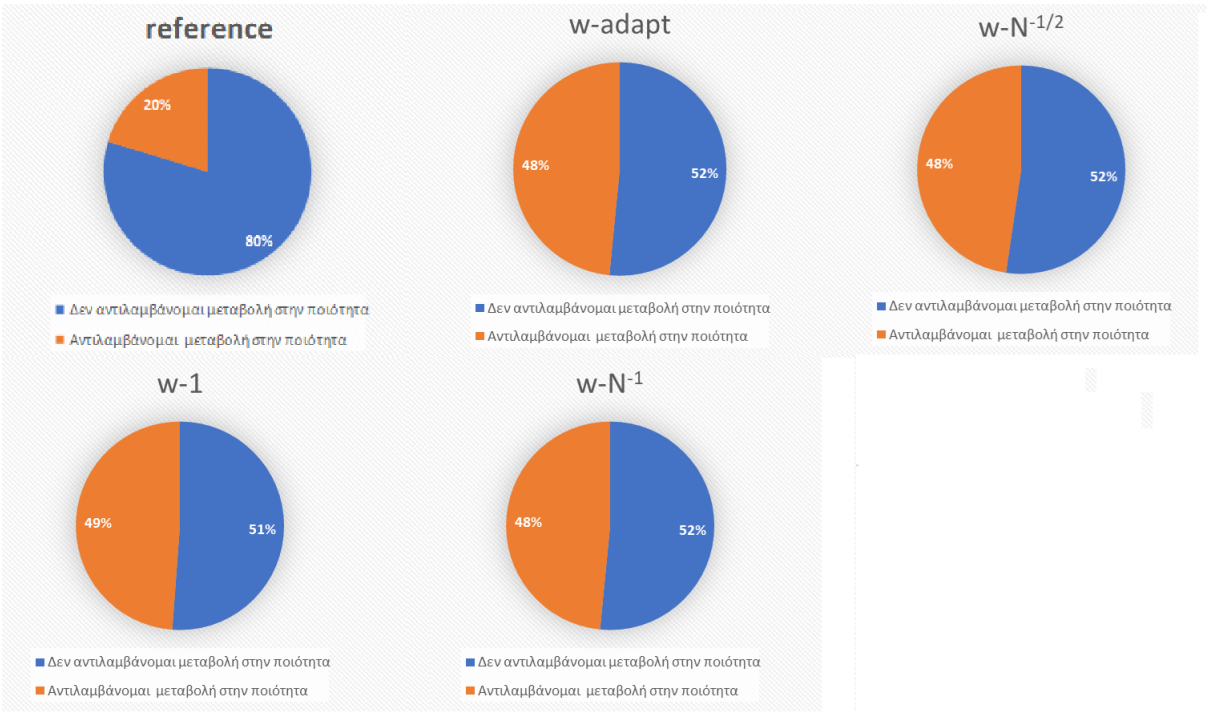
	Δεν αντιλαμβάνομαι μεταβολή στην ένταση (Ε0)	**Αντιλαμβάνομαι μεταβολή στην ένταση (Ε1)**	**Δεν αντιλαμβάνομαι μεταβολή στην ποιότητα (Π0)**	**Αντιλαμβάνομαι μεταβολή στην ποιότητα (Π1)**
reference	165	27	153	39
w-adapt	173	66	132	107
w-$N^{-1/2}$	185	71	134	122
w-1	74	182	131	125
w-N^{-1}	100	156	132	124

Πίνακας 2: Αποτελέσματα απαντήσεων ως προς κάθε μέθοδο

Στα Σχήματα 8 και 9 παρουσιάζεται η κατανομή των απαντήσεων ως προς κάθε μέθοδο σε μορφή pie charts.



<u>Σχήμα 8</u>: Κατανομή των απαντήσεων αναφορικά με τις υποθέσεις Ε0 και Ε1 (μεταβολή έντασης).



<u>Σχήμα 9</u>: Κατανομή των απαντήσεων αναφορικά με τις υποθέσεις (Π0 και Π1) (μεταβολές ποιότητας).

3.3.2 Ανάλυση αποτελεσμάτων

Όπως παρατηρούμε, στα αρχεία αναφοράς, οι περισσότεροι ακροατές δεν αντιλήφθηκαν μεταβολή στην ένταση. Πιο συγκεκριμένα, το 14% έναντι 86% των απαντήσεων δηλώνουν μεταβολή στην ένταση. Από αυτό συμπεραίνει κανείς πως τα αποτελέσματα είναι αξιόπιστα κατά βάση, και οι ακροατές ακολούθησαν σωστά τις οδηγίες. Οι ακροατές που απάντησαν ότι αντιλαμβάνονται μεταβολή στην ένταση μας δίνουν επίσης το συμπέρασμα ότι ίσως υπάρχει μια μορφή προκατάληψης στον άνθρωπο να παρατηρεί μεταβολή ενώ δεν υπάρχει, με βάση πιθανόν τον τρόπο που τον έχουν προϊδεάσει. Επιπλέον, παρόλο που τα ηχητικά αρχεία αναφοράς δεν φέρουν κάποια επεξεργασία, οι φυσικές διακυμάνσεις του ηχητικού περιεχομένου (μουσικές παύσεις/εξάρσεις, φωνές κ.τ.λ.) μπορούν σε ορισμένες περιπτώσεις να μπερδέψουν τον ακροατή.

Αναφορικά με τις μεθόδους μίξης που εφαρμόστηκαν πάνω σε μερικώς επικαλυπτόμενα αρχεία ήχου οι διαφορές σχετικά με την αντίληψη της έντασης είναι εμφανείς στο Σχήμα 8. Στις μεθόδους w-adapt και $w\text{-}N^{-1/2}$ το ποσοστό που επαληθεύει την υπόθεση Ε0 (όχι μεταβολή στην ένταση) είναι 68% και 72% αντίστοιχα. Από την άλλη, το ίδιο ποσοστό στις μεθόδους w-1 και $w\text{-}N^{-1}$ είναι 29% και 39% αντίστοιχα. Από αυτό φαίνεται ότι οι w-1 και $w\text{-}N^{-1}$ τεχνικές δεν αντισταθμίζουν σωστά τη διαδοχική μείωση ή αύξηση του πλήθους των UGR, οδηγώντας σε αισθητές μεταβολές της έντασης κατά τη διάρκεια της ακρόασης.

Όσον αφορά την μεταβολή στην ποιότητα, είναι αξιοσημείωτο ότι στο 20% των περιπτώσεων η υπόθεση Π1 επιβεβαιώνεται στα αρχεία αναφοράς, παρόλο που το ηχητικό απόσπασμα έχει εξαχθεί από μία μόνο συσκευή που καταγράφει με σταθερά στο χρόνο ποιοτικά χαρακτηριστικά (βλ. Σχήμα 9). Αυτό δείχνει πάλι μια προκατάληψη του ακροατή να εντοπίζει μεταβολές εκεί που δεν υπάρχει καμία. Σε όλες τις υπόλοιπες μεθόδους, ο αριθμός των απαντήσεων που δόθηκε είναι σχεδόν ίδιος και για τις δύο περιπτώσεις, δηλαδή η υπόθεση Π0 και Π1 επιβεβαιώνεται με ποσοστό κοντά στο 50%. Είναι αξιοσημείωτο ότι οι διαφοροποιήσεις από μέθοδο σε μέθοδο είναι αμελητέες. Ως ένα βαθμό αυτό επιβεβαιώνει την αξιοπιστία της έρευνας, καθότι οι μέθοδοι που εξετάζονται διαφοροποιούνται ως προς τον τρόπο που διαχειρίζονται τη στάθμη του σήματος και όχι άλλα ποιοτικά χαρακτηριστικά (πχ. διαφορές στη συχνοτική απόκριση λόγω διαφορετικών συσκευών ή θέσης στο χώρο).

4 Συμπεράσματα

Επιτυγχάνεται μια στατιστικά σημαντική διαφορά συγκρίνοντας την w-$N^{-1/2}$ με την w-N^{-1} και την w-1. Αυτό καταδεικνύει το γεγονός ότι υπάρχει σημαντική διαφοροποίηση μεταξύ των αποτελεσμάτων που παράγει η κάθε μέθοδος. Συγκεκριμένα, όπως έχει είδη αποδειχθεί στην ενότητα των αποτελεσμάτων, μεταβολές στην ένταση με τη μέθοδο w-$N^{-1/2}$ γίνονται σπανιότερα αντιληπτές από ότι με τις συμβατικές υλοποιήσεις των w-1 και w-N^{-1} τεχνικών. Ομοίως με τα προηγούμενα, υπάρχει μια στατιστικά σημαντική διαφορά μεταξύ της w-adapt και των w-1 και w-N^{-1} τεχνικών. Το γεγονός ότι οι ακροατές αντιλαμβάνονται ανύπαρκτες διαφορές στην ένταση στα αρχεία αναφοράς (με ποσοστό έως και 14%) καταδεικνύει ότι τα ποσοστά 72% και 68% με τα οποία επιβεβαιώνεται η υπόθεση Ε0 είναι στην πράξη αρκετά ικανοποιητικά.

Ενάντια στην αρχική μας διαίσθηση ωστόσο, δεν παρατηρείται κάποια στατιστικά σημαντική διαφορά συγκρίνοντας την w-$N^{-1/2}$ με την w-adapt. Οι δύο αυτές τεχνικές είναι εξίσου αποτελεσματικές στο να αντισταθμίσουν τις διαφορές στη στάθμη του σήματος, παρόλο που η w-adapt τεχνική περιλαμβάνει ένα επιπλέον στάδιο που αποσκοπεί στο να διορθώσει φαινόμενα καταστροφικής ή δημιουργικής συμβολής κατά την υπέρθεση των ηχοφραφήσεων λόγω μη ικανοποίησης της υπόθεσης της ανεξαρτησίας. Αξίζει να σημειωθεί βέβαια ότι η w-$N^{-1/2}$ τεχνική είναι υπολογιστικά λιγότερο απαιτητική από την w-adapt, γεγονός που της προσδίδει ένα επιπλέον πλεονέκτημα.

Δεν παρατηρείται κάποια στατιστικά σημαντική διαφορά συγκρίνοντας τις διαφορετικές τεχνικές ως προς την υπόθεση της μεταβολής της ποιότητας (Π0 και Π1). Συγκεκριμένα, μεταβολές ως προς την ποιότητα γίνονται αισθητές σε όλες τις τεχνικές με την ίδια συχνότητα και μάλιστα σε σχετικά υψηλά ποσοστά (κοντά στο 50%). Αυτό μπορεί να εξηγηθεί ως ένα βαθμό από το γεγονός ότι όσο αυξάνεται ή μειώνεται το πλήθος των ηχογραφήσεων που συμμετέχουν στη μίξη, η αντιληπτή ποιότητα του ήχου βελτιώνεται ή χειροτερεύει αντίστοιχα. Σε αυτό συνηγορεί η μελέτη που παρατίθεται στη δημοσίευση [10]. η οποία έδειξε ότι τα αρχεία που παράγονται με την υπέρθεση πολλών συγχρονισμένων UGR βελτιώνουν την εμπειρία ακρόασης συγκριτικά με την ακρόαση μεμονωμένων καταγραφών που προέρχονται από μία μόνο συσκευή.

5 Βιβλιογραφία

1. Kammerl, J., et al, “Temporal synchronization of multiple audio signals,” in Proceeding of Acoustics, Speech and Signal Processing (ICASSP), 2014

2. M. Muller, F. Kurth and M. Clausen, “Audio matching via chroma-based statistical features,” in 6th International Conference on Music Information Retrieval (ISMIR), 2005.

3. S. Cavallaro and A. Bano, “Discovery and organization of multi-camera user-generated

videos of the same event,” Information Sciences, vol. 302, p. 108–121, 2015.

4. Stefanakis, N., Chonianakis S., and Mouchtaris A., “Automatic matching and synchronization of user generated videos from a large scale sport event,” in Proceeding of Acoustics, Speech and Signal Processing (ICASSP), 2017.

5. C. Ellis and D. Cotton, “Audio fingerprinting to identify multiple videos of an event,” in Proceeding of Acoustics, Speech and Signal Processing (ICASSP), 2010.

6. H. Ozer, B. Sankur and N. Memon, “Robust audio hashing for audio identification,” in Proceedings of Eur. Signal Process. Conf. (EUSIPCO), 2004.

7. A. Wang, “The Shazam Music Recognition Service,” Commun. ACM, vol. 49, pp. 44-48, 2006.

8. Stefanakis, N. and Mouchtaris A., “Normalization of partly overlapping recordings from the same event based on relative signal powers,” in Proceeding of Acoustics, Speech and Signal Processing (ICASSP), 2018.

9. “Loudness normalisation and permitted maximum level of audio signals.,” [Online]. Available: https://tech.ebu.ch/docs/r/r128.pdf.

10. Stefanakis, N.,Viskadouros, M. and Mouchtaris, A., “A subjective evaluation on mixtures of crowdsourced audio recordings”, in Proceeding of European Signal Processing Conference (EUSIPCO), 2017.