



Σχολή Εφαρμοσμένων Τεχνών
Ακουστικός Σχεδιασμός και Ψηφιακός Ήχος

Διπλωματική Εργασία
Αξιολόγηση Εργαλείων Λογισμικού για την Αποθροφυβοποίηση
Ψηφιακού Ήχου με Έμφαση στις Σύγχρονες Τεχνικές Τεχνητής
Νοημοσύνης

Παναγιώτης Κουκοβέτσος

Επιβλέπων καθηγητής: Βρύσης Λάζαρος

Αταλάντη, Ιούνιος 2026

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή («συγγραφέας/δημιουργός») που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο ΕΑΠ, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογής, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσής τους διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο διάρκειας των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, αποθήκευση, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (downloading), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.



Αξιολόγηση Εργαλείων Λογισμικού για την Αποθρομβοποίηση
Ψηφιακού Ήχου με Έμφαση στις Σύγχρονες Τεχνικές Τεχνητής
Νοημοσύνης

Παναγιώτης Κουκοβέτσος

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής:

Λάζαρος Βρύσης

ΕΔΠ, Τμήμα Δημοσιογραφίας και
Μέσων Μαζικής Επικοινωνίας ΑΠΘ

Συν-Επιβλέπων Καθηγητής:

Χαράλαμπος Δημούλας

Καθηγητής, Τμήμα Δημοσιογραφίας και
Μέσων Μαζικής Επικοινωνίας ΑΠΘ

Αταλάντη, Ιούνιος 2026

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα, τον επιβλέποντα καθηγητή μου κ. Λάζαρο Βρύση για την ουσιαστική συμβολή του, την καθοδήγηση και την αμέριστη βοήθειά του κατά την διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.

Επίσης θα ήθελα να ευχαριστήσω θερμά τον β' επιβλέποντα καθηγητή μου κ. Χαράλαμπο Δημούλα, καθώς και τους καθηγητές των θεματικών ενοτήτων του προγράμματος σπουδών Ακουστικός Σχεδιασμός και Ψηφιακός Ήχος (ΑΣΠ) του ΕΑΠ, κ. Νικόλαο Μάρκα, Γεώργιο Καλλίρη και Παναγιώτη Χατζηαντωνίου, για όσα ο καθένας προσέφερε στα πλαίσια του μεταπτυχιακού αυτού προγράμματος.

Ευχαριστώ πολύ τον διδάκτορα κ. Ιορδάνη Θωΐδη για τις χρήσιμες συμβουλές του και την διάθεση δειγμάτων ομιλίας από το αρχείο του, μέρος των οποίων χρησιμοποιήθηκε κατά την υλοποίηση αυτής της διπλωματικής εργασίας.

Παράλειψή μου θα ήταν να μην ευχαριστήσω κάθε έναν και μία ξεχωριστά, όλα τα άτομα που συμμετείχαν στις κριτικές ακροάσεις και την αξιολόγηση της αποτελεσματικότητας των υπό εξέταση εργαλείων, για την πολύτιμη βοήθειά τους.

Καταλήγοντας θα ήθελα να ευχαριστήσω την σύζυγό μου Λίνα και την κόρη μου Κορίνα, για την συμπαράσταση και την κατανόηση τους, κατά την διάρκεια των σπουδών ενός ακόμα μεταπτυχιακού μου, το οποίο τους υπόσχομαι ότι θα είναι το τελευταίο.

*Αφιερώνεται σε όλους τους θορύβους
που κάποτε υπήρξαν ήχοι αγαπημένοι*

Περίληψη

Η παρούσα διπλωματική εργασία, έχει σαν αντικείμενό της τη διερεύνηση των τεχνικών που χρησιμοποιούνται για την αφαίρεση θορύβου από ψηφιακά ηχητικά σήματα, τόσο συμβατικών όσο και αυτών που βασίζονται σε τεχνητή νοημοσύνη.

Σκοπός της εργασίας είναι η μελέτη της απόδοσης και της πρακτικής χρησιμότητας κατάλληλα επιλεγμένων εφαρμογών, με στόχο τη βελτίωση της ακουστικής ποιότητας, σε ηχητικά σήματα στα οποία ενυπάρχει θόρυβος.

Η προσέγγιση που ακολουθείται, περιλαμβάνει συνδυασμό βιβλιογραφικής έρευνας και εφαρμοσμένης αξιολόγησης κατάλληλα επιλεγμένων εργαλείων λογισμικού, μέσω διενέργειας ακροάσεων και σχολιασμών από άτομα με σχετική εμπειρία στον ήχο.

Στα προσδοκώμενα αποτελέσματα, περιλαμβάνονται η συγκριτική αποτίμηση της αποτελεσματικότητας των επιλεγμένων εργαλείων, καθώς και η διερεύνηση των πιθανών εφαρμογών τους.

Λέξεις - Κλειδιά

Λογισμικό, θόρυβος, αποθορυβοποίηση, ψηφιακός ήχος, τεχνητή νοημοσύνη.

Evaluation of Software Tools for Digital Audio Denoising with an Emphasis on Modern Artificial Intelligence Techniques

Panagiotis Koukovetsos

Abstract

The objective of current thesis, is the investigation of the techniques used for noise reduction in digital audio content, including both conventional methods and approaches based on artificial intelligence.

The aim of the study is to examine the performance and practical usefulness of appropriately selected applications, with the objective of improving audio quality in signals affected by noise.

The adopted methodology combines a theoretical background review with an applied evaluation of selected software tools, conducted through listening tests and qualitative assessments by individuals with relevant experience in audio.

The anticipated results, include a comparative assessment of the effectiveness of the selected tools, as well as an exploration of their potential applications.

Keywords

Software, noise, denoising, digital audio, artificial intelligence.

Περιεχόμενα

Ευχαριστίες	iv
Περίληψη.....	v
Abstract	vi
Περιεχόμενα	vii
Κατάλογος Εικόνων / Σχημάτων	x
Κατάλογος Πινάκων	xiv
Συνοτομογραφίες & Ακρωνύμια.....	xv
1. Εισαγωγή.....	1
1.1 Σκοπός και στόχοι	1
1.2 Μεθοδολογία	2
1.3 Διάρθρωση περιεχομένου.....	3
2. Ακουστικός Θόρυβος και Τεχνικές Καταστολής.....	5
2.1 Ορισμός ακουστικού θορύβου	5
2.2 Χαρακτηριστικά θορύβου ως ήχου	5
2.2.1 Χρονικά χαρακτηριστικά	6
2.2.2 Συχνотικά χαρακτηριστικά	8
2.3 Κατηγορίες θορύβου	9
2.4 Παρεμβολή θορύβου σε ηχητικό σήμα	11
2.4.1 Ένταξη θορύβου κατά την δημιουργία - αποθήκευση του ήχου	11
2.4.2 Ένταξη θορύβου κατά την μετάδοση - αναπαραγωγή του ήχου	16
2.5 Περικοπή θορύβου (Trimming and Cutting).....	18
2.6 Αφαίρεση ψαλιδισμού (Declipping).....	21
2.7 Πύλη θορύβου (Noise Gate).....	23
2.8 Φασματική επεξεργασία (Spectral Editing)	26
2.9 Εφαρμογή φίλτρων (Filtering)	29
2.10 Ισοστάθμιση (EQ)	33
2.11 Ακύρωση θορύβου (Noise Cancellation - NC)	35
3. Χρήση Ψηφιακών Εργαλείων Λογισμικού για Αφαίρεση Θορύβου	39
3.1 Σταθμοί επεξεργασίας ψηφιακού ήχου (Digital Audio Workstations - DAW)	39
3.2 Πρόσθετα λογισμικών επεξεργασίας ψηφιακού ήχου (Plugins).....	43
3.3 Εξειδικευμένα εργαλεία αποθορυβοποίησης	46
3.4 Υλοποιήσεις «επί τούτω» (Ad Hoc).....	48
4. Μεθοδολογία και Σχεδιασμός Πειραματικής Διαδικασίας	51
4.1 Καθορισμός πλαισίου υλοποίησης	52
4.2 Επιλογή αρχείων ήχου	53
4.2.1 Καθορισμός κριτηρίων επιλογής	53

4.2.2 Σχεδιασμός δημιουργίας αρχείων	54
4.3 Επιλογή εργαλείων λογισμικού	55
4.3.1 Κατηγορίες διαθέσιμων λογισμικών	55
4.3.2 Κριτήρια επιλογής εργαλείων λογισμικού	57
4.4 Επεξεργασία αρχείων ήχου από τα επιλεγθέντα λογισμικά	58
4.5 Αξιολόγηση αποτελεσματικότητας λογισμικών	59
5. Υλοποίηση Πειραματικής Διαδικασίας	61
5.1 Επιλογή και διαμόρφωση αρχείων ήχου	61
5.1.1 Αρχεία ομιλίας	62
5.1.1.1 Φωνή 1 (V01)	64
5.1.1.2 Φωνή 2 (V02)	66
5.1.2 Αρχεία θορύβου	67
5.1.2.1 Θόρυβος 1 (N01)	68
5.1.2.2 Θόρυβος 2 (N02)	70
5.1.2.3 Θόρυβος 3 (N03)	71
5.1.2.4 Θόρυβος 4 (N04)	73
5.1.2.5 Θόρυβος 5 (N05)	74
5.1.2.6 Θόρυβος 6 (N06)	75
5.1.2.7 Θόρυβος 7 (N07)	77
5.1.2.8 Θόρυβος 8 (N08)	78
5.1.2.9 Θόρυβος 9 (N09)	80
5.1.2.10 Θόρυβος 10 (N10)	81
5.1.2.11 Θόρυβος 11 (N11)	82
5.1.2.12 Θόρυβος 12 (N12)	84
5.1.3 Μίξη ομιλίας με θόρυβο για δημιουργία ενθόρυβων αρχείων	85
5.2 Επιλογή εργαλείων λογισμικού	97
5.2.1 Λογισμικό 1 (S01) - Cleanvoice	99
5.2.2 Λογισμικό 2 (S02) - Enhance (Adobe Podcast)	102
5.2.3 Λογισμικό 3 (S03) - LALAL.AI	105
5.2.4 Λογισμικό 4 (S04) - OpenVINO (Audacity Plugin)	107
5.2.5 Λογισμικό 5 (S05) - VocalPrep (Antares AutoTune)	110
5.3 Αποθορυβοποίηση αρχείων από τα επιλεγθέντα λογισμικά	112
5.1 Επεξεργασία με το Cleanvoice	113
5.2 Επεξεργασία με το Enhance	115
5.3 Επεξεργασία με το LALAL.AI	116
5.4 Επεξεργασία με το OpenVINO	119
5.5 Επεξεργασία με το VocalPrep	120
5.4 Αξιολόγηση αποτελεσματικότητας λογισμικών	121
6. Αποτελέσματα Πειραματικής Διαδικασίας - Πιθανές Εφαρμογές	125
6.1 Αποτίμηση αποτελεσματικότητας επιλεγμένων εργαλείων	125
6.1.1 Ευρήματα κατά την διαδικασία αποθορυβοποίησης	126
6.1.2 Αποτελέσματα από την διενέργεια των ακροάσεων	127
6.1.3 Παρατηρήσεις - Σχόλια ακροατών	136
6.1.4 Αποτύπωση υποκειμενικότητας στις απαντήσεις	137

6.2 Πιθανές εφαρμογές εργαλείων αποθρομβοποίησης	138
7. Συμπεράσματα - Συνεισφορά - Μελλοντικές Επεκτάσεις	141
7.1 Συμπεράσματα.....	141
7.2 Συνεισφορά της Εργασίας	145
7.3 Μελλοντικές Επεκτάσεις.....	146
Βιβλιογραφικές Αναφορές	149
Παράρτημα Α: Συνοδευτικά αρχεία.....	153
Παράρτημα Β: Τεχνικά χαρακτηριστικά εξοπλισμού που χρησιμοποιήθηκε	154
Παράρτημα Γ: Ερωτηματολόγιο ακροάσεων - Οδηγίες	156

Κατάλογος Εικόνων / Σχημάτων

Εικόνα 2.1. Κυματομορφή ηχητικού σήματος σε γραμμική μορφή απεικόνισης πλάτους ..6	
Εικόνα 2.2. Χρήση αντικραδασμικής βάσης και φίλτρου για καταστολή θορύβου σε μικρόφωνο.....	12
Εικόνα 2.3. Επιλογή προκαθορισμένων τιμών ισοστάθμισης RIAA στο Graphic EQ του Audacity	15
Εικόνα 2.3. Έλεγχος ταχύτητας περιστροφής δίσκου βινυλίου με χρήση στροβοσκοπίου	17
Εικόνα 2.4. Στιγμιότυπο οθόνης χρήσης του Audacity για περικοπή θορύβου.....	19
Εικόνα 2.5. Στιγμιότυπο επιλογών περικοπής θορύβου στην πλατφόρμα Auphonic.....	20
Εικόνα 2.6. Στιγμιότυπο χρήσης εργαλείου εύρεσης σημείων ψαλιδισμού στο Audacity.	22
Εικόνα 2.7. Οθόνη του προσθέτου αφαίρεσης ψαλιδισμού ClipRemover AI της BORISFX	23
Εικόνα 2.8. Κυματομορφή σήματος ομιλίας στο οποίο υπάρχει στατικός θόρυβος	24
Εικόνα 2.9. Απεικόνιση σήματος ομιλίας με εφαρμογή πύλης θορύβου.....	24
Εικόνα 2.10. Αναλογικό σύστημα ραδιοεπικοινωνίας με δυνατότητα φίμωσης (Squelch)	25
Εικόνα 2.11. Πεντάλ μουσικού οργάνου για εφαρμογή πύλης θορύβου.....	26
Εικόνα 2.12. Στιγμιότυπο οθόνης επιλογής προφίλ θορύβου του Audacity.....	27
Εικόνα 2.13. Στιγμιότυπο οθόνης εργασίας του λογισμικού SpectraLayers της Steinberg	28
Εικόνα 2.14. Εφαρμογή φίλτρου αποκοπής χαμηλών συχνοτήτων στο μικρόφωνο ZOOM H1n.....	31
Εικόνα 2.15. Οθόνη φίλτρου εγκοπής του προσθέτου Renaissance DeEsser της Waves ..	32
Εικόνα 2.16. Στιγμιότυπο οθόνης κατά την χρήση του Curve EQ στο Audacity.....	34
Εικόνα 2.17. Συνάρτηση μεταφοράς παραμορφωμένου σήματος στο REW	35
Εικόνα 3.1. Στιγμιότυπο οθόνης εργασίας του σταθμού επεξεργασίας ήχου Audacity	40
Εικόνα 3.2. Στιγμιότυπο οθόνης χρήσης του προσθέτου DeNoise 2 της Acon Digital.....	44
Εικόνα 3.3. Στιγμιότυπο οθόνης χρήσης του προσθέτου Clarity Vx της Waves.....	45
Εικόνα 3.4. Στιγμιότυπο οθόνης χρήσης του προσθέτου Clear μέσα από το Audacity.....	46
Εικόνα 3.5. Στιγμιότυπο περιβάλλοντος εργασίας του λογισμικού RX12 Advanced της Izotope.....	47
Εικόνα 3.6. Στιγμιότυπο οθόνης ρύθμισης του εργαλείου Two-Way AI Noise Cancellation της ASUS	49
Εικόνα 3.7. ROG Clavis AI Noise Cancelling Mic Adapter της ASUS	50
Εικόνα 3.8. Στιγμιότυπο οθόνης χρήσης του λογισμικού Adrenaline της AMD.....	50
Εικόνα 5.1. Στιγμιότυπο παραγωγής ομιλίας από κείμενο στην πλατφόρμα Eleven Labs.	63
Εικόνα 5.2. Κυματομορφή ομιλίας γυναίκας - φωνή 1 (V01).....	65
Εικόνα 5.3. Φασματογράφημα ομιλίας γυναίκας - φωνή 1 (V01).....	65
Εικόνα 5.4. Κυματομορφή ομιλίας άνδρα - φωνή 2 (V02).....	66
Εικόνα 5.5. Φασματογράφημα ομιλίας άνδρα - φωνή 2 (V02)	67
Εικόνα 5.6. Κυματομορφή γαβγίσματος σκύλου - θόρυβος 1 (N01)	69
Εικόνα 5.7. Φασματογράφημα γαβγίσματος σκύλου - θόρυβος 1 (N01).....	69
Εικόνα 5.8. Κυματομορφή αλυσοπρίονου - θόρυβος 2 (N02).....	70
Εικόνα 5.9. Φασματογράφημα αλυσοπρίονου - θόρυβος 2 (N02)	71
Εικόνα 5.10. Κυματομορφή πουλιών με γρύλο - θόρυβος 3 (N03).....	72
Εικόνα 5.11. Φασματογράφημα πουλιών με γρύλο - θόρυβος 3 (N03)	72
Εικόνα 5.12. Κυματομορφή ελικοπτέρου - θόρυβος 4 (N04).....	73

Εικόνα 5.13. Φασματογράφημα ελικοπτέρου - θόρυβος 4 (N04)	74
Εικόνα 5.14. Κυματομορφή ρολογιού τοίχου - θόρυβος 5 (N05)	75
Εικόνα 5.15. Φασματογράφημα ρολογιού τοίχου - θόρυβος 5 (N05)	75
Εικόνα 5.16. Κυματομορφή ομιλίας πλήθους - θόρυβος 6 (N06)	76
Εικόνα 5.17. Φασματογράφημα ομιλίας πλήθους - θόρυβος 6 (N06)	76
Εικόνα 5.18. Κυματομορφή βουητό ανεμιστήρα - θόρυβος 7 (N07)	77
Εικόνα 5.19. Φασματογράφημα βουητό ανεμιστήρα - θόρυβος 7 (N07)	78
Εικόνα 5.20. Κυματομορφή μουσικής - θόρυβος 8 (N08)	79
Εικόνα 5.21. Φασματογράφημα μουσικής - θόρυβος 8 (N08)	79
Εικόνα 5.22. Κυματομορφή πληκτρολόγησης - θόρυβος 9 (N09)	80
Εικόνα 5.23. Φασματογράφημα πληκτρολόγησης - θόρυβος 9 (N09)	80
Εικόνα 5.24. Κυματομορφή σημάτων Μορς - θόρυβος 10 (N10)	81
Εικόνα 5.25. Φασματογράφημα σημάτων Μορς - θόρυβος 10 (N10)	82
Εικόνα 5.26. Κυματομορφή σειρήνας και κόρνας - θόρυβος 11 (N11)	83
Εικόνα 5.27. Φασματογράφημα σειρήνας και κόρνας - θόρυβος 11 (N11)	83
Εικόνα 5.28. Κυματομορφή δυνατής βροχής - θόρυβος 12 (N12)	84
Εικόνα 5.29. Φασματογράφημα δυνατής βροχής - θόρυβος 12 (N12)	85
Εικόνα 5.30. Κυματομορφή πουλιών και γρύλλου με γυναικεία ομιλία - ενθόρυβο 03 (N03V01)	89
Εικόνα 5.31. Κυματομορφή ομιλιών πλήθους με γυναικεία ομιλία - ενθόρυβο 06 (N06V01)	89
Εικόνα 5.32. Κυματομορφή βουητού ανεμιστήρα με γυναικεία ομιλία - ενθόρυβο 07 (N07V01)	90
Εικόνα 5.33. Κυματομορφή πληκτρολόγησης με γυναικεία ομιλία - ενθόρυβο 09 (N09V01)	90
Εικόνα 5.34. Κυματομορφή σειρήνας και κόρνας με γυναικεία ομιλία - ενθόρυβο 11 (N11V01)	90
Εικόνα 5.35. Κυματομορφή δυνατής βροχής με γυναικεία ομιλία - ενθόρυβο 12 (N12V01)	91
Εικόνα 5.36. Κυματομορφή γαβγίσματος σκύλου με αντρική ομιλία - ενθόρυβο 01 (N01V02)	91
Εικόνα 5.37. Κυματομορφή αλυσοπρίονου με αντρική ομιλία - ενθόρυβο 02 (N02V02)	91
Εικόνα 5.38. Κυματομορφή ελικοπτέρου με αντρική ομιλία - ενθόρυβο 04 (N04V02)	92
Εικόνα 5.39. Κυματομορφή ρολογιού τοίχου με αντρική ομιλία - ενθόρυβο 05 (N05V02)	92
Εικόνα 5.40. Κυματομορφή μουσικής με αντρική ομιλία - ενθόρυβο 08 (N08V02)	92
Εικόνα 5.41. Κυματομορφή σημάτων Μορς με αντρική ομιλία - ενθόρυβο 10 (N10V02)	93
Εικόνα 5.42. Φασματογράφημα πουλιών και γρύλλου με γυναικεία ομιλία - ενθόρυβο 03 (N03V01)	93
Εικόνα 5.43. Φασματογράφημα ομιλιών πλήθους με γυναικεία ομιλία - ενθόρυβο 06 (N06V01)	93
Εικόνα 5.44. Φασματογράφημα βουητού ανεμιστήρα με γυναικεία ομιλία - ενθόρυβο 07 (N07V01)	94
Εικόνα 5.45. Φασματογράφημα πληκτρολόγησης με γυναικεία ομιλία - ενθόρυβο 09 (N09V01)	94
Εικόνα 5.46. Φασματογράφημα σειρήνας και κόρνας με γυναικεία ομιλία - ενθόρυβο 10 (N11V01)	94

Εικόνα 5.47. Φασματογράφημα δυνατής βροχής με γυναικεία ομιλία - ενθόρυβο 12 (N12V01)	95
Εικόνα 5.48. Φασματογράφημα γαβγίσματος σκύλου με αντρική ομιλία - ενθόρυβο 01 (N01V02)	95
Εικόνα 5.49. Φασματογράφημα αλυσοπρίονου με αντρική ομιλία - ενθόρυβο 02 (N02V02)	95
Εικόνα 5.50. Φασματογράφημα ελικοπτέρου με αντρική ομιλία - ενθόρυβο 04 (N04V02)	96
Εικόνα 5.51. Φασματογράφημα ρολογιού τοίχου με αντρική ομιλία- ενθόρυβο 05 (N05V02)	96
Εικόνα 5.52. Φασματογράφημα μουσικής με αντρική ομιλία - ενθόρυβο 08 (N08V02)...	96
Εικόνα 5.53. Φασματογράφημα σημάτων Μορς με αντρική ομιλία - ενθόρυβο 10 (N10V02)	97
Εικόνα 5.54. Cleanvoice - Επιλογή προτύπου επεξεργασίας (Template).....	100
Εικόνα 5.55. Cleanvoice - Αρχική οθόνη εφαρμογής.....	101
Εικόνα 5.56. Enhance - Περιβάλλον εργασίας της εφαρμογής	103
Εικόνα 5.57. LALAL.AI - Διαδικτυακό περιβάλλον εργασίας της εφαρμογής	106
Εικόνα 5.58. OpenVINO - Στιγμιότυπο οθόνης κατά την χρήση του προσθέτου στο Audacity	108
Εικόνα 5.59. OpenVINO - Οθόνη ιστοτόπου μεταφόρτωσης της εφαρμογής	109
Εικόνα 5.60. VocalPrep - Αρχική οθόνη εφαρμογής.....	111
Εικόνα 5.61. Cleanvoice - Οθόνη δημιουργίας προτύπου (Custom Template).....	113
Εικόνα 5.62. Cleanvoice - Οθόνη από το στάδιο της επεξεργασίας.....	114
Εικόνα 5.63. Cleanvoice - Οθόνη επισκόπησης και εξαγωγής προϊόντων επεξεργασίας	114
Εικόνα 5.64. Enhance - Οθόνη από το στάδιο της επεξεργασίας με το μοντέλο v2.....	115
Εικόνα 5.65. Enhance - Επιλογές μοντέλου v2 διαθέσιμες από τον Μάρτιο του 2026....	116
Εικόνα 5.66. LALAL.AI - Οθόνη επιλογής καναλιών επεξεργασίας.....	117
Εικόνα 5.67. LALAL.AI - Οθόνη επιλογής καναλιών επεξεργασίας.....	118
Εικόνα 5.68. LALAL.AI - Οθόνη επιλογής παραμέτρων επεξεργασίας	118
Εικόνα 5.69. OpenVINO - Οθόνη επιλογής μοντέλου αποθρονοποίησης	119
Εικόνα 5.70. VocalPrep - Οθόνη προόδου κατά την αποθρονοποίηση.....	120
Εικόνα 5.71. VocalPrep - Οθόνη φασματογραφημάτων διαχωρισμένων ήχων	121
Εικόνα 5.72. Χρωματικός κώδικας περιγραφής λογισμικών.....	123
Εικόνα 5.73. Στιγμιότυπο από την διαδικασία των ακροάσεων	124
Σχήμα 2.1. Φάσματα λευκού και ροζ θορύβου μέσω αναλυτή και φίλτρων οκταβικού εύρους.....	10
Σχήμα 2.2. Καμπύλες ισοστάθμισης RIAA κατά την εγγραφή και αναπαραγωγή δίσκων βινυλίου.....	14
Σχήμα 2.3. Εύρος συχνοτήτων τυπικών ηχητικών πηγών	29
Σχήμα 2.4. Αποκρίσεις συχνότητας ιδανικών φίλτρων	30
Σχήμα 2.5. Μείωση θορύβου με τεχνική ενεργητικής ακύρωσης (ANC)	37
Σχήμα 2.6. Συχνότητες κάλυψης υβριδικής υλοποίησης για ακύρωση θορύβου ευρέως φάσματος.....	38
Σχήμα 4.1. Ακολουθία βημάτων σχεδιασμού πειραματικής διαδικασίας.....	51
Σχήμα 5.1. Δημογραφικά στοιχεία της ομάδας ακροατών	122
Σχήμα 6.1. Επίδοση λογισμικών στην αποθρονοποίηση του 1 ^{ου} δείγματος.....	128
Σχήμα 6.2. Επίδοση λογισμικών στην αποθρονοποίηση του 2 ^{ου} δείγματος.....	128

Σχήμα 6.3. Επίδοση λογισμικών στην αποθρομβοποίηση του 3 ^{ου} δείγματος.....	129
Σχήμα 6.4. Επίδοση λογισμικών στην αποθρομβοποίηση του 4 ^{ου} δείγματος.....	129
Σχήμα 6.5. Επίδοση λογισμικών στην αποθρομβοποίηση του 5 ^{ου} δείγματος.....	129
Σχήμα 6.5. Επίδοση λογισμικών στην αποθρομβοποίηση του 6 ^{ου} δείγματος.....	130
Σχήμα 6.7. Επίδοση λογισμικών στην αποθρομβοποίηση του 7 ^{ου} δείγματος.....	130
Σχήμα 6.8. Επίδοση λογισμικών στην αποθρομβοποίηση του 8 ^{ου} δείγματος.....	130
Σχήμα 6.9. Επίδοση λογισμικών στην αποθρομβοποίηση του 9 ^{ου} δείγματος.....	131
Σχήμα 6.10. Επίδοση λογισμικών στην αποθρομβοποίηση του 10 ^{ου} δείγματος.....	131
Σχήμα 6.11. Επίδοση λογισμικών στην αποθρομβοποίηση του 11 ^{ου} δείγματος.....	131
Σχήμα 6.12. Επίδοση λογισμικών στην αποθρομβοποίηση του 12 ^{ου} δείγματος.....	132
Σχήμα 6.13. Επίδοση λογισμικού Cleanvoice στην αποθρομβοποίηση των δειγμάτων...	132
Σχήμα 6.14. Επίδοση λογισμικού Enhance στην αποθρομβοποίηση των δειγμάτων.....	133
Σχήμα 6.15. Επίδοση λογισμικού LALAL.AI στην αποθρομβοποίηση των δειγμάτων ..	133
Σχήμα 6.15. Επίδοση λογισμικού OpenVINO στην αποθρομβοποίηση των δειγμάτων ..	133
Σχήμα 6.17. Επίδοση λογισμικού VocalPrep στην αποθρομβοποίηση των δειγμάτων....	134
Σχήμα 6.18. Συγκεντρωτικό γράφημα επίδοσης των λογισμικών τύπου ραντάρ	134
Σχήμα 6.19. Συγκεντρωτικό γράφημα επίδοσης των λογισμικών τύπου διασποράς.....	135
Σχήμα 6.20. Διασπορά απαντήσεων κατά την συμπλήρωση του ερωτηματολογίου.....	137

Κατάλογος Πινάκων

Πίνακας 3.1. Πίνακας διαδεδομένων λογισμικών επεξεργασίας ψηφιακού ήχου (DAW).....	41
Πίνακας 5.1. Ρυθμίσεις παραμέτρων φωνών για δημιουργία συνθετικής ομιλίας	62
Πίνακας 5.2. Κατανομή αντρικής και γυναικείας ομιλίας ανά εξεταζόμενο θόρυβο	86
Πίνακας 5.3. Μέσες τιμές στάθμης (RMS) και ρύθμιση κέρδους (Gain) σημάτων	87
Πίνακας 5.4. Λόγος σήματος προς θόρυβο (SNR) τελικών ενθόρυβων δειγμάτων	88
Πίνακας 5.5. Πίνακας εργαλείων TN που διερευνήθηκαν κατά την διαδικασία επιλογής λογισμικών	98
Πίνακας 5.6. Τυχαία διάταξη της σειράς ακρόασης των λογισμικών.....	123
Πίνακας B.1. Χαρακτηριστικά σταθερού σταθμού εργασίας	154
Πίνακας B.2. Χαρακτηριστικά φορητού σταθμού εργασίας για διενέργεια ακροάσεων .	154
Πίνακας B.3. Χαρακτηριστικά εξόδου κάρτας ήχου για διενέργεια ακροάσεων	155
Πίνακας B.4. Χαρακτηριστικά ακουστικών A, για διενέργεια ακροάσεων	155
Πίνακας B.5. Χαρακτηριστικά ακουστικών B, για διενέργεια ακροάσεων.....	155

Συντομογραφίες & Ακρωνύμια

ΑΠΘ	Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
ΔΕ	Διπλωματική Εργασία
ΕΑΠ	Ελληνικό Ανοικτό Πανεπιστήμιο
ΕΔΙΠ	Εργαστηριακό Διδακτικό Προσωπικό
ΘΕ	Θεματική Ενότητα
ΝΔ	Νευρωνικό Δίκτυο
ΠΣ	Πρόγραμμα Σπουδών
ΣΥΝ	Συντονιστής/Συντονίστρια
TN	Τεχνητή Νοημοσύνη

AAX	Avid Audio eXtension,
AI	Artificial Intelligence
ANC	Active Noise Cancellation
AU	Audio Unit
CNN	Convolutional Neural Networks
DAW	Digital Audio Workstation
DSP	Digital Signal Processing
dB	Decibel
dBFS	Decibel (relative to) Full Scale
EQ	Equalization
Hz	Hertz
ML	Machine Learning
MOS	Mean Opinion Score
NC	Noise Cancellation
NG	Noise Gate
NN	Neural Network
NR	Noise Reduction
OS	Operating System
PC	Personal Computer
RIAA	Recording Industry Association of America

RMS	Root Mean Square
RNN	Recurrent Neural Network
SaaS	Software as a Service
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level
SQNR	Signal to Quantization Noise Ratio
STI	Speech Transmission Index
UI	User Interface
VST	Virtual Studio Technology
WAV	Waveform Audio File (Format)
ZCR	Zero Crossing Rate

1. Εισαγωγή

Η παρούσα διπλωματική εργασία, φέρει τον τίτλο: «Αξιολόγηση Εργαλείων Λογισμικού για την Αποθορυβοποίηση Ψηφιακού Ήχου με Έμφαση στις Σύγχρονες Τεχνικές Τεχνητής Νοημοσύνης». Θα εκπονηθεί κατά το ακαδημαϊκό έτος 2025-2026, στα πλαίσια του μεταπτυχιακού προγράμματος σπουδών: «Ακουστικός Σχεδιασμός και Ψηφιακός Ήχος», που παρέχεται από το Ελληνικό Ανοικτό Πανεπιστήμιο (ΕΑΠ).

Στις υποενότητες που ακολουθούν, καταγράφεται ο σκοπός και οι στόχοι της εργασίας, παρουσιάζεται η μεθοδολογία που θα ακολουθηθεί κατά την υλοποίησή της και παρατίθεται συνοπτικά η διάρθρωση του περιεχομένου της.

1.1 Σκοπός και στόχοι

Σκοπό της παρούσας διπλωματικής εργασίας, αποτελεί η διερεύνηση των τεχνικών αφαίρεσης θορύβου από ψηφιακά ηχητικά σήματα, οι οποίες πραγματοποιούνται με την χρήση σχετικών μεθόδων, δίνοντας ιδιαίτερη έμφαση σε εκείνες που βασίζονται στην Τεχνητή Νοημοσύνη, καθώς και η μελέτη της απόδοσης και της πρακτικής χρησιμότητας κατάλληλα επιλεγμένων εφαρμογών Τεχνητής Νοημοσύνης, που στοχεύουν στη βελτίωση της ακουστικής ποιότητας, ενθόρυβων σημάτων ήχου.

Στους στόχους της εργασίας περιλαμβάνεται η καταγραφή και η μελέτη των βασικών τεχνικών αποθορυβοποίησης ψηφιακού ήχου, τόσο συμβατικών όσο και αυτών που βασίζονται στην Τεχνητή Νοημοσύνη.

Στόχο της εργασίας, αποτελεί επίσης η αποτίμηση της αποτελεσματικότητας επιλεγμένων εργαλείων, μέσα από την διενέργεια ακουστικής αξιολόγησης, που θα πραγματοποιηθεί με συμμετοχή ομάδας ατόμων, που έχουν σχετική εμπειρία στον ήχο.

Στα προσδοκώμενα αποτελέσματα, περιλαμβάνονται η συγκριτική αποτίμηση της αποτελεσματικότητας των επιλεγμένων εργαλείων, καθώς και η διερεύνηση των πιθανών εφαρμογών τους στην καθημερινή πρακτική.

1.2 Μεθοδολογία

Κατά την εκπόνηση της διπλωματικής εργασίας θα χρησιμοποιηθούν δύο κύριοι μεθοδολογικοί άξονες. Αυτοί περιλαμβάνουν αφενός το σκέλος της θεωρητικής προσέγγισης του θέματος και αφ' ετέρου το σκέλος της αναλυτικής - πειραματικής προσέγγισης και διαδικασίας υλοποίησής της.

Το θεωρητικό μέρος αφορά σε δεδομένα που θα συγκεντρωθούν έπειτα από εκτενή βιβλιογραφική έρευνα και επισκόπηση, με περιεχόμενο τα έως τώρα επιστημονικά στοιχεία και ευρήματα, με παράλληλη ενδελεχή αναζήτηση στο διαδίκτυο των διαθέσιμων τεχνικών και εργαλείων, που χρησιμοποιούνται για την αποθορυβοποίηση του ψηφιακού ήχου. Η σύνθεση των βιβλιογραφικών ευρημάτων, συντελείται κυρίως στα αρχικά θεωρητικά κεφάλαια της παρούσας διπλωματικής εργασίας, ενώ αξιοποιείται με την χρήση αναφορών, σε όλη την έκταση της εργασίας.

Η εργασία δίνει ιδιαίτερη έμφαση στις σύγχρονες τεχνικές Τεχνητής Νοημοσύνης (TN). Με βάση το γεγονός αυτό, θα επιλεγθεί μεθοδολογικά, η καταγραφή και παράθεση των συμβατικών μεθόδων αποθορυβοποίησης να καλυφθεί σε θεωρητικό - βιβλιογραφικό επίπεδο, ενώ η αναλυτική προσέγγιση της έρευνας, θα επικεντρωθεί στην πειραματική διερεύνηση της αποτελεσματικότητας εργαλείων και μεθόδων Τεχνητής Νοημοσύνης.

Η αναλυτική διαδικασία που θα ακολουθηθεί, περιγράφεται διεξοδικά ως προς την μεθοδολογία της, σε επόμενο κεφάλαιο της παρούσας εργασίας. Σε αυτή συμπεριλαμβάνεται η επιλογή και δημιουργία πρωτογενών αντιπροσωπευτικών ηχητικών δεδομένων και η χρήση κατάλληλων λογισμικών τεχνητής νοημοσύνης, επιλεγμένων μετά από εκτεταμένη έρευνα.

Η πειραματική μεθοδολογία που θα ακολουθηθεί, εμπεριέχει επίσης την διαδικασία εφαρμοσμένης αξιολόγησης των επιλεγμένων εργαλείων λογισμικού, μέσω διενέργειας ακροάσεων και σχολιασμών από άτομα με σχετική εμπειρία στον ήχο.

Για την ποσοτική αποτύπωση των αποτελεσμάτων των ακροάσεων, θα γίνει χρήση ερωτηματολογίου που θα δημιουργηθεί για την συλλογή των πειραματικών δεδομένων που προέκυψαν. Για την επεξεργασία και την γραφική απεικόνιση των αποτελεσμάτων, θα

χρησιμοποιηθούν δομημένα εργαλεία λογιστικών φύλων, ενώ η ποιοτική αποτίμηση θα γίνει μέσα από την καταγραφή σχολίων και παρατηρήσεων.

Από τα ευρήματα της ακολουθούμενης μεθοδολογίας, θα προκύψουν και τα σχετικά συμπεράσματα, τα οποία θα αποτυπωθούν στο τελευταίο κεφάλαιο της εργασίας.

1.3 Διάρθρωση περιεχομένου

Για την επίτευξη της καλύτερης οργάνωσής της, η παρούσα διπλωματική εργασία θα δομηθεί με τη μορφή κεφαλαίων, η ακολουθία των οποίων θα διαμορφωθεί με τρόπο ώστε να συμβάλει στην επίτευξη της εννοιολογικής ροής του περιεχομένου της.

Το παρόν κεφάλαιο -που αποτελεί και το πρώτο κεφάλαιο της διπλωματικής εργασίας- περιλαμβάνει την καταγραφή του σκοπού και των στόχων της εργασίας, την παρουσίαση της ακολουθούμενης μεθοδολογίας υλοποίησής της και την παράθεση της διάρθρωσης του περιεχομένου της.

Το δεύτερο κεφάλαιο περιλαμβάνει την θεωρητική προσέγγιση του θορύβου ως έννοια και περιεχόμενο, με βάση την βιβλιογραφία και τις διαθέσιμες πηγές. Ορίζεται ο ακουστικός θόρυβος και καταγράφονται τα χαρακτηριστικά και οι κατηγορίες στις οποίες αυτός διαχωρίζεται. Παράλληλα περιγράφονται οι τρόποι με τους οποίους αυτός παρεμβάλλεται σε ένα ηχητικό σήμα και γίνεται η παρουσίαση των κύριων συμβατικών μεθόδων και τεχνικών που ακολουθούνται για την αποθορυβοποίηση του ήχου.

Το τρίτο κεφάλαιο εστιάζει στην αξιοποίηση των διαθέσιμων εξειδικευμένων εργαλείων λογισμικού, τα οποία χρησιμοποιούνται για την αφαίρεση θορύβου, διαχωρίζοντάς τα σε βασικές κατηγορίες ανάλογα με το είδος και την χρήση τους.

Το τέταρτο κεφάλαιο περιλαμβάνει την καταγραφή του σχεδιασμού του αναλυτικού σκέλους της διπλωματικής εργασίας, με περιεχόμενό του τον καθορισμό του πλαισίου της υλοποίησης, της επιλογής αρχείων και λογισμικών, καθώς και τον σχεδιασμό της επεξεργασίας και του τρόπου αξιολόγησης της αποτελεσματικότητας της διαδικασίας.

Η υλοποίηση της πειραματικής διαδικασίας αποτελεί το περιεχόμενο του πέμπτου κεφαλαίου. Εδώ περιλαμβάνεται η καταγραφή της επιλογής και διαμόρφωσης των αρχείων

ήχου που θα χρησιμοποιηθούν, η αντίστοιχη επιλογή των λογισμικών και η διαδικασία αποθορυβοποίησης που θα ακολουθηθεί σε κάθε ένα από αυτά.

Στο έκτο κεφάλαιο περιέχονται τα αποτελέσματα της πειραματικής διαδικασίας που θα προκύψουν, με την αποτίμηση της αποτελεσματικότητας των επιλεγμένων εργαλείων, όσον αφορά στην επιδιωκόμενη αποθορυβοποίηση. Παράλληλα καταγράφονται οι πιθανές εφαρμογές και χρήσεις των εργαλείων αποθορυβοποίησης.

Το τελευταίο έβδομο κεφάλαιο, περιλαμβάνει την καταγραφή των συμπερασμάτων, τα οποία που θα εξαχθούν κατά την υλοποίηση της παρούσας διπλωματικής εργασίας. Το κεφάλαιο αυτό ακολουθείται από την παράθεση της Βιβλιογραφίας και των Παραρτημάτων της εργασίας.

Σημείωση: Στην συνέχεια της παρούσας διπλωματικής εργασίας, παρατίθενται για λόγους πληρότητας του περιεχομένου, αναφορές, σύνδεσμοι ή ενδεικτικά στιγμιότυπα από οθόνες λογισμικών ή προϊόντων διάφορων εταιρειών. Επισημαίνεται ότι δεν υπονοείται προτίμηση όσων παρατίθενται, ή ένδειξη υπεροχής τους έναντι άλλων εναλλακτικών επιλογών, οι οποίες δεν αναφέρονται ή δεν συμπεριλαμβάνονται στις όποιες αναφορές και απεικονίσεις.

2. Ακουστικός Θόρυβος και Τεχνικές Καταστολής

Το περιεχόμενο του παρόντος κεφαλαίου, περιλαμβάνει την θεωρητική προσέγγιση των κύριων στοιχείων του ακουστικού θορύβου. Δίνεται ο ορισμός του θορύβου και καταγράφονται τα σημαντικότερα χαρακτηριστικά του, καθώς και οι κατηγορίες στις οποίες αυτός εντάσσεται. Επιπρόσθετα, γίνεται αναφορά στους τρόπους με τους οποίους αυτός παρεισφρεί σε ένα ηχητικό σήμα, στα διάφορα στάδια της δημιουργίας, της αποθήκευσης, της μετάδοσης και της αναπαραγωγής του, και το κεφάλαιο ολοκληρώνεται με την παράθεση των βασικών μεθόδων και τεχνικών, που χρησιμοποιούνται για την εξάλειψη του θορύβου, σε σήματα με ηχητικό περιεχόμενο.

2.1 Ορισμός ακουστικού θορύβου

Σύμφωνα με τον Ελληνικό Οργανισμό Τυποποίησης (ΕΛΟΤ), θόρυβος ονομάζεται «...κάθε απεριοδικός σύνθετος ήχος, που η στιγμιαία τιμή του αυξομειώνεται γενικά με τυχαίο τρόπο (Σκαρλάτος, 2021)». Ο θόρυβος ορίζεται και ως κάθε ήχος που είναι ανεπιθύμητος από τον αποδέκτη, καθώς θεωρείται δυσάρεστος, ενοχλητικός, επιβλαβής ή παρεμβάλλεται σε άλλες ακουστικές δραστηριότητες (ΕΛΟΤ, 1996).

Καθώς αποτελεί μορφή ήχου, ο θόρυβος σύμφωνα με τον ορισμό του ήχου (ΕΛΟΤ 263.1 / 1.184), αποτελεί μηχανική διαταραχή που διαδίδεται με ορισμένη ταχύτητα μέσα σε ένα μέσο (στερεό, υγρό ή αέριο) το οποίο μπορεί να αναπτύξει εσωτερικές δυνάμεις (π.χ. ελαστικότητας, εσωτερικής τριβής) και έχει τέτοιο χαρακτήρα ώστε να μπορεί να διεγείρει το αισθητήριο της ακοής. Στην συνέχεια της παρούσας διπλωματικής εργασίας, οποιαδήποτε αναφορά σε θόρυβο, αφορά αποκλειστικά σε ακουστικό θόρυβο, εκτός αν αναφέρεται διαφορετικά.

2.2 Χαρακτηριστικά θορύβου ως ήχου

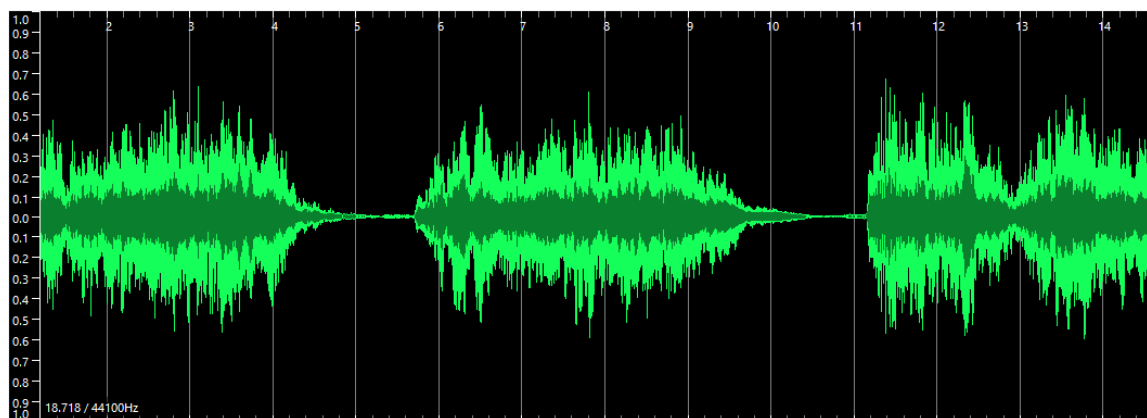
Ο θόρυβος ως μορφή ήχου, διατηρεί όλες τις ιδιότητες που χαρακτηρίζουν κάθε ηχητικό σήμα. Με βάση τα χαρακτηριστικά αυτά, μπορεί να περιγραφεί και να καταγραφεί

με τον ίδιο τρόπο, όπως κάθε άλλος ωφέλιμος ήχος, όταν εμπεριέχεται σε ένα σήμα που θέλουμε να επεξεργασθούμε. Παράλληλα όμως, ο θόρυβος παρουσιάζει διαφοροποιήσεις σε σχέση με τον ωφέλιμο ήχο, ως προς τα χαρακτηριστικά αυτά, οι οποίες αξιοποιούνται στις διάφορες διαδικασίες και τεχνικές αποθορυβοποίησης.

2.2.1 Χρονικά χαρακτηριστικά

Τα χαρακτηριστικά αυτής της μορφής, περιγράφονται από τις μεταβολές στην στάθμη του σήματος, προϊόντος του χρόνου. Έτσι, οι γραφικές απεικονίσεις των σημάτων περιλαμβάνουν την χρονική εξέλιξη στον οριζόντιο άξονα, με παράλληλη κάθετη απεικόνιση του ενεργειακού πλάτους των σημάτων.

Επειδή κατά την διάρκεια της επεξεργασίας τους από τα υπολογιστικά συστήματα, τα εξεταζόμενα ηχητικά σήματα είναι πλέον ψηφιοποιημένα, η απεικόνιση τους κατά πλάτος αναφέρεται σε συνήθως σε μονάδες **dBFS (Decibel [relative to] Full Scale)**, όπου το μέγιστο πλάτος, αντιστοιχεί στην τιμή 0, όταν έχουμε λογαριθμική απεικόνιση, ή στην τιμή 1, όταν η απεικόνιση είναι γραμμική, με τιμές από μηδέν έως ένα, όπως αποτυπώνεται στο παράδειγμα της Εικόνας 2.1 που ακολουθεί.



Εικόνα 2.1. Κυματομορφή ηχητικού σήματος σε γραμμική μορφή απεικόνισης πλάτους

Εκτός από τις απεικονίσεις σε **dB** ή **dBFS**, άλλες μονάδες που χρησιμοποιούνται ανά περίπτωση, είναι ενδεικτικά οι **dBu**, **dBTP**, **LU** και **LUFs** (Mason, 2011).

Όταν ένα ηχητικό σήμα είναι σταθερό και αμετάβλητο, τότε εμφανίζει σταθερό πλάτος σε όλη την χρονική του εξέλιξη. Στην πράξη βέβαια, δύσκολα θα εμφανιστεί μία τέτοια συμπεριφορά, σε ένα σήμα. Ακόμη και στα σήματα θορύβου, η εμφάνιση σταθερότητας συναντάται σε ορισμένες μόνο περιπτώσεις, όπου οι πηγές θορύβου παραμένουν αμετάβλητες.

Για τον λόγο αυτό, χρησιμοποιείται συνήθως η μέτρηση της μέσης τετραγωνικής τιμής (**Root Mean Square - RMS**), σαν μέθοδος για τον υπολογισμό της μέσης ηχητικής στάθμης των σημάτων. Ο υπολογισμός της RMS τιμής των σημάτων πραγματοποιείται εύκολα, με την χρήση εργαλείων που διαθέτουν για τον σκοπό αυτό τα λογισμικά επεξεργασίας ήχου (**Digital Audio Workstations - DAW**).

Ο λόγος ανάμεσα στην στάθμη του ωφέλιμου σήματος και την στάθμη θορύβου που παρεισφρεί σε αυτό (**Signal to Noise Ratio - SNR**), μετράται σε **dB**, και καθορίζει σε μεγάλο βαθμό την ποιότητα του αποτελέσματος που φτάνει στα αυτιά του ακροατή. Είναι προφανές ότι ο λόγος αυτός, πρέπει να είναι όσο το δυνατόν μεγαλύτερος, για την επίτευξη βέλτιστων αποτελεσμάτων.

Με δεδομένο ότι οι στάθμες αποτελούν λογαριθμικά μεγέθη, στην περίπτωση όπου είναι γνωστές οι RMS τιμές των σταθμών σήματος και θορύβου, ο υπολογισμός του SNR μπορεί να πραγματοποιηθεί με αριθμητική αφαίρεση των τιμών αυτών, καθώς σύμφωνα με τις ιδιότητες των λογαρίθμων, ο λογάριθμος του λόγου δύο μεγεθών ισούται με την διαφορά των λογαρίθμων τους (Σκαρλάτος, 2013α).

Ένα χρονικά εξελισσόμενο χαρακτηριστικό, το οποίο μπορεί να αξιοποιηθεί για τον διαχωρισμό του θορύβου από το χρήσιμο ηχητικό σήμα, στην περίπτωση που αυτό είναι σήμα ομιλίας, εντοπίζεται στο γεγονός ότι στα επιμέρους στοιχεία και το τονικό ύψος της ομιλίας παρατηρούνται χρονικές διακυμάνσεις, οι οποίες αναφέρονται σαν πληροφορίες περιοδικότητας, σε αντίθεση με τον θόρυβο, ο οποίος εμφανίζεται με στοχαστικά μη αρμονικά χαρακτηριστικά. Οι πληροφορίες αυτές συμβάλουν στην αναγνώριση της ομιλίας και τον διαχωρισμό της από τους υπόλοιπους ανεπιθύμητους ήχους (Θωϊδης, 2024).

Ένα ακόμη χαρακτηριστικό, που αποτυπώνεται χρονικά στην διάρκεια εξέλιξης του σήματος, είναι ο ρυθμός μηδενικών διελεύσεων (**Zero Crossing Rate - ZCR**). Ο ρυθμός αυτός αποτυπώνει το πώς το σήμα μεταβαίνει από θετικές σε αρνητικές τιμές και αντίστροφα στην πάροδο του χρόνου. Ο δείκτης ZCR χρησιμοποιείται ιδιαίτερα στις

τεχνικές διαχωρισμού του σήματος ομιλίας από θόρυβο, καθώς έχει παρατηρηθεί ότι, λόγω της ανοργάνωτης μορφής του, ο θόρυβος παρουσιάζει υψηλότερο ZCR απ' ό,τι τα σήματα ομιλίας, που συνυπάρχουν στο ίδιο ενθόρυβο σήμα (Bachu et al., 2010).

2.2.2 Συχνотικά χαρακτηριστικά

Τα συχνотικά χαρακτηριστικά του ήχου, αποτυπώνονται στο φασματικό του περιεχόμενο, όπως αυτό κατανέμεται στις διάφορες συχνотικές συνιστώσες, οι οποίες συνθέτουν το ηχητικό σήμα. Με βάση τα χαρακτηριστικά αυτά, επιχειρείται η αξιοποίηση ιδιαιτεροτήτων, οι οποίες διαφοροποιούν τον θόρυβο από τα υπόλοιπα «ωφέλιμα» ηχητικά σήματα -και κυρίως τα σήματα ομιλίας- ώστε οι διαφορές αυτές να συμβάλουν στην αποθορυβοποίηση των σημάτων, με την χρήση κατάλληλα εφαρμοσμένων τεχνικών.

Μία διαφοροποίηση του θορύβου από τα «καθαρά» ηχητικά σήματα, εντοπίζεται στην έννοια της φασματικής επιπεδότητας (**Spectral Flatness**), η οποία χρησιμοποιείται και στην ψηφιακή επεξεργασία ακουστικού σήματος. Αυτή περιγράφει τον βαθμό συγκέντρωσης και κατανομής της ενέργειας του σήματος στο πεδίο της συχνότητας, και βοηθά στον διαχωρισμό ενός ενθόρυβου σήματος από ένα σήμα με λιγότερο ή καθόλου θόρυβο (Boashash et al., 2015). Εμφάνιση υψηλής τιμής επιπεδότητας, υποδηλώνει ότι η ενέργεια κατανέμεται ομοιόμορφα, κάτι που συνήθως συμβαίνει στην περίπτωση του θορύβου, ενώ αντίθετα μία χαμηλή τιμή φανερώνει ενεργειακή συγκέντρωση του σήματος σε συγκεκριμένη συχνотική περιοχή, κάτι που αποτελεί γνώρισμα πιο οργανωμένων μορφών ήχου, όπως είναι η ανθρώπινη ομιλία (Petters, 2004).

Οι παραδοσιακές μέθοδοι και τεχνικές, αξιοποιούν τις φασματικές διαφοροποιήσεις μεταξύ «καθαρού» ήχου και θορύβου, ώστε να πετύχουν την επιθυμητή αποθορυβοποίηση ενός ηχητικού σήματος. Περισσότερα για τις τεχνικές αυτές, περιλαμβάνονται στο επόμενο κεφάλαιο της παρούσας διπλωματικής εργασίας. Εκτός όμως από τις παραδοσιακές προσεγγίσεις, σημαντικό ρόλο πλέον διαδραματίζει και η χρήση της Τεχνητής Νοημοσύνης (TN), στο πεδίο της αφαίρεσης του θορύβου από ένα ηχητικό σήμα.

Σε αντίθεση με τις παραδοσιακές τεχνικές, οι οποίες χρησιμοποιούν κατά κύριο λόγο αφαιρετικούς αλγορίθμους για την απομόνωση και ενίσχυση του «κυρίαρχου» ήχου,

οι μέθοδοι που βασίζονται στην TN, χρησιμοποιούν αλγορίθμους Μηχανικής Μάθησης (**Machine Learning - ML**), οι οποίοι με την βοήθεια «εκπαίδευσης» μαθαίνουν από προηγούμενα ηχητικά δεδομένα, ενώ η αξιοποίηση ακόμα μεγαλύτερης ποσότητας δεδομένων συντελείται με αλγορίθμους Βαθιάς Μάθησης (**Deep Learning - DL**), κατ' επέκταση των προηγούμενων αλγορίθμων (Brittain, 2024).

Οι αλγόριθμοι Βαθιάς Μάθησης έχουν επιδείξει ότι διαθέτουν μεγάλες προοπτικές σε διεργασίες αποθορυβοποίησης ήχου, αξιοποιώντας Νευρωνικά Δίκτυα (**Neural Networks - NN**), για την εκμάθηση μοντέλων που μπορούν να διαχωρίσουν ενθόρυβα σήματα από το αντίστοιχο «καθαρό» τους περιεχόμενο, με αποτέλεσμα είναι ικανά να εκτιμήσουν πότε ένα σήμα ομιλίας είναι απαλλαγμένο από θόρυβο (Dhuriya, 2023).

Η αξιολόγηση εργαλείων λογισμικού τα οποία βασίζονται στην TN, για την επίτευξη αποθορυβοποίησης ηχητικών σημάτων, αποτελεί αντικείμενο της πειραματικής διαδικασίας που εφαρμόστηκε στα πλαίσια της παρούσας διπλωματικής εργασίας, το περιεχόμενο της οποίας περιλαμβάνεται σε επόμενο τμήμα αυτής της εργασίας.

2.3 Κατηγορίες θορύβου

Το πρότυπο ΕΛΟΤ 263.1 (1.210) αναφέρεται στην ακουστική και, ειδικότερα, στον ορισμό και τη μέτρηση του θορύβου βάθους (background noise). Σύμφωνα με το πρότυπο αυτό, ως θόρυβος βάθους ορίζεται ο θόρυβος που προέρχεται από όλες τις πηγές, εξαιρουμένου του συγκεκριμένου ήχου που μας ενδιαφέρει να διατηρήσουμε. Μέρος του θορύβου βάθους αποτελεί και ο περιβαλλοντικός θόρυβος.

Ως προς το είδος της πηγής προέλευσης του θορύβου, οι θόρυβοι μπορούν να διαχωρισθούν σε αυτούς που προέρχονται από σημειακές, γραμμικές και επίπεδες πηγές. Οι επίπεδες πηγές με την σειρά τους μπορεί να διακριθούν σε οριζόντιες, κατακόρυφες ή πλάγιες (Κοντομίχος et al, 2013β). Ο διαχωρισμός αυτός δεν αφορά στο περιεχόμενο του θορύβου, αλλά σχετίζεται με τον τρόπο υπολογισμού της στάθμης του, σε δεδομένη απόσταση από την πηγή του, ο οποίος είναι διαφορετικός σε κάθε περίπτωση.

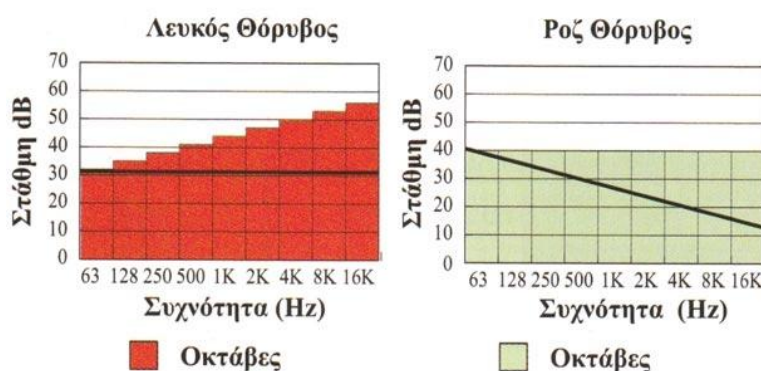
Η υιοθέτηση της ψηφιακής τεχνολογίας, οδήγησε στην εμφάνιση νέων πηγών θορύβου, όπως είναι ενδεικτικά ο θόρυβος που οφείλεται στο σφάλμα κβάντισης ή ο

θόρυβος που οφείλεται στην κωδικοποίηση και συμπίεση των ηχητικών δεδομένων (Ζαρούχας et al, 2013).

Ως προς το περιεχόμενό τους, οι θόρυβοι μπορούν ενδεικτικά να κατηγοριοποιηθούν σε θορύβους που προέρχονται από το περιβάλλον, από διάφορα αντικείμενα που βρίσκονται στον χώρο, από τον ίδιο τον άνθρωπο, από ηλεκτρικές συσκευές, και σε διάφορους άλλους θορύβους.

Έτσι έχουμε θορύβους από επίγειες, θαλάσσιες και εναέριες συγκοινωνίες και μέσα μεταφοράς, από κατασκευαστικές ή άλλες εξωτερικές εργασίες, βιομηχανικούς θορύβους, θορύβους από καιρικά φαινόμενα, όπως από αέρα, βροχή ή αστραπές, από ήχους πουλιών και ζώων, από φωνές πλήθους, κλάματα μωρών, θόρυβο από βήχα ή φτάρνισμα, από οικιακές συσκευές σε λειτουργία, όπως ανεμιστήρες, κλιματιστικά ή πλυντήρια, από αντικείμενα που μετακινούνται, πέφτουν ή σπάνε, από μηχανικούς θορύβους κατά την εργασία σε εσωτερικούς χώρους με υπολογιστές, όπως είναι οι ήχοι πληκτρολόγησης ή το κλικ του ποντικιού, ή ακόμα και ο ήχος του τηλεφώνου που χτυπάει.

Ειδική κατηγορία θορύβων, αποτελούν συγκεκριμένοι ειδικά διαμορφωμένοι θόρυβοι, με ιδιαίτερα φασματικά χαρακτηριστικά, οι οποίοι χρησιμοποιούνται σε ειδικές εφαρμογές, που σχετίζονται κυρίως με διενέργειες μετρήσεων. Κύριοι εκπρόσωποι των θορύβων αυτών είναι ο λευκός και ο ροζ θόρυβος. Στο Σχήμα 2.1 που ακολουθεί, αποτυπώνονται γραφικά τα φάσματα των δύο αυτών ειδικών θορύβων.



Σχήμα 2.1. Φάσματα λευκού και ροζ θορύβου μέσω αναλυτή και φίλτρων οκταβικού εύρους
(Πηγή: Σκαρλάτος, 2021, σ.82, Σχήμα 3.1.1)

Σύμφωνα με τους ορισμούς τους, λευκός ονομάζεται ο θόρυβος ο οποίος έχει συνεχές ηχητικό φάσμα με σταθερή φασματική πυκνότητα, στην περιοχή ενδιαφέροντος,

ενώ αντίστοιχα ο ροζ θόρυβος, ο οποίος έχει και αυτός συνεχές ηχητικό φάσμα, εμφανίζει σταθερή ηχητική ισχύ σε κάθε μία από τις φασματικές ζώνες με σταθερό σχετικό εύρος ζώνης (Σκαρλάτος, 2021).

Στις περιπτώσεις της ανθρώπινης ομιλίας, ο θόρυβος μπορεί να εισαχθεί και κατά την εκφορά του λόγου. Σαν τέτοια μορφή θορύβου, νοείται η αθέλητη προσθήκη φωνημάτων ή η επανάληψη λέξεων, οι οποίες οφείλονται σε δισταγμό ή αμηχανία του ομιλητή. Στην περίπτωση αυτή, μιλάμε για «λέξεις-γεμίσματα» (**filler words**), ή αλλιώς «λέξεις-παράσιτα» (**parasite words**), όπως είναι τα ενδεικτικά επιφωνήματα «...εμμ», «...εεε» και άλλα. Οι παραδοσιακές μέθοδοι αποθορυβοποίησης σε αυτή την περίπτωση, χρησιμοποιούν κυρίως απλές τεχνικές αποκοπής του θορύβου (*trimming and cutting*), ή αντικατάστασής του με ισόχρονα διάστημα σιωπής, όπως καταγράφεται και στο επόμενο κεφάλαιο της παρούσας διπλωματικής εργασίας.

Στην συνέχεια εξετάζονται ορισμένες συνηθισμένες περιπτώσεις εμφάνισης θορύβων, οι οποίοι μπορούν να παρεμβληθούν επιμολύνοντας ένα ηχητικό σήμα, σε διάφορες εκφάνσεις της ύπαρξής του.

2.4 Παρεμβολή θορύβου σε ηχητικό σήμα

Ο θόρυβος μπορεί να παρεισφρήσει σε οποιοδήποτε στάδιο της ύπαρξης ενός ηχητικού σήματος. Έτσι μπορεί να προστεθεί σε αυτό κατά την δημιουργία, την αποθήκευση, την μετάδοση ή την αναπαραγωγή του ήχου. Οι δυνατοί τρόποι, με τους οποίους ο θόρυβος μπορεί παρεμβληθεί, επιμολύνοντας τα ηχητικά σήματα σε κάθε περίπτωση, εξετάζονται σε αυτό το σημείο.

2.4.1 Ένταξη θορύβου κατά την δημιουργία - αποθήκευση του ήχου

Κατά τον σχεδιασμό ενός κλειστού χώρου, που προορίζεται για χρήση ακουστικών εφαρμογών, όπως είναι για παράδειγμα ένας χώρος ηχογραφήσεων ή σύνθεσης και παραγωγής ήχου, επιβάλλεται να διασφαλίζονται οι προϋποθέσεις, που να οδηγούν σε όσο το δυνατόν υψηλότερη τιμή λόγου σήματος προς θόρυβο (SNR), (Κοντομίχος et al, 2013α).

Η δημιουργία ενός όσο κατά το δυνατόν αθόρυβου περιβάλλοντος, αποτελεί βασική επιδίωξη και καθοριστικό παράγοντα παραγωγής βέλτιστων αποτελεσμάτων, για κάθε χώρο ηχογραφήσεων. Έτσι πρέπει να διασφαλισθεί ότι ο χώρος θα είναι απομονωμένος από το εξωτερικό περιβάλλον, με την χρήση κατάλληλων ηχομονωτικών υλικών, τα οποία παράλληλα πρέπει να περιορίζουν φαινόμενα όπως αυτά των ανακλάσεων, που συμβάλλουν στην δημιουργία συντονισμών, αντηχήσεων και άλλων αλλοιώσεων του ήχου.

Βασικές πηγές θορύβου κατά την διενέργεια των ηχογραφήσεων, ή των μεταδόσεων σε πραγματικό χρόνο, αποτελούν και εκείνες που σχετίζονται με την χρήση του μικροφώνου. Στον τομέα αυτό, απαραίτητη είναι η χρήση αντικραδασμικών βάσεων στήριξης, και μικροφώνων υψηλής ποιότητας με κατευθυντικά χαρακτηριστικά, τα οποία να εστιάζουν και απομονώνουν την ωφέλιμη ηχητική πηγή, από άλλες ανεπιθύμητες ηχητικές προελεύσεις.

Ταυτόχρονα, η χρήση προστατευτικών φίλτρων που παρεμβάλλονται ανάμεσα στο στόμα των ομιλητών και τα μικρόφωνα, όπως είναι τα αντιανέμια καλύμματα (**Windshields**) και τα φίλτρα «ποπ» (**Pop Filters**), εμποδίζουν τους θορύβους που προκαλούνται από τον αέρα ή άλλους θορύβους που προέρχονται από το στόμα των ομιλητών να παρεισφρήσουν στο σήμα. Στο στιγμιότυπο της Εικόνας 2.2 που ακολουθεί, εμφανίζεται η χρήση ενός τέτοιου φίλτρου, σε μικρόφωνο που στηρίζεται πάνω σε ειδική αντικραδασμική βάση.



Εικόνα 2.2. Χρήση αντικραδασμικής βάσης και φίλτρου για καταστολή θορύβου σε μικρόφωνο
(Πηγή: <https://www.diyvideostudio.com/what-is-a-pop-filter-for-a-microphone/>)

Άλλες πηγές θορύβου μπορεί να προέρχονται από τα ηλεκτρικά και ηλεκτρονικά συστήματα, που εμπλέκονται στην παραγωγή του ήχου, και ιδιαίτερα από τις αναλογικές συσκευές που χρησιμοποιούνται. Ενδεικτικές τέτοιες συσκευές αποτελούν οι προενισχυτές (**Preamplifiers**), οι ενισχυτές (**Amplifiers**), οι μίκτες (**Mixers**), τα καταγραφικά (**Recorders**), τα διάφορα μηχανικά εφέ (**Effects**) και άλλες σχετικές διατάξεις.

Ηλεκτρική προέλευση έχει και ο θόρυβος που προκαλείται από κακές γειώσεις, προβληματικές καλωδιώσεις οι οποίες δεν διαθέτουν την απαραίτητη θωράκιση, ελλειπτικές τροφοδοσίες και χαλαρές συνδέσεις του εξοπλισμού, ή ακόμα και ο θόρυβος που μπορεί να προκαλείται από βόμβο (**Hum**), ο οποίος μπορεί να προέρχεται από την συχνότητα λειτουργίας του ηλεκτρικού δικτύου. (Η συχνότητα του εναλλασσόμενου ρεύματος του ηλεκτρικού δικτύου για την χώρα μας είναι αυτή των **50 Hz**).

Κάθε μορφή παραμόρφωσης, η οποία μπορεί να προκληθεί σε ένα ηχητικό σήμα, μπορεί να νοηθεί ως θόρυβος. Μία τέτοια περίπτωση αποτελεί φαινόμενο του ψαλιδισμού (**clipping**). Αυτός εμφανίζεται στην περίπτωση που κατά την διαδικασία παραγωγής συντελείται μίξη διαφορετικών ηχητικών καναλιών, όπου ενώ οι επιμέρους στάθμες τους δεν ξεπερνούν το αποδεκτό όριο στάθμης, το άθροισμά τους το υπερβαίνει, με αποτέλεσμα την πρόκληση παραμόρφωσης και αλλοίωσης της ηχητικής πληροφορίας. Για την αντιμετώπιση του ψαλιδισμού, είναι συνήθης πρακτική από τους επαγγελματίες του ήχου, να μην κανονικοποιούν τα επιμέρους σήματα σε πλήρες πλάτος στάθμης (**0 dB**), αλλά σε μικρότερες τιμές, συνήθως στα **-3 dB** (Soren et al., 2000).

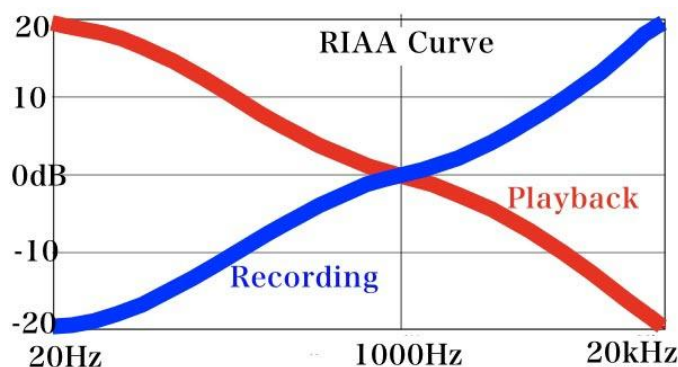
Θόρυβος μπορεί να εισαχθεί και κατά την αποθήκευση του ήχου σε αναλογικά αποθηκευτικά μέσα. Εκεί μπορεί να προκληθεί αλλοίωση της πληροφορίας, όταν για παράδειγμα επιδράσει μαγνητικό πεδίο σε μέσα όπως μαγνητικές ταινίες που παλαιότερα είχαν ευρύτατη χρήση σαν μέσα αποθήκευσης, ή ακόμα και με την μορφή σκόνης, στατικού ηλεκτρισμού ή δημιουργία χαράξεων στις περιπτώσεις που πρόκειται για δίσκους βινυλίου.

Στην τελευταία περίπτωση, ο θόρυβος αυτός μπορεί σε ορισμένες περιπτώσεις να είναι επιθυμητός, καθώς αποτελεί μέρος της ιστορίας που συνοδεύει το μέσο, το οποίο για κάποιους όπως και ο γράφοντας αποτελεί κάτι παραπάνω από ένα απλό αντικείμενο αποθήκευσης ηχητικής πληροφορίας. Αυτό εξηγεί σε μεγάλο βαθμό και την αναβίωση της χρήσης των δίσκων βινυλίου που παρατηρείται έντονα εδώ και αρκετό χρονικό διάστημα (Taylor, 2025).

Μία περίπτωση εσκεμμένης παραμόρφωσης του ηχητικού περιεχομένου, αποτελεί η περίπτωση υιοθέτησης του προτύπου **RIAA (Recording Industry Association of America)**, το οποίο άρχισε να χρησιμοποιείται κατά την χάραξη των δίσκων βινυλίου την δεκαετία του 1950 (Howard, 2009).

Το πρότυπο αυτό δημιουργήθηκε για να αυξηθεί η χωρητικότητα των δίσκων, με βάση την παρατήρηση ότι σύμφωνα με το μήκος κύματος που αντιστοιχεί σε κάθε συχνότητα, οι χαμηλές συχνότητες απαιτούν για την ίδια ηχητική στάθμη μεγαλύτερο πλάτος στην αυλάκωση του δίσκου, απ' ότι οι υψηλές. Έτσι κατά την εγγραφή υποβαθμίζονταν ενεργειακά οι χαμηλότερες περιοχές, με ταυτόχρονη ενίσχυση των υψηλών, κάτι που επανορθωνόταν με εφαρμογή αντίστοιχης ισοστάθμισης κατά την αναπαραγωγή του ήχου.

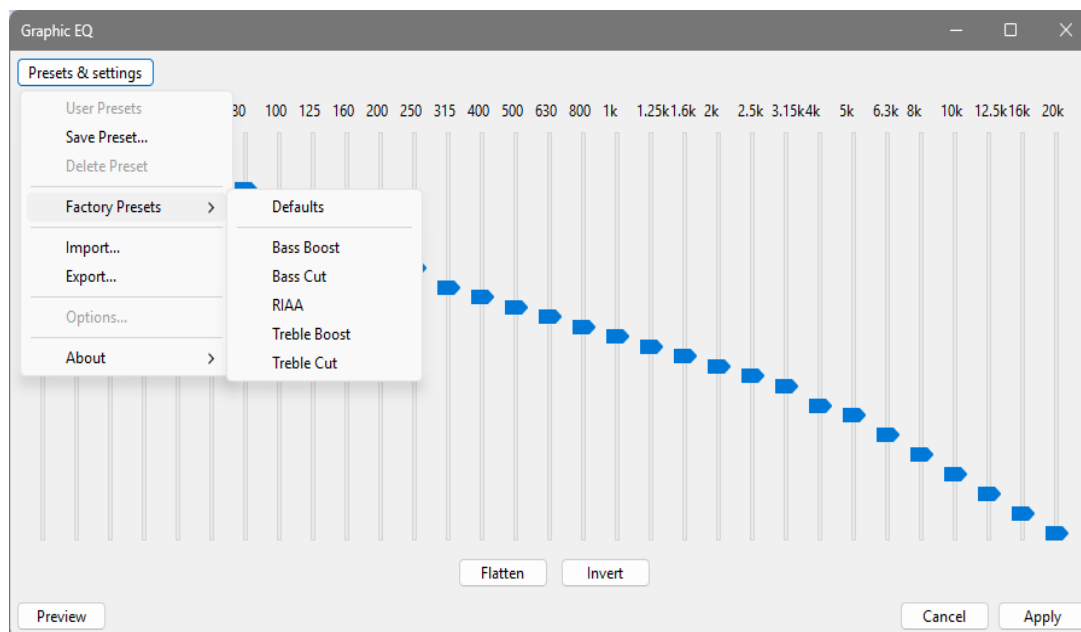
Στο Σχήμα 2.2 που ακολουθεί, αποτυπώνεται η προτυποποιημένη καμπύλη ισοστάθμισης RIAA, στα στάδια εγγραφής (μπλε καμπύλη) και αναπαραγωγής (κόκκινη καμπύλη) των ηχογραφήσεων σε δίσκους βινυλίου.



Σχήμα 2.2. Καμπύλες ισοστάθμισης RIAA κατά την εγγραφή και αναπαραγωγή δίσκων βινυλίου
(Πηγή: <https://jamaicatimeline.com/music/hitchins-riaa.html>)

Τυχόν αστοχία στην ισοστάθμιση του προτύπου RIAA, είχε σαν συνέπεια την διατήρηση των παραμορφώσεων αυτών, οι οποίες αποτελούν μία ακόμα μορφή θορύβου. Χαρακτηριστικό είναι το παράδειγμα, όπου στις δεκαετίες του '60 και του '70 οι τοπικές παραγωγές ρέγκε μουσικής στην Τζαμάικα, θεωρούνταν κατώτερης ηχητικής ποιότητας, σε σχέση με άλλα είδη εισαγόμενων ηχογραφήσεων, και γι' αυτό δεν μεταδίδονταν ιδιαίτερα από το ραδιόφωνο, ακριβώς επειδή δεν λάμβαναν υπόψη τους το πρότυπο ισοστάθμισης καμπύλης RIAA (Hitchins, n.d.).

Για την υλοποίηση της ισοστάθμισης RIAA, αρκετά λογισμικά ψηφιακής επεξεργασίας ήχου (DAW), ενσωματώνουν έτοιμες τις προκαθορισμένες τιμές (presets), όπως ενδεικτικά αποτυπώνεται στο στιγμιότυπο της Εικόνας 2.3, από την χρήση του γραφικού ισοσταθμιστή (**Graphic EQ**) στο ελεύθερο λογισμικό ανοιχτού κώδικα **Audacity** (<https://www.audacityteam.org>).



Εικόνα 2.3. Επιλογή προκαθορισμένων τιμών ισοστάθμισης RIAA στο Graphic EQ του Audacity

Οι προκαθορισμένες τιμές της συγκεκριμένης ισοστάθμισης, είναι διαθέσιμες στα λογισμικά αυτά, καθώς σε αρκετές περιπτώσεις χρειάζεται να εφαρμοστεί στην επεξεργασία μαγνητοφωνημένου αρχειακού υλικού, το οποίο περιλαμβάνει αντίγραφα παλιών δίσκων που παράχθηκαν ακολουθώντας το πρότυπο RIAA.

Παραμόρφωση που μπορεί να εμφανισθεί κατά την ψηφιοποίηση του ήχου αποτελεί και ο θόρυβος κβαντισμού (ή σφάλμα κβαντισμού). Ο θόρυβος αυτός, είναι αποτέλεσμα των αναγκαστικών στρογγυλοποιήσεων των τιμών που παίρνει το σήμα, κατά την μετατροπή του από αναλογική σε ψηφιακή μορφή.

Σε αυτή την περίπτωση, το ποσό του θορύβου που παρεισφρεί στο σήμα εξαρτάται από την απόδοση του κβαντιστή, η οποία σε αναλογία με το SNR, μετράται από **τον λόγο σήματος προς θόρυβο κβαντισμού (Signal to Quantization Noise Ratio - SQNR)** σε dB. Προσεγγιστικά το SQNR υπολογίζεται ως το εξαπλάσιο του μήκους της κωδικής λέξης,

εκφρασμένης σε bits (Ζαρούχας et al, 2013). Έτσι για κάθε επιπλέον bit στην ευκρίνεια του κβαντιστή, συντελείται αύξηση του SQNR κατά περίπου **6 dB**.

2.4.2 Ένταξη θορύβου κατά την μετάδοση - αναπαραγωγή του ήχου

Κατά την μετάδοση και αναπαραγωγή του ψηφιακού ήχου, ο θόρυβος προστίθεται κυρίως στις ενδιάμεσες μετατροπές του σήματος από αναλογικό σε ψηφιακό και αντίστροφα. Καθώς το περιεχόμενο ενός ψηφιοποιημένου ήχου δεν μπορεί πρακτικά να αλλοιωθεί, η όποια ένταξη θορύβου συντελείται στα ενδιάμεσα στάδια, όπου μεσολαβούν αναλογικές συσκευές και διατάξεις. Στην κατηγορίες αυτές ενδεικτικά περιλαμβάνονται αναλογικοί ενισχυτές, ισοσταθμιστές και φίλτρα συμπίεσης (**compressors**), μεγάφωνα και φίλτρα διαχωρισμού συχνοτήτων (**crossover**) ηχείων, συνδέσεις και καλωδιώσεις, που εμπλέκονται κατά την αναπαραγωγή και μετάδοση του ήχου.

Για την αποφυγή του θορύβου κατά την μετάδοση του ήχου σε αναλογικά κανάλια επικοινωνίας, έχουν χρησιμοποιηθεί μέθοδοι αποθορυβοποίησης οι οποίοι παραμορφώνουν εσκεμμένα με γνωστό τρόπο το σήμα στο σημείο εκπομπής, ώστε να το καταστήσουν λιγότερο ευπρόσβλητο, με παράλληλη αναίρεση της παραμόρφωσης και αποκατάσταση του σήματος στο σημείο της λήψης.

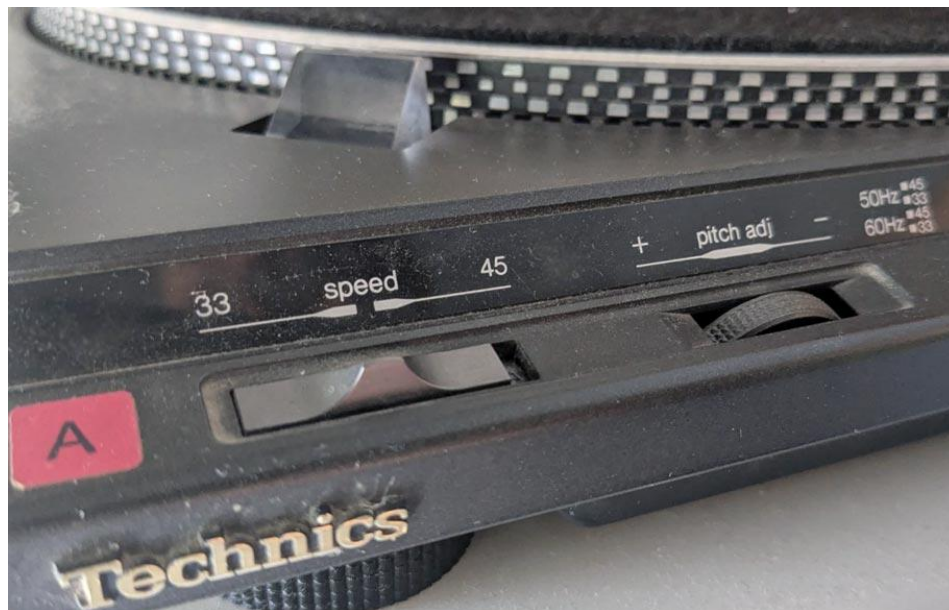
Τέτοια περίπτωση αποτελούν οι μέθοδοι προέμφασης / αποέμφασης (**pre-emphasis** / **de-emphasis**) του σήματος, όπου ο θόρυβος παρεμβαίνει στο κανάλι επικοινωνίας. Σύμφωνα με τις μεθόδους αυτές, οι οποίες μπορούν να εφαρμοστούν μόνο στην περίπτωση που είναι διαθέσιμο το «καθαρό» σήμα (Ζαρούχας et al, 2013), στην φάση της εκπομπής ενισχύονται (προέμφαση) οι υψηλές συχνότητες, οι οποίες είναι και περισσότερο ευάλωτες στον θόρυβο, με συνεπακόλουθη ανάκτηση του σήματος στον δέκτη, με αντίστοιχη μείωση (αποέμφαση) της ισχύος των ίδιων συχνοτήτων.

Παρεμβολές ηλεκτρομαγνητικής φύσεως, μπορούν να αποτελέσουν μία ακόμα ανεπιθύμητη πηγή θορύβου στην περίπτωση της ασύρματης μετάδοσης του ήχου. Στους ψηφιακούς τρόπους μετάδοσης, έχουν υλοποιηθεί δικλίδες ασφαλούς μεταφοράς της πληροφορίας, με την οργάνωση των δεδομένων σε δομημένα «πακέτα», τα οποία μπορούν να επαναπροωθούνται σε περίπτωση αποτυχίας και το περιεχόμενό τους παραμένει

αναλλοίωτο με την προσθήκη ψηφίων ελέγχου στο σήμα. Αντίθετα, το αναλογικό σήμα είναι ευπρόσβλητο σε παρεμβολές οι οποίες μπορεί υποβαθμίσουν σημαντικά την ποιότητα του σήματος, καθώς δεν είναι εφαρμόσιμες αντίστοιχες υλοποιήσεις.

Ο θόρυβος μπορεί παρεισφρήσει και με άλλες μορφές, όπως με την εμφάνιση στάσιμων κυμάτων, εξαιτίας του περιβάλλοντος χώρου, ή από την σχετική θέση και προσανατολισμό των μεγαφώνων στον χώρο ακρόασης. Άλλη μορφή αλλοίωσης, η οποία μπορεί να χαρακτηριστεί θόρυβος με την ευρύτερη έννοια, συντελείται στην χρονική παράμετρο, όπως για παράδειγμα συμβαίνει στον αναλογικό ήχο, όταν η αναπαραγωγή δίσκων βινυλίου γίνεται με αποσυγχρονισμένη περιστροφή του πλατό του πικάπ. Η αντιμετώπιση στην περίπτωση αυτή, γίνεται με μηχανικό τρόπο, μέσω της χρήσης στροβοσκοπίου.

Στην Εικόνα 2.3 που ακολουθεί, αποτυπώνεται φωτογραφικό στιγμιότυπο από σύστημα μηχανικού τρόπου ελέγχου και διόρθωσης της ταχύτητας περιστροφής των δίσκων βινυλίου, σε πικάπ που διαθέτει ενσωματωμένο στροβοσκόπιο.



Εικόνα 2.3. Έλεγχος ταχύτητας περιστροφής δίσκου βινυλίου με χρήση στροβοσκοπίου

Αξίζει τέλος να σημειωθεί ότι η εμφάνιση νέων μοντέλων ψηφιακής διάθεσης της ηχητικής πληροφορίας -και ιδιαίτερα στον τομέα της μουσικής- έχει οδηγήσει στην τάση συγκέντρωσης μεγάλου όγκου ηχητικού περιεχομένου, με αντίστοιχη υποβάθμιση της

ποιότητας του ήχου. Έτσι παρατηρείται «...στροφή των καταναλωτών σε ζήτηση υψηλής χωρητικότητας (**High Capacity**) εις βάρος της τελικής ποιότητας (**High Fidelity - HiFi**), αντιστρέφοντας την απαίτηση για ποιότητα (Φλώρος, 2013)», η οποία υπήρχε τις παλαιότερες δεκαετίες.

Η υποβάθμιση αυτή, που ισοδυναμεί με ένταξη θορύβου στον ψηφιακό ήχο, οφείλεται στην απωλεστική συμπίεση η οποία χρησιμοποιείται συνήθως στις κωδικοποιήσεις, η οποία με την σειρά της προκαλείται από την ανάγκη για περιορισμό του απαιτούμενου αποθηκευτικού χώρου, αλλά και την απαίτηση για επίτευξη καλύτερου ρυθμού μεταφοράς των δεδομένων (bitrate), κατά την μετάδοση και αναπαραγωγή του ήχου μέσα από τα διάφορα κανάλια επικοινωνίας.

2.5 Περικοπή θορύβου (Trimming and Cutting)

Η περικοπή θορύβου (**Trimming and Cutting**), είναι μία από τις πιο απλές τεχνικές επεξεργασίας ηχητικών σημάτων. Εφαρμόζεται κυρίως για την αφαίρεση των ενθόρυβων τμημάτων ενός αποθηκευμένου ήχου, τα οποία δεν συμπίπτουν χρονικά με τα «ωφέλιμα» τμήματά του. Περιλαμβάνει μεταξύ άλλων το κόψιμο αποσπασμάτων αθέλητης ομιλίας και ήχων που προκαλούνται από τον ομιλητή κατά την διάρκεια της εκφοράς του λόγου.

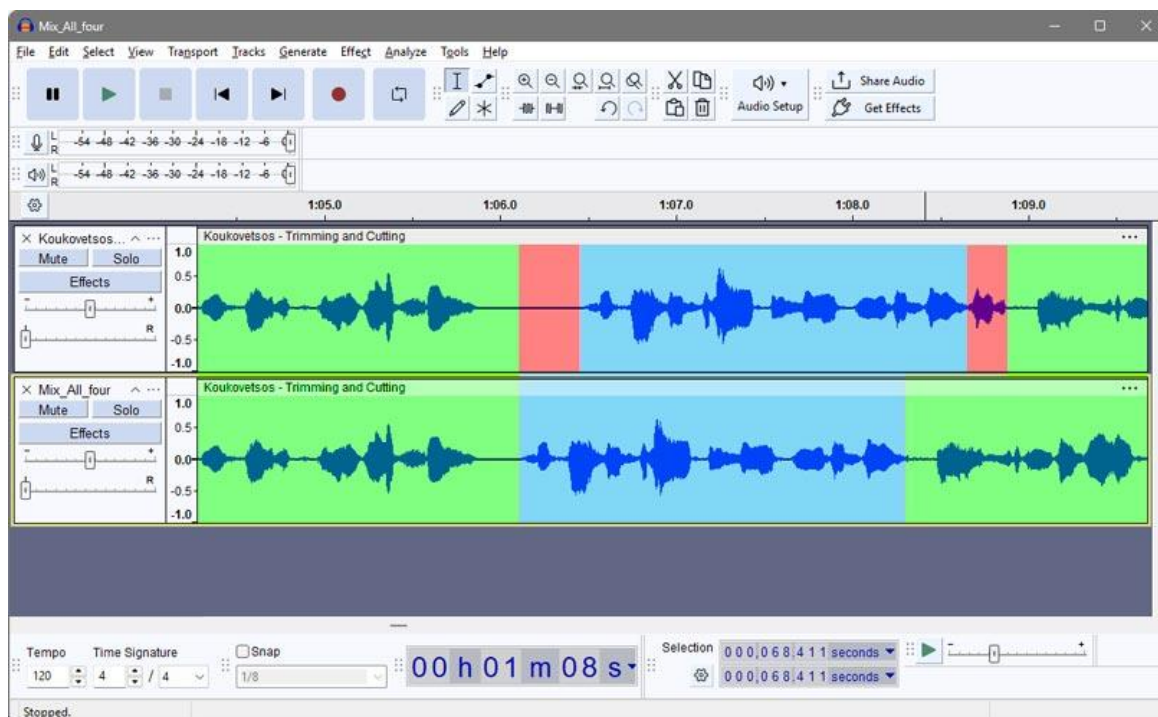
Σε αυτήν εντάσσονται οι «λέξεις-γεμίσματα» (**filler words**), όπως είναι ενδεικτικά τα επιφωνήματα δισταγμού «...εμμ» και «...εεε», οι επαναλήψεις λέξεων, η εμφάνιση βήχα, και γενικά η παρουσία θορύβου σε παύσεις ανάμεσα σε διαστήματα ομιλίας. Επιπρόσθετα, με την μέθοδο αυτή αντιμετωπίζεται η ύπαρξη μεγάλων παύσεων, με την περικοπή τους στην επιθυμητή μικρότερη διάρκεια, για την αισθητική βελτίωση του περιεχομένου.

Η μέθοδος δεν εφαρμόζεται σε σήματα που μεταδίδονται σε πραγματικό χρόνο, ενώ αποτελεί διαδικασία που αλλάζει την συνολική χρονική διάρκεια του ηχητικού σήματος. Στην περίπτωση της επεξεργασίας αναλογικών μέσων αποθήκευσης, η υλοποίηση της μεθόδου πραγματοποιείται με την φυσική αποκοπή και συρραφή των τμημάτων από το μέσο, το οποίο αποτελείται κυρίως από μαγνητικές ταινίες. Η διαδικασία βασίζεται στην κοπή της ταινίας με ξυράφι και την ένωσή της με την χρήση ειδικής μονωτικής ταινίας

(**splicing tape**), ώστε να αφαιρεθούν τα ανεπιθύμητα τμήματα της ηχογράφησης (McGowan, 2023).

Η ψηφιοποίηση της επεξεργασίας του ήχου, συνέβαλε ώστε η τεχνική αυτή να γίνει απλούστερη, μέσα από την οπτικοποίηση του ήχου με την χρήση των DAWs. Έτσι ο υπεύθυνος επεξεργασίας, μπορούσε πλέον να δει τα ενθόρυβα τμήματα, ώστε η επιλογή και η αφαίρεσή τους να γίνει με απλές κινήσεις του ποντικιού στο υπολογιστικό του σύστημα. Ταυτόχρονα η όλη διαδικασία έγινε μη απωλεστική (**non-destructive**), καθώς το αρχικό ενθόρυβο σήμα διατηρείται ψηφιακά και δεν περικόπτεται με φυσικό τρόπο.

Στην Εικόνα 2.4 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο οθόνης από επεξεργασία περικοπής θορύβου με χρήση του λογισμικού **Audacity**.



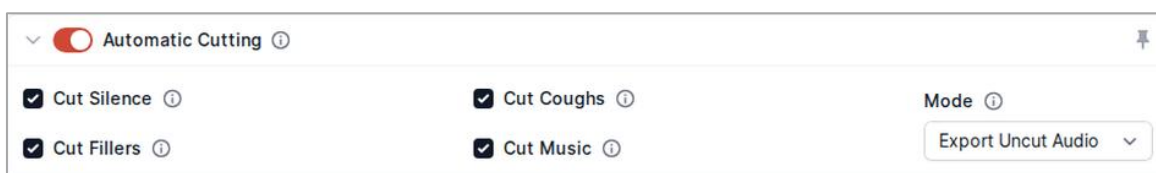
Εικόνα 2.4. Στιγμιότυπο οθόνης χρήσης του Audacity για περικοπή θορύβου

Στην παραπάνω Εικόνα, εμφανίζεται στο πάνω κανάλι το ηχητικό σήμα πριν την εφαρμογή περικοπής, ενώ στο κάτω κανάλι φαίνεται το ίδιο επεξεργασμένο σήμα, το οποίο έχει προφανώς υποστεί μετατόπιση στον χρονικό άξονα (κεντρική μπλε και δεξιά πράσινη περιοχή). Από το συγκεκριμένο εικονιζόμενο σήμα, έχουν αφαιρεθεί δύο ανεπιθύμητα τμήματα «θορύβου», τα οποία αποτυπώνονται με κόκκινο χρώμα στην αρχική πάνω

κυματομορφή. Το πρώτο αφαιρέθηκε για μείωση της χρονικής διάρκειας σε παύση, ενώ το δεύτερο για να αποκοπεί λέξη η οποία επαναλαμβάνεται εκ παραδρομής.

Η μέθοδος αυτή περιλαμβάνει διαδικασία υλοποίησης, η οποία μπορεί να πραγματοποιηθεί σχετικά απλά με την χρήση συμβατικών τεχνικών. Εντούτοις η βοήθεια της Τεχνητής Νοημοσύνης έχει συμβάλει στην αυτοματοποίησή της, καθιστώντας την όλη διαδικασία πολύ πιο σύντομη, μειώνοντας τον συνολικό φόρτο εργασίας, ιδιαίτερα σε περιπτώσεις που το ηχητικό περιεχόμενο είναι ιδιαίτερα μεγάλο.

Έτσι, η χρήση TN απλοποιεί την διαδικασία, ιδιαίτερα για επεξεργασία σημάτων μεγάλης χρονικής διάρκειας, ή εκεί όπου η ταχύτητα στην παραγωγή αποτελέσματος είναι επιβεβλημένη. Για τον λόγο αυτό, πολλά από τα σύγχρονα εργαλεία TN, ενσωματώνουν στους αλγόριθμους τους και την δυνατότητα αποκοπής θορύβου μαζί με τις άλλες διαθέσιμες διεργασίες αποθορυβοποίησης που παρέχουν. Στην Εικόνα 2.5 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από επιλογές περικοπής θορύβου με χρήση TN, στην διαδικτυακή πλατφόρμα **Auphonic**.



Εικόνα 2.5. Στιγμιότυπο επιλογών περικοπής θορύβου στην πλατφόρμα Auphonic
(Πηγή: <https://auphonic.com/help/algorithms/singletrack.html#automatic-cutting>)

Όπως αναφέρεται στην πλατφόρμα, ο αλγόριθμος TN που χρησιμοποιείται από αυτήν, έχει ήδη εκπαιδευτεί ώστε να λειτουργεί αποτελεσματικά σε πολλές από τις ομιλούμενες γλώσσες, ενώ προβλέπεται η εκπαίδευση του αλγόριθμου με χρήση μοντέλων σε ακόμα περισσότερες γλώσσες στο μέλλον, εφόσον κριθεί αναγκαίο (Auphonic, n.d.)

Για την πληρότητα σημειώνεται ότι σε αρκετές περιπτώσεις, αντί για την περικοπή των ενθόρυβων τμημάτων προτιμάται η αντικατάστασή τους με σιωπή (σήμα μηδενικής στάθμης) ίσης χρονικής διάρκειας, τόσο σε περιπτώσεις όπου πρέπει για λόγους φυσικότητας στην ομιλία να παραμείνουν οι παύσεις, όπως στα σημεία των αναπνοών, όσο και σε περιπτώσεις όπου η χρονική εξέλιξη του σήματος δεν πρέπει να τροποποιηθεί για λόγους συγχρονισμού, κάτι το οποίο συμβαίνει κυρίως σε επεξεργασία ήχου που περιλαμβάνεται σε αρχεία βίντεο.

2.6 Αφαίρεση ψαλιδισμού (Declipping)

Μία μορφή θορύβου, η οποία εμφανίζεται με την μορφή παραμόρφωσης, είναι ο ψαλιδισμός (**clipping**) του ηχητικού σήματος. Ο ψαλιδισμός εμφανίζεται όταν η στάθμη του σήματος ξεπεράσει το μέγιστο επιτρεπτό όριο, το οποίο ο εμπλεκόμενος εξοπλισμός, υλικό (hardware) ή λογισμικό (software), μπορεί να διαχειριστεί.

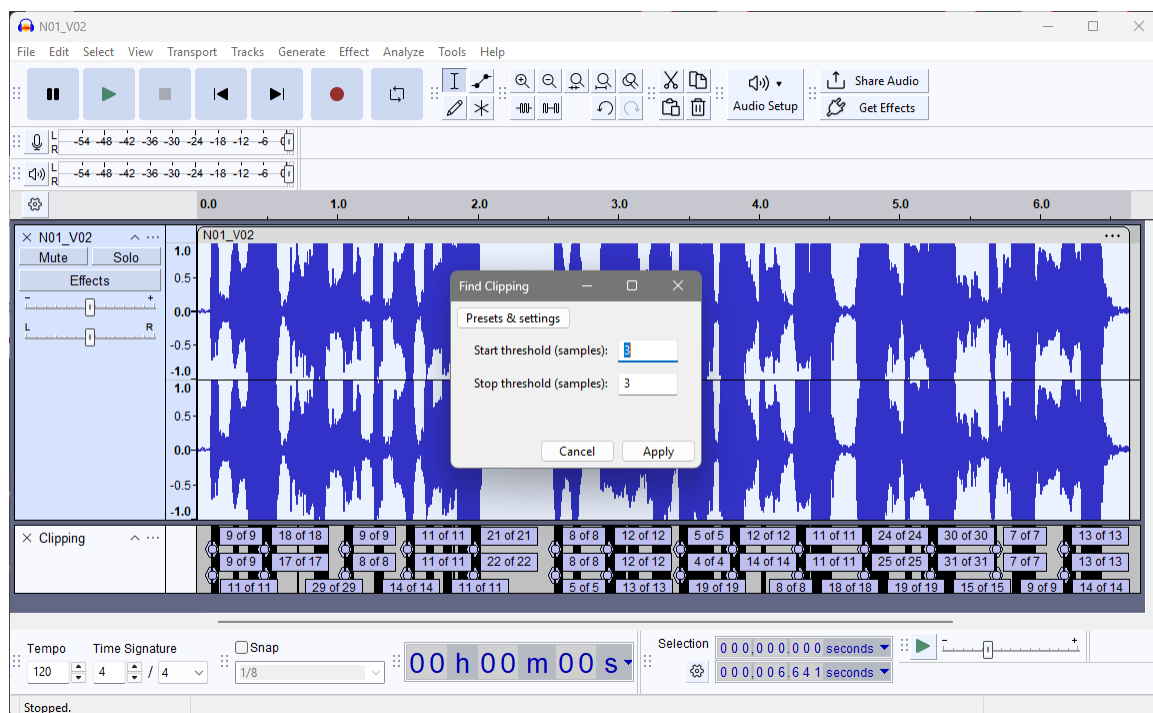
Στην περίπτωση της αναλογικής καταγραφής και επεξεργασίας, ο ψαλιδισμός συντελείται με την υπέρβαση των φυσικών ορίων των ηλεκτρονικών κυκλωμάτων από τα οποία διέρχεται ο ήχος, ενώ όταν η επεξεργασία γίνεται ψηφιακά ταυτίζεται με την υπέρβαση του ορίου του μετατροπέα από αναλογικό σήμα σε ψηφιακό (**Analog to Digital Converter - DAC**) στην φάση της μετατροπής ή στην υπέρβαση που μπορεί να συμβεί συνήθως κατά την διάρκεια μίξεων διαφορετικών καναλιών, που το καθένα έχει ήδη υψηλή ηχητική στάθμη. Το τμήμα της ηχητικής πληροφορίας που υπερβαίνει το όριο αποκόπτεται, προκαλώντας μη αναστρέψιμη παραμόρφωση, με αποτέλεσμα στην απεικόνιση του ήχου, να εμφανίζονται κυματομορφές με επίπεδες περιοχές στο άνω όριο των σημάτων.

Κατά την αναπαραγωγή του ηχητικού σήματος, ο ψαλιδισμός προκαλεί αρμονική παραμόρφωση, η οποία οφείλεται στις οριακές συνθήκες κάτω από τις οποίες λειτουργεί το σύστημα ενίσχυσης του ήχου. Αυτή η αρμονική παραμόρφωση συντελείται επειδή η ηλεκτρική ισχύς εξόδου υπερβαίνει την τιμή για την οποία είναι σχεδιασμένο το συνδεδεμένο ηχείο, και εντοπίζεται συνήθως στην περιοχή των υψηλών συχνοτήτων (Τσιλφίδης, 2013).

Η αποφυγή του φαινομένου απαιτεί προσεκτικό έλεγχο και διαρκή επιτήρηση της ηχητικής στάθμης, κατά την εκτέλεση όλων των σταδίων που εμπλέκονται στην διενέργεια των ηχογραφήσεων και της συνεπακόλουθης παραγωγικής διαδικασίας. Για την αποφυγή δημιουργίας ψαλιδισμού, ιδιαίτερα σε ηχητικά σήματα πραγματικού χρόνου (real-time), όπως για παράδειγμα είναι ραδιοφωνικές εκπομπές, συναυλίες και ζωντανές μεταδόσεις, χρησιμοποιούνται ειδικές διατάξεις, οι οποίες ονομάζονται περιοριστές (Limiters), οι οποίοι επιτηρούν και επεμβαίνουν άμεσα όταν χρειάζεται, εμποδίζοντας την ηχητική στάθμη να ξεπεράσει το μέγιστο επιτρεπτό όριο που οδηγεί στον ψαλιδισμό του σήματος (QSC, 2024).

Σε αρκετές περιπτώσεις, ο ήχος μπορεί να φτάνει αλλά να μην ξεπερνά το άνω επιτρεπτό όριο, κάτι το οποίο μπορεί να δώσει λανθασμένη οπτική εντύπωση κατά την επεξεργασία από τα διαθέσιμα ψηφιακά εργαλεία (DAWs). Έτσι αυτά ενσωματώνουν

ειδικά εργαλεία ανάλυσης και εντοπισμού των σημείων όπου μπορεί να εμφανίζεται ο ψαλιδισμός, όπως ενδεικτικά αποτυπώνεται στο στιγμιότυπο της Εικόνας 2.6 που ακολουθεί, κατά την χρήση του λογισμικού **Audacity**.



Εικόνα 2.6. Στιγμιότυπο χρήσης εργαλείου εύρεσης σημείων ψαλιδισμού στο Audacity

Ο τρόπος αποκατάστασης του ήχου στην περίπτωση που ήδη έχει επιδράσει ο ψαλιδισμός στο σήμα, γίνεται με αξιοποίηση εργαλείων τα οποία με την χρήση ειδικών αλγορίθμων αξιοποιούν την πληροφορία που υπάρχει στις γειτονικές περιοχές του σήματος, αναλύοντάς τες και ανασυνθέτοντας το χαμένο σήμα κατά προσέγγιση. Τέτοια εργαλεία αποτελούν ενδεικτικά το εργαλείο «**De-clip**», που βρίσκεται ενσωματωμένο στο εξειδικευμένο λογισμικό «**RX**» της εταιρίας **Izotope** (<https://www.izotope.com>), ή το εργαλείο «**DeClipper**», το οποίο περιλαμβάνεται στο λογισμικό «**Audition**» της εταιρίας **Adobe** (https://www.adobe.com/gr_en/products/audition.html).

Η εμπλοκή της Τεχνητής Νοημοσύνης στην συγκεκριμένη μορφή αποθρονοποίησης, έχει οδηγήσει στην δημιουργία νέων εργαλείων λογισμικού, τα οποία είναι κυρίως διαθέσιμα με την μορφή προσθέτων (plugins). Στην Εικόνα 2.7 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από επιλογές περικοπής θορύβου με χρήση TN, στην

οθόνη του προσθέτου λογισμικού με το όνομα «**ClipRemover**», της εταιρίας **BORISFX** (<https://borisfx.com/products/clipremover/>).



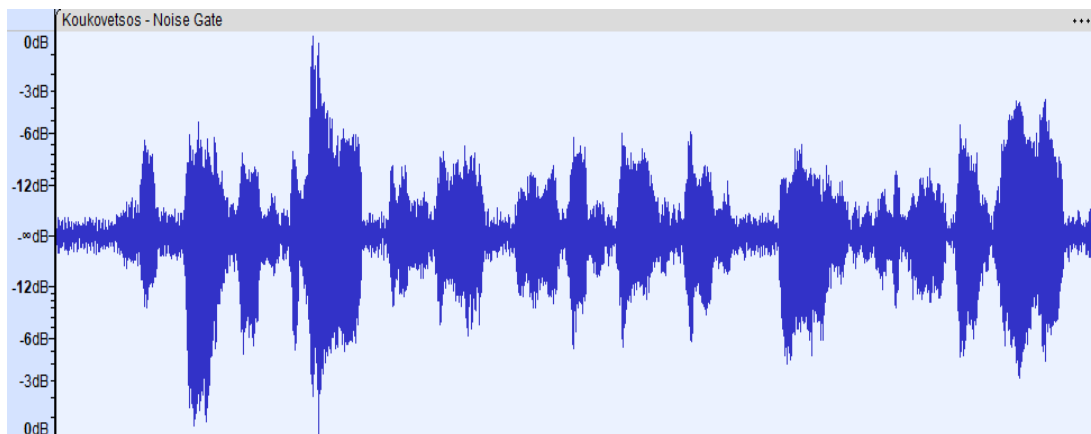
Εικόνα 2.7. Οθόνη του προσθέτου αφαίρεσης ψαλιδισμού ClipRemover AI της BORISFX
(Πηγή: <https://borisfx.com/products/clipremover/>, επεξεργασμένη)

Όπως αναφέρεται στον ιστότοπο του λογισμικού, ο αλγόριθμος TN που χρησιμοποιείται από αυτό, αναγνωρίζει τις ψαλιδισμένες περιοχές της κυματομορφής του σήματος και τις ανακατασκευάζει με αυτόματο τρόπο, περιορίζοντας τις παραμορφώσεις, επαναφέροντας την φυσική χροιά στον ήχο του σήματος (BORISFX, n.d.).

2.7 Πύλη θορύβου (Noise Gate)

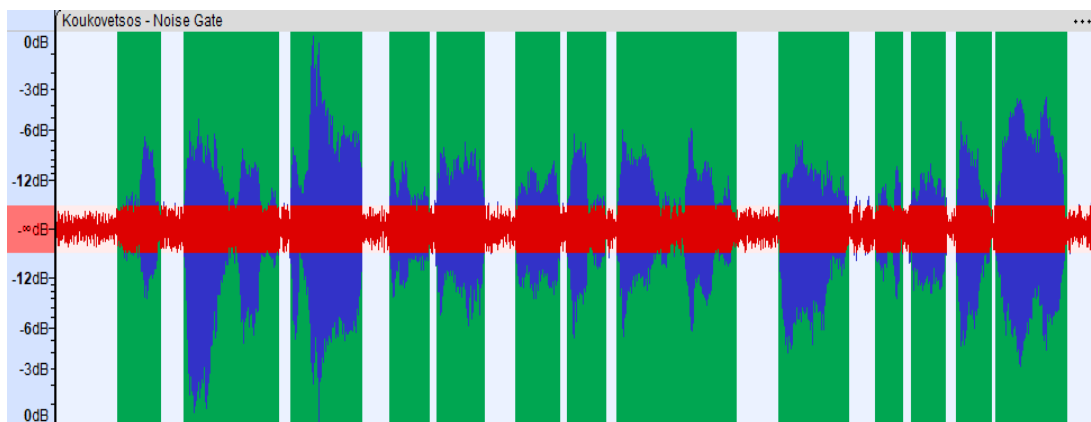
Μέθοδος εφαρμογής πύλης θορύβου (**noise gate**) είναι μία ιδιαίτερη μορφή τεχνικής αποθορυβοποίησης. Εφαρμόζεται στις περιπτώσεις που ο θόρυβος είναι στατικός και έχει σημαντικά χαμηλότερη στάθμη από την στάθμη του ωφέλιμου ηχητικού σήματος. Στην πράξη δεν πρόκειται για αποθορυβοποίηση με την στενή έννοια, καθώς ο θόρυβος δεν αφαιρείται από το ηχητικό σήμα, αλλά μέσω ειδικής διάταξης ο ήχος αποκόπτεται εντελώς στα χρονικά σημεία που η στάθμη του ήχου βρίσκεται κάτω από μία ορισμένη τιμή, η οποία αντιστοιχεί στην ανεκτή τιμή του θορύβου. Η τιμή αυτή αποτελεί το ονομαζόμενο «**κατώφλι θορύβου**» (**noise threshold**).

Στην Εικόνα 2.8, που ακολουθεί, εμφανίζεται ενδεικτική περίπτωση κυματομορφής σήματος ομιλίας, στο οποίο υπάρχει στατικός θόρυβος, μέγιστης στάθμης περίπου ίσης με **-18 dB**. Ο θόρυβος διακρίνεται στα σημεία όπου κάτω ιδανικές συνθήκες θα έπρεπε να υπάρχει μηδενική στάθμη, λόγω των παύσεων στην εκφορά του λόγου ανάμεσα στις λέξεις.



Εικόνα 2.8. Κυματομορφή σήματος ομιλίας στο οποίο υπάρχει στατικός θόρυβος

Στην Εικόνα 2.9, βλέπουμε την ίδια κυματομορφή χρωματικά επεξεργασμένη για καλύτερη εποπτεία της μεθόδου. Με κόκκινο χρώμα αποτυπώνεται η ζώνη στάθμης στην οποία εμφανίζεται ο θόρυβος. Η εφαρμογή πύλης θορύβου (**gating**) στο άκρο της περιοχής αυτής, έχει σαν συνέπεια να αποκόπτεται πλήρως ο ήχος στα σημεία όπου η στάθμη του είναι κάτω από αυτό το όριο (κόκκινη περιοχή) και να επιτρέπεται μόνο όταν το ξεπερνά. Έτσι τα σημεία στα οποία επιτρέπεται η διέλευση και ακούγεται ο ήχος είναι αυτά που ανήκουν στις πράσινες ενότητες της κυματομορφής.

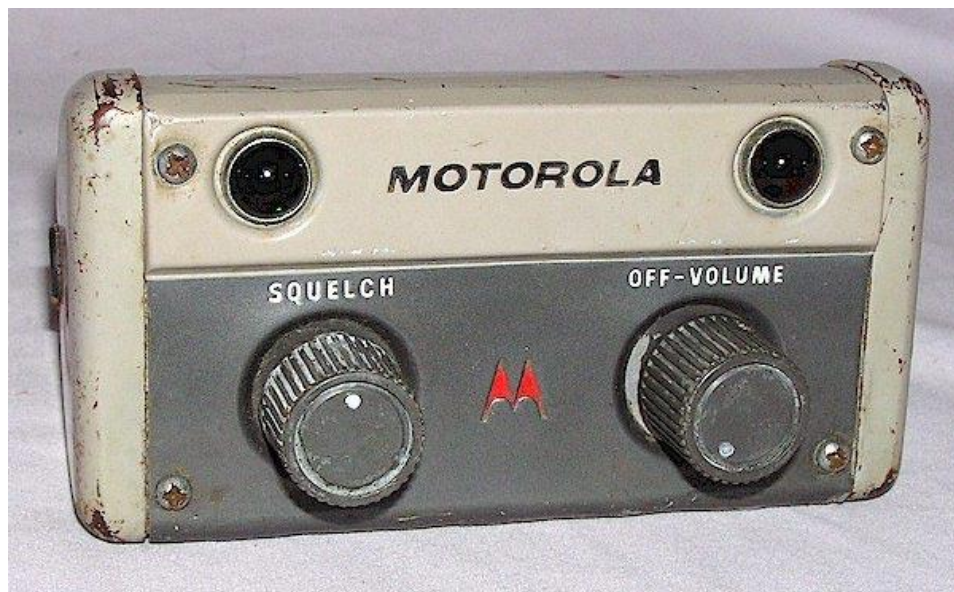


Εικόνα 2.9. Απεικόνιση σήματος ομιλίας με εφαρμογή πύλης θορύβου

Από την προηγούμενη Εικόνα, γίνεται και οπτικά φανερό ότι ο μαζί με το ωφέλιμο σήμα μεταδίδεται αμετάβλητος και θόρυβος, τις χρονικές στιγμές που επιτρέπεται να ακουστεί ο ήχος (πράσινες ζώνες). Έτσι η μέθοδος αυτή χρησιμοποιείται επικουρικά σε περιπτώσεις όπου το SNR του σήματος είναι σχετικά μεγάλο, ή όταν δεν ενδιαφέρει σε σημαντικό βαθμό η ποιότητα του τελικού σήματος.

Τέτοια περίπτωση αποτελούν τα αναλογικά συστήματα ραδιοεπικοινωνίας, στα οποία τη τεχνική αυτή βρήκε εφαρμογή από τα πρώτα βήματα της ανάπτυξής τους. Μάλιστα στα συστήματα αυτά, στη θέση του όρου «πύλη θορύβου», χρησιμοποιείται ευρέως και ο όρος «**φίμωση**» (**Squelch**), για να περιγράψει την ίδια λειτουργία.

Στην Εικόνα 2.10 που ακολουθεί, αποτυπώνεται για ιστορικούς λόγους, το αναλογικό σύστημα ραδιοεπικοινωνίας **TWIN-V** της εταιρείας **Motorola** της περιόδου 1955-61, εξοπλισμένο με δυνατότητα φίμωσης (squelch).



Εικόνα 2.10. Αναλογικό σύστημα ραδιοεπικοινωνίας με δυνατότητα φίμωσης (Squelch)
(Πηγή: <https://www.wb6nvh.com/Moto42/M42Page2.htm>)

Ένα άλλο πεδίο στο οποίο εφαρμόζεται η τεχνική της πύλης θορύβου, είναι στην χρήση των ηλεκτρικών μουσικών οργάνων. Στην περίπτωση αυτή η υλοποίηση γίνεται με την χρήση ειδικών ηλεκτρονικών διατάξεων, οι οποίες ονομάζονται «**πεντάλ ή πετάλια**» (**pedals**), και σκοπό τους έχουν να επέμβουν στο σήμα που παράγουν τα όργανα, ώστε να προκύψει το επιθυμητό ηχητικό αποτέλεσμα.

Ενδεικτικό παράδειγμα μιας τέτοιας ηλεκτρονικής διάταξης, αποτελεί το πεντάλ «Zuul+» της εταιρίας Fortin Amps (<https://fortinamps.com/>), το οποίο εμφανίζεται στην Εικόνα 2.11, που ακολουθεί.



Εικόνα 2.11. Πεντάλ μουσικού οργάνου για εφαρμογή πύλης θορύβου
(Πηγή: <https://fortinamps.com/products/zuul-plus>)

Στο συγκεκριμένο πεντάλ, εκτός από το κεντρικό ποτενσιόμετρο ρύθμισης του κατωφλίου με το όνομα «GATE», υπάρχουν δύο επιπλέον ποτενσιόμετρα ρυθμίσεων, με τα ονόματα «HOLD» και «RELEASE», τα οποία παρέχουν μεγαλύτερο βαθμό ελέγχου, καθώς με την χρήση τους μπορούν αντίστοιχα να καθοριστούν οι χρόνοι που η πύλη μένει ανοιχτή (hold), κάτι που είναι χρήσιμο για μικρές παύσεις, όπως και το πόσο γρήγορα γίνεται η μετάβαση από λειτουργία διάβασης σε λειτουργία αποκοπής (release) του σήματος που διέρχεται από το πεντάλ.

2.8 Φασματική επεξεργασία (Spectral Editing)

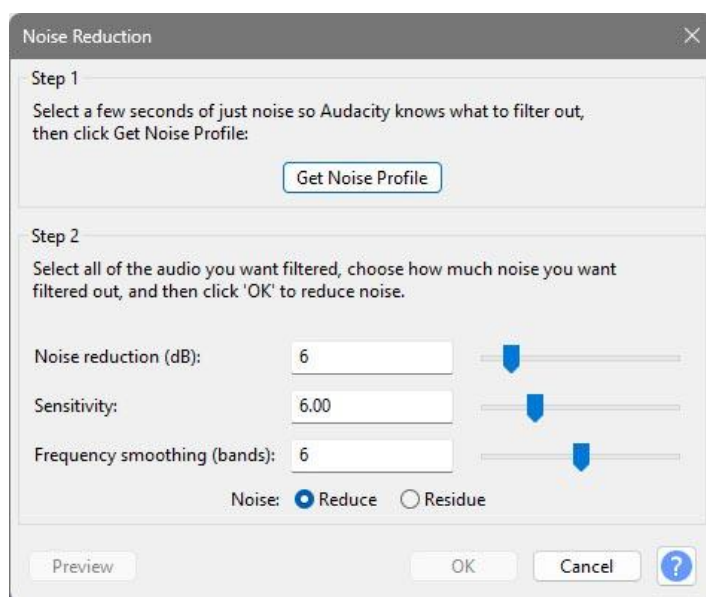
Με τον όρο φασματική επεξεργασία (**Spectral Editing**), για την αποθορυβοποίηση ενός ενθόρυβου σήματος, κυρίως αναφερόμαστε σε μεθόδους φασματικής αφαίρεσης (**spectral subtraction methods**) του θορύβου. Οι μέθοδοι αυτές υλοποιούνται σχετικά

εύκολα και εδράζονται στην βασική αρχή ότι όσο ο θόρυβος εμφανίζεται προσθετικά, μπορεί να εκτιμηθεί φασματικά εκ των προτέρων, σε σημεία στα οποία το «καθαρό» ηχητικό σήμα απουσιάζει, ώστε να μπορέσει να αφαιρεθεί στην συνέχεια από το ενθόρυβο σήμα (Ζαρούχας et al, 2013).

Η δημιουργία ενός κατάλληλου προφίλ θορύβου (**noise profile**), από χρονικά σημεία στα οποία δεν υπάρχουν άλλα ηχητικά δεδομένα, αποτελεί προϋπόθεση για την επιτυχή διενέργεια της όλης διαδικασίας. Για τον λόγο αυτό, η συγκεκριμένη επεξεργασία αποθρονοποίησης, συχνά ονομάζεται και απλά σαν «**Noise Profiling**».

Παρόλο που η εφαρμογή της ανήκει στις σχετικά εύκολα υλοποιήσιμες, η μέθοδος αυτή εμφανίζει δύο σημαντικές αδυναμίες. Η πρώτη, εντοπίζεται στην απαίτηση να υπάρχει ο καθορισμός ενός σιωπηλού διαστήματος, ο οποίος θα πρέπει να γίνει από τον χρήστη. Συνεπώς δεν πρόκειται για μία αυτοματοποιημένη διαδικασία αποθρονοποίησης. Η δεύτερη αδυναμία έγκειται στο γεγονός ότι μπορεί το δείγμα θορύβου αναφοράς που θα επιλεγεί να μην είναι επαρκές, ιδιαίτερα στην περίπτωση που ο θόρυβος έχει μη στατικό περιεχόμενο (Xu et al., 2020).

Στην Εικόνα 2.12 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από το σχετικό παράθυρο επιλογής προφίλ θορύβου, για επεξεργασία αποθρονοποίησης με την χρήση του λογισμικού **Audacity**.

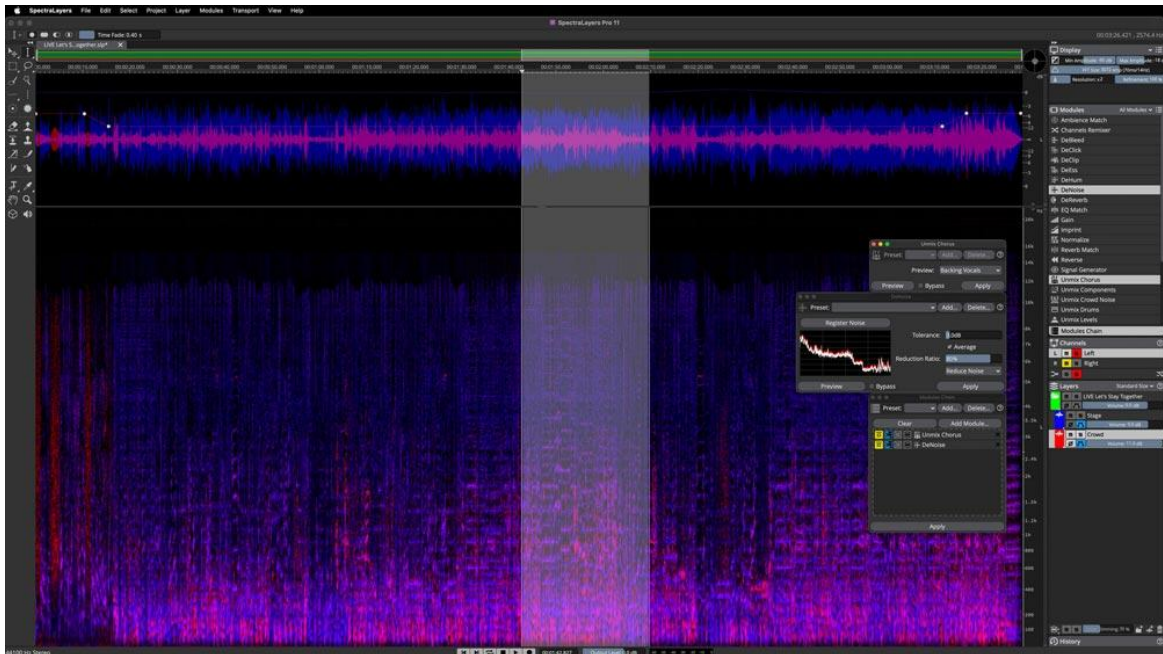


Εικόνα 2.12. Στιγμιότυπο οθόνης επιλογής προφίλ θορύβου του Audacity

Μία άλλη μορφή φασματικής επεξεργασίας, εφαρμόζεται όταν πρέπει να αφαιρεθεί ο θόρυβος που μεταβάλλεται προϊόντος του χρόνου, ή εμφανίζεται αποσπασματικά σε πολύ συγκεκριμένα σημεία του ηχητικού σήματος, όπως χαρακτηριστικά τέτοια παραδείγματα αποτελούν οι σποραδικοί θόρυβοι, που συνήθως έχουν πολύ μικρή διάρκεια.

Στην περίπτωση αυτή, χάρη στις σύγχρονες δυνατότητες που παρέχει η ψηφιακή επεξεργασία, η επιλογή και αφαίρεση του θορύβου συντελείται με οπτικό τρόπο, μέσα από την επεξεργασία του φασματογραφήματος των ηχητικών σημάτων. Έτσι ο θόρυβος εντοπίζεται, επιλέγεται και αποκόπτεται, με διαδικασία όπως αυτή που χρησιμοποιείται όταν πρόκειται για επεξεργασία εικόνας.

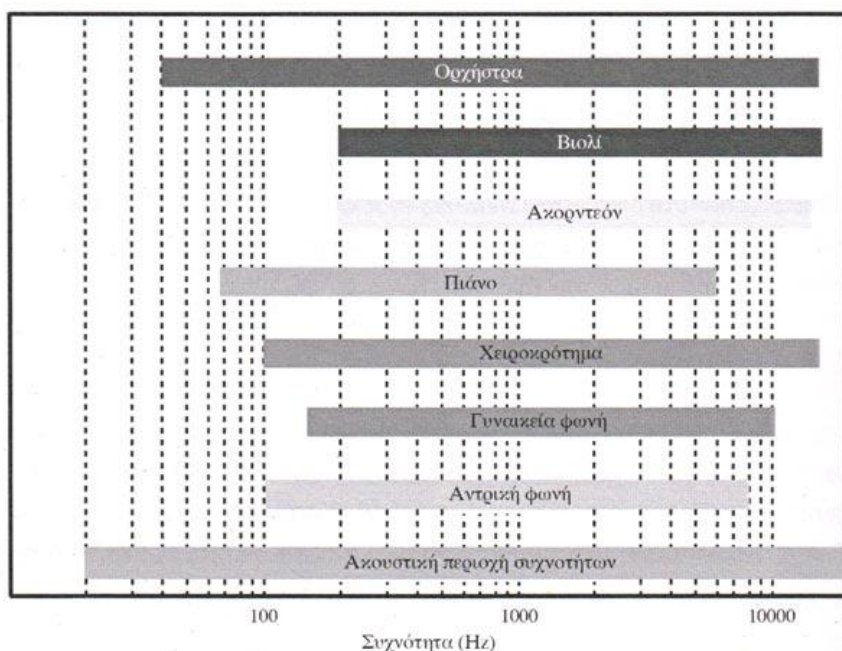
Για τον σκοπό αυτό, έχουν αναπτυχθεί εξειδικευμένα λογισμικά, τα οποία χρησιμοποιούνται από επαγγελματίες για φασματική επεξεργασία που συντελείται με στόχο την αποθρομβοποίηση και βελτίωση του ήχου. Στην Εικόνα 2.13 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από την οθόνη εργασίας του λογισμικού «SpectraLayers» της εταιρίας Steinberg. (<https://www.steinberg.net/spectralayers/>), το οποίο αποτελεί ένα τέτοιο ψηφιακό εργαλείο.



Εικόνα 2.13. Στιγμιότυπο οθόνης εργασίας του λογισμικού SpectraLayers της Steinberg
(Πηγή: <https://www.steinberg.net/spectralayers/>)

2.9 Εφαρμογή φίλτρων (Filtering)

Οι τυπικές ηχητικές πηγές, όπως είναι τα διάφορα γνωστά μουσικά όργανα και η ανθρώπινη φωνή, παράγουν ήχους που έχουν περιορισμένο εύρος, σε σχέση με την ακουστική περιοχή συχνοτήτων, η οποία εκτείνεται από **20 Hz** έως **20000 Hz**. Τυπικές τιμές των ζωνών που καλύπτονται από τις συνηθέστερες περιπτώσεις, αποτυπώνονται στο Σχήμα 2.3, που ακολουθεί.



Σχήμα 2.3. Εύρος συχνοτήτων τυπικών ηχητικών πηγών
(Πηγή: Φλώρος, 2013, σ.50, Σχήμα 1.12)

Για την ανθρώπινη ομιλία, έχει παρατηρηθεί ότι στην περίπτωση του άνδρα αυτή συνήθως έχει μέσο όρο τιμής θεμελιώδους συχνότητας η οποία κυμαίνεται ανάμεσα στα **90 Hz** και τα **1000 Hz**, ενώ στην περίπτωση της γυναίκας αυτή τοποθετείται συχνотικά λίγο υψηλότερα (Φλώρος, 2013). (Το άνω όριο που απεικονίζεται στο Σχήμα 3.1, αφορά στην περίπτωση τραγουδιού ή παραγωγής άλλων ήχων φωνής και όχι την τυπική μορφή ομιλίας).

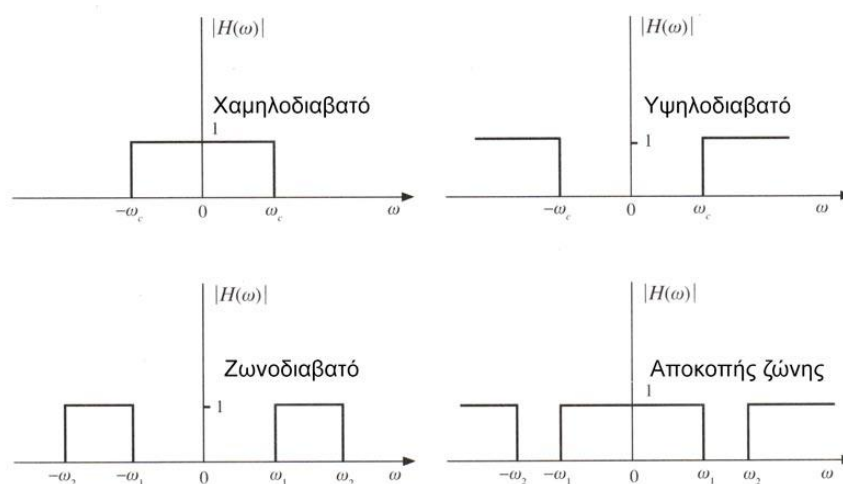
Είναι επίσης γνωστό, ότι τα διάφορα φωνήματα της ομιλίας, όπως και οι φωνητικοί φθόγγοι και τα σύμφωνα, «...αντιστοιχούν σε συγκεκριμένες συχνότητες συντονισμού και προστίθενται στο ακουστικό σήμα... (Φλώρος, 2013)», για κάθε ανθρώπινη φωνή.

Τα παραπάνω δεδομένα αξιοποιούνται στην μέθοδο αποθρομβοποίησης του ήχου με την χρήση φίλτρων. Στην περίπτωση αυτή, γνωρίζοντας την περιοχή συχνοτήτων που πρέπει να διατηρηθεί, εφαρμόζονται τα κατάλληλα φίλτρα, ώστε να αποκοπούν οι συχνότητες που βρίσκονται έξω από τα συχνοτικά αυτά όρια. Για τον λόγο αυτό, η συχνότητα στην οποία το φίλτρο επιδρά ονομάζεται «**συχνότητα αποκοπής**» του φίλτρου.

Κάτι ανάλογο συμβαίνει και εντός των συχνοτικών ορίων, όταν με την χρήση των φίλτρων επιχειρείται να ελεγχθεί μία μικρή προβληματική ζώνη συχνοτήτων, η οποία περιλαμβάνει συγκεκριμένη πηγή θορύβου, όπως είναι για παράδειγμα ο περιορισμός των έντονων συριστικών συμφώνων σε μία ηχογράφιση.

Τα φίλτρα είναι φυσικά ηλεκτρονικά κυκλώματα, ή υλοποιήσεις λογισμικού, που επιτρέπουν την διέλευση συγκεκριμένων συχνοτήτων, αποκόπτοντας όλες τις υπόλοιπες. Ανάλογα με την περιοχή των συχνοτήτων που επιτρέπουν, τα φίλτρα διακρίνονται σε **χαμηλοδιαβατά** ή **χαμηλοπερατά (Low-Pass Filters)**, σε **υψηλοπερατά** ή **υψιπερατά (High-Pass Filters)**, σε **ζωνοδιαβατά** ή **ζωνοπερατά (Band-Pass Filters)** και σε φίλτρα **αποκοπής ζώνης (Band-Stop Filters)** (Ζαρούχας et al, 2013). Επειδή τα φίλτρα αποκόπτουν τις υπόλοιπες συχνότητες, τα χαμηλοπερατά συναντώνται και με τον όρο «**Hi-Cut Filter**», ενώ αντίστοιχα τα υψιπερατά με τον όρο «**Lo-Cut Filter**».

Οι αποκρίσεις των παραπάνω φίλτρων, στην ιδανική περίπτωση λειτουργίας τους, αποτυπώνονται στο Σχήμα 2.4 που ακολουθεί.

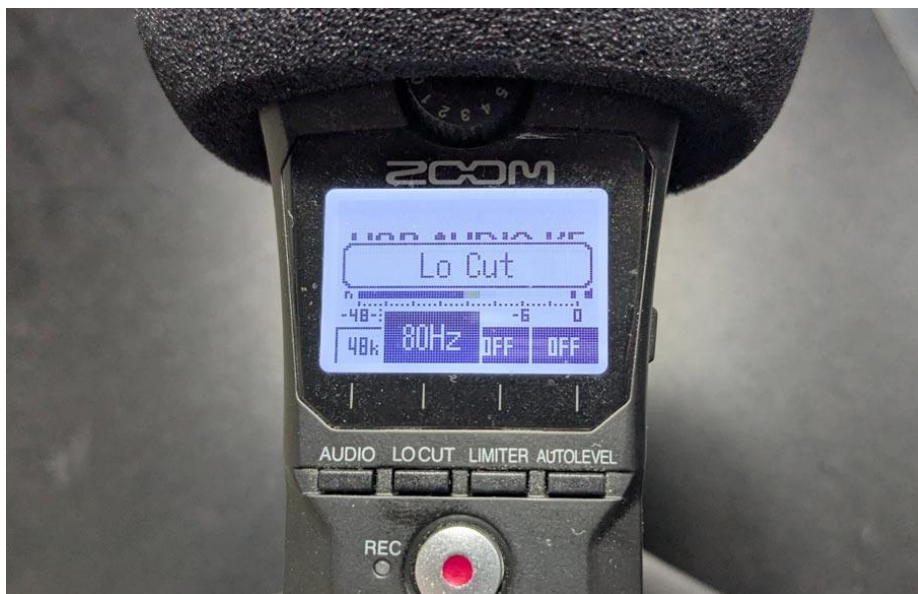


Σχήμα 2.4. Αποκρίσεις συχνότητας ιδανικών φίλτρων
(Πηγή: Ζαρούχας et al, 2013, σ.35, Εικόνα 1.8, επεξεργασμένη)

Παράδειγμα χαμηλοδιαβατού φίλτρου, που χρησιμοποιείται κατά την διαδικασία ψηφιοποίησης του ήχου, αποτελεί η περίπτωση του **αντι-αναδιπλωτικού** φίλτρου (**Anti-aliasing Filter**). Αυτό αποκόπτει τις συχνότητες του αναλογικού σήματος που υπερβαίνουν τα **22050 Hz**, που είναι το μισό της πρότυπης συχνότητας δειγματοληψίας ποιότητας CD, η οποία διατηρείται στα **44100 Hz** στις περισσότερες περιπτώσεις, πριν από την μετατροπή του σήματος σε ψηφιακό.

Ανάλογο φίλτρο, το οποίο ονομάζεται φίλτρο **ανακατασκευής ή εξομάλυνσης** (**Reconstruction / Smoothing Filter**), χρησιμοποιείται για την διαδικασία μετατροπής από ψηφιακό σε αναλογικό, για να αποφευχθούν ανεπιθύμητες παραμορφώσεις όπως είναι η ενδοδιαμόρφωση, που προκαλείται σε υποσυστήματα μετάδοσης του ήχου, με παράδειγμα τα κυκλώματα των προενισχυτών και των ενισχυτών (Ζαρούχας et al, 2013).

Στην Εικόνα 2.14 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από παράδειγμα υπερβατού φίλτρου αποκοπής χαμηλών συχνοτήτων (**Lo Cut**), για περιορισμό του θορύβου περιβάλλοντος, το οποίο εφαρμόζεται στο δημοσιογραφικό μικρόφωνο **H1n** της εταιρίας **ZOOM** για την βελτίωση της ποιότητας της ηχογράφησης.

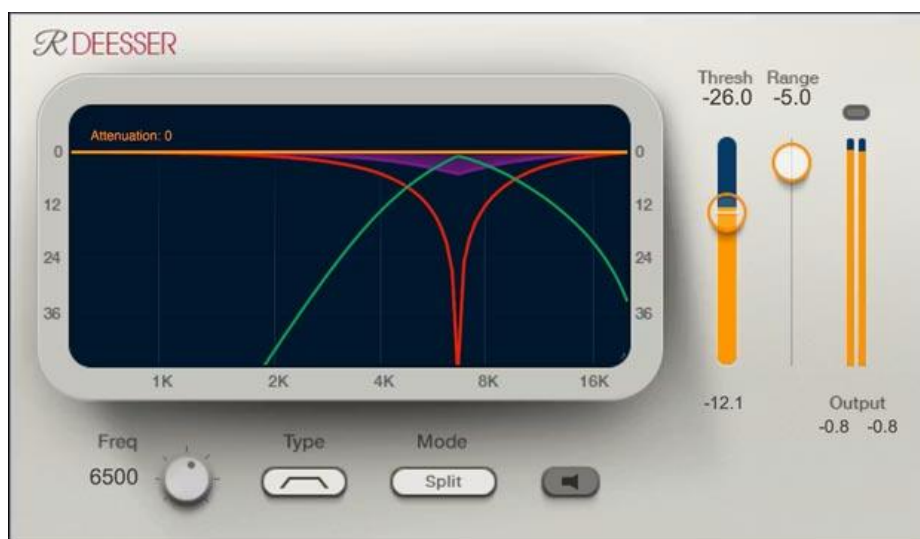


Εικόνα 2.14. Εφαρμογή φίλτρου αποκοπής χαμηλών συχνοτήτων στο μικρόφωνο ZOOM H1n

Στην εικονιζόμενη περίπτωση, επιλέχθηκε να αποκοπούν οι συχνότητες κάτω από **80 Hz**, καθώς η ανθρώπινη ομιλία έχει ωφέλιμο φασματικό περιεχόμενο πάνω από αυτή την συχνότητα.

Στην προσπάθεια απομόνωσης ανεπιθύμητων ήχων, οι οποίοι έχουν φασματικό περιεχόμενο που εκτείνεται σε μία πολύ στενή ζώνη συχνοτήτων, ιδιαίτερη εφαρμογή έχει μία κατηγορία φίλτρων αποκοπής ζώνης, τα οποία ονομάζονται φίλτρα **εγκοπής (Notch Filters)**. Χαρακτηριστική περίπτωση εφαρμογής των φίλτρων εγκοπής, αποτελεί η χρήση τους για την μείωση των έντονων συριστικών συμφώνων, όπως είναι τα «σ», «ξ» και «ψ», γνωστή και ως **αποσυρισμός (De-Essing)**, σε μία ηχογράφιση φωνής.

Στην Εικόνα 2.15 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από την οθόνη ορισμού παραμέτρων του φίλτρου εγκοπής «**Renaissance DeEsser**» της εταιρείας **Waves** (<https://www.waves.com/plugins/renaissance-deesser>), το οποίο είναι διαθέσιμο σαν λογισμικό με την μορφή προσθέτου, για τον σκοπό αυτό.



Εικόνα 2.15. Οθόνη φίλτρου εγκοπής του προσθέτου **Renaissance DeEsser** της **Waves**
(Πηγή: <https://www.waves.com/plugins/renaissance-deesser>)

Φίλτρο εγκοπής μπορεί να εφαρμοστεί και στην περίπτωση αρμονικών, θορύβων όπως είναι η βόμβος του ηλεκτρικού δικτύου (**mains hum**). Στην περίπτωση αυτή, παρά το γεγονός ότι η θεμελιώδης συχνότητα του θορύβου, (που είναι **50 Hz** για το ελληνικό δίκτυο), μπορεί να αποκοπεί με την χρήση ενός υπερυπερατού φίλτρου, υπάρχουν αρμονικές στα **100Hz**, στα **150Hz** ή και υψηλότερα, οι οποίες συχνотικά συμπίπτουν με το ωφέλιμο ηχητικό σήμα. Με την χρήση των φίλτρων εγκοπής, μειώνεται αισθητά η επίδραση των αρμονικών, συμβάλλοντας στην αποθορυβοποίηση του ήχου (Audacity, n.d.).

2.10 Ισοστάθμιση (EQ)

Η μέθοδος της ισοστάθμισης, είναι μία ακόμα τεχνική η οποία εφαρμόζεται για την αποθορυβοποίηση του ήχου. Χρησιμοποιείται για την «διόρθωση» των παραμορφώσεων που προκαλούνται από την απόκριση των συστημάτων που εμπλέκονται στην δημιουργία, μετάδοση και αναπαραγωγή του ηχητικού σήματος.

Κύρια διαφορά της ισοστάθμισης από την εφαρμογή των φίλτρων, είναι ότι σε αντίθεση με εκείνα, δεν μεταβάλλεται το φασματικό περιεχόμενο της εισόδου, ανάλογα με την συχνότητα αποκοπής του φίλτρου, αλλά εφαρμόζεται ενίσχυση ή εξασθένηση του σήματος σε συγκεκριμένες περιοχές συχνοτήτων, αφήνοντας τις υπόλοιπες περιοχές ανεπηρέαστες (Ζαρούχας et al, 2013).

Αρχικά η μέθοδος της ισοστάθμισης υλοποιήθηκε με την δημιουργία ειδικών αναλογικών συσκευών, οι οποίες ονομάζονται «**αναλογικοί ισοσταθμιστές**» (**Equalizers**). Οι ισοσταθμιστές αυτοί επιδίωκαν να αναιρέσουν τις γραμμικές παραμορφώσεις που προκαλούνται στο πεδίο της συχνότητας, με την χρήση κατάλληλων αναλογικών φίλτρων, τα οποία ρυθμίζονται εμπειρικά.

Παρατηρήθηκε όμως ότι τα φίλτρα αυτά δημιουργούσαν παραμορφώσεις που εισάγονταν με την μεταβολή του πλάτους των σημάτων, και τις καθυστερήσεις στην φάση των επιμέρους ζωνών συχνοτήτων που ρύθμιζαν. Το πρόβλημα αυτό αντιμετωπίστηκε σε μεγάλο βαθμό με την έλευση της ψηφιακής ισοστάθμισης, καθιστώντας την διόρθωση των παραμορφώσεων περισσότερο εφικτή σε σχέση με την ισοστάθμιση που γίνονταν αναλογικά. (Ζαρούχας et al, 2013).

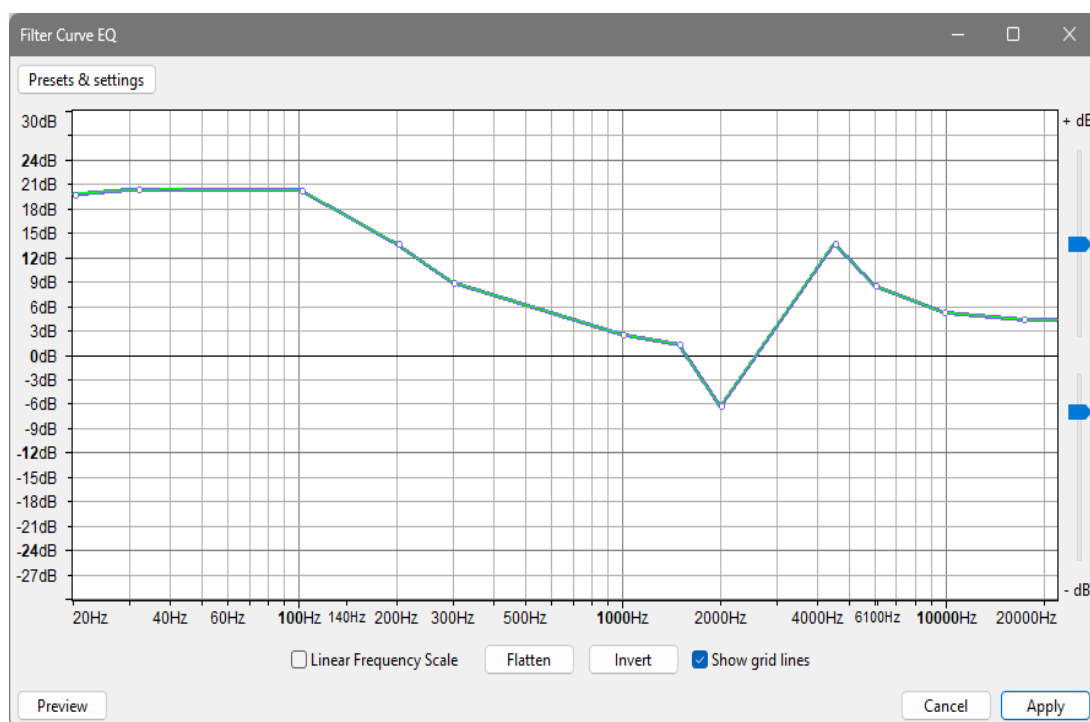
Η ψηφιακή ισοστάθμιση μπορεί να διαχωριστεί σε τρεις κατηγορίες: α) την ισοστάθμιση των γραμμικών παραμορφώσεων που εισάγουν οι αποκρίσεις ηλεκτρονικών συστημάτων, όπως είναι οι ενισχυτές, οι προενισχυτές και οι μίκτες, β) την ισοστάθμιση των ηλεκτρακουστικών συστημάτων, όπως είναι τα ηχεία, και γ) την ισοστάθμιση των ακουστικών συστημάτων, όπως είναι οι χώροι στους οποίους γίνεται η ηχογράφιση και η αναπαραγωγή του ήχου (Ζαρούχας et al, 2013).

Με την χρήση ψηφιακής ισοστάθμισης, μπορεί να γίνει ακόμα και σε πραγματικό χρόνο, ο περιορισμός των ηχητικών παραμορφώσεων που εισάγει η επίδραση ενός χώρου

ακρόασης, συμβάλλοντας στην βελτίωση της ποιότητας του ήχου που φτάνει στον ακροατή (Κοντομίχος et al, 2013α).

Η λογική της ψηφιακής ισοστάθμισης βασίζεται στην ιδέα ότι ο σχεδιασμός και εφαρμογή ενός ψηφιακού φίλτρου το οποίο «...θα έχει αντίστροφη συνάρτηση μεταφοράς από αυτή του αρχικού συστήματος ...θα οδηγήσει το συνολικό σύστημα να αποκριθεί με ιδανικό τρόπο (Ζαρούχας et al, 2013)». Η συνάρτηση αυτή περιγράφεται σαν «**καμπύλη εξισορρόπησης**» και μπορεί να σχεδιαστεί γραφικά με την βοήθεια εργαλείων που ενσωματώνονται στα σύγχρονα DAWs.

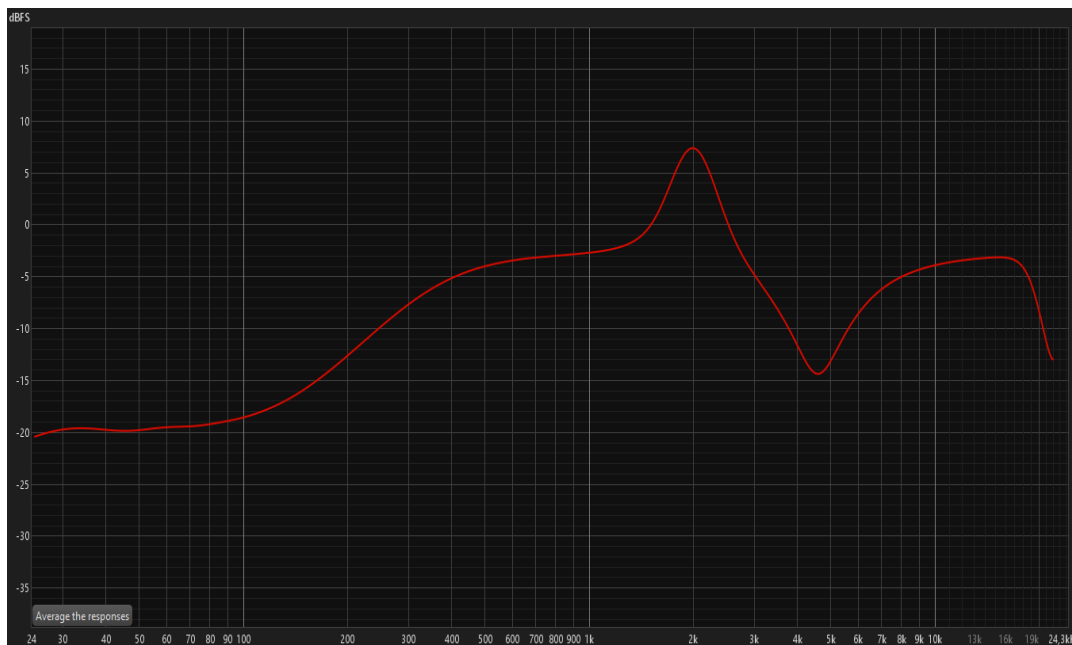
Στην Εικόνα 2.16 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από την χρήση του ισοσταθμιστή καμπύλης εξισορρόπησης (**Curve EQ**) στο **Audacity**.



Εικόνα 2.16. Στιγμιότυπο οθόνης κατά την χρήση του Curve EQ στο Audacity

Η συγκεκριμένη ισοστάθμιση, που περιλαμβάνεται εδώ σαν παράδειγμα, εφαρμόστηκε για να αφαιρεθεί ο θόρυβος από την κρουστική παραμόρφωση που προκαλείται σε σήμα, του οποίου η συνάρτηση μεταφοράς απεικονίζεται στην Εικόνα 2.17, που ακολουθεί στην επόμενη σελίδα. Η λήψη του στιγμιότυπου του σήματος,

πραγματοποιήθηκε μέσα από την χρήση του εξειδικευμένου λογισμικού REW (<https://www.roomeqwizard.com/>).



Εικόνα 2.17. Συνάρτηση μεταφοράς παραμορφωμένου σήματος στο REW

Όπως είναι εμφανές από την αντιπαραβολή των δύο προηγούμενων εικόνων, στόχος της επιβολής της ισοστάθμισης ήταν το τελικό σήμα να αποκτήσει ξανά γραμμική φασματική συμπεριφορά πλάτους **0 dB**, εξαλείφοντας έτσι την όποια παραμόρφωση είχε εισαχθεί σε αυτό.

2.11 Ακύρωση θορύβου (Noise Cancelation - NC)

Οι μέθοδοι και τεχνικές, που αναφέρθηκαν μέχρι αυτό το σημείο, έχουν σαν κοινό τους στόχο την εξάλειψη του θορύβου. Σε αυτές αναφέρεται ότι με την χρήση τους ο θόρυβος ακυρώνεται σε μικρότερο ή μεγαλύτερο βαθμό. Ο καθεαυτός όμως όρος «**ακύρωση θορύβου**» (**Noise Cancelation**), αφορά σε συγκεκριμένη τεχνική αποθορυβοποίησης, η οποία εφαρμόζεται στην περίπτωση που το ηχητικό σήμα μεταδίδεται σε πραγματικό χρόνο. Ανάλογα με τον τρόπο υλοποίησης της, υπάρχουν δύο

κατηγορίες ακύρωσης θορύβου, η **παθητική ακύρωση θορύβου (Passive Noise Cancellation)** και η **ενεργή ακύρωση θορύβου (Active Noise Cancellation - ANC)**.

Η παθητική ακύρωση, η οποία ονομάζεται διαφορετικά και **απομόνωση θορύβου (Noise Isolation)**, εστιάζει στην αντιμετώπιση των πηγών θορύβου με υλικά τα οποία έχουν απορροφητικό ή / και ηχομονωτικό χαρακτήρα. Οτιδήποτε εμποδίζει κατασκευαστικά τον θόρυβο να παρεισφρήσει στο σήμα μπορεί να χρησιμοποιηθεί για τον σκοπό αυτό.

Έτσι στην περίπτωση που η ακύρωση εφαρμόζεται σε ένα χώρο όπου γίνεται η ακρόαση, αυτή μπορεί να περιλαμβάνει ενίσχυση τοιχωμάτων και ανοιγμάτων κουφωμάτων με χρήση με διπλών υαλοπινάκων και ηχομονωμένων θυρών, σιγαστήρες, αποσβεστήρες ταλαντώσεων και ειδικούς κλωβούς για απομόνωση μηχανημάτων.

Με ανάλογο τρόπο, στην περίπτωση της χρήσης ακουστικών, η παθητική ακύρωση του θορύβου συντελείται με την ενίσχυση του κελύφους των ακουστικών και την χρήση σπογγωδών υλικών, τα οποία απομονώνουν τα αυτιά του ακροατή από τους εξωτερικούς θορύβους. Λόγω της κατασκευής τους, ιδιαίτερη συμβολή στην περίπτωση της παθητικής ακύρωσης του θορύβου παρέχουν τα ακουστικά κλειστού τύπου (over ear phones), τα οποία καλύπτουν εντελώς τα αυτιά του ακροατή.

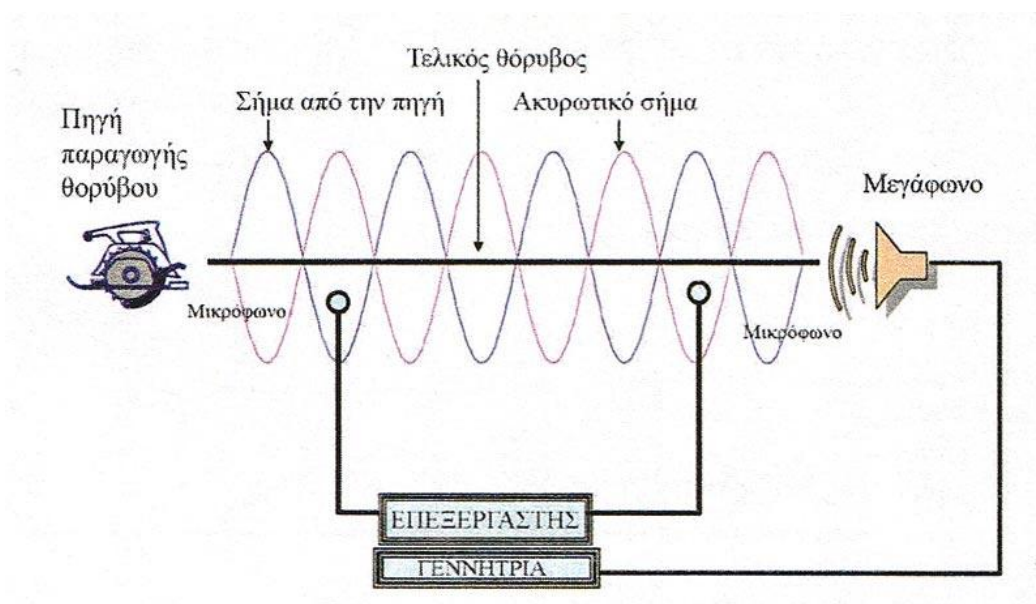
Η αποτελεσματικότητα των απορροφητικών επιφανειών είναι μεγαλύτερη στις υψηλές συχνότητες, καθώς εκεί συντελείται μετατροπή μεγαλύτερου ποσοστού της μηχανικής ενέργειας του ήχου σε θερμική, εξαιτίας της επαφής του ηχητικού κύματος με την επιφάνεια των μέσων (Κοντομίχος et al, 2013α). Έτσι η παθητική ακύρωση θορύβου είναι καταλληλότερη για καταστολή θορύβων με υψίσυχο περιεχόμενο.

Στην περίπτωση της ενεργής ακύρωσης θορύβου (ANC), η αντιμετώπιση είναι διαφορετική. Με την μέθοδο αυτή, η οποία είναι γνωστή από την δεκαετία του 1930, έχει σαν βασική αρχή της, την εξουδετέρωση που συμβαίνει σε ένα σήμα, «...από την συμβολή δύο όμοιων ήχων που έχουν διαφορά φάσης 180 μοίρες (Σκαρλάτος, 2013β)».

Η λειτουργία της μεθόδου υλοποιείται με την χρήση ειδικής διάταξης, η οποία σε πρώτη φάση περιλαμβάνει ένα μικρόφωνο (ή και περισσότερα), το οποίο «συλλαμβάνει» τον εξωτερικό θόρυβο. Στην συνέχεια ο θόρυβος οδηγείται σε ειδικό επεξεργαστή, ο οποίος αναλύει την συχνότητα και το πλάτος του θορύβου και στην συνέχεια με την χρήση ειδικής

γεννήτριας ακουστικών συχνοτήτων, δημιουργείται ένα κατά το δυνατόν πανομοιότυπο σήμα με αυτόν, σε αναστραμμένη φάση (διαφορά φάσης 180°).

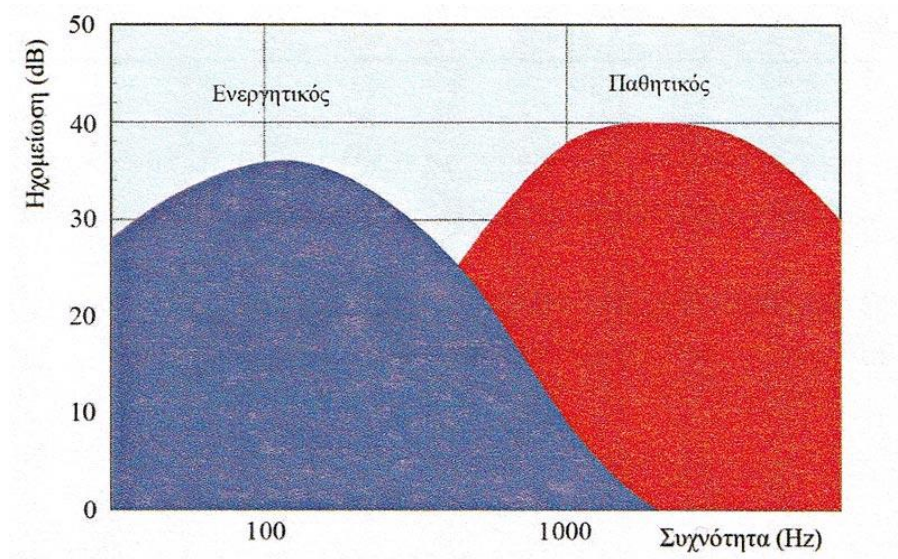
Το ανεστραμμένο αυτό σήμα, στην περίπτωση που η ακρόαση γίνεται σε κάποιο χώρο, εκπέμπεται από ειδικά τοποθετημένο μεγάφωνο (ή διάταξη μεγαφώνων) και προστίθεται ηχητικά στο αρχικό σήμα θορύβου, μηδενίζοντάς το -στην ιδανική περίπτωση- και κατά συνέπεια ακυρώνοντάς το με τον τρόπο αυτό. Στην περίπτωση που η ακρόαση γίνεται με την χρήση ακουστικών το μικρόφωνο είναι συνήθως προσαρμοσμένο πάνω τους και η ακύρωση συντελείται πριν ο ήχος μεταδοθεί από τα ακουστικά. Η διάταξη που χρησιμοποιείται για την ενεργή ακύρωση θορύβου αποδίδεται στο Σχήμα 2.5 που ακολουθεί.



Σχήμα 2.5. Μείωση θορύβου με τεχνική ενεργητικής ακύρωσης (ANC)

(Πηγή: Σκαρλάτος, 2013β, σ.159, Εικόνα 1.23)

Λόγω του ότι η ενεργητική ακύρωση θορύβου επιτυγχάνει καλές επιδόσεις στις χαμηλές συχνότητες, στην πράξη όταν πρέπει να εξουδετερωθεί θόρυβος που εκτείνεται σε ευρεία περιοχή συχνοτήτων, χρησιμοποιούνται ταυτόχρονα υβριδικές υλοποιήσεις και των δύο τρόπων (Σκαρλάτος, 2013β), όπως φαίνεται στο σχήμα 2.6, που ακολουθεί στην επόμενη σελίδα.



Σχήμα 2.6. Συχνότητες κάλυψης υβριδικής υλοποίησης για ακύρωση θορύβου ευρέως φάσματος
(Πηγή: Σκαρλάτος, 2013β, σ.153, Εικόνα 1.26)

Με τον τρόπο αυτό, επιτυγχάνεται η αντιμετώπιση του θορύβου σε ευρύτερο φάσμα συχνοτήτων, καθιστώντας την μέθοδο περισσότερο αποτελεσματική.

3. Χρήση Ψηφιακών Εργαλείων Λογισμικού για Αφαίρεση Θορύβου

Με την μετάβαση από την αναλογική στην ψηφιακή επεξεργασία του ήχου, αναπτύχθηκε μία ολόκληρη κατηγορία εργαλείων λογισμικού, που σκοπό της είχε την παραγωγή όσο το δυνατόν αρτιότερων ηχητικών αποτελεσμάτων. Η διαρκής βελτίωση των επιδόσεων των υπολογιστικών συστημάτων στα οποία βρίσκονται εγκατεστημένα, συνέβαλε στην μεγάλη διάδοση των λογισμικών αυτών, τα οποία πλέον χρησιμοποιούνται ευρύτατα. Με την ψηφιακή επεξεργασία, μπορούμε να αφαιρέσουμε τον θόρυβο και να εξαλείψουμε παραμορφώσεις, κάτι που ήταν «...πολύ δύσκολο έως ακατόρθωτο μερικά χρόνια πριν (Ζαρούχας et al, 2013)».

Εκτός από τις εφαρμογές λογισμικού, οι οποίες παραδοσιακά εκτελούνταν με τοπικές εγκαταστάσεις και λειτουργούσαν χωρίς την ανάγκη σύνδεσης με το διαδίκτυο, εδώ και αρκετό καιρό, έχουν βρει χρήση και αρχίζουν να διαδίδονται με ολοένα αυξημένο ρυθμό, αντίστοιχες διαδικτυακές εφαρμογές, οι οποίες έχουν σαν κύρια πλεονεκτήματά τους, την κατάργηση της απαίτησης για κατοχή υπολογιστικού εξοπλισμού από τον χρήστη και την ανεξαρτησία τους από συγκεκριμένα λειτουργικά συστήματα.

Από την άλλη πλευρά, οι τοπικές υλοποιήσεις, παρέχουν μεγαλύτερη ασφάλεια, καθώς τα ηχητικά δεδομένα προς επεξεργασία δεν βρίσκονται στο διαδίκτυο, ενώ προσφέρουν ανεξαρτησία, καθώς η λειτουργία τους δεν απαιτεί ο χρήστης να είναι συνδεδεμένος με αυτό.

Οι βασικές κατηγορίες λογισμικών, που χρησιμοποιούνται για την επεξεργασία του ήχου και κατ' επέκταση για την αφαίρεση του θορύβου, καταγράφονται στην συνέχεια.

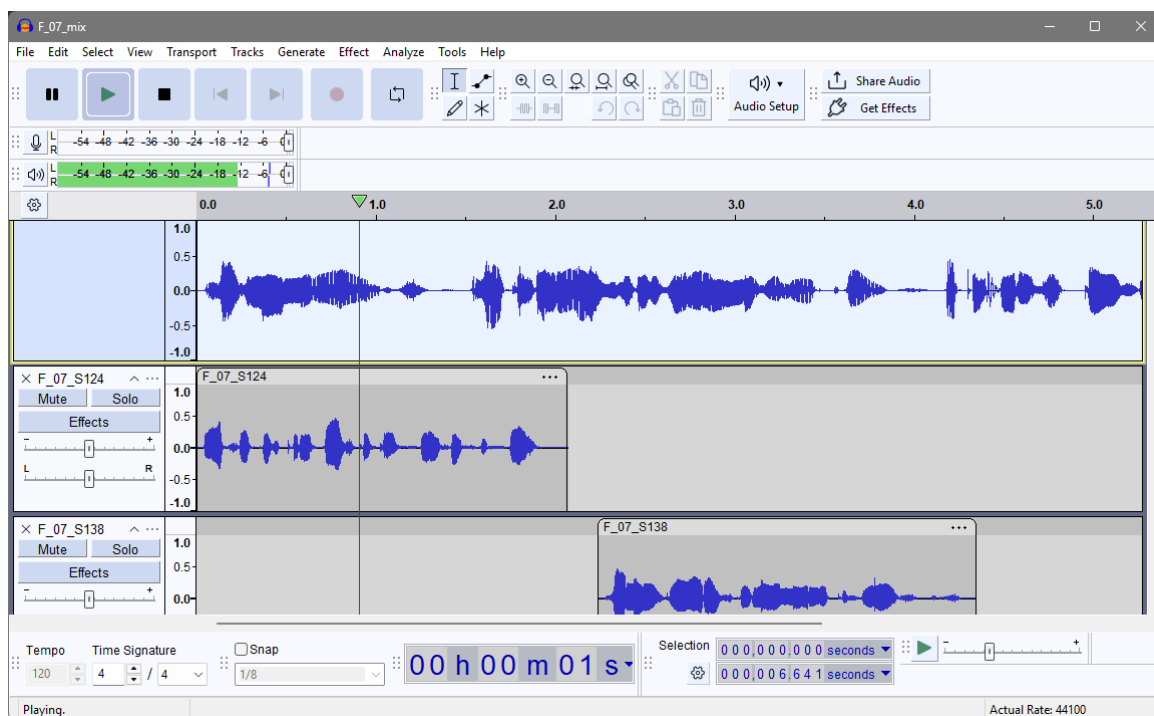
3.1 Σταθμοί επεξεργασίας ψηφιακού ήχου (Digital Audio Workstations - DAW)

Η ευρεία εξέλιξη και διάδοση των ψηφιακών συστημάτων, έχει σηματοδοτήσει την μετάβαση από την διαχείριση και επεξεργασία των ηχητικών δεδομένων με συστήματα που

βασίζονταν σε εξειδικευμένο υλικό, στην χρήση λογισμικού για τις ίδιες εργασίες, καθιστώντας τον υπολογιστή σταθμό επεξεργασίας του ήχου (**Digital Audio Workstation - DAW**), με ουσιαστικά απεριόριστες δυνατότητες επέμβασης στον ήχο (Φλώρος, 2013).

Σύμφωνα με τον ορισμό του, ένας σταθμός ψηφιακής επεξεργασίας ήχου αποτελείται από ένα σύνολο ηλεκτρονικών συσκευών και λογισμικών, τα οποία χρησιμοποιούνται για ηχογράφηση, επεξεργασία και αναπαραγωγή μουσικής (Ζαρούχας et al, 2013). Ο ορισμός αυτός μπορεί να επεκταθεί, συμπεριλαμβάνοντας διεργασίες, όπως αυτή της καταστολής του θορύβου σε ένα ηχητικό σήμα.

Οι σταθμοί ψηφιακής επεξεργασίας ήχου, διαθέτουν ένα σύνολο εργαλείων, με τα οποία μπορούν να πραγματοποιηθούν οι επιθυμητές επεξεργασίες, μέσα από γραφικά περιβάλλοντα με οπτικό τρόπο. Στην Εικόνα 3.1 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από την οθόνη εργασίας, του σταθμού ψηφιακής επεξεργασίας ήχου **Audacity**.



Εικόνα 3.1. Στιγμιότυπο οθόνης εργασίας του σταθμού επεξεργασίας ήχου Audacity

Τα τελευταία χρόνια, έχουν εμφανιστεί αρκετές περιπτώσεις σταθμών επεξεργασίας ψηφιακού ήχου, οι οποίες είναι διαθέσιμες για online χρήση, μέσω διαδικτυακών πλατφορμών. Αντιπροσωπευτικά δείγματα τέτοιων εργαλείων, αποτελούν τα: **Amped Studio**, **Audiosauna Studio**, **Audiotool**, **BandLab**, **openDAW**, **Soundation**, **Soundtrap**

και **Viktor NV-1**. (Οι διευθύνσεις των ιστοτόπων των εργαλείων αυτών, συμπεριλαμβάνονται στον Πίνακα 3.1 που ακολουθεί στην συνέχεια).

Οι online πλατφόρμες, είναι ιδανικές για παραγωγή μουσικής σε υπολογιστικά συστήματα χαμηλής επεξεργαστικής ισχύος, όπως παλαιότερα λάπτοπ ή ακόμα και κινητές συσκευές (Mendenhall, 2026). Ακόμα μπορούν να χρησιμοποιηθούν χωρίς την ανάγκη να αγορασθεί μία λύση βασισμένη σε συγκεκριμένη πλατφόρμα (Vanacoro, 2020). Αυτά καθιστούν την επιλογή των online πλατφορμών δελεαστική, ιδιαίτερα για όσους θέλουν να έρθουν για πρώτη φορά σε επαφή με την επεξεργασία ήχου (Albanese, 2026).

Στον Πίνακα 3.1, περιλαμβάνονται με αλφαβητική σειρά, διαδεδομένα offline και online λογισμικά επεξεργασίας ψηφιακού ήχου (DAW), με τους αντίστοιχους συνδέσμους από τους οποίους αυτά είναι διαθέσιμα. (Ο τελευταίος έλεγχος ορθής λειτουργίας των συνδέσμων, πραγματοποιήθηκε στις 28/05/2026).

Πίνακας 3.1. Πίνακας διαδεδομένων λογισμικών επεξεργασίας ψηφιακού ήχου (DAW)

Όνομα	Σύνδεσμος
Amped Studio	https://ampedstudio.com/
Ardour	https://ardour.org/
Audacity	https://www.audacityteam.org
Audiosauna	https://www.audiosauna.com/studio/
Audiotool	https://www.audiotool.com/
Audition (Adobe)	https://www.adobe.com/gr_en/products/audition.html
BandLab	https://www.bandlab.com
Bitwig	https://www.bitwig.com/
Cubase (Steinberg)	https://www.steinberg.net/cubase/
Digital Performer (Motu)	https://motu.com/en-us/products/software/dp/
FL Studio (Image-Line)	https://www.image-line.com/
Garage Band (Apple)	https://www.apple.com/mac/garageband/
Live (Ableton)	https://www.ableton.com/
LMMS	https://github.com/LMMS/

Logic Pro (Apple)	https://www.apple.com/logic-pro/
Luna (Universal Audio)	https://www.uaudio.com/products/luna
mixcraft (Acoustica)	https://acoustica.com/products/mixcraft
MuLab (MuTools)	https://www.mutools.com/
Nuendo (Steinberg)	https://www.steinberg.net/nuendo/
openDAW	https://opendaw.org/
Pro Tools (Avid)	https://www.avid.com/pro-tools
Reaper (Cockos)	https://www.reaper.fm/
Reason (Propellerhead)	https://www.reasonstudios.com/reason
Serato	https://serato.com/studio
Sonar (Cakewalk)	https://www.cakewalk.com/sonar
Soundation	https://soundation.com/
Soundtrap	https://www.soundtrap.com/musicmakers
Studio Pro (Fender)	https://eu.fender.com/pages/fender-studio-pro
Victor NV-1	https://microto.github.io/viktor/
Waveform Free (Tracktion)	https://www.tracktion.com/products/waveform-free
Wavelab (Steinberg)	https://www.steinberg.net/wavelab/

Παρά το γεγονός, ότι οι τοπικές εγκαταστάσεις εξακολουθούν να είναι η κύρια επιλογή για επαγγελματική χρήση, δυνατότητες όπως η συνεργασία ανάμεσα σε δημιουργούς που βρίσκονται σε εντελώς διαφορετικούς χώρους, η δημιουργία κοινοτήτων για διαφήμιση και η προώθηση σε διαδικτυακά κανάλια αναπαραγωγής και διάθεσης μουσικής, η αποθήκευση των προϊόντων, καθώς και η πρόσβαση σε διαθέσιμες βάσεις ηχητικών δεδομένων, οδηγούν στην δημιουργία υβριδικών υλοποιήσεων, όπου η διασύνδεση των τοπικά εγκατεστημένων εφαρμογών με το διαδίκτυο, αποτελεί βασικό τμήμα της λειτουργίας των λογισμικών.

3.2 Πρόσθετα λογισμικών επεξεργασίας ψηφιακού ήχου (Plugins)

Η μετατροπή του υπολογιστή σε σταθμό επεξεργασίας ψηφιακού ήχου, δημιούργησε την ανάγκη ενσωμάτωσης των διάφορων εφέ των μουσικών οργάνων και άλλων μονάδων επεξεργασίας, όπως είναι τα διάφορα φίλτρα, των οποίων η υλοποίηση γίνονταν πριν με φυσικό εξοπλισμό (hardware). Τα παραπάνω, όπως και οι διεπαφές των υποσυστημάτων της μουσικής παραγωγής, άρχισαν πλέον να υλοποιούνται σαν ιδεατές επεξεργασίες (**virtual effect processors**) μέσω λογισμικού (Steinberg, n.d.), το οποίο ενσωματώθηκε στους σταθμούς επεξεργασίας (DAW) με την μορφή προσθέτων (plugins).

Για τα πρόσθετα αυτά δημιουργήθηκαν συγκεκριμένα πρότυπα (formats), ανάλογα με την πλατφόρμα για την οποία προορίζονταν, ώστε να είναι συμβατά τόσο με τα διάφορα διαθέσιμα λειτουργικά συστήματα, όσο και με το συμπληρωματικό υλικό στο οποίο βρίσκουν εφαρμογή, έτσι ώστε να ανταποκρίνονται στο είδος χρήσης τους, ανάλογα με το αν προορίζονται ή όχι για επεξεργασία σε πραγματικό χρόνο (real-time). Μερικά από τα γνωστότερα πρότυπα, αποτελούν ενδεικτικά τα **VST**, **VST2**, **VST3**, **AU**, **AAX**, **AAX DSP**, **AAX Native**, **ARA** και **AU ARA**.

Το πρότυπο **VST (Virtual Studio Technology)**, δημιουργήθηκε από τη εταιρία **Steinberg** (<https://www.steinberg.net/>), το 1996. Τα πρόσθετα αυτής της μορφής, περιλάμβαναν λογισμικό εξομοίωσης ακουστικών εφέ και μουσικών οργάνων, ενώ επηρέασαν σε μεγάλο βαθμό τον τρόπο με τον οποίο η μουσική παράγεται σήμερα (Zlatic, 2026). Η εξέλιξη του προτύπου οδήγησε το 1999 στην έκδοση **VST2**, ενώ η πιο πρόσφατη έκδοση **VST3** υλοποιήθηκε το 2008, εισάγοντας σημαντικές βελτιώσεις στην επεξεργασία, όπως είναι η καλύτερη αξιοποίηση υπολογιστικών πόρων και η περαιτέρω αυτοματοποίηση της μουσικής παραγωγής (Alessi, 2024).

Τα πρόσθετα **AU (Audio Unit)**, αποτελούν πρότυπο υλοποίησης της εταιρίας **Apple** (<https://www.apple.com/>) και σχεδιάστηκαν για να συνεργάζονται με λογισμικά όπως είναι το **GarageBand** ή το **Logic**, που εκτελούνται στα δικά της λειτουργικά συστήματα. Τα πρόσθετα AU, λειτουργούν πλέον και με άλλα λογισμικά DAW, όπως για παράδειγμα το Audacity, αλλά προορίζονται κυρίως για χρήση στα συστήματα Mac της εταιρίας (Alessi, 2024).

Το πρότυπο **AAX (Avid Audio eXtension)**, δημιουργήθηκε από την εταιρεία **Avid** (<https://www.avid.com/>). Αποτελεί ενοποίηση δύο παραλλαγών του προτύπου, το **AAX DSP** και το **AAX Native**. Η υλοποίηση αυτή προέκυψε όταν η εταιρεία δημιούργησε την έκδοση 64 bit του λογισμικού **Pro Tools**, για να καλύψει την ανάγκη επεξεργασίας με την αρχιτεκτονική των 64 bit (Greatbatch, 2026).

Ένας μεγάλος αριθμός προσθέτων, έχει αναπτυχθεί ειδικά για την επεξεργασία του ήχου πάνω στο θέμα της αποθορυβοποίησης. Για το σκοπό αυτό υπάρχουν διαθέσιμα εξειδικευμένα πρόσθετα, όπως ενδεικτικά αναφέρεται το plugin «**DeNoise 2**», της εταιρείας **Acon Digital** (<https://acondigital.com/>), οθόνη από την χρήση του οποίου φαίνεται στην Εικόνα 3.2 που ακολουθεί.



Εικόνα 3.2. Στιγμιότυπο οθόνης χρήσης του προσθέτου DeNoise 2 της Acon Digital
(Πηγή: <https://acondigital.com/products/restoration-suite>)

Το λογισμικό αυτό περιλαμβάνεται σε μία ομάδα προσθέτων αποθορυβοποίησης, τα οποία διατίθενται με την εμπορική ονομασία «**Restoration Suite 2**». Σε αυτήν υπάρχουν συνολικά τέσσερα εξειδικευμένα πρόσθετα αποθορυβοποίησης με τις ονομασίες: «**DeNoise 2**», «**DeHum 2**», «**DeClick 2**», και «**DeClip 2**».

Το DeNoise υλοποιεί ειδικούς αλγορίθμους, σχεδιασμένους έτσι ώστε να ανταποκριθούν σε δυναμικά μεταβαλλόμενους θορύβους, όπως μπορεί να είναι ο θόρυβος του αέρα που αλλάζει χρονικά, ή ο θόρυβος που μπορεί να προκληθεί από ένα μικρόφωνο πέτου (lavalier ή lav microphone) κατά την επαφή του με τα ρούχα (Acon Digital, n.d.).

Στο σημείο αυτό πρέπει να αναφερθεί, ότι η υλοποίηση έξυπνων αλγορίθμων σε πρόσθετα όπως το παραπάνω, δεν πρέπει να δημιουργεί την εντύπωση ότι το λογισμικό κάνει απαραίτητα χρήση Τεχνητής Νοημοσύνης. Έτσι περιπτώσεις όπως και η προηγούμενη, αποτελούν συμβατικά εργαλεία αποθορυβοποίησης. Τα τελευταία χρόνια βέβαια, έχουν εμφανιστεί αρκετές περιπτώσεις προσθέτων αποθορυβοποίησης, τα οποία ενσωματώνουν πλήρως την χρήση ΤΝ στην λειτουργία τους.

Στην Εικόνα 3.3 που ακολουθεί, εμφανίζεται ενδεικτικό στιγμιότυπο από την οθόνη λειτουργίας ενός τέτοιου προσθέτου. Πρόκειται για το plugin «**Clarity Vx**», της εταιρείας **Waves** (<https://www.waves.com/>).

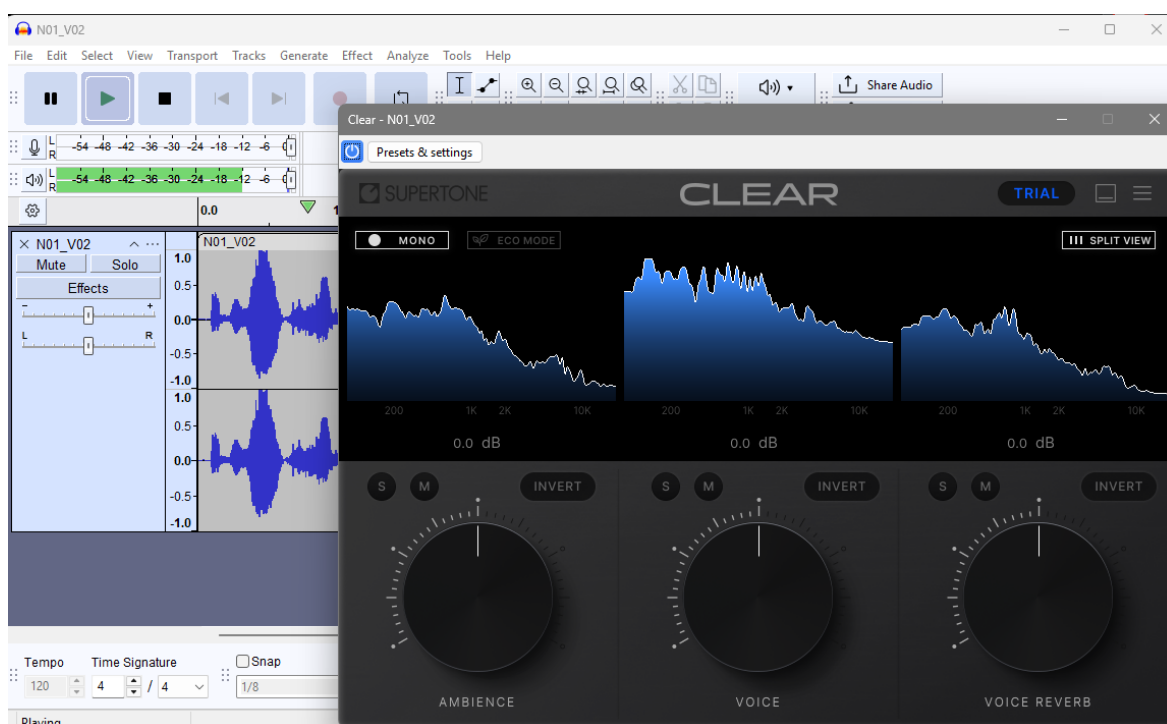


Εικόνα 3.3. Στιγμιότυπο οθόνης χρήσης του προσθέτου Clarity Vx της Waves
(Πηγή: <https://www.waves.com/plugins/clarity-vx>)

Το πρόσθετο αυτό, όπως αναφέρεται στον ιστότοπο της εταιρείας (<https://www.waves.com/plugins/clarity-vx>), χρησιμοποιεί Νευρωνικά Δίκτυα, τα οποία έχουν ειδικά εκπαιδευτεί για να αναγνωρίζουν φωνές και να τις διαχωρίζουν από τον θόρυβο. Για τον λόγο αυτό, όπως επίσης αναφέρεται στον ίδιο ιστότοπο, έχουν αναλυθεί εκατομμύρια ωρών ήχου, ώστε τα μοντέλα να μπορούν να ξεχωρίσουν ήχους ομιλίας από

άλλους ήχους, σε πραγματικό χρόνο, χωρίς να χρειάζεται να «μάθουν» το προφίλ του κάθε καναλιού.

Ένα ακόμη διαδεδομένο πρόσθετο αποθορυβοποίησης, το οποίο κάνει χρήση ΤΝ, αποτελεί το «Clear» της εταιρίας **Supertone** (<https://www.supertone.ai/en/clear>). Το λογισμικό αυτό, με άδεια δοκιμαστικής χρήσης, προστέθηκε από τον γράφοντα στο **Audacity**, στο πλαίσιο των εκτενών δοκιμών που πραγματοποιήθηκαν στην διάρκεια υλοποίησης της παρούσας διπλωματικής εργασίας. Στην Εικόνα 3.4 που ακολουθεί, εμφανίζεται ενδεικτικό στιγμιότυπο από την οθόνη ελέγχου του προσθέτου.



Εικόνα 3.4. Στιγμιότυπο οθόνης χρήσης του προσθέτου Clear μέσα από το Audacity

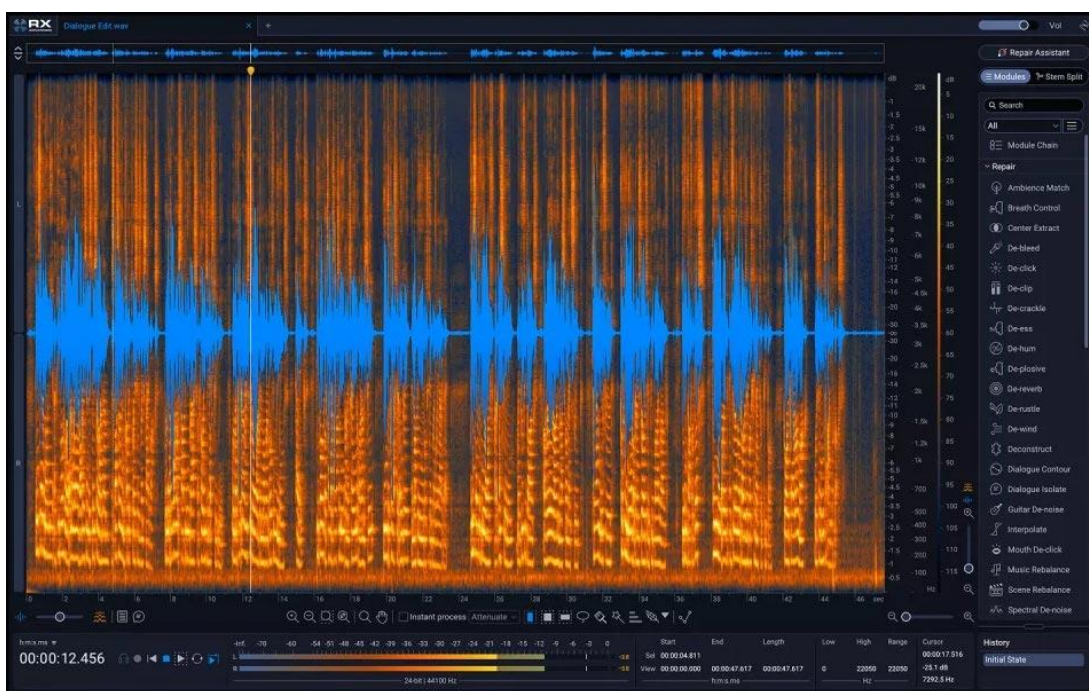
Το Clear, εγκαταστάθηκε στο Audacity με έκδοση του προτύπου VST3, και μπορεί να λειτουργεί με έλεγχο μεταβολής παραμέτρων σε πραγματικό χρόνο.

3.3 Εξειδικευμένα εργαλεία αποθορυβοποίησης

Εκτός από τους σταθμούς επεξεργασίας ψηφιακού ήχου (DAW), των οποίων η χρήση μπορεί καλύψει γενικότερες απαιτήσεις στην μουσική διαδικασία και παραγωγή,

έχουν αναπτυχθεί και επαγγελματικά εργαλεία λογισμικού, που κύριο σκοπό τους έχουν την βελτίωση του ήχου και την απαλλαγή του από τον θόρυβο.

Ενδεικτικό τέτοιο λογισμικό, το οποίο αποτελεί και σημείο αναφοράς στα επαγγελματικά ψηφιακά εργαλεία αποκατάστασης του ήχου (audio restoration software), αποτελεί το «RX», της εταιρείας **Izotope** (<https://www.izotope.com/>), στιγμιότυπο από το περιβάλλον εργασίας του οποίου, φαίνεται στην Εικόνα 3.5 που ακολουθεί.



Εικόνα 3.5. Στιγμιότυπο περιβάλλοντος εργασίας του λογισμικού RX12 Advanced της Izotope (Πηγή: <https://www.izotope.com/en/shop/rx-12-advanced/>)

Το RX διαθέτει μία σειρά από εξειδικευμένα πρόσθετα, όπως είναι τα «RX De-noise», «RX De-reverb», «RX De-clip», «RX De-click», «RX De-hum», και «RX De-ess», τα οποία είναι ενοποιημένα και διαθέσιμα κάτω από μία δέσμη με την ονομασία «Repair Assistant» (<https://www.izotope.com/en/products/rx/features/repair-assistant>), η οποία είναι διαθέσιμη και στην βασική έκδοση «RX Elements» (Metlay, 2022).

Ανάλογο εξειδικευμένο λογισμικό αποθρομβοποίησης και βελτίωσης ήχου, αποτελεί και το «SpectraLayers» της εταιρείας **Steinberg**, το οποίο ήδη έχει αναφερθεί σε προηγούμενο σημείο της παρούσας διπλωματικής εργασίας. Όπως αναφέρεται στον ιστότοπο του προϊόντος (<https://www.steinberg.net/spectralayers/>), το λογισμικό χρησιμοποιεί το πρότυπο ARA 2, για την συνεργασία του με τα υπόλοιπα λογισμικά DAW

της ίδιας αλλά και τρίτων εταιριών, παρέχοντας μη καταστρεπτική οπτική επεξεργασία (non-destructive visual editing) του ήχου που πρέπει να βελτιωθεί. Στον ίδιο ιστότοπο αναφέρεται επίσης, ότι η τρέχουσα έκδοση «**SpectraLayers Pro 12**», χρησιμοποιεί Τεχνητή Νοημοσύνη δεύτερης γενιάς, για να διαχωρίσει αυτόματα την φωνή από τους υπόλοιπους ήχους, σε διαφορετικά κανάλια.

Ο διαχωρισμός της ομιλίας από τους υπόλοιπους ήχους, είναι αντικείμενο της λειτουργίας και πολλών σύγχρονων εξειδικευμένων εφαρμογών, οι οποίες είναι διαθέσιμες μέσω διαδικτύου. Έτσι, εκτός από τα λογισμικά που προσφέρονται για τοπικές εγκαταστάσεις, εμφανίζονται πλέον όλο και συχνότερα online εργαλεία, τα οποία προσφέρονται είτε δωρεάν, είτε έναντι αμοιβής, με την μορφή **SaaS (Software as a Service)**.

Δύο από τα online αυτά εργαλεία, το **Enhance (Adobe Podcast)** και το **LALAL.AI**, μελετήθηκαν διεξοδικά για την αποτελεσματικότητά τους στην αποθορυβοποίηση ψηφιακού ήχου, κατά την διάρκεια του πειραματικού σκέλους της της παρούσας διπλωματικής εργασίας. Η όλη διαδικασία καταγράφεται αναλυτικά σε επόμενο στάδιο, στα κεφάλαια που ακολουθούν.

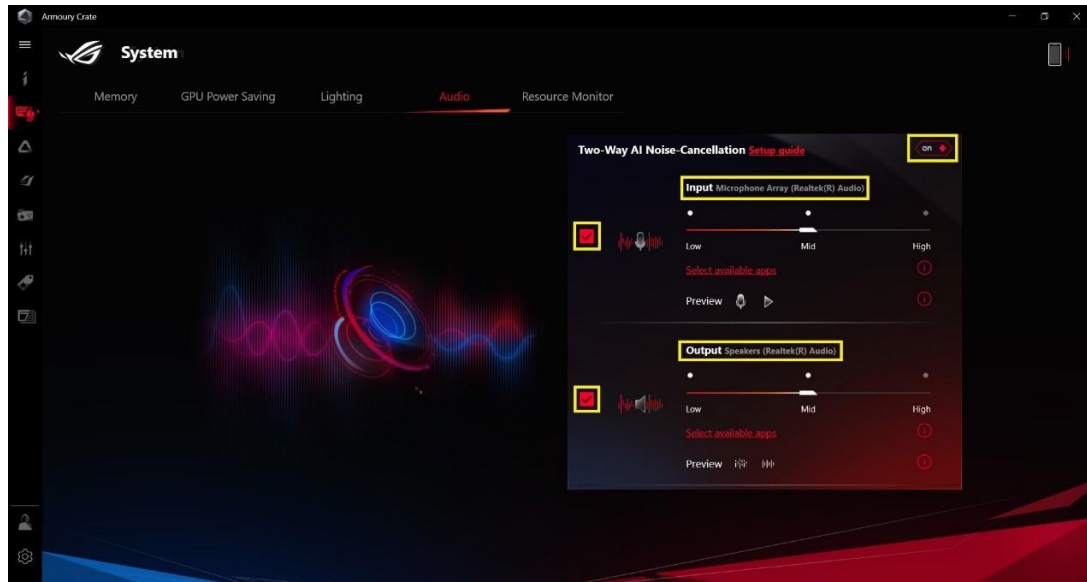
3.4 Υλοποιήσεις «επί τούτω» (Ad Hoc)

Εκτός από τις κατηγορίες που αναφέρθηκαν πριν, πολύ συχνά εμφανίζονται και λογισμικά αποθορυβοποίησης από διάφορες πολύ γνωστές εταιρίες, οι οποίες αποτελούν κατά βάση παραδοσιακούς κατασκευαστές υλικού (hardware), για να βελτιώσουν την ποιότητα στην εμπειρία χρήσης των προϊόντων τους.

Έτσι έχουν αναπτυχθεί βοηθητικά εργαλεία (utilities), τα οποία στοχεύουν στην καταστολή του θορύβου, που εμφανίζεται κατά την χρήση μικροφώνων, ποντικών και πληκτρολογίων, ή από την λειτουργία καρτών γραφικών και άλλων εξαρτημάτων υλικού, όπως για παράδειγμα αποτελούν οι ανεμιστήρες των υπολογιστών που χρησιμοποιούνται.

Στην Εικόνα 3.6 που ακολουθεί, εμφανίζεται ενδεικτικό στιγμιότυπο από την οθόνη λειτουργίας ενός τέτοιου εργαλείου. Πρόκειται για την εφαρμογή καταστολής θορύβου με

το όνομα «**Two-Way AI Noise Cancellation**», που περιλαμβάνεται στο λογισμικό «**Armoury Crate**» της εταιρείας **ASUS** (<https://armoury-crate.com/>).



Εικόνα 3.6. Στιγμιότυπο οθόνης ρύθμισης του εργαλείου Two-Way AI Noise Cancellation της ASUS (Πηγή: <https://www.asus.com/gr/support/faq/1044576/>)

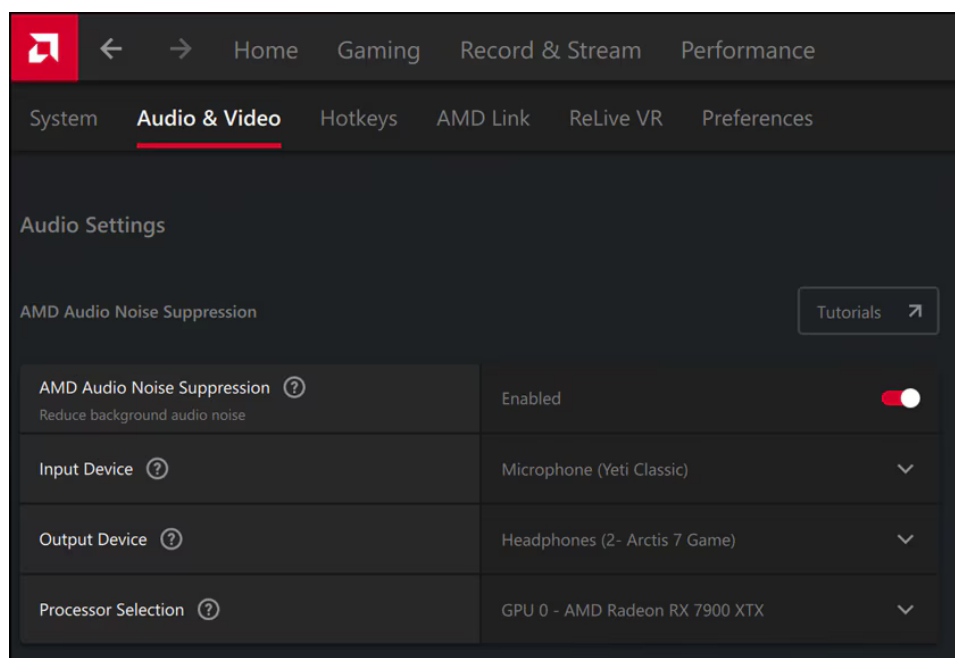
Το εργαλείο Two-Way AI Noise Cancellation, προορίζεται για τις σειρές **ROG** και **TUF** των gaming υπολογιστών της εταιρείας, και μέσα από μία εκτεταμένη βάση δεδομένων βαθιάς μάθησης επιχειρεί να απομονώσει θορύβους του περιβάλλοντος με την χρήση Τεχνητής Νοημοσύνης (ASUS, 2024). Περισσότερες πληροφορίες για το εργαλείο αυτό υπάρχουν στην διεύθυνση: <https://www.asus.com/microsite/motherboard/Intelligent-motherboard/Two-Way-AI-Noise-Cancelation.html>.

Πολλές φορές, οι λύσεις που προσφέρονται για καταστολή του θορύβου συμπεριλαμβάνουν και εξειδικευμένο hardware, όπως είναι ο αντάπτορας μικροφώνου «**ROG Clavis AI Noise Cancelling Mic Adapter**» της ίδιας εταιρείας, (<https://rog.asus.com/no/headsets-audio/accessories/rog-clavis-model/>), που εμφανίζεται στην Εικόνα 3.7, η οποία ακολουθεί στην επόμενη σελίδα. Και ο αντάπτορας αυτός, όπως και το λογισμικό Two-Way AI Noise Cancellation που αναφέρθηκε πριν, χρησιμοποιεί TN για την επίτευξη καταστολής του θορύβου.



Εικόνα 3.7. ROG Clavis AI Noise Cancelling Mic Adapter της ASUS
(Πηγή: <https://rog.asus.com/no/headsets-audio/accessories/rog-clavis-model/>)

Καταλήγοντας, αναφέρεται ότι ανάλογες λύσεις στην προσπάθεια καταστολής του θορύβου με χρήση TN, έχουν διαθέσιμες και άλλες εταιρίες του χώρου, όπως ενδεικτικά αναφέρεται η εφαρμογή «**Adrenalin**» της εταιρίας **AMD**. Στιγμιότυπο από την οθόνη λειτουργίας της εφαρμογής εμφανίζεται στην Εικόνα 3.8 που ακολουθεί.



Εικόνα 3.8. Στιγμιότυπο οθόνης χρήσης του λογισμικού Adrenalin της AMD
(Πηγή: <https://www.amd.com/en/products/software/adrenalin/amd-noise-suppression.html>)

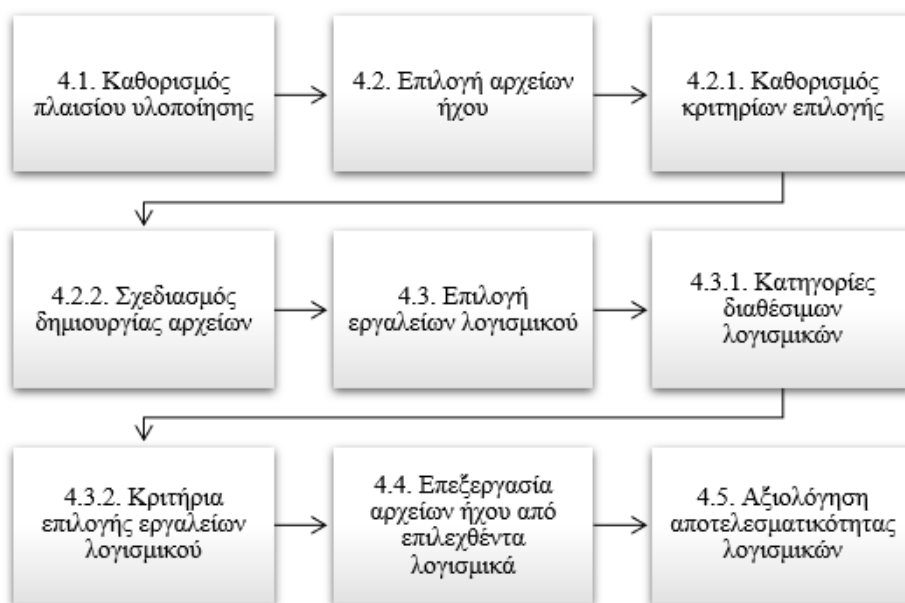
4. Μεθοδολογία και Σχεδιασμός Πειραματικής Διαδικασίας

Στα κεφάλαια που προηγήθηκαν, παρατέθηκαν τα ευρήματα της βιβλιογραφικής έρευνας σε θεωρητικό επίπεδο, σε ότι αφορά στα πεδία του ακουστικού θορύβου, των μεθόδων και τεχνικών αφαίρεσης θορύβου και των διαθέσιμων εργαλείων αποθορυβοποίησης.

Στην συνέχεια της εργασίας, δίνοντας έμφαση στις σύγχρονες τεχνικές Τεχνητής Νοημοσύνης, η έρευνα θα προσεγγιστεί αναλυτικά, μελετώντας πειραματικά την απόδοση των εργαλείων που χρησιμοποιούν Τεχνητή Νοημοσύνη για την αποθορυβοποίηση του ψηφιακού ήχου.

Στο κεφάλαιο αυτό, περιλαμβάνεται η καταγραφή της ανάλυσης και του μεθοδολογικού σχεδιασμού, που θα πραγματοποιηθεί για την διενέργεια της πειραματικής διαδικασίας, η οποία συμπληρώνει το περιεχόμενο της παρούσας διπλωματικής εργασίας, συμβάλλοντας στην πληρότητα της κάλυψης του ερευνώμενου θέματος.

Στο Σχήμα 4.1 που ακολουθεί, παρουσιάζονται τα βήματα της διαδικασίας που θα ακολουθηθεί, με την μορφή μπλοκ διαγράμματος.



Σχήμα 4.1. Ακολουθία βημάτων σχεδιασμού πειραματικής διαδικασίας

Όπως αποτυπώνεται και σχηματικά, πρώτο και ουσιαστικότερο βήμα για την διασφάλιση της ορθής υλοποίησης της διαδικασίας, θα αποτελέσει ο καθορισμός του πλαισίου της έρευνας. Παράλληλα θα καθοριστούν οι παράμετροι δημιουργίας και επιλογής των κατάλληλων αρχείων ήχου, σε συνδυασμό με την επιλογή των εργαλείων λογισμικού που θα εμπλακούν στην διαδικασία. Τέλος θα σχεδιαστεί ο τρόπος επεξεργασίας των αρχείων από τα επιλεγμένα εξεταζόμενα λογισμικά, και ο τρόπος αξιολόγησης της διαδικασίας. Τα βήματα που θα ακολουθηθούν, παρουσιάζονται αναλυτικά στην συνέχεια.

4.1 Καθορισμός πλαισίου υλοποίησης

Ο καθορισμός του πλαισίου υλοποίησης, αποτελεί ιδιαίτερη πρόκληση, καθώς θα πρέπει να περιλαμβάνει όσο το δυνατόν περισσότερες περιπτώσεις, με όσο το δυνατόν αντιπροσωπευτικότερα δείγματα, τόσο σε επίπεδο εφαρμογών χρήσης, όσο και σε επίπεδο ηχητικών σημάτων και εξεταζόμενων λογισμικών, φροντίζοντας παράλληλα το εύρος των προϊόντων να μην είναι απαγορευτικό για την διενέργεια της διαδικασίας αξιολόγησης.

Έτσι, θα πρέπει να δοθεί απάντηση σε μία σειρά από ερωτήματα, σχετικά με το πού θα επικεντρωθεί η μελέτη, όπως το αν τα εξεταζόμενα ηχητικά δείγματα θα περιλαμβάνουν μόνο ομιλία ή και μουσικό περιεχόμενο, εάν θα εξεταστεί μόνο ηχογραφημένος ήχος ή θα περιλαμβάνεται και διερεύνηση μετάδοσης ήχου σε πραγματικό χρόνο, και τέλος το ποιος θα πρέπει να είναι ο αριθμός εξεταζόμενων θορύβων και λογισμικών για επαρκή κάλυψη του πεδίου της έρευνας, φροντίζοντας παράλληλα αυτός να διατηρηθεί σε χαμηλά επίπεδα.

Στην προσπάθεια να απαντηθούν τα ερωτήματα αυτά, θα πραγματοποιηθεί αναλυτική έρευνα, συνοδευόμενη από διενέργεια επαναλαμβανόμενων πειραματικών ελέγχων, ώστε να καθοριστεί το βέλτιστο πλαίσιο διεξαγωγής της πειραματικής διαδικασίας.

Σημαντικές κατευθυντήριες γραμμές θα δοθούν και μέσα από τις τακτικές προγραμματισμένες διαδικτυακές συναντήσεις, που έχει ο γράφοντας με τον επιβλέποντα καθηγητή κ. Βρύση, στα πλαίσια της εκπόνησης της παρούσας διπλωματικής εργασίας.

Ακολουθούν οι επιλογές που έγιναν για κάθε περίπτωση, οι οποίες τελικά αποτέλεσαν και την περίμετρο της πειραματικής διαδικασίας.

4.2 Επιλογή αρχείων ήχου

Μία από τις πρώτες επιλογές που θα γίνουν, αφορά στο είδος των ηχητικών δεδομένων τα οποία πρέπει να αποθορυβοποιηθούν. Εδώ κρίνεται αναγκαίο να υπάρχει αναφορά σε σταθερά ηχητικά δείγματα, ώστε η επακόλουθη αξιολόγηση της διαδικασίας να μην επηρεαστεί από την τυχαιότητα που εισάγουν στα δεδομένα, οι μεταβολές στις κρατούσες συνθήκες, κατά τη μετάδοση του ήχου σε πραγματικό χρόνο.

Έτσι, αποφασίζεται να χρησιμοποιηθούν αποθηκευμένα ψηφιακά αρχεία ήχου, δημιουργημένα και επιλεγμένα για την υλοποίηση της διαδικασίας. Ο καθορισμός των κριτηρίων επιλογής και ο σχεδιασμός της δημιουργίας τους περιγράφεται στην συνέχεια.

4.2.1 Καθορισμός κριτηρίων επιλογής

Βασικό κριτήριο επιλογής θα αποτελέσει το περιεχόμενο των αρχείων ήχου. Διαπιστώνοντας ότι η πλειοψηφία των καθημερινών εφαρμογών στοχεύει στην αποθορυβοποίηση ομιλίας, θα προτιμηθεί να γίνει στόχευση της διαδικασίας σε ομιλία και όχι σε μουσική. Ένας ακόμα λόγος που οδηγεί στην απόφαση αυτή είναι το γεγονός ότι η επεξεργασία της μουσικής γίνεται συνήθως κάτω από ιδιαίτερες συνθήκες, (σε στούντιο από ειδικούς), κάτι που βρίσκεται έξω από την περίμετρο της παρούσας μελέτης.

Μία άλλη παράμετρος που θα εξετασθεί, είναι η γλώσσα ομιλίας που θα πρέπει να περιλαμβάνεται στα εξεταζόμενα αρχεία ήχου. Με δεδομένο ότι εξετάζονται εφαρμογές βασισμένες σε Τεχνητή Νοημοσύνη, πρέπει αρχικά να διερευνηθεί το κατά πόσο τα εμπλεκόμενα γλωσσικά μοντέλα είναι εκπαιδευμένα με επαρκή δείγματα ελληνικής ομιλίας, καθώς είναι επιθυμητό, για λόγους κατανόησης του περιεχομένου από τους αξιολογητές, αυτό να είναι στην ελληνική γλώσσα.

Μετά από διεξαγωγή διαδικτυακής έρευνας και την πραγματοποίηση εκτεταμένων δοκιμών διαπιστώνεται ότι οι ανάλογες εφαρμογές υποστηρίζουν επαρκώς την επεξεργασία ελληνικής γλώσσας, συνεπώς τα αρχεία ήχου θα περιλαμβάνουν ελληνικές ομιλίες. Με τον τρόπο αυτό, στα προϊόντα προς αξιολόγηση, ενσωματώνεται η επίδοση των λογισμικών στην αποθορυβοποίηση ελληνικού περιεχομένου.

Ως προς τα δείγματα ομιλίας, θα επιλεγεί αυτά να περιορισθούν σε δύο (2) αντιπροσωπευτικά διαφορετικά δείγματα, ένα ανδρικής και ένα γυναικείας φωνής, καθώς δεν είναι στόχος να ελεγχθεί η συμπεριφορά των λογισμικών για διαφορετικούς ομιλητές, αλλά για όσο το δυνατό περισσότερους διαφορετικούς θορύβους, ώστε έχοντας μεγαλύτερο σχετικό μέγεθος λίστας, να είναι πιο ισορροπημένο και επαρκέστερο στατιστικά, το αποτέλεσμα της αποτίμησης της συνολικής διαδικασίας.

4.2.2 Σχεδιασμός δημιουργίας αρχείων

Για πρακτικούς λόγους θα επιλεγεί τα ηχητικά δείγματα να μην προέλθουν από την διενέργεια νέων ηχογραφήσεων ομιλίας σε ενθόρυβα περιβάλλοντα, αλλά αυτά να προέλθουν από συνθετική μίξη ομιλιών και θορύβων από έτοιμες ηχογραφήσεις, έτσι ώστε να υπάρξει δυνατότητα διενέργειας δοκιμών και ελέγχου της συμπεριφοράς των λογισμικών, ελέγχοντας κατάλληλα τους σχετικούς λόγους σήματος προς θόρυβο (SNR).

Για την συλλογή των ήχων με σκοπό την σύνθεση του σχετικού περιεχομένου, θα πραγματοποιηθεί διαδικτυακή έρευνα με αναζήτηση σε εξειδικευμένες διαθέσιμες βάσεις ηχητικών δεδομένων, φροντίζοντας ώστε αυτά να έχουν τα απαραίτητα δικαιώματα χρήσης για την διεξαγωγή της διαδικασίας.

Σημαντική είναι και η συμβολή του διδάκτορα κ. Ιορδάνη Θωΐδη, με την διάθεση δειγμάτων ομιλίας από το αρχείο του, μέρος των οποίων θα χρησιμοποιηθεί κατά την υλοποίηση αυτής της διπλωματικής εργασίας.

Η διάρκεια των δειγμάτων θα διατηρηθεί μικρή, ώστε ο συνολικά μεγάλος αριθμός δειγμάτων προς ακρόαση, να μην δημιουργήσει προβλήματα κόπωσης στους ακροατές κατά την διάρκεια των ακροάσεων, φροντίζοντας παράλληλα να είναι επαρκής για την εξαγωγή ασφαλών συμπερασμάτων.

Παράλληλα θα επιλεγεί η διάρκεια να είναι η ίδια για όλα τα δείγματα. Αυτό θα συμβάλλει στην καλύτερη συγκριτική εποπτεία των γραφικών απεικονίσεων των χαρακτηριστικών των σημάτων, ενώ ταυτόχρονα θα διασφαλίσει, η χρονική έκταση των δειγμάτων να μην εμπλακεί διαμορφωτικά, στην διεξαγωγή των αποθορυβοποιήσεων.

4.3 Επιλογή εργαλείων λογισμικού

Η επιλογή των λογισμικών, θα λάβει υπόψη της το γεγονός ότι υπάρχει διαθέσιμη μία ευρύτατη γκάμα λογισμικών τα οποία απαντώνται σε πολλές διαφορετικές χρήσεις αποθορυβοποίησης και εμφανίζουν αντίστοιχα χαρακτηριστικά.

Έτσι, αρχικά θα συντελεστεί ενδελεχής διαδικτυακή έρευνα, με βάση την οποία θα προκύψει η κατηγοριοποίηση των λογισμικών που θα αναπτυχθούν στην συνέχεια. Σύμφωνα με τα ιδιαίτερα επιμέρους χαρακτηριστικά τους, θα τεθούν τα κριτήρια επιλογής των λογισμικών, φροντίζοντας η επιλογή να καλύπτει όσο το δυνατόν πληρέστερα αυτές τις κατηγορίες. Ακολουθεί στην συνέχεια, η παράθεση των κατηγοριών και των κριτηρίων επιλογής τους.

4.3.1 Κατηγορίες διαθέσιμων λογισμικών

Μια πρώτη παράμετρος διαχωρισμού των εξεταζόμενων λογισμικών σε κατηγορίες, είναι η κύρια χρήση για την οποία αυτά προορίζονται. Εδώ τα προγράμματα διαχωρίζονται, ανάλογα με το αν θα χρησιμοποιηθούν για βελτίωση του ήχου που μεταδίδεται σε πραγματικό χρόνο (real-time), ή επεμβαίνουν σε δεύτερο χρόνο πάνω σε ήδη αποθηκευμένα ηχητικά δεδομένα. Όπως ήδη έχει αναφερθεί, το πλαίσιο υλοποίησης της έρευνας, εξετάζει την δεύτερη περίπτωση.

Τα αποθηκευμένα ηχητικά δεδομένα, χωρίζονται με την σειρά τους σε δύο επιμέρους κατηγορίες, με βάση το αν αυτά περιλαμβάνουν αποκλειστικά σήματα ομιλίας, ή περικλείουν μουσικό περιεχόμενο. Για την περίπτωση αυτή, έχει θα επιλεγεί η διερεύνηση της αποτελεσματικότητας των λογισμικών, στην εξέταση ενθόρυβων σημάτων, που εμπεριέχουν αποκλειστικά ομιλίες.

Μία τρίτη ουσιαστική διαφοροποίηση μεταξύ των λογισμικών, αποτελεί η δυνατότητα παρέμβασης από τον χρήστη, κατά την διαδικασία αποθορυβοποίησης των ηχητικών δεδομένων. Οι βασικές κατηγορίες περιλαμβάνουν τόσο λογισμικά με αυξημένες δυνατότητες επέμβασης, όσο και εκείνα που στον αντίποδα λειτουργούν εντελώς αυτόματα, με αδιαφανή τρόπο ως προς την επεξεργασία που επιτελούν. Αναφερόμενοι μάλιστα σε

λογισμικά που χρησιμοποιούν Τεχνητή Νοημοσύνη, ιδιαίτερα διαδεδομένα για την απλότητά τους, είναι αυτά στα οποία απλά τοποθετείς το αρχείο και το αποτέλεσμα παράγεται χωρίς αλληλεπίδραση με τον χρήστη (one-click tools).

Σε σχέση με την άδεια και το κόστος κτήσης / χρήσης, τα λογισμικά διακρίνονται σε αυτά που παρέχουν δωρεάν επεξεργασία (free), με πλήρεις ή περιορισμένες δυνατότητες (demo license), σε συνδρομητικά, με μηνιαία ή ετήσια συνδρομή (subscribed), σε αυτά για τα οποία μπορεί να προπληρωθεί επεξεργασία αρχείων συγκεκριμένης συνολικής χρονικής έκτασης (με tokens ή credits), και σε αυτά τα οποία μπορείς να αποκτήσεις με μόνιμη άδεια, που πληρώνεται εφάπαξ (perpetual license).

Η τάση που διαμορφώνεται στην αγορά λογισμικού τα τελευταία χρόνια, δείχνει την μετατόπιση από την πώληση μόνιμων αδειών, στην πώληση συνδρομητικών αδειών χρήσης, με την μορφή «λογισμικού ως υπηρεσία» (Software as a Service - SaaS), ιδιαίτερα σε εφαρμογές διαθέσιμες μέσω υπολογιστικού νέφους (Haario et al., 2024).

Σχετικά με τον βαθμό ενσωμάτωσης Τεχνητής Νοημοσύνης, τα λογισμικά μπορούν να κατανεμηθούν σε συμβατικά (χωρίς χρήση TN), ημισυμβατικά, όπου η TN χρησιμοποιείται σε κάποιο βαθμό επικουρικά στην λειτουργία τους (AI Assisted), και λογισμικά στα οποία η αποθρομβοποίηση γίνεται αποκλειστικά με τεχνικές TN.

Ένα ακόμα στοιχείο διαφοροποίησης, αποτελεί η μορφή με την οποία αυτά είναι διαθέσιμα για χρήση. Υπάρχουν τα αυτόνομα (standalone) λογισμικά με μορφή τοπικής εγκατάστασης, τα λογισμικά με την μορφή αρθρωμάτων / επεκτάσεων (modules / plugins), τα οποία εφαρμόζονται συνδυαστικά με προγράμματα επεξεργασίας ήχου (DAWs), τα λογισμικά που φιλοξενούνται σε υποδομές υπολογιστικού νέφους και είναι προσπελάσιμα από το διαδίκτυο (web based), και -πιο πρόσφατα- τα λογισμικά με την μορφή επεκτάσεων (extensions) για προγράμματα περιήγησης διαδικτύου.

Τα τελευταία, αξιοποιούνται στην βελτίωση της ποιότητας του ήχου, στις διαδικτυακές συναντήσεις, που πραγματοποιούνται κυρίως για λόγους τηλεεργασίας, μέσω φυλλομετρητών (web browsers). Οι κατάταξη των λογισμικών με βάση τη μορφή τους, έχει αναπτυχθεί διεξοδικά στο προηγούμενο κεφάλαιο της παρούσας διπλωματικής εργασίας.

Τα λογισμικά με μορφή τοπικής εγκατάστασης, επίσης κατηγοριοποιούνται με βάση το λειτουργικό σύστημα (OS) για το οποίο έχουν εκδοθεί, ανάλογα με το υπολογιστικό

σύστημα, στο οποίο και χρησιμοποιούνται. Κύρια λειτουργικά συστήματα, για τα οποία υπάρχουν διαθέσιμες εκδόσεις, είναι τα Windows της εταιρίας Microsoft, το macOS που χρησιμοποιούν τα υπολογιστικά συστήματα της εταιρίας Apple, οι διανομές λειτουργικών συστημάτων ανοιχτού κώδικα Linux, και τέλος τα λειτουργικά συστήματα των φορητών συσκευών, με κυριότερα τα Android και iOS / iPadOS, των εταιριών Google και Apple αντίστοιχα.

Η παράμετρος αυτή, μπορεί να είναι από περιοριστική έως καθοριστική για την επιλογή ενός λογισμικού, καθώς δεν υποστηρίζονται όλα τα λειτουργικά συστήματα από όλα τα λογισμικά που υπάρχουν. Έτσι αποτελεί σημαντικό κριτήριο για την επιλογή μίας εφαρμογής έναντι μίας άλλης, ανάλογα με την διαθεσιμότητά της ή μη, για το λειτουργικό σύστημα των υπολογιστικών συστημάτων του χρήστη..

4.3.2 Κριτήρια επιλογής εργαλείων λογισμικού

Βασικό κριτήριο στην επιλογή των εργαλείων λογισμικού που θα χρησιμοποιηθούν, αποτελεί η απαίτηση αυτά να χρησιμοποιούν τεχνικές Τεχνητής Νοημοσύνης. Θα επιλεγεί να μην εξεταστεί κάποιο λογισμικό που χρησιμοποιεί συμβατικές μεθόδους, καθώς η χρήση τέτοιων εργαλείων βασίζεται καθαρά στην ικανότητα και την εμπειρία του χειριστή τους, κάτι το οποίο επηρεάζει σε μεγάλο βαθμό το τελικό αποτέλεσμα, και μπορεί να οδηγήσει σε λανθασμένη αξιολόγηση, με συνεπακόλουθη εξαγωγή λανθασμένων συμπερασμάτων.

Για την επιλογή των αντιπροσωπευτικότερων και ευρύτερα διαδεδομένων εργαλείων, θα πραγματοποιηθεί λεπτομερής διαδικτυακή έρευνα, μέσα από την οποία θα εξετασθούν και αξιοποιηθούν δεδομένα από εξειδικευμένους ιστοτόπους, φόρουμ και ιστολόγια, ώστε μετά από τις απαραίτητες δοκιμές των λογισμικών, να γίνει η επιλογή αυτών που τελικά θα χρησιμοποιηθούν στην παρούσα μελέτη.

Θα εξετασθεί ένας ικανός αριθμός από τα λογισμικά, διερευνώντας για κάθε ένα τους, την ευκολία χρήσης, την ποιότητα του παραγόμενου αποθρονοποιημένου ήχου, την συμβατότητα με το είδος του ήχου θα πρέπει να αποθρονοποιηθεί, το κόστος τους, τις πιθανές επιπλέον δυνατότητές τους, τις διαθέσιμες στο διαδίκτυο απόψεις, κριτικές και

αξιολογήσεις των χρηστών, τον βαθμό παραμετροποίησης και επέμβασης κατά την χρήση των εφαρμογών, την ταχύτητα εκτέλεσης και την απαίτησή τους σε υπολογιστικούς πόρους.

Αναγκαίο για την διενέργεια της όλης διαδικασίας, είναι ο συνολικός αριθμός των εξεταζόμενων λογισμικών να κρατηθεί αρκετά χαμηλά, με επιλογή αντιπροσωπευτικών εκπροσώπων, ώστε η αξιολόγηση και αποτίμηση της αποτελεσματικότητας να μην απαιτήσει εύρος που την καθιστά προβληματική.

Έτσι, και με βάση τα κριτήρια που αναφέρθηκαν ήδη, θα επιλεγεί ο αριθμός των εφαρμογών, να μην υπερβαίνει τα πέντε, καθώς ο αριθμός αυτός, παρότι μικρός, παρέχει αντιπροσωπευτικότητα, με παραμέτρους την επεξεργασία αποθηκευμένου ήχου, με χρήση αποκλειστικά Τεχνητής Νοημοσύνης, την αποθρομβοποίηση δεδομένων ομιλίας και όχι μουσικής, με όσο γίνεται ευκολότερο χειρισμό και επιλέγοντας την οικονομικότερη κατά το δυνατόν χρήση.

4.4 Επεξεργασία αρχείων ήχου από τα επιλεγθέντα λογισμικά

Η επεξεργασία των αρχείων, θα γίνει από τον γράφοντα, φροντίζοντας για την επίτευξη του βέλτιστου δυνατού αποτελέσματος, για κάθε ένα από αυτά. Εδώ, με εξαίρεση την κατηγορία των one-click λογισμικών, που όπως ήδη αναφέρθηκε, δεν παρέχουν την δυνατότητα οποιασδήποτε παρέμβασης κατά την αποθρομβοποίηση, η εισαγωγή του υποκειμενικού παράγοντα στο αποτέλεσμα, είναι σε κάποιο βαθμό αναπόφευκτη, καθώς πρέπει να γίνουν επιλογές παραμέτρων, ώστε να βρεθεί η «χρυσή τομή», ανάμεσα στην επίτευξη της μεγαλύτερης δυνατής εξάλειψης του θορύβου, με την ταυτόχρονη διασφάλιση όσο το δυνατόν καλύτερης ποιότητας της παραμένουσας φωνής στο ηχητικό αποτέλεσμα.

Για τον σκοπό αυτό, σε όποια από τα εξεταζόμενα λογισμικά υπάρχει αυτή η δυνατότητα, θα πραγματοποιηθεί επαναληπτικά η διαδικασία για κάθε αρχείο, εφαρμόζοντας κάθε φορά διαφορετικό μοντέλο TN, από τα διαθέσιμα της εφαρμογής, ενώ οι εξαγωγές των αρχείων θα πραγματοποιηθούν με διαφορετικές, εναλλακτικές ρυθμίσεις παραμέτρων, με κυριότερη αυτή που επιδρά στην αναλογία εξάλειψης θορύβου προς ποιότητα.

Από την αναδρομική αυτή διαδικασία, καταβάλλοντας προσπάθεια περιορισμού κατά το δυνατόν του υποκειμενικού στοιχείου, θα προκύψουν τα τελικά αποθρομβημένα αρχεία, που θα χρησιμοποιηθούν στις ακροάσεις αξιολόγησης της επίδοσης των λογισμικών.

4.5 Αξιολόγηση αποτελεσματικότητας λογισμικών

Η πειραματική αξιολόγηση της αποτελεσματικότητας των εξεταζόμενων εργαλείων, θα υλοποιηθεί με την διενέργεια κριτικών ακροάσεων (critical listening). Τα ποσοτικά αποτελέσματα των ακροάσεων, θα αποτυπωθούν μέσω έντυπου ερωτηματολογίου, που θα δημιουργηθεί για την συλλογή των δεδομένων που θα προκύψουν. Η επεξεργασία και η γραφική απεικόνιση των αποτελεσμάτων, θα γίνει με την χρήση δομημένων εργαλείων λογιστικών φύλων, ενώ η ποιοτική τους αποτίμηση θα γίνει μέσα από την καταγραφή σχολίων και παρατηρήσεων.

Στα πλαίσια της διαδικασίας ακρόασης και αξιολόγησης (experimental setup), αρχικά θα καθοριστεί ο τρόπος οργάνωσης των αρχείων ώστε να είναι βέβαιο ότι δεν θα υπάρξει μεροληπτική κρίση υπέρ κάποιου συγκεκριμένου λογισμικού. Αυτό θα διασφαλιστεί με το να διατηρηθεί άγνωστη η ταυτότητα των λογισμικών στους ακροατές, ενώ επιπρόσθετα ο τρόπος διάταξης των δειγμάτων κατά την ακρόαση, θα γίνει με τυχαίο εναλλασσόμενο τρόπο, ώστε αυτά να μην αντιστοιχούν πάντα στο ίδιο λογισμικό.

Όσον αφορά στην επιλογή του πρωτοκόλλου ακρόασης, θα επιλεγεί σαν καταλληλότερη η 5βαθμη κλίμακα αξιολόγησης **MOS (Mean Opinion Score)**, η οποία αποτελεί τυπική κλίμακα στις ακουστικές αξιολογήσεις (Streijl et al., 2016), και επιτρέπει ευκολότερη σύγκριση μεταξύ των δειγμάτων. Για την επίτευξη αντιπροσωπευτικότερων αποτελεσμάτων, θα μεριμνηθεί ώστε να η ομάδα ακρόασης να αποτελείται από ικανό αριθμό ακροατών, ώστε το επιλεγμένο πρωτόκολλο να παράξει καλύτερα αποτελέσματα.

Θα καταβληθεί προσπάθεια, οι συμμετέχοντες ακροατές να έχουν όσο το δυνατόν μεγαλύτερη σχέση με τον ήχο, σε επαγγελματικό επίπεδο ή μέσω σχετικών σπουδών. Με τον τρόπο αυτό, όπως έχει παρατηρηθεί (Man & Reiss, 2017), λόγω της αυξημένης

εμπειρίας τους με τον ήχο, οι αξιολογήσεις τους θα είναι πιο στοχευμένες και η διάκριση των λεπτομερειών πιο αποτελεσματική.

Σχετικά με την δομή των ερωτήσεων, θα επιλεγεί να υπάρχει μία ερώτηση ανά αρχείο, η οποία να αφορά στην συνολική ποιότητα / καθαρότητα του αποτελέσματος, όπου θα αξιολογούνται ταυτόχρονα τόσο ο βαθμός εξάλειψης θορύβου, όσο και η ποιότητα της αποθορυβοποιημένης ομιλίας. Έτσι θα είναι αναγκαία μόνο μία βαθμολόγηση ανά δείγμα, περιορίζοντας έτσι τον συνολικό αριθμό απαντήσεων, ο οποίος λόγω και των πολλών προϊόντων αρχείων, θα είναι ήδη μεγάλος.

Η διεξαγωγή των ακροάσεων θα πραγματοποιηθεί σε ελεγχόμενο περιβάλλον, στο οποίο ήταν διασφαλισμένες οι συνθήκες της απαραίτητης ησυχίας, με την χρήση ενσύρματων ακουστικών κλειστού τύπου, για την απομόνωση των εξωτερικών θορύβων του χώρου ακρόασης. Στους ακροατές θα δοθούν οι απαραίτητες οδηγίες και τους επισημάνθηκε δυνατότητα επανάληψης της ακρόασης πριν τη βαθμολόγηση.

Ακολουθεί η λεπτομερής αποτύπωση των βημάτων της υλοποίησης, η οποία θα πραγματοποιηθεί σύμφωνα με την ανάλυση και τον σχεδιασμό που αναπτύσσεται στο παρόν κεφάλαιο της εργασίας.

5. Υλοποίηση Πειραματικής Διαδικασίας

Μετά από την διενέργεια της ανάλυσης και του σχεδιασμού της πειραματικής διαδικασίας, θα υλοποιηθούν τα βήματα που θα καταγραφούν σε αυτό το κεφάλαιο της παρούσας διπλωματικής εργασίας.

Σαν πρώτο βήμα, θα πραγματοποιηθεί η επιλογή και διαμόρφωση των αρχείων ομιλίας και θορύβου, τα οποία θα χρησιμοποιηθούν για τον έλεγχο της επίδοσης αποθορυβοποίησης. Η υλοποίηση θα συμπεριλάβει την μίξη ομιλίας με θόρυβο για την δημιουργία των ενθόρυβων αρχείων.

Μετά από αυτή την διαδικασία, θα γίνει η παράθεση των αντιπροσωπευτικών εργαλείων λογισμικού, που θα χρησιμοποιηθούν για την αποθορυβοποίηση των αρχείων, ο τρόπος με τον οποίο θα γίνουν οι αποθορυβοποιήσεις από το καθένα τους, και η συνεπακόλουθη καταγραφή της διαδικασίας αξιολόγησης της αποτελεσματικότητάς τους.

5.1 Επιλογή και διαμόρφωση αρχείων ήχου

Όπως έχει ήδη αναπτυχθεί στην φάση του σχεδιασμού, για την υλοποίηση θα πραγματοποιηθούν πρωτογενείς ηχογραφήσεις από τον γράφοντα, αλλά θα αξιοποιηθούν δείγματα, τα οποία θα ληφθούν από βάσεις ηχητικών δεδομένων που είναι διαθέσιμες στο διαδίκτυο, με δωρεάν δικαιώματα χρήσης των αρχείων τους.

Για τα δεδομένα θορύβου, θα χρησιμοποιηθεί η βάση ηχητικών εφέ του βρετανικού δικτύου **BBC** (<https://sound-effects.bbcwind.co.uk/>), η βάση δειγμάτων ήχου **Sample Focus** (<https://samplefocus.com/>) και η βάση μουσικών αρχείων **Free Music Archive - FMA** (<https://freemusicarchive.org/>).

Η χρονική διάρκεια του κάθε ηχητικού δείγματος, μετά από διενέργεια δοκιμών και σχετικών ελέγχων, θα περιοριστεί στα **~6,64 sec**, καθώς κρίνεται επαρκής για την εξαγωγή ασφαλών συμπερασμάτων, και είναι η ίδια για όλα τα δείγματα.

Η επιλογή και διαμόρφωση των αρχείων φωνής και θορύβου, περιγράφεται στην συνέχεια.

5.1.1 Αρχεία ομιλίας

Με γνώμονα την επιθυμία το αρχικό περιεχόμενο των σημάτων ομιλίας να είναι απαλλαγμένο από πιθανές επιδράσεις θορύβου, αρχικά θα διερευνηθεί η περίπτωση αυτά να δημιουργηθούν με συνθετικό τρόπο, κάνοντας χρήση εργαλείων βασισμένων σε Τεχνητή Νοημοσύνη.

Για τον σκοπό αυτό θα χρησιμοποιηθεί η διαδικτυακή πλατφόρμα **Eleven Labs**, και συγκεκριμένα η δυνατότητα μετατροπής κειμένου σε ομιλία (Text to Speech) της πλατφόρμας (<https://elevenlabs.io/app/speech-synthesis/text-to-speech>). Η πρόσβαση στην υπηρεσία αυτή, παρέχεται και δωρεάν για την επεξεργασία σημαντικού όγκου δεδομένων, με την δημιουργία λογαριασμού στην πλατφόρμα.

Για την σύνθεση θα χρησιμοποιηθεί το κείμενο: «**Δείγμα ομιλίας, για αξιολόγηση εργαλείων λογισμικού, αποθρονοποίησης ψηφιακού ήχου**», το οποίο εκτείνεται στην επιθυμητή χρονική διάρκεια των έξι δευτερολέπτων, ενώ δομικά αποτελείται από αντιπροσωπευτικά φωνήματα και ενδιάμεσες παύσεις.

Για την παραγωγή των ήχων θα επιλεγεί η έκδοση 2 (**Eleven Multilingual v2**), από τα διαθέσιμα γλωσσικά μοντέλα της πλατφόρμας, το οποίο υποστηρίζει πολυγλωσσικό περιεχόμενο, στο οποίο περιλαμβάνεται και η ελληνική γλώσσα. Σημειώνεται ότι κατά την χρονική στιγμή που δημιουργούνται τα αρχεία ομιλίας, στα τέλη Ιανουαρίου 2026, η διαθέσιμη πλέον έκδοση 3, είναι ακόμα σε φάση ανάπτυξης Alpha και δεν θα επιλεγεί.

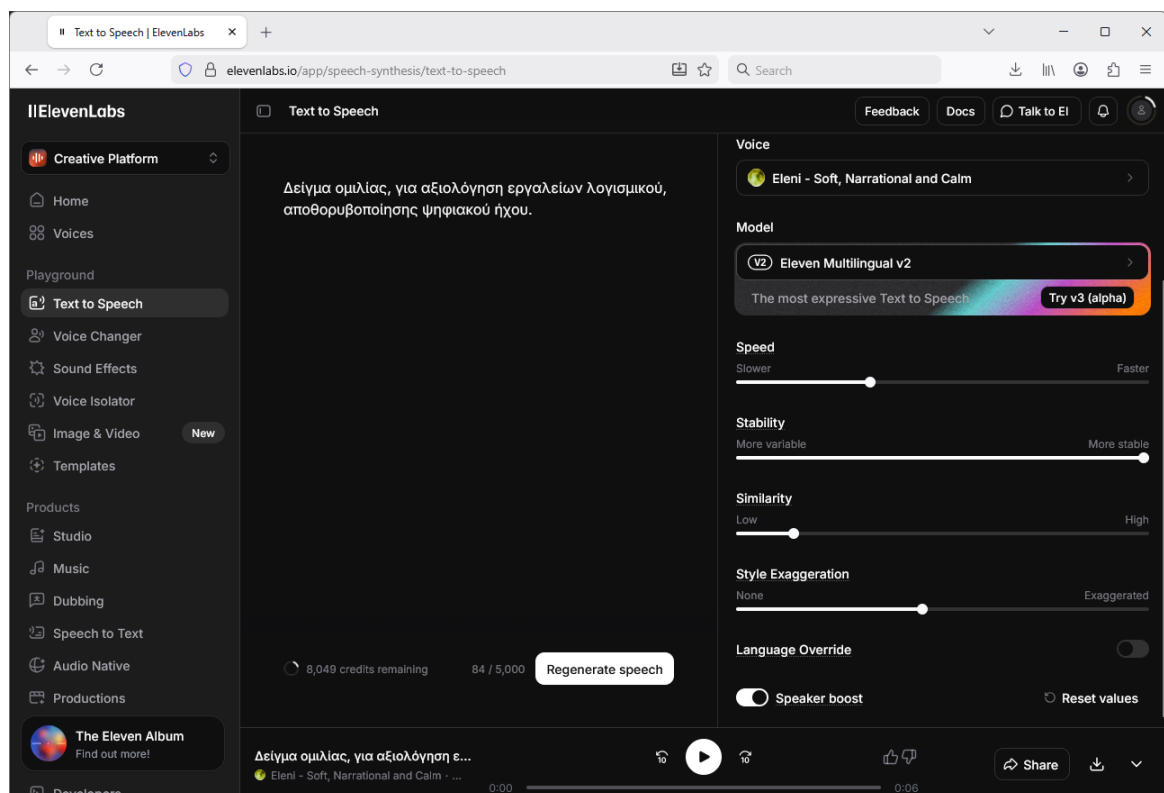
Σε αυτό το γλωσσικό μοντέλο, θα χρησιμοποιηθούν τέσσερις διαφορετικές ελληνικές φωνές, δύο αντρικές και δύο γυναικείες, από τις διαθέσιμες στην βάση δειγμάτων ομιλίας της πλατφόρμας. Μετά από διαδοχικές δοκιμές απόδοσης, θα επιλεγούν οι φωνές «**Ελένη**», «**Στέφανος**», «**Χρύσα**» και «**Άτλας**». Οι τελικές επιμέρους ρυθμίσεις των παραμέτρων των φωνών, που επιφέρουν το καλύτερο αποτέλεσμα, φαίνονται στον ακόλουθο Πίνακα 5.1.

Πίνακας 5.1. Ρυθμίσεις παραμέτρων φωνών για δημιουργία συνθετικής ομιλίας

Παράμετρος	Ελένη	Στέφανος	Χρύσα	Άτλας
Ταχύτητα ομιλίας (Speed)	36%	83%	77%	100%

Σταθερότητα (Stability)	100%	83%	89%	50%
Ομοιότητα (Similarity Enhancement)	13%	100%	58%	75%
Έμφαση ύφους (Style Exaggeration)	45%	36%	8%	0%
Ενίσχυση ομιλητή (Speaker Boost)	Enabled	Enabled	Enabled	Enabled

Στην Εικόνα 5.1 που ακολουθεί, αποτυπώνεται ενδεικτικό στιγμιότυπο από την διαδικασία παραγωγής ομιλίας στην πλατφόρμα Eleven Labs, για την φωνή της «Ελένης».



Εικόνα 5.1. Στιγμιότυπο παραγωγής ομιλίας από κείμενο στην πλατφόρμα Eleven Labs

Από τις ακροάσεις των παραχθέντων συνθετικών αρχείων ομιλίας, διαπιστώνεται ότι το περιεχόμενό τους παρουσιάζει εμφανή στοιχεία επεξεργασίας βελτίωσης, όπως συμπίεση, φιλτράρισμα και περιορισμό στάθμης, με παράλληλη απουσία της επίδρασης ενός φυσικού χώρου κατά την διάρκεια μίας ηχογράφησης, με αποτέλεσμα η κάθε φωνή να διαφέρει από τη φυσική ομιλία τόσο στην αντίληψη όσο και ως είσοδος στα εργαλεία αποθρονοποίησης. Σαν συνέπεια του γεγονότος αυτού, θα πραγματοποιηθεί αναθεώρηση

της διαδικασίας σχετικά με το περιεχόμενο των δειγμάτων ομιλίας, με υιοθέτηση της επιλογής τα αρχεία φωνής να προέρχονται από φυσική ηχογραφημένη ομιλία αντί για συνθετική.

Έτσι, τα συνθετικά δείγματα ομιλίας που έχουν δημιουργηθεί στην πλατφόρμα ElevenLabs, θα αντικατασταθούν με νέα βαθμονομημένα δείγματα ελληνικής ομιλίας, τα οποία θα προέλθουν από το αρχείο του διδάκτορα κ. Ιορδάνη Θωΐδη, και περιλαμβάνουν φυσικές ομιλίες, κατάλληλες για τους σκοπούς της παρούσας διπλωματικής εργασίας.

Από τις λεπτομερείς δοκιμές που πραγματοποιούνται κατά την διάρκεια της συγγραφής της εργασίας, διαπιστώνεται ότι η ύπαρξη διαφορετικών ομιλιών στα δείγματα δεν είναι τόσο κρίσιμη για το τελικό αποθορυβοποιημένο αποτέλεσμα, όσο το ίδιο το περιεχόμενο του θορύβου προς εξέταση. Το συμπέρασμα αυτό προκύπτει πειραματικά, παρατηρώντας ότι στα αποθορυβημένα αποτελέσματα που προκύπτουν, για τα διαφορετικά λογισμικά που εξετάζονται, οι όποιες «αδυναμίες» στην αποθορυβοποίηση, εμφανίζονται έχοντας σχεδόν αποκλειστική προέλευση το διαφορετικό περιεχόμενο θορύβου, ενώ είναι σχεδόν ανεξάρτητες από τις διαφορετικές φωνές που περιλαμβάνονται στα δείγματα.

Συνεπώς ο περιορισμός του αριθμού των δειγμάτων ομιλίας σε δύο, (ένα αντρικό και ένα γυναικείο), κρίνεται επαρκής και αντιπροσωπευτικός για την υλοποίηση της διαδικασίας, καθώς το να υπάρχουν περισσότερα νέα δημιουργημένα δείγματα δεν προσφέρει κάτι ουσιαστικό, ενώ μεγαλώνει τον συνολικό αριθμό των προϊόντων, την διάρκεια των τελικών ακροάσεων και την κόπωση των αξιολογητών.

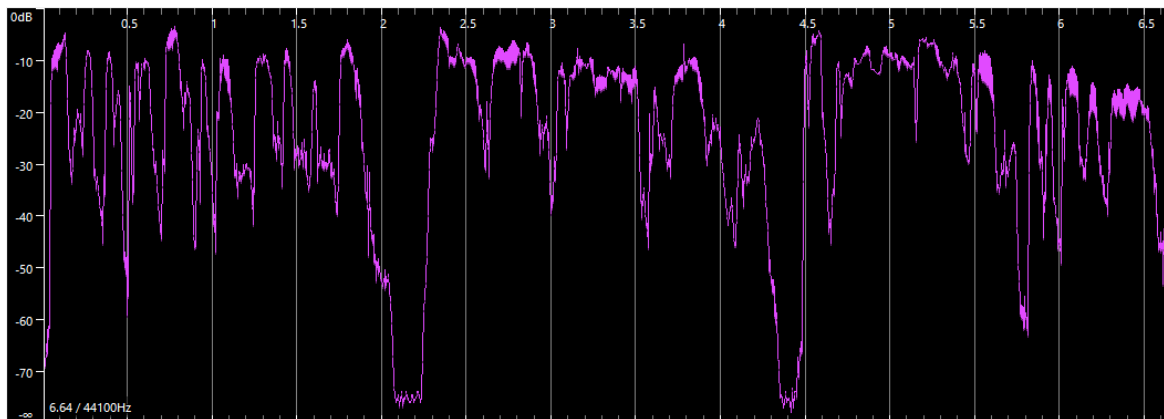
Για την δημιουργία των αρχείων αντρικής και γυναικείας ομιλίας, θα γίνει προσθετική μίξη των διαθέσιμων αρχείων, με στόχο να προκύψουν ήχοι στην επιθυμητή χρονική διάρκεια των έξι περίπου δευτερολέπτων, με παράλληλη διασφάλιση της ύπαρξης αντιπροσωπευτικού αριθμού φωνημάτων και παύσεων κατά την διάρκεια της ομιλίας.

5.1.1.1 Φωνή 1 (V01)

Για την δημιουργία του αρχείου γυναικείας φωνής, θα συνενωθούν τρία διαφορετικά αρχεία ομιλίας. Αυτά περιλαμβάνουν τις φράσεις «**Έτσι όπως κρατάς τον καφέ θα χυθεί**», «**Δίνουν δώρα στα νέα μέλη της λέσχης**» και «**Πρέπει να είναι ωραίο**

να γράφεις παιδικά βιβλία». Η επιλογή των προτάσεων θα πραγματοποιηθεί έτσι ώστε στο τελικό αποτέλεσμα να συμπεριλαμβάνεται αντιπροσωπευτικός αριθμός φωνημάτων.

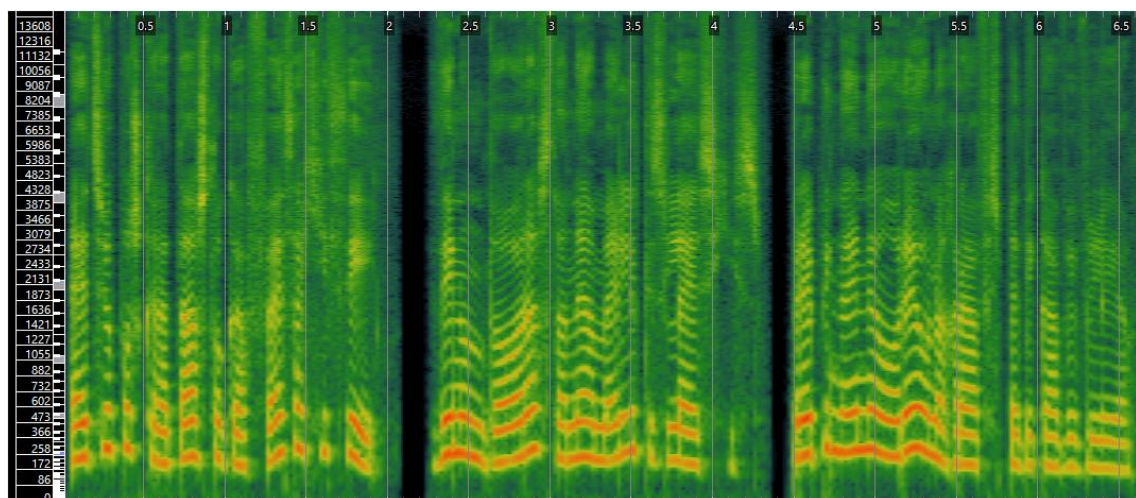
Στην Εικόνα 5.2 που ακολουθεί, αποτυπώνεται ως η κυματομορφή του πλάτους της γυναικείας ομιλίας. Οι έντονες βυθίσεις στο σήμα, αποτυπώνουν τα σημεία των παύσεων.



Εικόνα 5.2. Κυματομορφή ομιλίας γυναίκας - φωνή 1 (V01)

Η κυματομορφή εμφανίζει την συνδυαστική (mean) τιμή των δύο καναλιών του στερεοφωνικού ήχου. Οι τιμές του κατακόρυφου άξονα είναι εκφρασμένες σε κλίμακα dBFS, με μέγιστη τιμή τα 0 dB. Για την διευκόλυνση του αναγνώστη, στην συνέχεια της εργασίας, οι απεικονίσεις γυναικείου ηχητικού σήματος θα εμφανίζονται με μωβ χρώμα.

Στην επόμενη Εικόνα 5.3, εμφανίζεται το φασματογράφημα του ίδιου ήχου.



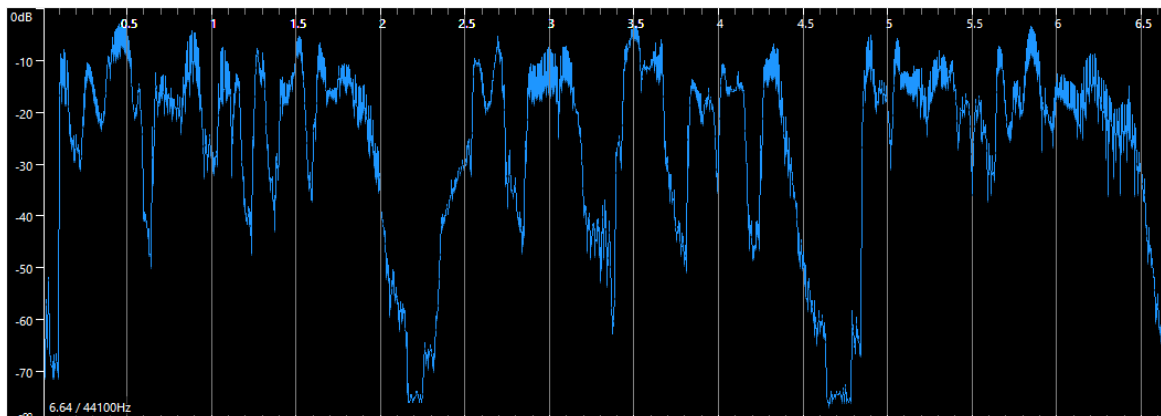
Εικόνα 5.3. Φασματογράφημα ομιλίας γυναίκας - φωνή 1 (V01)

Όπως φαίνεται από τον κάθετο άξονα στο προηγούμενο φασματογράφημα, η ομιλία της γυναίκας έχει σαν κύρια συχνοτική περιοχή, αυτή των **~200Hz**, ενώ καταγράφονται τρεις τουλάχιστον αρμονικές της, με αρκετά αξιόλογο ενεργειακό περιεχόμενο. Στο φασματογράφημα είναι επίσης ορατός ο -εσκεμμένος- διαχωρισμός των προτάσεων που συνθέτουν την τελική ομιλία. Αυτός αποτυπώνεται στις μαύρες περιοχές, που χρονικά τοποθετούνται στα **2,1 με 2,3 sec** και **4,3 με 4,5 sec**.

5.1.1.2 Φωνή 2 (V02)

Για την δημιουργία του αρχείου ανδρικής φωνής, θα συνενωθούν τρία διαφορετικά αρχεία ομιλίας, όπως και στην περίπτωση της γυναικείας ομιλίας. Αυτά περιλαμβάνουν τις φράσεις «**Αντί για άσπρη ζάχαρη προτιμώ κανέλα**», «**Φύγε τώρα και έλα πίσω μετά**» και «**Όσα μου λες δε βγάζουν νόημα**». Η επιλογή των προτάσεων, θα πραγματοποιηθεί πάλι έτσι ώστε στο αποτέλεσμα να εμπεριέχεται αντιπροσωπευτικός αριθμός φωνημάτων.

Στην Εικόνα 5.4 που ακολουθεί, αποτυπώνεται η κυματομορφή του πλάτους της ανδρικής ομιλίας. Και εδώ, οι έντονες βυθίσεις αποτυπώνουν τα σημεία των παύσεων.

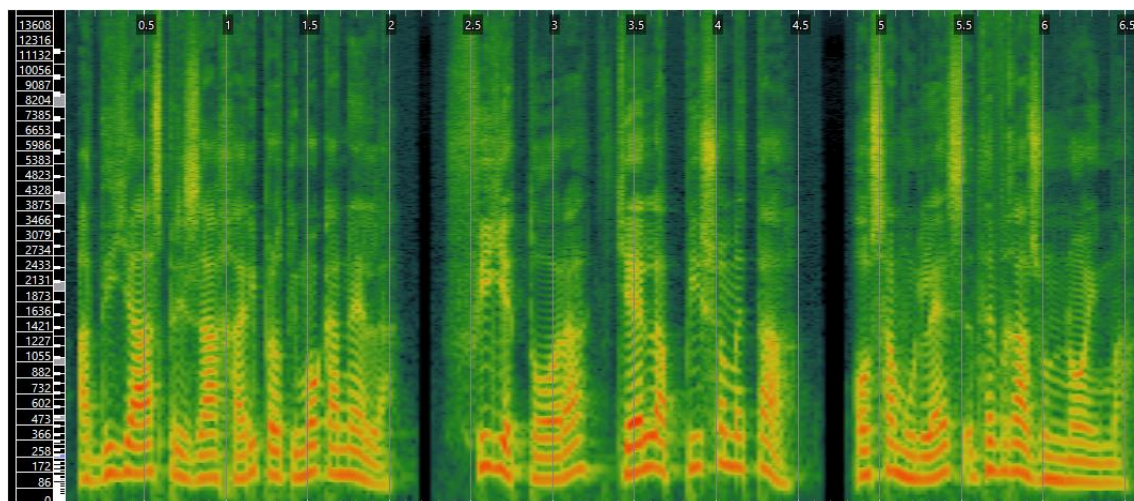


Εικόνα 5.4. Κυματομορφή ομιλίας άνδρα - φωνή 2 (V02)

Όπως και πριν, η κυματομορφή εμφανίζει την συνδυαστική (mean) τιμή των δύο καναλιών του στερεοφωνικού ήχου. Οι τιμές του κατακόρυφου άξονα είναι και πάλι εκφρασμένες σε κλίμακα dBFS, με μέγιστη τιμή τα 0 dB. Οι ρυθμίσεις αυτές διατηρούνται για την απεικόνιση όλων των κυματομορφών, που θα αποτυπώνονται στην συνέχεια της

παρούσας διπλωματικής εργασίας, ενώ οι απεικονίσεις κυματομορφών που περιγράφουν το ανδρικό ηχητικό σήμα θα εμφανίζονται με γαλάζιο χρώμα όπως εδώ.

Στην Εικόνα 5.5, εμφανίζεται το φασματογράφημα της ανδρικής ομιλίας.



Εικόνα 5.5. Φασματογράφημα ομιλίας άνδρα - φωνή 2 (V02)

Αποτυπώνεται στον κάθετο άξονα συχνοτήτων, πως η ανδρική ομιλία όπως είναι αναμενόμενο (Φλώρος, 2013), τοποθετείται σε χαμηλότερη συχνότητα, έχοντας σαν κύρια συχνοτική περιοχή, αυτή των $\sim 100\text{Hz}$. Εδώ είναι εμφανείς περισσότερες αρμονικές σε σχέση με την περίπτωση της γυναικείας ομιλίας. Και στο φασματογράφημα αυτό, είναι ορατός ο διαχωρισμός των προτάσεων που συνθέτουν την τελική ομιλία, όπως αποτυπώνεται στις μαύρες περιοχές, που στην δεδομένη περίπτωση τοποθετούνται χρονικά στα 2,1 με 2,3 sec και 4,6 με 4,8 sec αντίστοιχα. Αυτό μπορεί εύκολα να το παρατηρήσει κανείς, από τις χρονικές διαγραμμίσεις που περιλαμβάνονται στα εικονιζόμενα δεδομένα, και αποτυπώνουν εμφανώς την χρονική εξέλιξη των σημάτων, στο οριζόντιο πάνω μέρος των εικόνων.

5.1.2 Αρχεία θορύβου

Βασική επιδίωξη για τα αρχεία θορύβου, είναι αυτά να καλύπτουν επαρκώς τις κατηγορίες που καταγράφηκαν ήδη, σε προηγούμενο σημείο της παρούσας διπλωματικής εργασίας. Παράλληλα, επιδίωξη αποτελεί και η ικανοποίηση της απαίτησης για περιορισμό

του αριθμού τους, ώστε ενώ αυτά να είναι αντιπροσωπευτικά, να είναι ταυτόχρονα και όσο γίνεται λιγότερα, ώστε η όλη διαδικασία να μην καταστεί δυσανάλογα μεγάλη.

Μετά από εκτεταμένη έρευνα και διεξαγωγή αντίστοιχων δοκιμών, θα επιλεγεί η δημιουργία **δώδεκα (12)** συνολικά δειγμάτων θορύβου. Για να επιτευχθεί ο αριθμός αυτός, θα χρειαστεί σε κάποιες περιπτώσεις να γίνει μίξη θορύβων, που προέρχονται από το ίδιο περιβάλλον σε κανονικές συνθήκες παρατήρησης, ώστε να δημιουργηθούν αρχεία με συνδυαστικό περιεχόμενο. Η τελική επιλογή θα πραγματοποιηθεί μετά από έλεγχο παραμέτρων, όπως φασματικές ιδιαιτερότητες, τυχόν περιοδικότητες και διακυμάνσεις σε στάθμες ή συχνότητες.

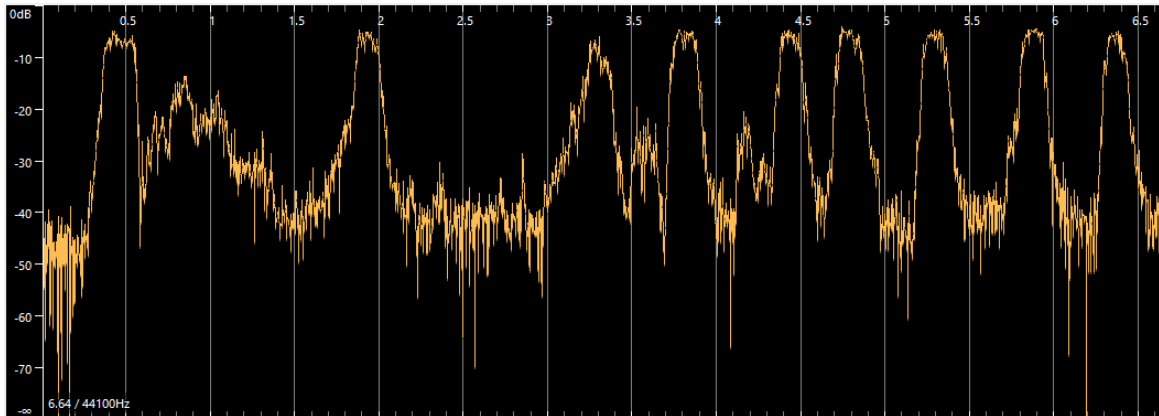
Για την επιλογή των συγκεκριμένων θορύβων, θα συνεκτιμηθεί το γεγονός ότι αυτοί θα πρέπει στην συνέχεια να εξαλειφθούν με την χρήση εργαλείων Τεχνητής Νοημοσύνης. Έτσι συμπεριλαμβάνονται αρχεία, που ενώ φαινομενικά περικλείουν επικαλυπτόμενους ή «συγγενείς» θορύβους, εντούτοις πρέπει να εξεταστούν στην μελέτη, ώστε η πιθανότητα δείγματα συγκεκριμένων θορύβων να περιλαμβάνονται ή όχι στα δεδομένα με τα οποία έχει εκπαιδευθεί το μοντέλο TN της κάθε εξεταζόμενης εφαρμογής, να συμβάλει στο τελικό αποτέλεσμα της αποθορυβοποίησης.

Οι 12 επιλεγμένοι θόρυβοι θα περιλαμβάνουν αποσπάσματα από: γαύγισμα σκύλου, αλυσοπρίονο, μίξη ήχου πουλιών με γρύλο (τριζόνι), ελικόπτερο, χτύπο λειτουργίας ρολογιού τοίχου, πλήθος σε γεμάτη καφετέρια (με αγγλικές ομιλίες σε δεύτερο πλάνο), θόρυβο ανεμιστήρα, μουσικό κομμάτι στο παρασκήνιο, ήχο πληκτρολογίου, σήματα μορς, μίξη σειρήνας αστυνομίας με κορναρίσματα και έντονη βροχή με απόμακρους κεραυνούς.

5.1.2.1 Θόρυβος 1 (N01)

Ο **θόρυβος 1 (N01)**, περιλαμβάνει γαύγισμα σκύλου και θα προέλθει από την βάση ηχητικών δεδομένων του **BBC**. Συγκεκριμένα πρόκειται για το αποθηκευμένο αρχείο της βάσης, με ονομασία «**Animals - Airedale barking angrily**» και κωδικό αριθμό **07037461**. Το αρχείο αυτό, με ελεύθερη άδεια χρήσης για μη εμπορική χρήση, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://sound-effects.bbcrewind.co.uk/search?q=07037461>.

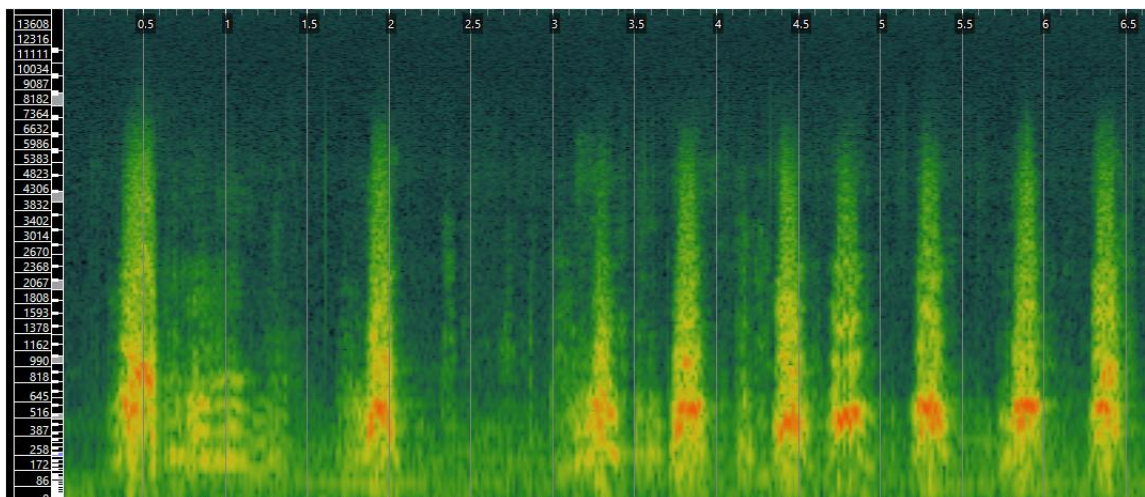
Ο θόρυβος αυτός, θα επιλεγεί σαν ιδιαίτερα συνηθισμένος στην περίπτωση εξωτερικών δραστηριοτήτων. Στην Εικόνα 5.6 που ακολουθεί, αποτυπώνεται η κυματομορφή του πλάτους του θορύβου. Οι έντονες εξάρσεις αποτυπώνουν τα διαφορετικά διακριτά γαυγίσματα, ενώ η παραμένουσα στάθμη του περιβάλλοντος θορύβου σε στιγμές ηρεμίας είναι επίσης εμφανής.



Εικόνα 5.6. Κυματομορφή γαυγίσματος σκύλου - θόρυβος 1 (N01)

Σημειώνεται ότι οι απεικονίσεις κυματομορφών που περιγράφουν ηχητικά σήματα θορύβου, θα εμφανίζονται στην συνέχεια της εργασίας με πορτοκαλί χρώμα όπως εδώ.

Στην Εικόνα 5.7, εμφανίζεται το φασματογράφημα του ίδιου θορύβου.



Εικόνα 5.7. Φασματογράφημα γαυγίσματος σκύλου - θόρυβος 1 (N01)

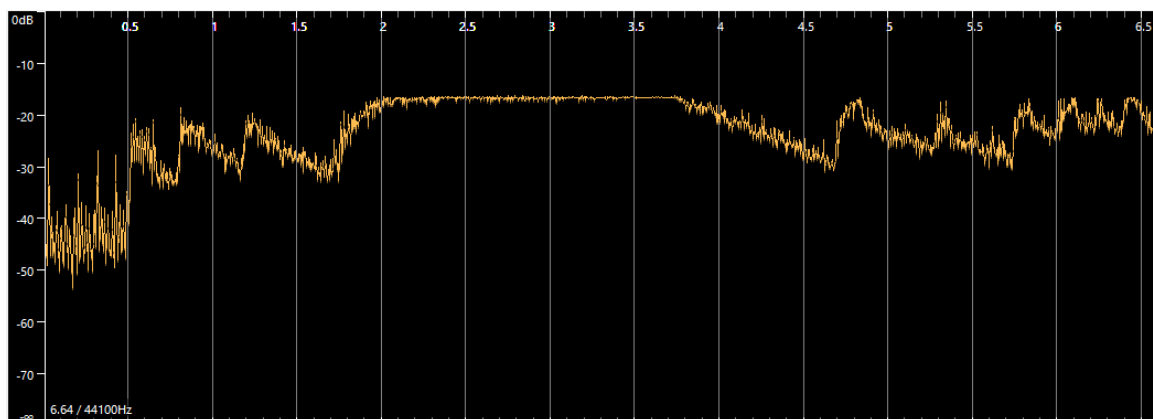
Όπως φαίνεται στην προηγούμενη εικόνα, ο θόρυβος είναι εντονότερος στην περιοχή που εκτείνεται από τα **250Hz** έως τα **1000Hz**, ενώ εκτείνεται προς τα πάνω σε συχνότητες που ξεπερνούν τα **6000 Hz**, γεγονός που σημαίνει ότι υπάρχει επίδραση κυρίως στις αρμονικές συνιστώσες των σημάτων ομιλίας.

5.1.2.2 Θόρυβος 2 (N02)

Ο **θόρυβος 2 (N02)**, περιλαμβάνει θόρυβο από αλυσοπρίονο και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Chainsaw FX**», από τον χρήστη **J** (ή **Oldie Ash**). Το αρχείο αυτό, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/chainsaw-fx>, με ελεύθερη άδεια χρήσης.

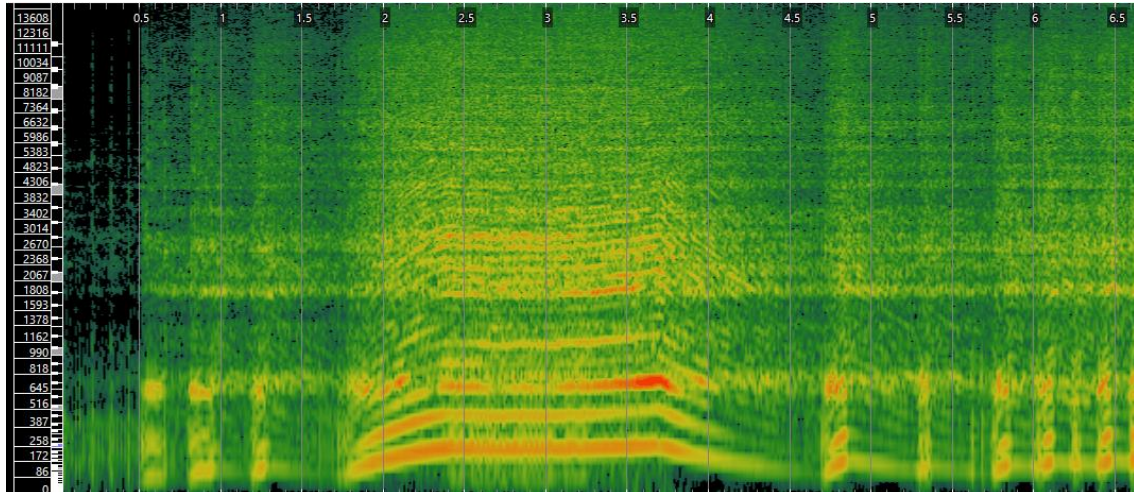
Στην Εικόνα 5.8 που ακολουθεί, αποτυπώνεται η κυματομορφή του πλάτους του θορύβου. Εδώ παρατηρείται, ότι οι ρυθμίσεις εισόδου κατά την ηχογράφηση του θορύβου, σε συνδυασμό με την ιδιαίτερα μεγάλη ηχητική ένταση που προέρχεται από μία ηχητική πηγή όπως αυτή, προκάλεσαν την εμφάνιση έντονου ψαλιδισμού στον αποθηκευμένο ήχο.

Αυτό σε άλλη περίπτωση θα ήταν απαγορευτικό για την αξιοποίηση του συγκεκριμένου ήχου. Επειδή όμως στην περίπτωσή μας, ενδιαφέρει σαν θόρυβος, το δείγμα κρίνεται αποδεκτό και θα συμπεριληφθεί στην λίστα των εξεταζόμενων θορύβων.



Εικόνα 5.8. Κυματομορφή αλυσοπρίονου - θόρυβος 2 (N02)

Ο συγκεκριμένος θόρυβος, θα επιλεγεί επειδή όπως αποτυπώνεται και στο φασματογράφημα της Εικόνας 5.9, εμφανίζει ενδιαφέρουσα χρονική διαφοροποίηση στο περιεχόμενο, με ύπαρξη μικρών παύσεων και παρουσία ηχητικής ενέργειας σε συχνότητες, ακόμα και λίγο κάτω από τα **100Hz**.



Εικόνα 5.9. Φασματογράφημα αλυσοπρίονου - θόρυβος 2 (N02)

Από το φασματογράφημα φαίνεται ότι υπάρχουν αρμονικές συνιστώσες με κυριότερη αυτή στην περιοχή των **720Hz**.

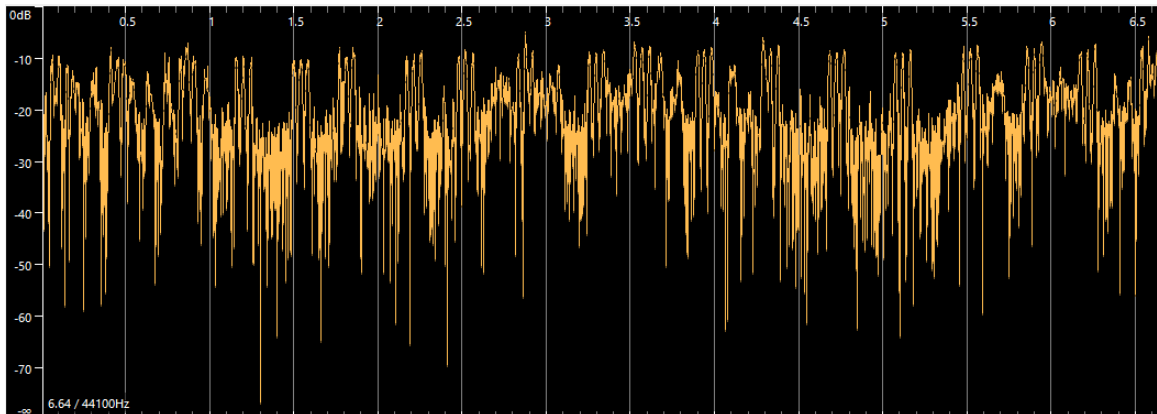
5.1.2.3 Θόρυβος 3 (N03)

Ο **θόρυβος 3 (N03)**, θα προκύψει από την μίξη δύο διαφορετικών ηχητικών σημάτων, που συχνά συνυπάρχουν στο εξοχικό φυσικό περιβάλλον. Το τιτίβισμα των πουλιών θα προέρχεται από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Birds Chirping - Sound effects**», από τον χρήστη **Kieran Ripoll**. Το αρχείο αυτό, με άδεια ελεύθερης χρήσης, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/birds-chirping-sound-effects>.

Ο ήχος του γρύλλου (τριζονιού) που θα ακούγεται, θα προέρχεται επίσης από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αποθηκευμένο αρχείο της βάσης, με ονομασία «**Higher Cricket Chirp**», από τον χρήστη **Jim John Marks**. Το αρχείο

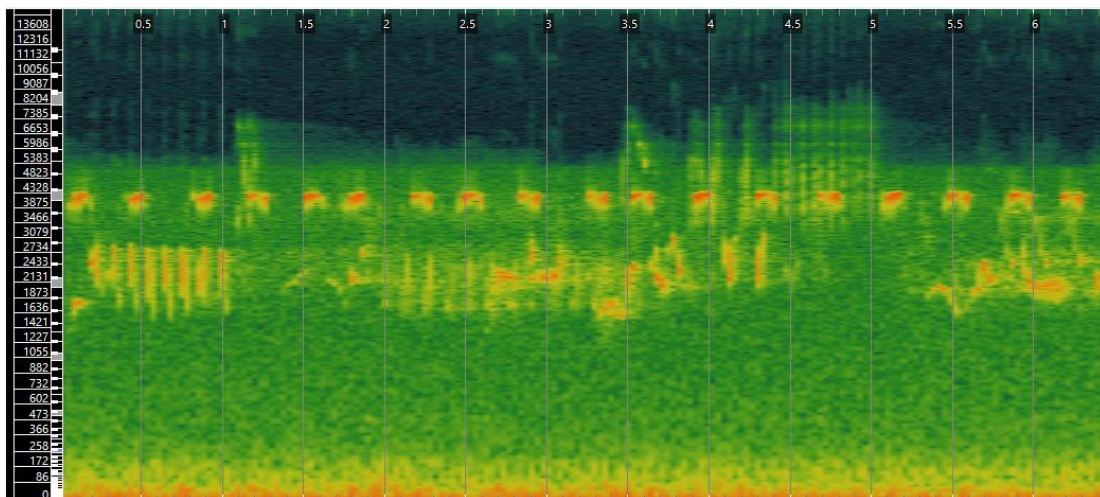
αυτό, επίσης με άδεια ελεύθερης χρήσης, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/higher-cricket-chirp>.

Η κυματομορφή του πλάτους αυτού του θορύβου, αποτυπώνεται στην Εικόνα 5.10 που ακολουθεί, στην οποία παρατηρούνται διακυμάνσεις της ηχητικής στάθμης, γύρω από έναν οριζόντιο νοητό άξονα, σε όλη την χρονική διάρκεια του ηχητικού περιεχομένου.



Εικόνα 5.10. Κυματομορφή πουλιών με γρύλο - θόρυβος 3 (N03)

Η επιλογή του συγκεκριμένου θορύβου έγινε με βάση την ιδιαίτερη φασματική συμπεριφορά του, που αποτυπώνεται στο φασματογράφημα της Εικόνας 5.11.



Εικόνα 5.11. Φασματογράφημα πουλιών με γρύλο - θόρυβος 3 (N03)

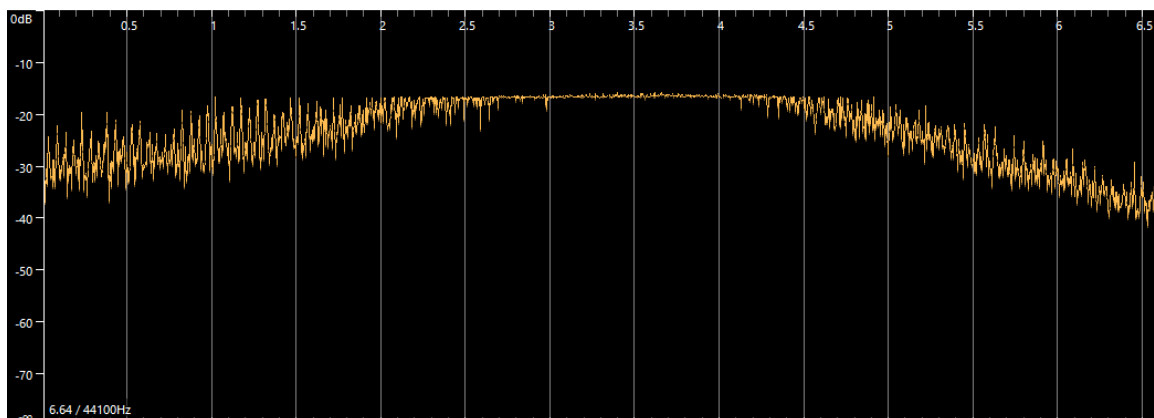
Εδώ ο θόρυβος είναι εντονότερος σε τρεις διακριτές συχνотικές περιοχές. Η πρώτη περιοχή χαρακτηρίζεται από ύπαρξη διαρκούς θορύβου σε εξαιρετικά χαμηλές συχνότητες.

Η δεύτερη περιοχή, που εκτείνεται από τα **1600Hz** έως τα **3000Hz**, περικλείει το κελάϊδισμα των πουλιών, ενώ η τρίτη περιοχή από τα **3900Hz** έως τα **4200Hz**, περιλαμβάνει το σχεδόν περιοδικό ηχητικό αποτύπωμα του γρύλου. Στο φασματογράφημα παρατηρείται επίσης μία αισθητή μείωση της ενέργειας σε συχνότητες πάνω από τα **5300 Hz**, κάτι που μπορεί να αποδοθεί σε επίδραση κάποιας μορφής επεξεργασίας στα στάδια που προηγήθηκαν πριν την αποθήκευση του θορύβου.

5.1.2.4 Θόρυβος 4 (N04)

Ο **θόρυβος 4 (N04)**, θα περιλαμβάνει ήχο ελικοπτέρου, και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Chopper Helicopter Sound - Recording**», από τον χρήστη **lou coeckelberghs**. Το αρχείο αυτό, με άδεια ελεύθερης χρήσης, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/chopper-helicopter-sound-recording>.

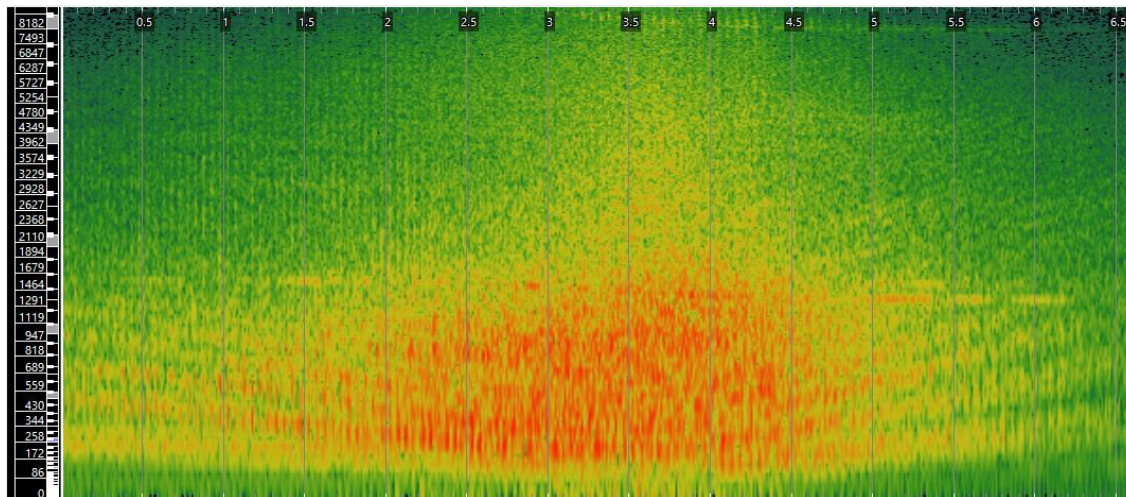
Στην Εικόνα 5.12 που ακολουθεί, αποτυπώνεται η κυματομορφή του πλάτους του συγκεκριμένου θορύβου. Όπως και στην περίπτωση του αλυσοπρίονου, έτσι και εδώ παρατηρείται εμφάνιση έντονου ψαλιδισμού στον αποθηκευμένο ήχο, που προκαλείται προφανώς για παρόμοιους λόγους, όπως και στην προηγούμενη περίπτωση.



Εικόνα 5.12. Κυματομορφή ελικοπτέρου - θόρυβος 4 (N04)

Ο λόγος για τον οποίο θα επιλεγεί συγκεκριμένος θόρυβος, γίνεται οπτικά φανερός στο φασματογράφημα της Εικόνας 5.13. Εδώ η ηχητική ενέργεια καταλαμβάνει μία

ολόκληρη περιοχή συχνοτήτων, και μεταβάλλεται χρονικά, με επίκεντρο της τον χρόνο από τα 2 sec έως τα 4,5 sec, όπου εμφανίζεται και ο ψαλιδισμός.



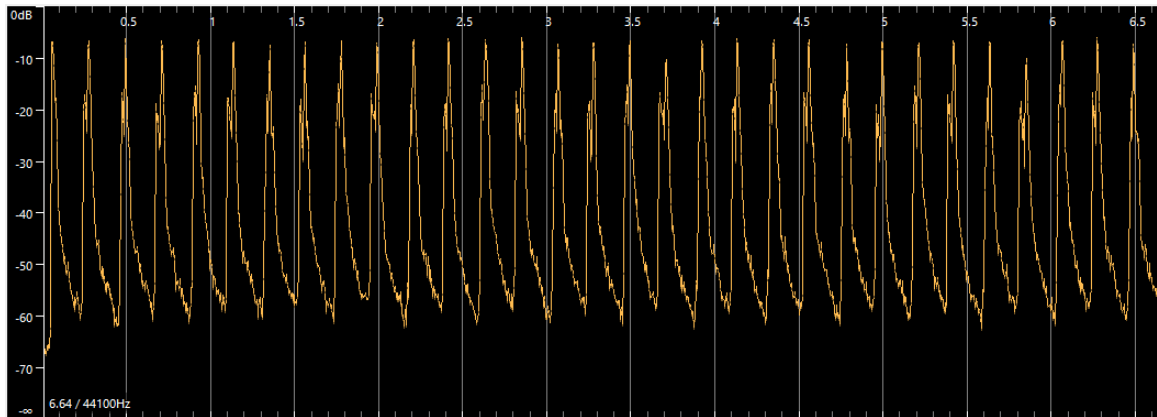
Εικόνα 5.13. Φασματογράφημα ελικοπτέρου - θόρυβος 4 (N04)

Λόγω αυτών των χαρακτηριστικών του, ο θόρυβος αυτός κρίνεται ως ένας από τους πλέον κατάλληλους για να δοκιμάσει τις «αντοχές» των λογισμικών στις διενέργειες των αποθορυβοποιήσεων.

5.1.2.5 Θόρυβος 5 (N05)

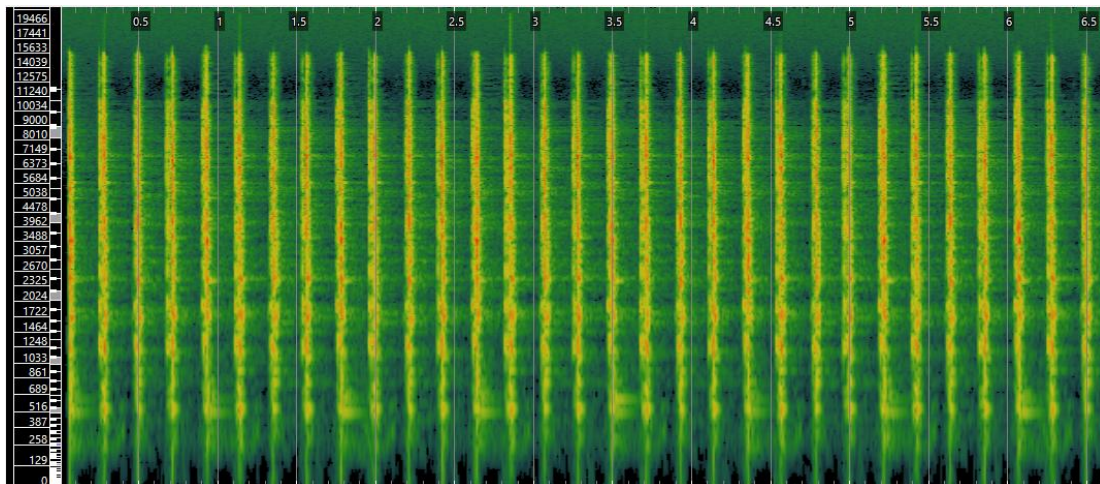
Ο **θόρυβος 5 (N05)**, περιλαμβάνει ήχο από κτύπο ρολογιού τοίχου και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο με ονομασία «**Fast Old Clock - Ticking Sound Recording**», από τον χρήστη **Garrett Williams**. Το αρχείο αυτό, με ελεύθερη άδεια χρήσης, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/fast-old-clock-ticking-sound-recording>.

Η επιλογή του θορύβου γίνεται για την περιοδικότητα που τον διακρίνει, κάτι που αποτυπώνεται εμφανώς στην κυματομορφή του πλάτους του θορύβου, στην Εικόνα 5.14 που ακολουθεί. Ένας ακόμα παράγοντας, που συντελεί στην επιλογή του ήχου, είναι κάλυψη της απαίτησης για συμπερίληψη ήχων από όλα τα περιβάλλοντα, με τον θόρυβο αυτό να αποτελεί αντιπροσωπευτικό δείγμα που προέρχεται από εσωτερικό χώρο.



Εικόνα 5.14. Κυματομορφή ρολογιού τοίχου - θόρυβος 5 (N05)

Ένας ακόμα λόγος για την επιλογή του συγκεκριμένου ήχου, είναι η επίδειξη κρουστικής συμπεριφοράς, με κάθε τον κάθε χτύπο να εκτείνεται συχνοτικά σε μεγάλα ύψη συχνοτήτων, όπως αποτυπώνεται και στο φασματογράφημα της Εικόνας 5.15.



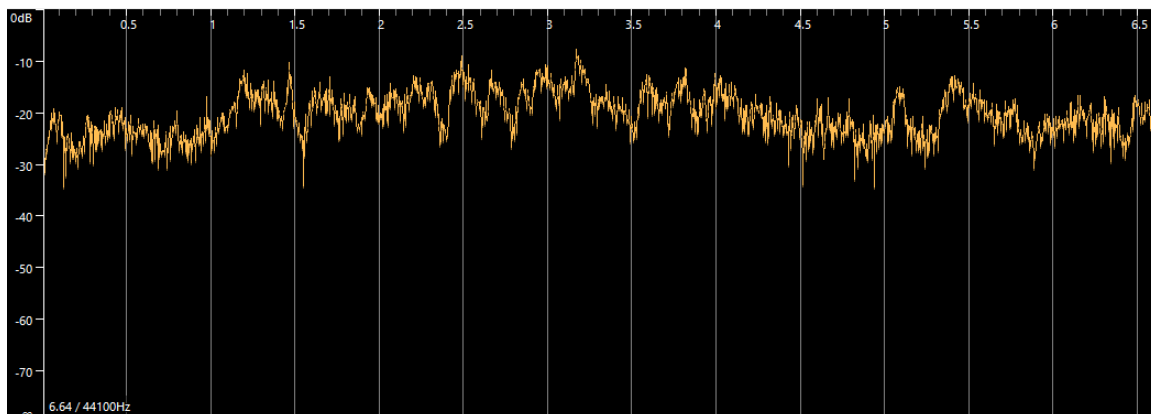
Εικόνα 5.15. Φασματογράφημα ρολογιού τοίχου - θόρυβος 5 (N05)

5.1.2.6 Θόρυβος 6 (N06)

Ο **θόρυβος 6 (N06)**, θα περιλαμβάνει ομιλίες πλήθους και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Small Crowd Talking**», από τον χρήστη **Antoine DURON**. Το αρχείο αυτό, θα

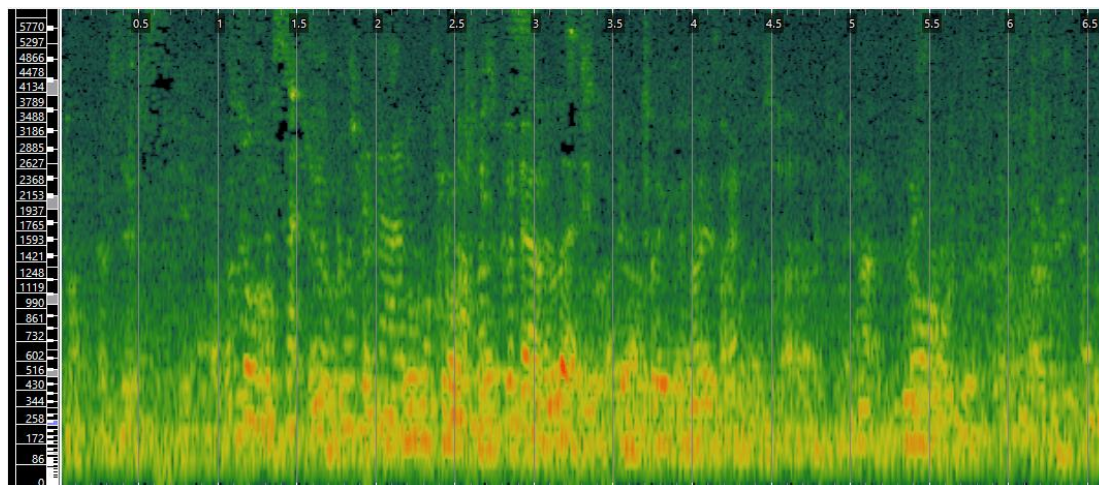
μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/small-crowd-talking>, με ελεύθερη άδεια χρήσης.

Όπως φαίνεται στην Εικόνα 5.16, που αποτυπώνει την κυματομορφή του πλάτους του θορύβου, στην περίπτωση αυτή υπάρχει σχετική σταθερότητα στο επίπεδο της ηχητικής στάθμης, σε όλη την χρονική διάρκεια του δείγματος.



Εικόνα 5.16. Κυματομορφή ομιλίας πλήθους - θόρυβος 6 (N06)

Το φασματογράφημα του θορύβου αποτυπώνεται στην επόμενη Εικόνα.



Εικόνα 5.17. Φασματογράφημα ομιλίας πλήθους - θόρυβος 6 (N06)

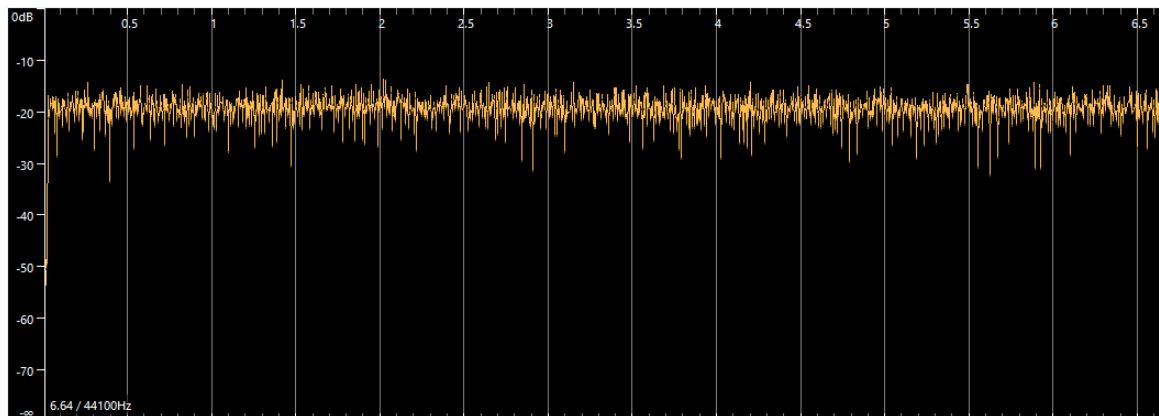
Η επιλογή του συγκεκριμένου θορύβου, κρίνεται αναγκαία για να εξετασθεί η επίδοση των λογισμικών TN, στην περίπτωση που στο σήμα συμπεριλαμβάνονται φωνές τρίτων προσώπων. Με δεδομένο ότι οι ομιλίες έχουν παρόμοιο φασματικό περιεχόμενο, η

διερεύνηση της ικανότητας διαχωρισμού και απομόνωσης της κύριας από τις δευτερεύουσες φωνές αποκτά ιδιαίτερο ενδιαφέρον, σαν μέτρο αποτελεσματικότητας των εξεταζόμενων εργαλείων. Παράλληλα, η συμπερίληψη του συγκεκριμένου είδους θορύβου στα δείγματα προς αποθορυβοποίηση, επιβάλλεται από το γεγονός ότι η απαίτηση για εξάλειψή του είναι ιδιαίτερα μεγάλη, καθώς απαντάται ευρύτατα σε διάφορους χώρους.

5.1.2.7 Θόρυβος 7 (N07)

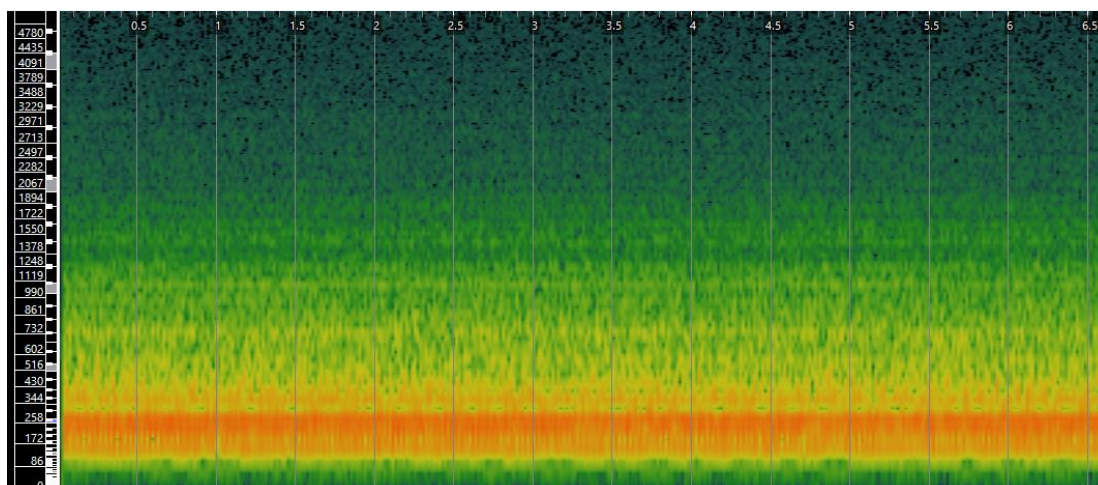
Ο **θόρυβος 7 (N07)**, θα περιλαμβάνει βουητό ανεμιστήρα και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Fan Sound - Reording FX**», από τον χρήστη **Hax**. Το αρχείο αυτό, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/fan-sound-reording-fx>, με άδεια ελεύθερης χρήσης.

Η κυματομορφή του θορύβου, αποτυπώνεται στην Εικόνα 5.18. Εδώ, όπως είναι αναμενόμενο, δεν παρατηρούνται διακυμάνσεις της ηχητικής στάθμης, η οποία παραμένει σταθερή σε όλη την χρονική διάρκεια του ηχητικού αποσπάσματος.



Εικόνα 5.18. Κυματομορφή βουητό ανεμιστήρα - θόρυβος 7 (N07)

Όσον αφορά στην συχνοτική συμπεριφορά του θορύβου, και αυτή εμφανίζει σταθερότητα στην εξέλιξη του χρόνου, όπως αποτυπώνεται στο φασματογράφημα της Εικόνας 5.11 που ακολουθεί.



Εικόνα 5.19. Φασματογράφημα βουητό ανεμιστήρα - θόρυβος 7 (N07)

Ο συγκεκριμένος θόρυβος εκτείνεται από τα **100Hz** έως τα **300Hz**, επικαλύπτοντας το περιεχόμενο των επιλεγμένων φωνών, στην συχνοτική αυτή ζώνη.

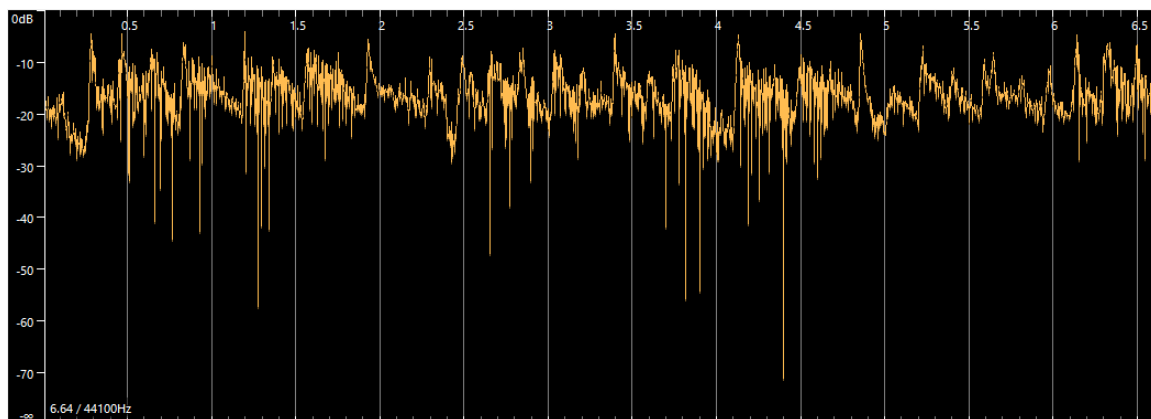
Η επιλογή του θορύβου πραγματοποιείται για να εξετασθεί η απόδοση των τεχνικών TN σε ένα είδος θορύβου, που η συμβατική του αντιμετώπιση γίνεται κυρίως με την χρήση κατάλληλων φίλτρων αποκοπής ζώνης.

5.1.2.8 Θόρυβος 8 (N08)

Ο **θόρυβος 8 (N08)**, θα περιλαμβάνει μουσικό κομμάτι και θα προέλθει από την βάση μουσικών αρχείων **Free Music Archive - FMA**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Latin Life**», από το άλμπουμ «**Upbeat Street**» του χρήστη **Ketsa**. Το αρχείο αυτό, θα μεταφορτωθεί με ελεύθερη άδεια χρήσης, από την ηλεκτρονική διεύθυνση <https://freemusicarchive.org/music/Ketsa/upbeat-street/latin-life/>.

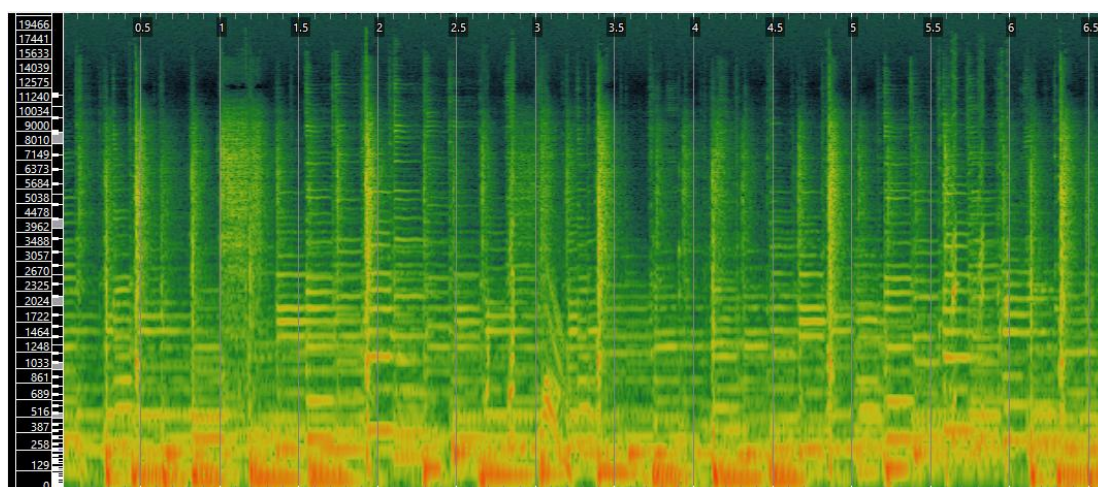
Η λόγος για τον οποίο θα επιλεγεί το συγκεκριμένο είδος θορύβου, στηρίζεται στο ότι πολύ συχνά εμφανίζεται η ανάγκη για απομόνωση της φωνής, καταστέλλοντας τον ήχο από τα υπόλοιπα όργανα που συμμετέχουν σε μία μουσική σύνθεση. Οι συμβατικές τεχνικές όμως, στοχεύουν συνήθως στο αντίθετο αποτέλεσμα, δηλαδή την εξάλειψη της φωνής, και την διατήρηση της μουσικής, όπως συμβαίνει με την μέθοδο καραόκε (Φλώρος, 2013). Έτσι η επίδοση της TN παρουσιάζει εδώ ιδιαίτερο ενδιαφέρον.

Η κυματομορφή του θορύβου, αποτυπώνεται στην Εικόνα 5.20. Εδώ, παρατηρούνται περιορισμένες διακυμάνσεις της ηχητικής στάθμης, σε όλη την χρονική διάρκεια του ηχητικού αποσπάσματος.



Εικόνα 5.20. Κυματομορφή μουσικής - θόρυβος 8 (N08)

Το φασματογράφημα του θορύβου αποτυπώνεται στην Εικόνα 5.21.



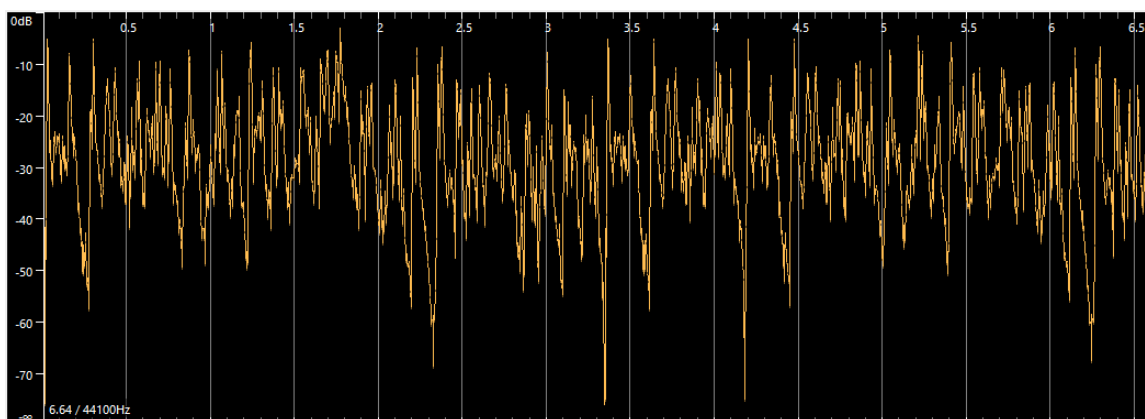
Εικόνα 5.21. Φασματογράφημα μουσικής - θόρυβος 8 (N08)

Συχνотικά ό θόρυβος εμφανίζεται από συχνότητες τοποθετημένες ιδιαίτερα χαμηλά και εκτείνεται έως τα **500Hz**, χωρίς να λείπουν και οι χαρακτηριστικές αρμονικές των μουσικών οργάνων, που εμφανίζονται σε υψηλότερες περιοχές, όπως αποτυπώνεται στο προηγούμενο φασματογράφημα.

5.1.2.9 Θόρυβος 9 (N09)

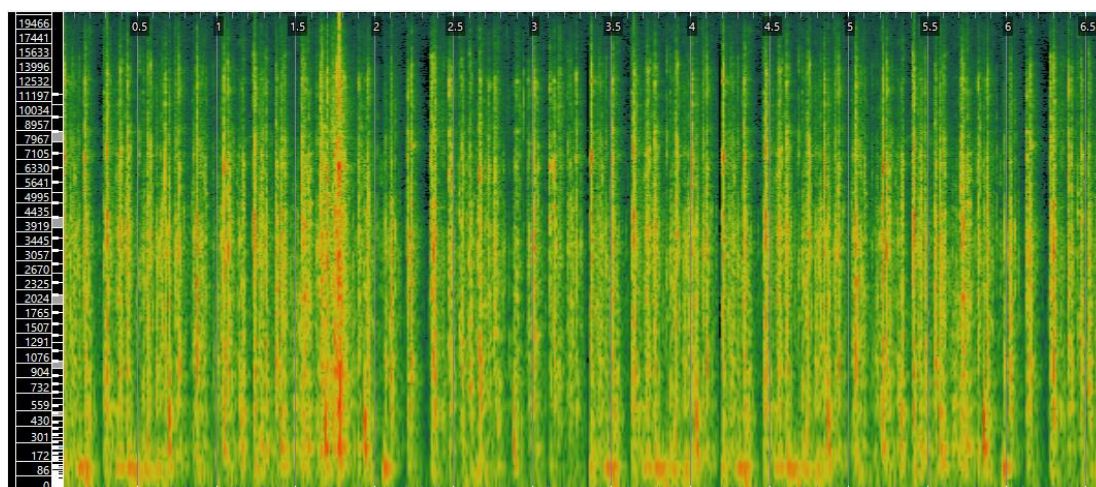
Ο **θόρυβος 9 (N09)**, θα περιλαμβάνει θόρυβο πληκτρολόγησης και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Real Life Recording - Keyboard Typing FX - Recording**», από τον χρήστη **user2891067436466**. Το αρχείο αυτό, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/real-life-recording-keyboard-typing-fx-recording>, με άδεια ελεύθερης χρήσης.

Η κυματομορφή του θορύβου, αποτυπώνεται στην Εικόνα 5.22.



Εικόνα 5.22. Κυματομορφή πληκτρολόγησης - θόρυβος 9 (N09)

Το φασματογράφημα του θορύβου αποτυπώνεται στην Εικόνα 5.23 που ακολουθεί.



Εικόνα 5.23. Φασματογράφημα πληκτρολόγησης - θόρυβος 9 (N09)

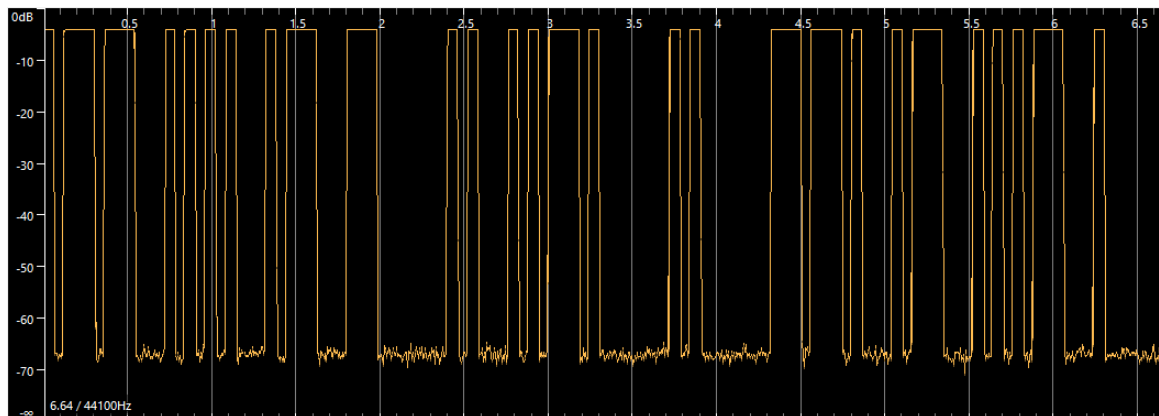
Στην περίπτωση αυτή παρουσιάζονται έντονες διακυμάνσεις της ηχητικής στάθμης, η οποίες παρατηρούνται αδιάληπτα, σε όλη την χρονική διάρκεια του θορύβου. Παράλληλα ο θόρυβος εμφανίζει κρουστική συμπεριφορά, που εκτείνεται συχνοτικά σχεδόν σε όλο το ακουστικό φάσμα.

Ο συγκεκριμένος θόρυβος θα επιλεγεί, επειδή είναι χαρακτηριστικός σε περιβάλλον εσωτερικών χώρων, όπου παρουσιάζει ιδιαίτερα συχνή εμφάνιση, σε εργασίες που γίνονται με την χρήση υπολογιστή.

5.1.2.10 Θόρυβος 10 (N10)

Ο **θόρυβος 10 (N10)**, περιλαμβάνει σήματα Μορς και θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Synth Ride Morse Code**», από τον χρήστη **Alt Account**. Το αρχείο αυτό, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/synth-ride-morse-code>, με ελεύθερη άδεια χρήσης.

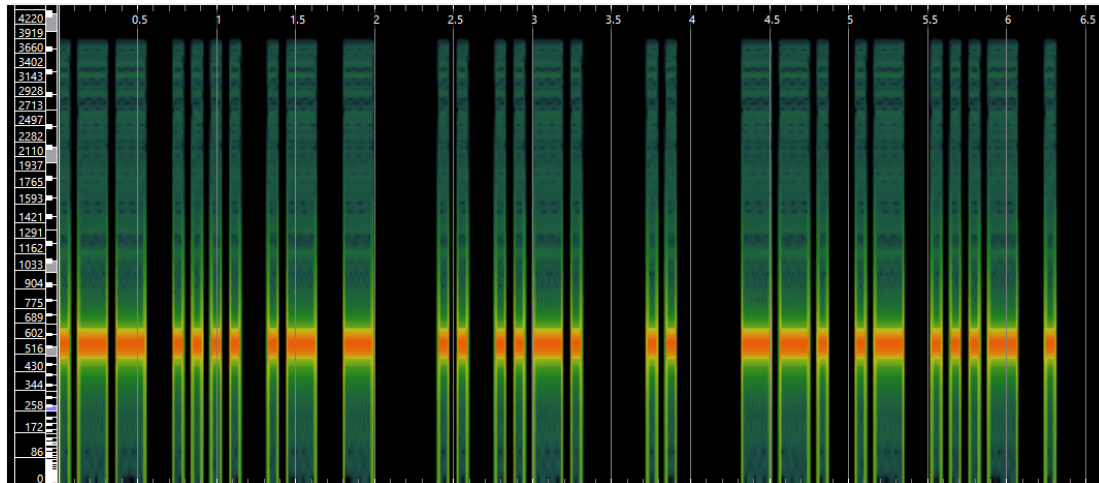
Η ιδιαίτερης μορφής κυματομορφή του θορύβου, αποτυπώνεται στην Εικόνα 5.18.



Εικόνα 5.24. Κυματομορφή σημάτων Μορς - θόρυβος 10 (N10)

Εδώ, εμφανίζεται σταθερή μέγιστη τιμή της ηχητικής στάθμης όπου υπάρχει τονικό σήμα, εναλλασσόμενη με απότομες βυθίσεις και ελαχιστοποίηση των τιμών κατά την διάρκεια των παύσεων, σε όλη την χρονική διάρκεια του ηχητικού αποσπάσματος.

Ο συγκεκριμένος θόρυβος είναι δημιουργημένος συνθετικά, δεν είναι δηλαδή ηχογραφημένος. Για τον λόγο αυτό, στο φασματογράφημα που αποτυπώνεται στην Εικόνα 5.25, εμφανίζεται παντελής έλλειψη περιεχομένου στα σημεία των παύσεων, ενώ όπως επίσης φαίνεται, ο θόρυβος έχει φιλτραριστεί για συχνότητες πάνω από τα **3850Hz**, με τον τόνο να εμφανίζεται στην περιοχή των **550Hz**.



Εικόνα 5.25. Φασματογράφημα σημάτων Μορς - θόρυβος 10 (N10)

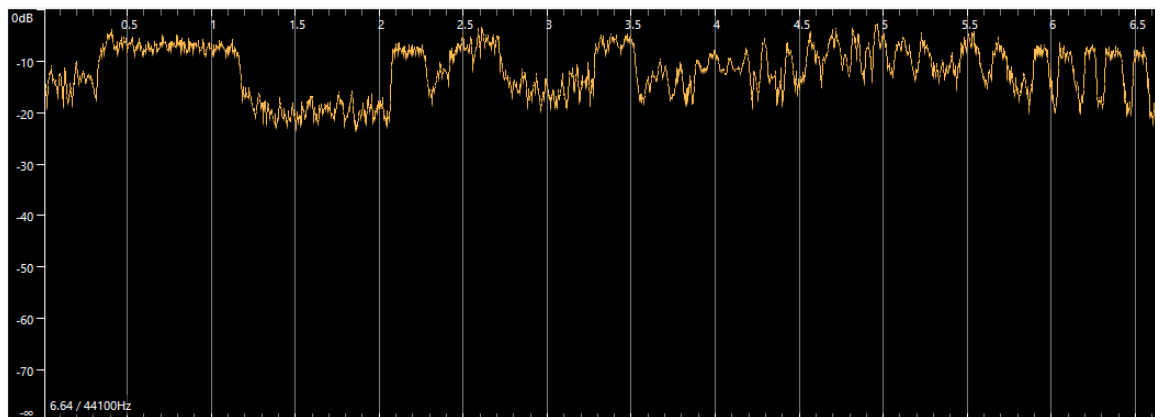
Όπως και στην περίπτωση του ανεμιστήρα, η επιλογή πραγματοποιήθηκε για να εξετασθεί η απόδοση των τεχνικών TN σε ένα είδος θορύβου, που η συμβατική του αντιμετώπιση γίνεται κυρίως με την χρήση κατάλληλων φίλτρων αποκοπής ζώνης, με διαφορετικά όμως χαρακτηριστικά αυτή την φορά, καθώς εδώ ο ήχος είναι διακοπτόμενος και σε διαφορετική συχνотική περιοχή.

5.1.2.11 Θόρυβος 11 (N11)

Ο **θόρυβος 11 (N11)**, θα προέλθει από την μίξη δύο διαφορετικών ηχητικών σημάτων, ώστε το αποτέλεσμα να είναι πιο περιεκτικό. Η σειρά που ακούγεται θα προέλθει από την βάση ηχητικών δεδομένων **Sample Focus**. Πρόκειται για το αρχείο της βάσης με ονομασία «**Police Siren SFX**», από τον χρήστη **Yael Trujillo**. Το αρχείο αυτό, θα μεταφορτωθεί με ελεύθερη άδεια χρήσης, από την ηλεκτρονική διεύθυνση <https://samplefocus.com/samples/police-siren-sfx>.

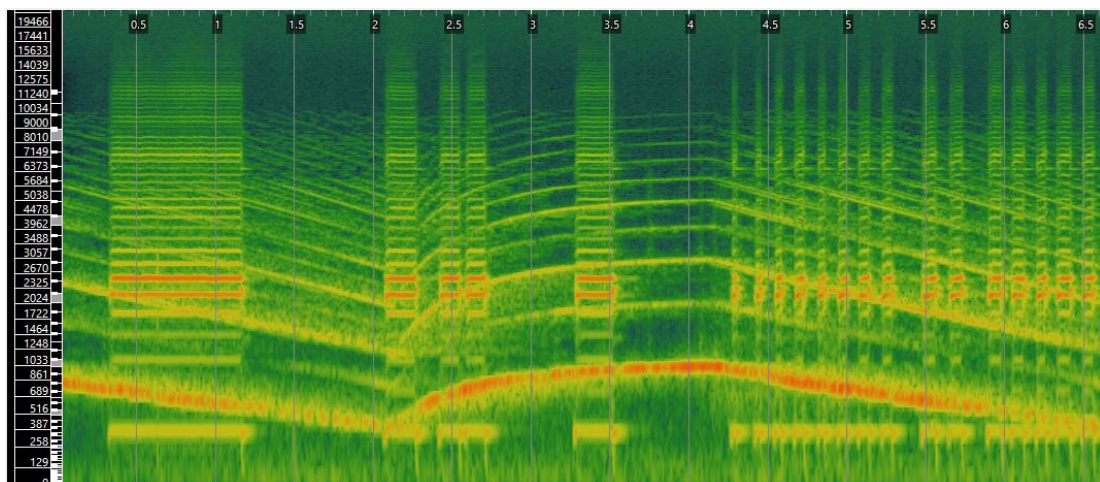
Οι κόρνες που ακούγονται, θα προέλθουν από την βάση ηχητικών δεδομένων του BBC. Πρόκειται για το αποθηκευμένο αρχείο της βάσης, με ονομασία «**Motor Horns - Single Tone Motor Horn sounded, long impatient blasts. (Electric motor horn.)**» και κωδικό αριθμό **07037290**. Το αρχείο αυτό, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://sound-effects.bbcrewind.co.uk/search?q=07037290>, με ελεύθερη άδεια χρήσης για μη εμπορική χρήση.

Η κυματομορφή του θορύβου, αποτυπώνεται στην Εικόνα 5.26. Εδώ η ηχητική στάθμη παρουσιάζει αυξημένες τιμές στις χρονικές στιγμές που ακούγεται ο ήχος του κορναρίσματος.



Εικόνα 5.26. Κυματομορφή σειρήνας και κόρνας - θόρυβος 11 (N11)

Το φασματογράφημα του θορύβου αποτυπώνεται στην Εικόνα 5.27 που ακολουθεί.



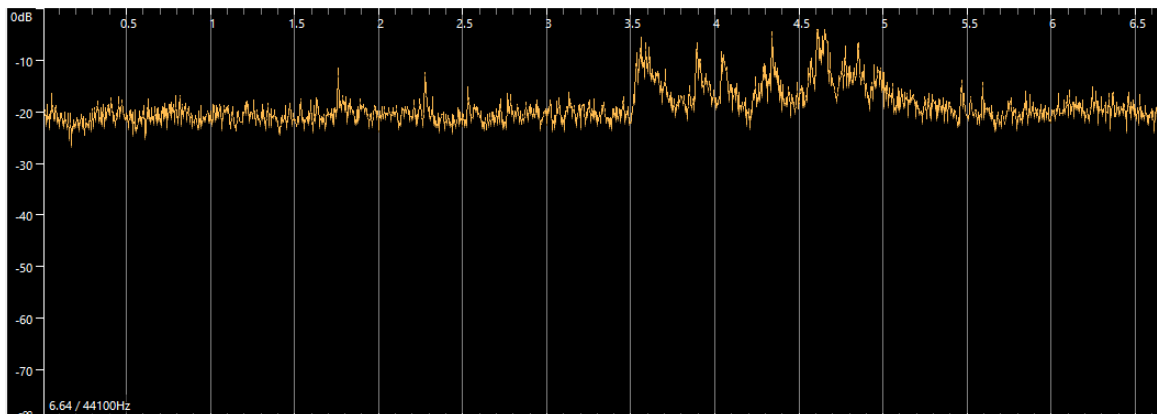
Εικόνα 5.27. Φασματογράφημα σειρήνας και κόρνας - θόρυβος 11 (N11)

Ο τύπος αυτού του μεικτού θορύβου, θα επιλεγεί ως αντιπροσωπευτικός, για την εμφάνισή του στους εξωτερικούς χώρους ενός αστικού περιβάλλοντος. Συνδυάζει δύο συχνοτικά εστιασμένους επιμέρους θορύβους, όπου ο ένας (σειρήνα) μεταβάλλεται χρονικά, ενώ ο άλλος (κόρνα) διατηρεί σταθερή την συχνότητά του. Και στις δύο περιπτώσεις παρατηρείται η συμβολή των αρμονικών στο τελικό αποτέλεσμα.

5.1.2.12 Θόρυβος 12 (N12)

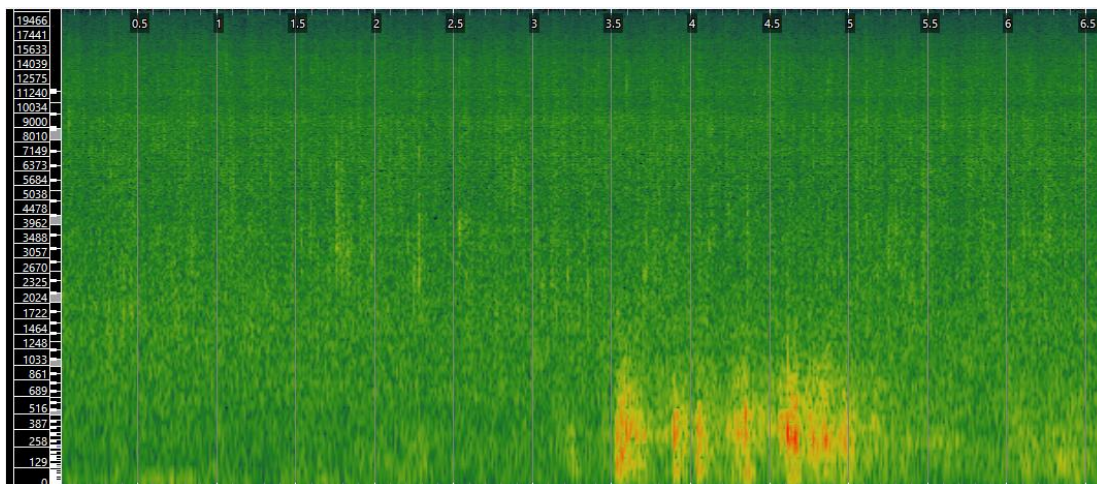
Ο **θόρυβος 12 (N12)**, θα περιλαμβάνει τον ήχο έντονης βροχής με κεραυνούς και θα προέλθει από την βάση ηχητικών δεδομένων του **BBC**. Πρόκειται για το αποθηκευμένο αρχείο της βάσης, με ονομασία **«Rain Atmosphere - constant rain, with occasional thunder in background (generic)»** και κωδικό αριθμό **NHU10352250**. Το αρχείο αυτό, με ελεύθερη άδεια χρήσης για μη εμπορική χρήση, θα μεταφορτωθεί από την ηλεκτρονική διεύθυνση <https://sound-effects.bbcrewind.co.uk/search?q=NHU10352250>.

Η κυματομορφή του θορύβου, αποτυπώνεται στην Εικόνα 5.28. Δεν παρατηρούνται διακυμάνσεις της στάθμης, η οποία παραμένει σχετικά σταθερή, σε όλη την χρονική διάρκεια του ηχητικού αποσπάσματος, με εξαίρεση τις μικρές εξάρσεις που συμπίπτουν με τις στιγμές που ακούγονται οι απόμακροι κερανοί.



Εικόνα 5.28. Κυματομορφή δυνατής βροχής - θόρυβος 12 (N12)

Σχετικά την συχνотική κατανομή του θορύβου, αυτή απλώνεται ομοιόμορφα σε όλο το φάσμα των συχνотήτων, όπως αποτυπώνεται και στο φασματογράφημα της Εικόνας 5.29 που ακολουθεί.



Εικόνα 5.29. Φασματογράφημα δυνατής βροχής - θόρυβος 12 (N12)

Εδώ διακρίνεται η επίδραση των κεραυνών, στο χρονικό διάστημα από τα **3 sec** έως τα **5 sec**. Ο συγκεκριμένος θόρυβος εκτείνεται από τα **100Hz** έως τα **300Hz**, επικαλύπτοντας το περιεχόμενο των επιλεγμένων φωνών, στην συχνотική αυτή ζώνη.

Ο θόρυβος αυτός, θα επιλεγεί ως αντιπροσωπευτικός των καιρικών συνθηκών που επικρατούν συχνά στο εξωτερικό περιβάλλον. Παράλληλα, τα χαρακτηριστικά λευκού θορύβου που διακρίνουν τον ήχο της βροχής, θα συμβάλλουν ώστε να εξετασθεί η απόδοση των λογισμικών TN και σε αυτό το είδος θορύβου.

5.1.3 Μίξη ομιλίας με θόρυβο για δημιουργία ενθόρυβων αρχείων

Μετά την επιλογή των δειγμάτων ομιλίας και θορύβου, θα πραγματοποιηθεί η μίξη τους, ώστε να προκύψουν τα τελικά ενθόρυβα αρχεία. Από κάθε θόρυβο θα επιλεγούν τα κατάλληλα τμήματα, ίδιας χρονικής διάρκειας για όλα, ίσης με **6,6 sec**.

Η διαδικασία θα υλοποιηθεί με την χρήση του σταθμού επεξεργασίας ψηφιακού ήχου (DAW) **Audacity**.

Τα ψηφιακά αρχεία ήχου θα δημιουργηθούν στερεοφωνικά, σε μορφή «**wav**» και ακουστική ποιότητα προτύπου **CD** (Ζαρούχας et al, 2013), ρυθμού δειγματοληψίας **44100 Hz** και ευκρίνειας **16 bit**. Στο σημείο αυτό πρέπει να σημειωθεί, ότι η διαφοροποίηση χαρακτηριστικών, όπως είναι ο ρυθμός δειγματοληψίας ενός δείγματος, μπορεί να επηρεάσει την εκπαίδευση των αλγορίθμων, με αποτέλεσμα να επηρεαστεί η τελική απόδοση των εργαλείων Τεχνητής Νοημοσύνης (Βρύσης, 2019).

Για αντιπροσωπευτικότητα στο αποτέλεσμα, οι δύο διαφορετικές εξεταζόμενες ομιλίες θα διαχωρισθούν ισομερώς για την μίξη τους με τους επιλεγμένους θορύβους με τυχαίο τρόπο. Έτσι θα δημιουργηθούν **έξι (6)** ενθόρυβα αρχεία που θα περιλαμβάνουν την αντρική φωνή και **έξι (6)** που θα περιλαμβάνουν την γυναικεία ομιλία. Η κατανομή των φωνών στους θορύβους αποτυπώνεται στον Πίνακα 5.2 που ακολουθεί.

Πίνακας 5.2. Κατανομή αντρικής και γυναικείας ομιλίας ανά εξεταζόμενο θόρυβο

Περιγραφή θορύβου	V01 - Ομιλία γυναίκας	V02 -Ομιλία άνδρα
N01 - Γάβγισμα σκύλου		✓
N02 - Αλυσσοπρίονο		✓
N03 - Πουλιά και γρύλλος	✓	
N04 - Ελικόπτερο		✓
N05 - Ρολόι τοίχου		✓
N06 - Ομιλίες πλήθους	✓	
N07 - Βουητό ανεμιστήρα	✓	
N08 - Μουσική		✓
N09 - Πληκτρολόγηση	✓	
N10 - Σήματα Μορς		✓
N11 - Σειρήνα και κόρνες	✓	
N12 - Δυνατή βροχή	✓	

Η μίξη της κάθε ομιλίας με τον αντίστοιχο θόρυβο, θα γίνει φροντίζοντας οι σχετικές ηχητικές στάθμες (SNR) να είναι ρεαλιστικές, αντικατοπτρίζοντας τις συνήθειες

πραγματικές συνθήκες. Παράλληλα όμως θα γίνει η επιλογή να υπάρχει αρκετός θόρυβος, ώστε να δοκιμαστούν οι «αντοχές» των εργαλείων αποθορυβοποίησης. Έτσι, θα γίνουν διεξοδικές δοκιμές με διάφορα κέρδη (Gain) στα δείγματα ομιλίας-θορύβου, για αντιπροσωπευτικά αποτελέσματα, επιτρέποντας την ύπαρξη αυξημένης στάθμης θορύβου, ώστε να διερευνηθούν οι επιδόσεις της TN κάτω από συνθήκες έντονου θορύβου.

Από τις δοκιμές αποθορυβοποίησης που πραγματοποιούνται την τρέχουσα στιγμή της συγγραφής της εργασίας, προκύπτει ότι δεν αντιμετωπίζεται ιδιαίτερο πρόβλημα, αφού για παράδειγμα ο θόρυβος του πληκτρολογίου ή ο κτύπος του ρολογιού εξαλείφεται και ας έχει στάθμη σαν το μικρόφωνο να βρίσκεται ακριβώς δίπλα του.

Για την καταγραφή του σηματοθορυβικού λόγου (SNR) των ενθόρυβων δειγμάτων, θα υπολογισθεί η ολική μέση τιμή στάθμης (total RMS), για όλα τα σήματα ομιλίας και θορύβου. Τα αποτελέσματα θα αποτυπωθούν στον Πίνακα 5.3 που ακολουθεί. Στον ίδιο πίνακα, αποτυπώνονται και οι τελικές ρυθμίσεις στο κέρδος των σημάτων, που κρίνεται αναγκαίο να γίνουν για την απόδοση πιο αντιπροσωπευτικού αποτελέσματος.

Πίνακας 5.3. Μέσες τιμές στάθμης (RMS) και ρύθμιση κέρδους (Gain) σημάτων

Ηχητικό σήμα	RMS (dB)	Gain (dB)	RMS + Gain (dB)
V01 - Ομιλία γυναίκας	-22,1416	+3,0000	-19,1416
V02 - Ομιλία άνδρα	-23,7958	+3,0000	-20,7958
N01 - Γάβγισμα σκύλου	-18,0578	-	-18,0578
N02 - Αλυσοπρίονο	-24,2835	-3,0000	-27,2835
N03 - Πουλιά και γρύλλος	-20,7649	-	-20,7649
N04 - Ελικόπτερο	-22,5598	-3,0000	-25,5598
N05 - Ρολόι τοίχου	-29,0994	-	-29,0994
N06 - Ομιλίες πλήθους	-24,3427	-	-24,3427
N07 - Βουητό ανεμιστήρα	-23,7661	-	-23,7661
N08 - Μουσική	-19,5255	-	-19,5255
N09 - Πληκτρολόγηση	-28,3847	-	-28,3847
N10 - Σήματα Μορς	-11,0992	-	-11,0992

N11 - Σειρήνα και κόρνες	-16,1132	-	-16,1132
N12 - Δυνατή βροχή	-21,3683	-3,0000	-24,3683

Όπως φαίνεται στον προηγούμενο πίνακα, έγιναν παρεμβάσεις στις στάθμες των φωνών και ορισμένων θορύβων, μετά από επαναληπτική διαδικασία μίξεων και ελέγχων SNR, δίνοντας τιμές κέρδους **+3 dB** στις ομιλίες και **-3 dB** στους θορύβους: **N02 - Αλυσοπρίονο**, **N04 - Ελικόπτερο** και **N12 - Δυνατή βροχή**, ώστε να αντικατοπτρίζονται καλύτερα και οι πραγματικές συνθήκες. Έτσι το SNR των ενθόρυβων δειγμάτων βελτιώνεται συνολικά κατά **+3 dB** ή **+6 dB** κατά περίπτωση. Οι τελικές SNR τιμές των ενθόρυβων ηχητικών σημάτων καταγράφονται στον Πίνακα 5.4, που ακολουθεί.

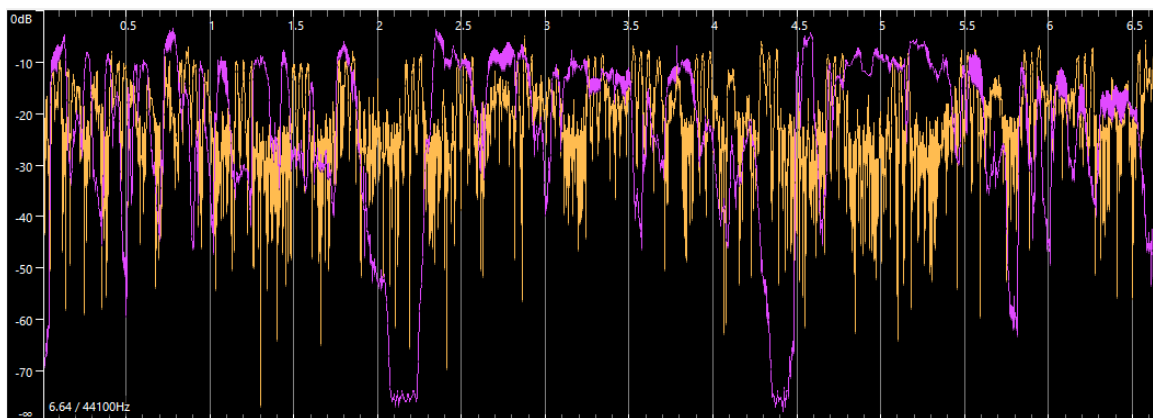
Πίνακας 5.4. Λόγος σήματος προς θόρυβο (SNR) τελικών ενθόρυβων δειγμάτων

Ενθόρυβο δείγμα	SNR (dB)
N01V02 - Γάβγισμα σκύλου με αντρική ομιλία	-2,7380
N02V02 - Αλυσοπρίονο με αντρική ομιλία	6,4877
N03V01 - Πουλιά και γρύλλος με γυναικεία ομιλία	1,6233
N04V02 - Ελικόπτερο με αντρική ομιλία	4,7640
N05V02 - Ρολόι τοίχου με αντρική ομιλία	8,3036
N06V01 - Ομιλίες πλήθους με γυναικεία ομιλία	5,2011
N07V01 - Βουητό ανεμιστήρα με γυναικεία ομιλία	4,6245
N08V02 - Μουσική με αντρική ομιλία	-1,2703
N09V01 - Πληκτρολόγηση με γυναικεία ομιλία	9,2431
N10V02 - Σήματα Μορς με αντρική ομιλία	-9,6966
N11V01 - Σειρήνα και κόρνες με γυναικεία ομιλία	-3,0284
N12V01- Δυνατή βροχή με γυναικεία ομιλία	5,2267

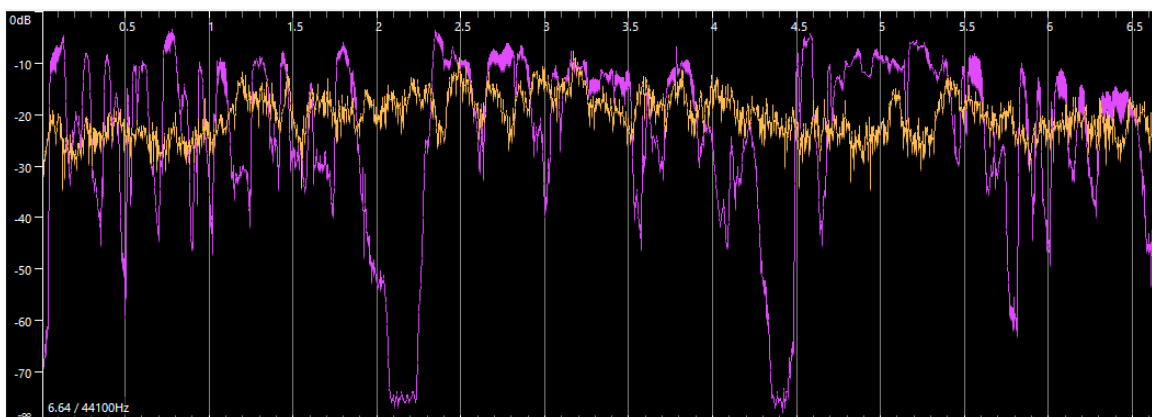
Στον προηγούμενο πίνακα, με κόκκινο αποτυπώνονται οι αρνητικές τιμές SNR, στις περιπτώσεις όπου ο θόρυβος υπερಿಸχύει του «ωφέλιμου» σήματος της ομιλίας. Όπως διαπιστώνεται με βάση αυτές τις μετρήσεις, τα εργαλεία λογισμικού Τεχνητής Νοημοσύνης

επιτυγχάνουν ιδιαίτερα καλά αποτελέσματα κατά την υλοποίηση των αποθρομβοποιήσεων, σε ιδιαίτερα χαμηλές ή ακόμα και σε αρνητικές τιμές SNR, όπως οι παραπάνω.

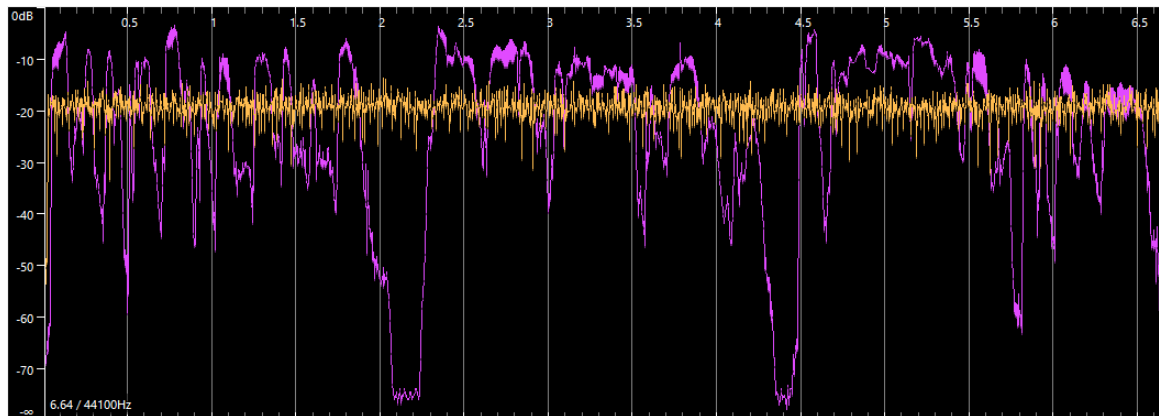
Στις επόμενες Εικόνες (από 5.30 έως 5.53), ακολουθούν συνδυαστικές απεικονίσεις των κυματομορφών και των φασματογραφημάτων των ενθόρυβων ηχητικών δειγμάτων. Οι κυματομορφές εμφανίζονται με χρωματική κωδικοποίηση, όπως και πριν, με το μωβ να αντιστοιχεί στο γυναικείο σήμα, το μπλε στο αντρικό σήμα και το πορτοκαλί στο σήμα θορύβου. Για καλύτερη συγκριτική εποπτεία, στην σειρά εμφάνισης προηγούνται αυτές με την γυναικεία ομιλία και έπονται εκείνες που περιλαμβάνουν την αντρική ομιλία.



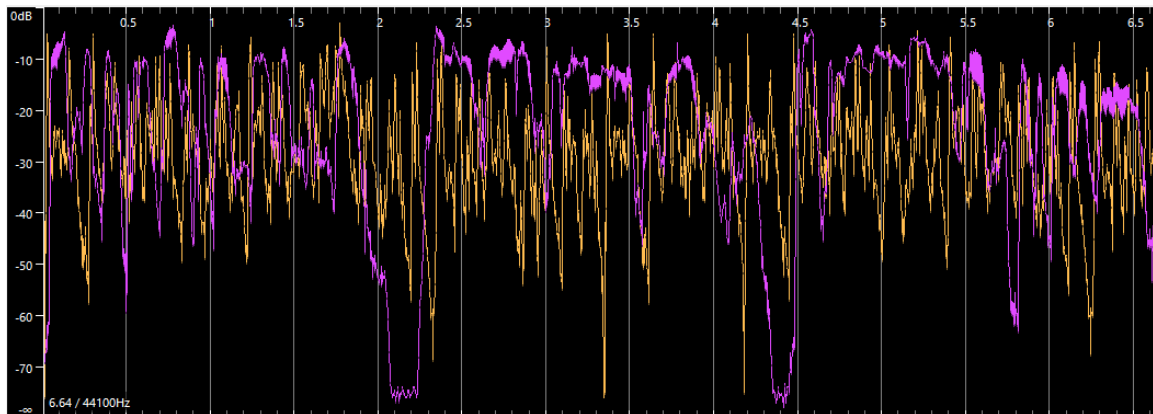
Εικόνα 5.30. Κυματομορφή πουλιών και γρύλλου με γυναικεία ομιλία - ενθόρυβο 03 (N03V01)



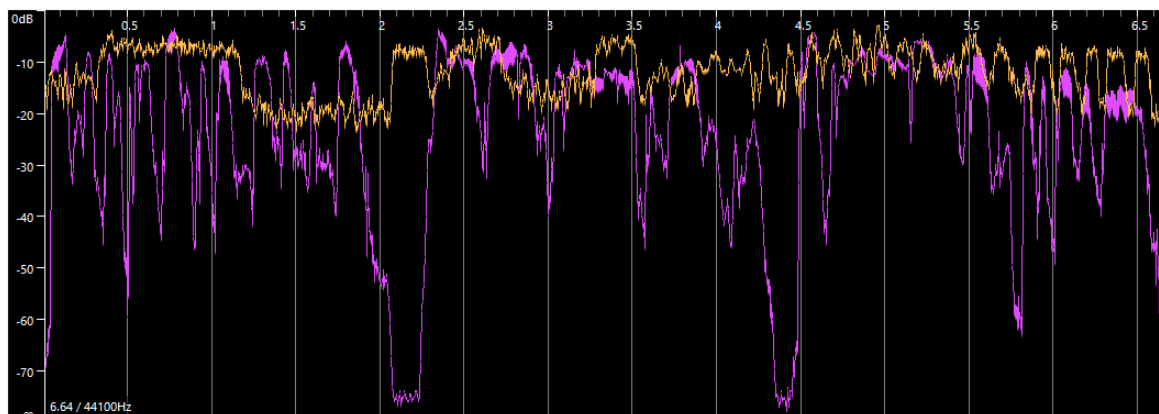
Εικόνα 5.31. Κυματομορφή ομιλιών πλήθους με γυναικεία ομιλία - ενθόρυβο 06 (N06V01)



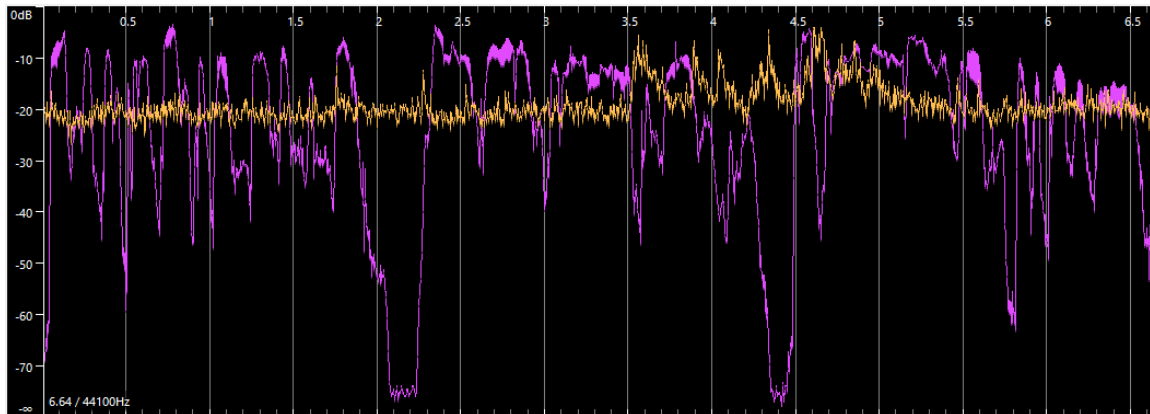
Εικόνα 5.32. Κυματομορφή βοηθού ανεμιστήρα με γυναικεία ομιλία - ενθόρυβο 07 (N07V01)



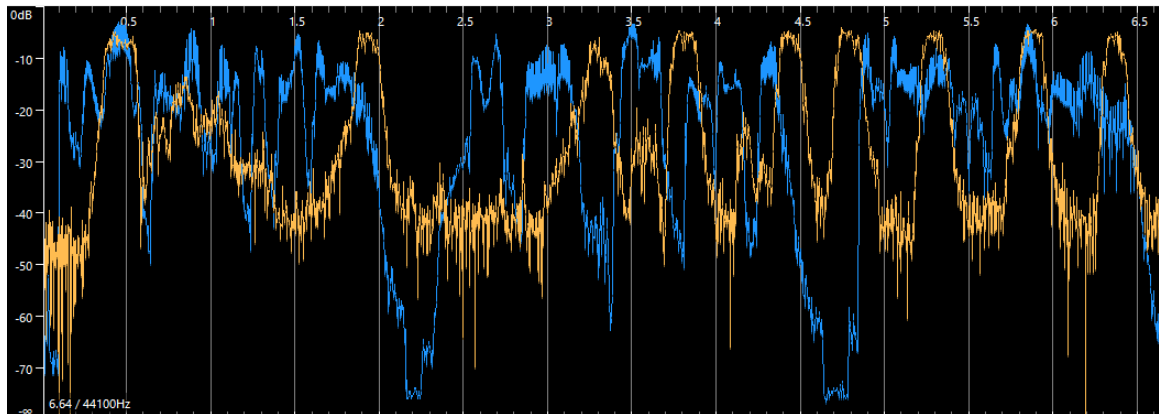
Εικόνα 5.33. Κυματομορφή πληκτρολόγησης με γυναικεία ομιλία - ενθόρυβο 09 (N09V01)



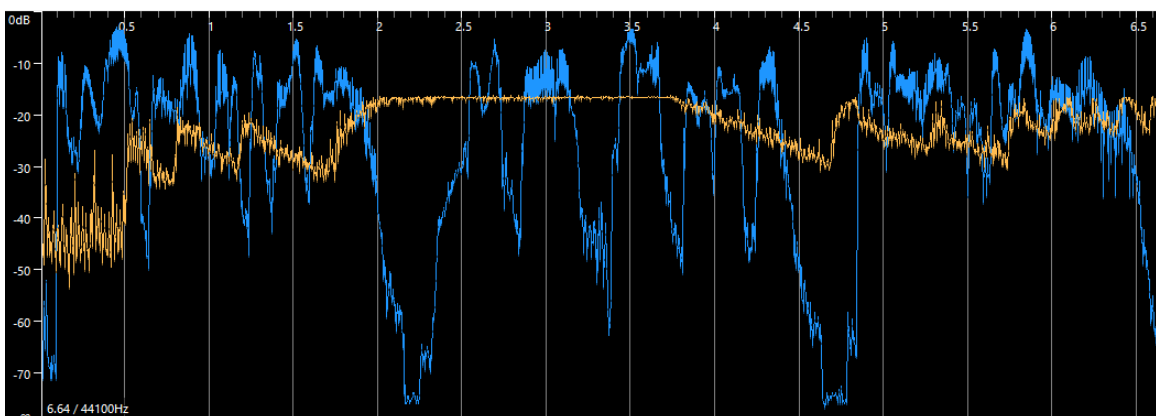
Εικόνα 5.34. Κυματομορφή σειρήνας και κόρνας με γυναικεία ομιλία - ενθόρυβο 11 (N11V01)



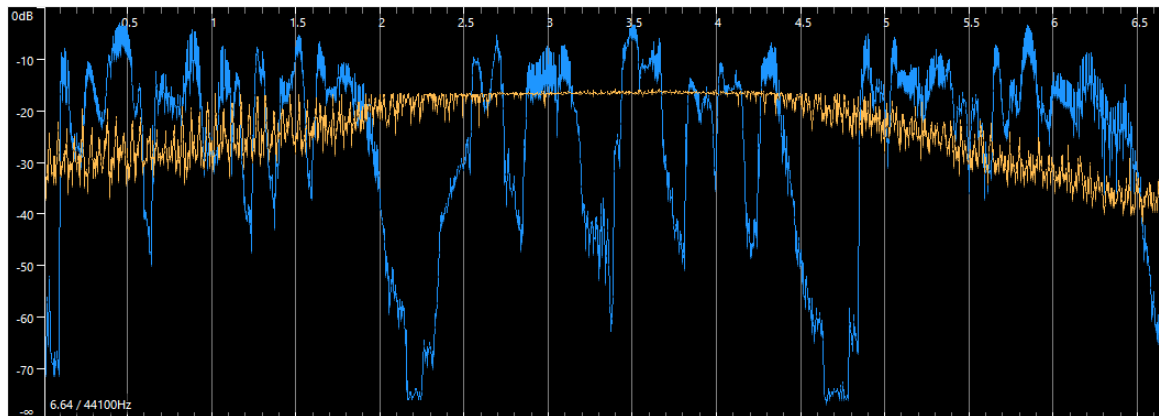
Εικόνα 5.35. Κυματομορφή δυνατής βροχής με γυναικεία ομιλία - ενθόρυβο 12 (N12V01)



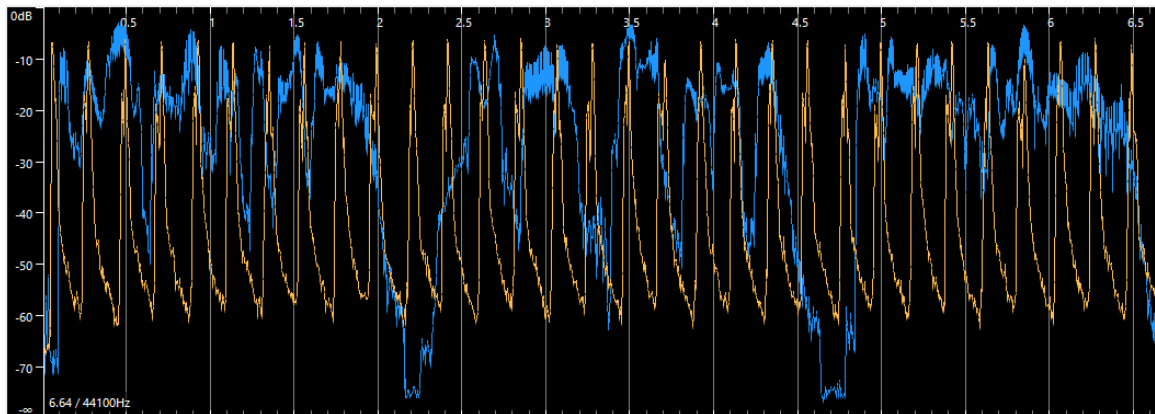
Εικόνα 5.36. Κυματομορφή γαβγίσματος σκύλου με αντρική ομιλία - ενθόρυβο 01 (N01V02)



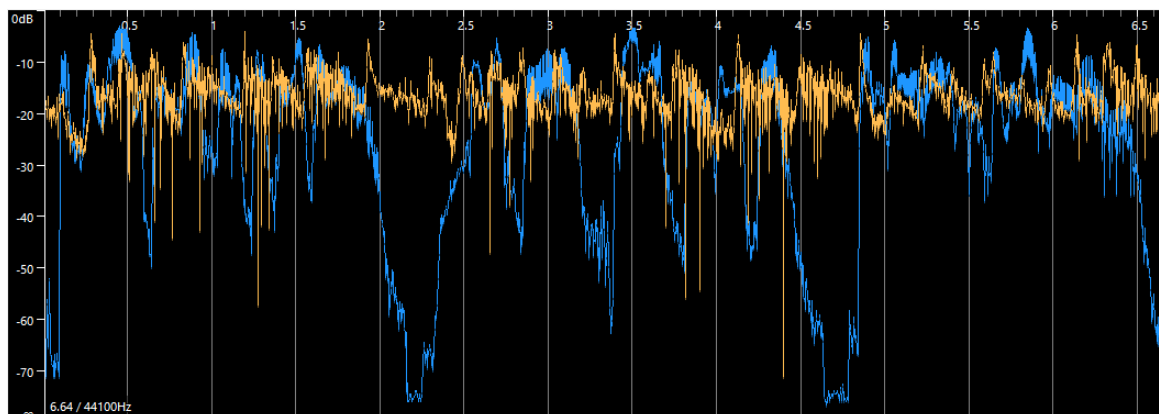
Εικόνα 5.37. Κυματομορφή αλυσοπρίονου με αντρική ομιλία - ενθόρυβο 02 (N02V02)



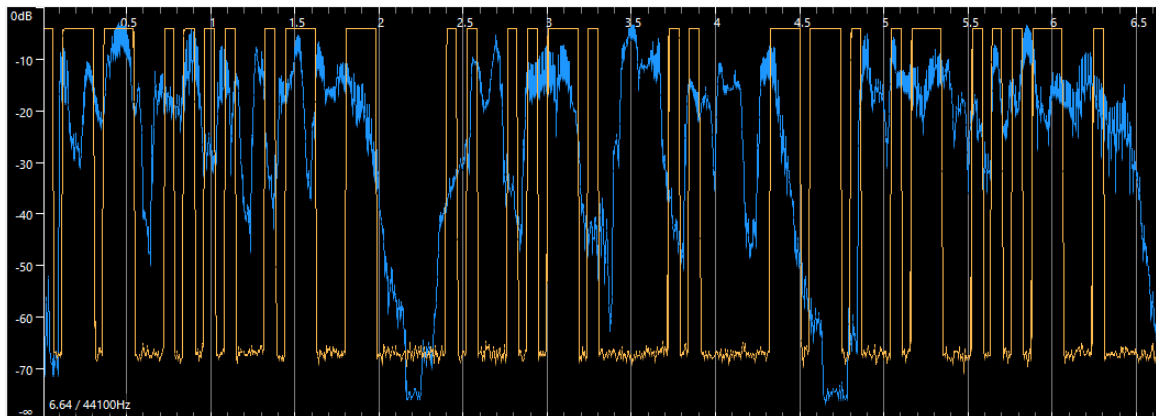
Εικόνα 5.38. Κυματομορφή ελικοπτερου με αντρική ομιλία - ενθόρυβο 04 (N04V02)



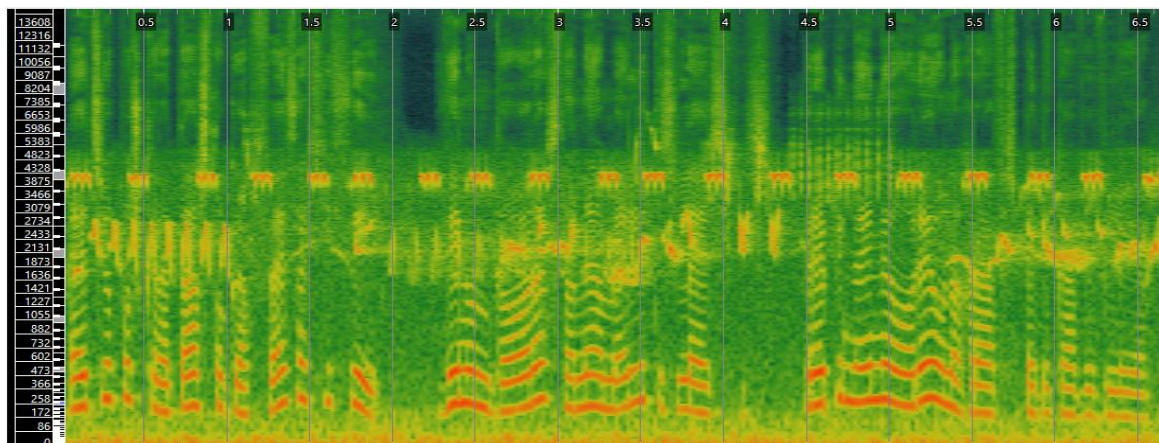
Εικόνα 5.39. Κυματομορφή ρολογιού τοίχου με αντρική ομιλία - ενθόρυβο 05 (N05V02)



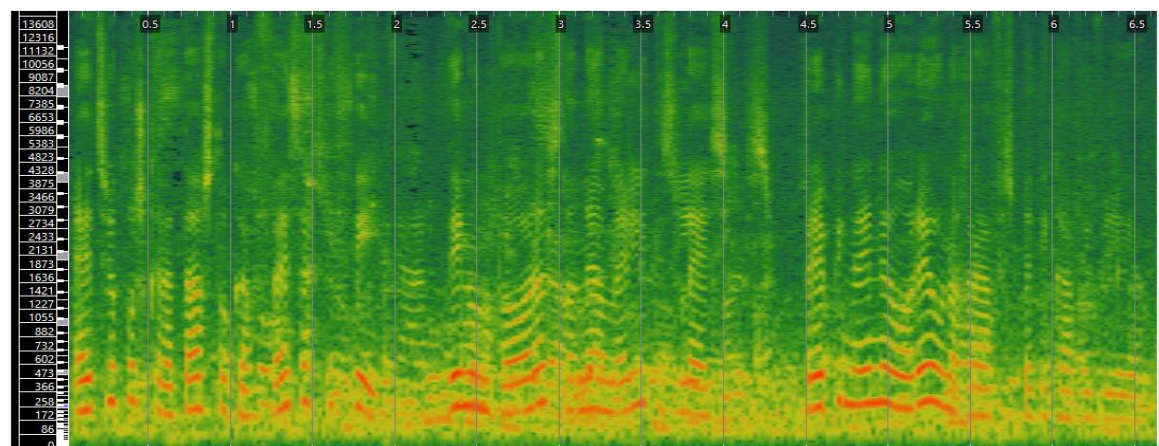
Εικόνα 5.40. Κυματομορφή μουσικής με αντρική ομιλία - ενθόρυβο 08 (N08V02)



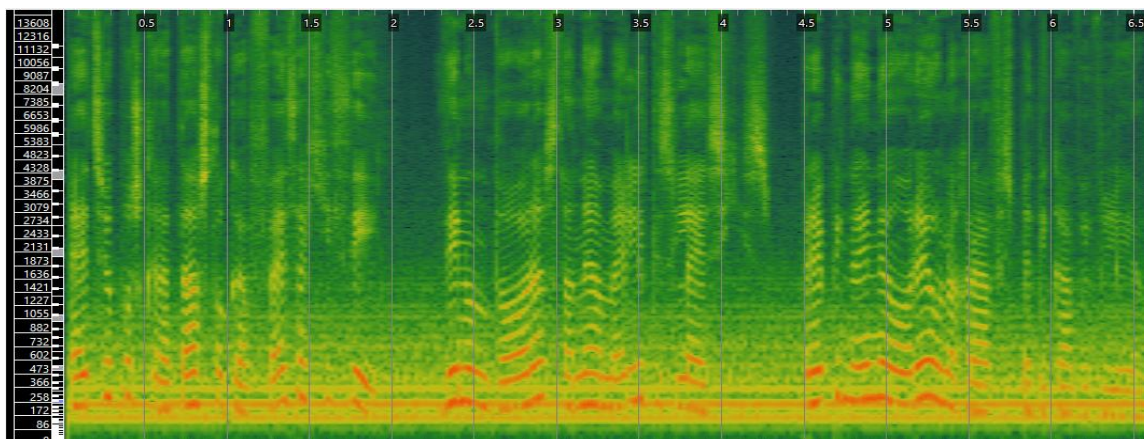
Εικόνα 5.41. Κυματομορφή σημάτων Μορς με αντρική ομιλία - ενθόρυβο 10 (N10V02)



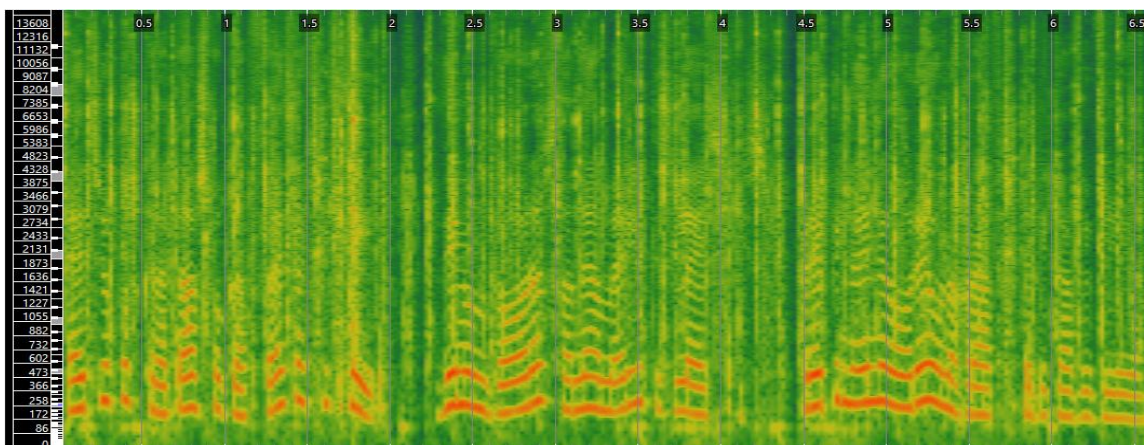
Εικόνα 5.42. Φασματογράφημα πουλιών και γρύλλου με γυναικεία ομιλία - ενθόρυβο 03 (N03V01)



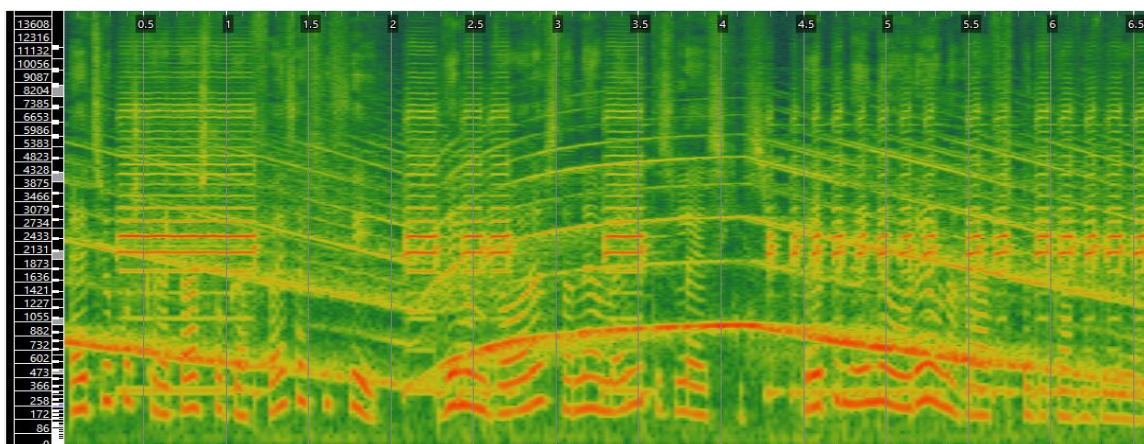
Εικόνα 5.43. Φασματογράφημα ομιλιών πλήθους με γυναικεία ομιλία - ενθόρυβο 06 (N06V01)



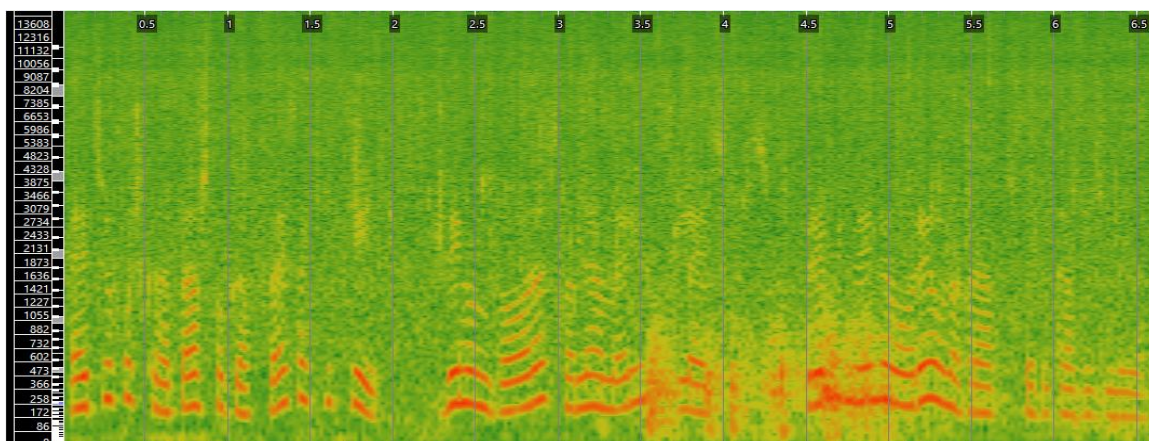
Εικόνα 5.44. Φασματογράφημα βουητού ανεμιστήρα με γυναικεία ομιλία - ενθόρυβο 07 (N07V01)



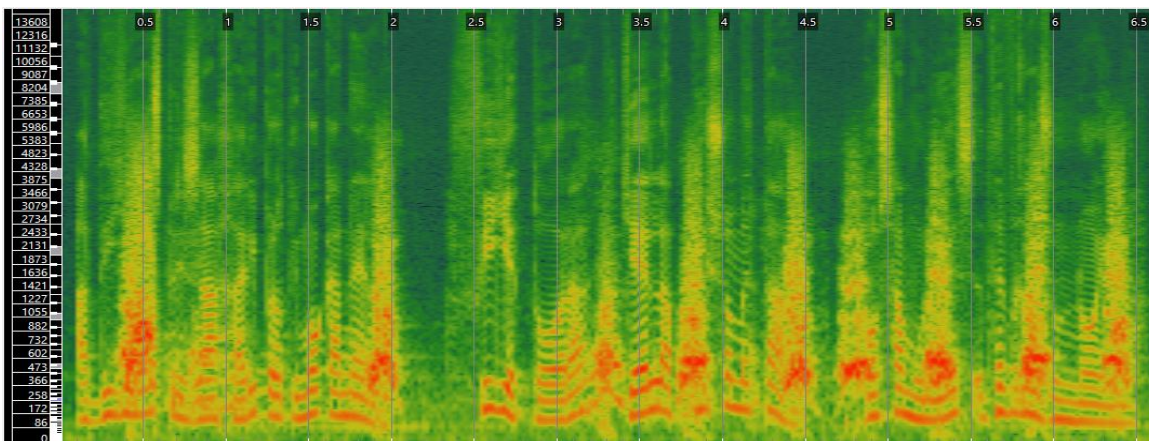
Εικόνα 5.45. Φασματογράφημα ηλεκτρολόγησης με γυναικεία ομιλία - ενθόρυβο 09 (N09V01)



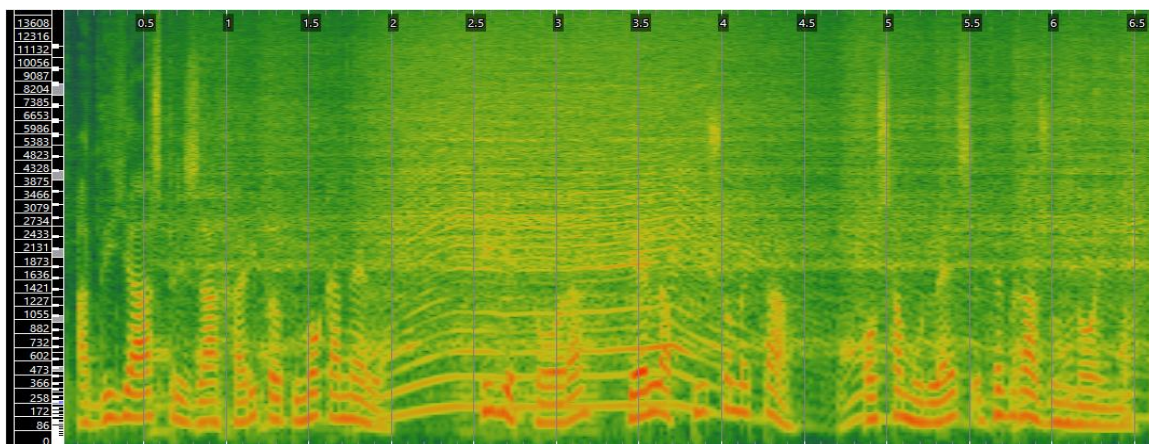
Εικόνα 5.46. Φασματογράφημα σειρήνας και κόρνας με γυναικεία ομιλία - ενθόρυβο 10 (N11V01)



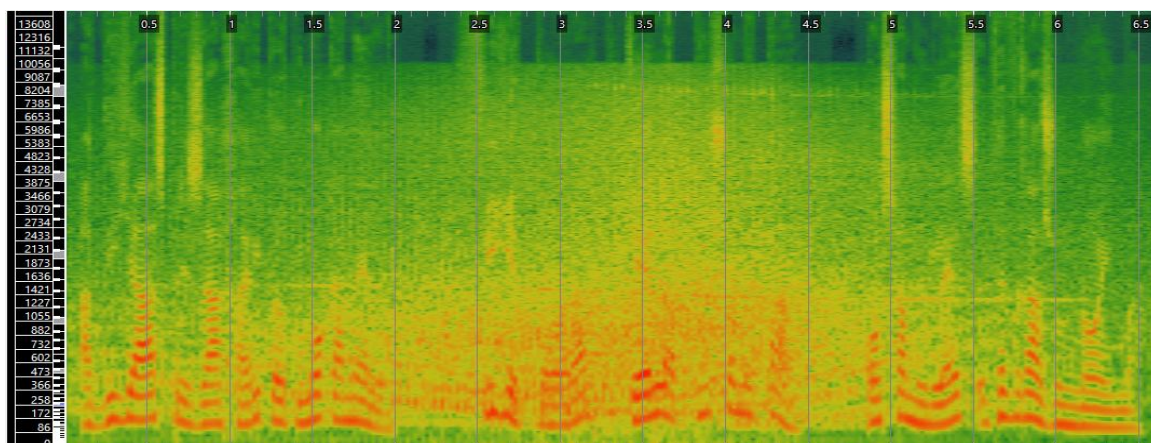
Εικόνα 5.47. Φασματογράφημα δυνατής βροχής με γυναικεία ομιλία - ενθόρυβο 12 (N12V01)



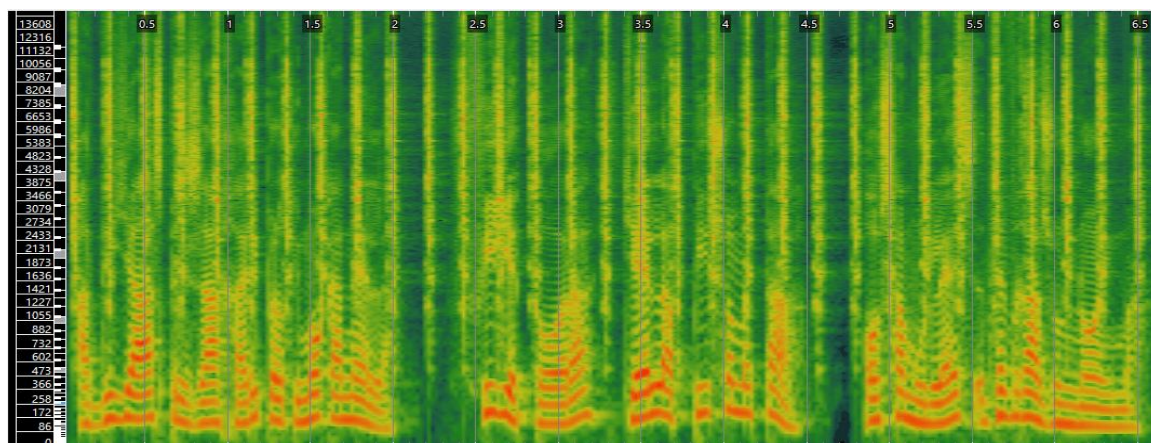
Εικόνα 5.48. Φασματογράφημα γαβγίσματος σκύλου με αντρική ομιλία - ενθόρυβο 01 (N01V02)



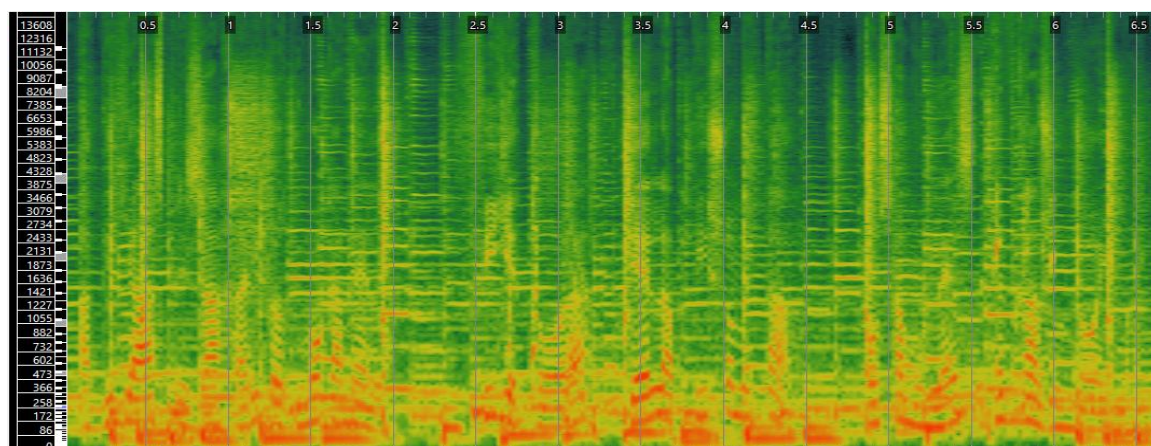
Εικόνα 5.49. Φασματογράφημα αλυσοπρίονου με αντρική ομιλία - ενθόρυβο 02 (N02V02)



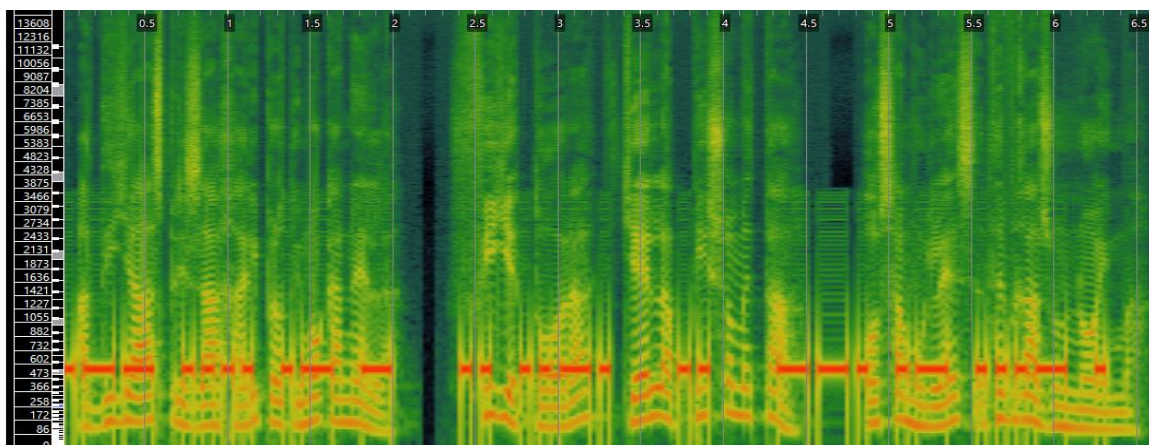
Εικόνα 5.50. Φασματογράφημα ελικοπτέρου με αντρική ομιλία - ενθόρυβο 04 (N04V02)



Εικόνα 5.51. Φασματογράφημα ρολογιού τοίχου με αντρική ομιλία- ενθόρυβο 05 (N05V02)



Εικόνα 5.52. Φασματογράφημα μουσικής με αντρική ομιλία - ενθόρυβο 08 (N08V02)



Εικόνα 5.53. Φασματογράφημα σημάτων Μορς με αντρική ομιλία - ενθόρυβο 10 (N10V02)

5.2 Επιλογή εργαλείων λογισμικού

Στο επόμενο στάδιο της πειραματικής διαδικασίας, θα πραγματοποιηθεί η επιλογή των εργαλείων λογισμικού Τεχνητής Νοημοσύνης, που θα συμπεριληφθούν στην υλοποίηση της παρούσας διπλωματικής εργασίας. Για την επιλογή των επικρατέστερων, θα γίνει λεπτομερής έρευνα στο διαδίκτυο, κατά την διάρκεια της οποίας θα εξετασθεί ένας μεγάλος αριθμός από αυτά, ώστε η τελική επιλογή να είναι αντιπροσωπευτική.

Κριτήριο της επιλογής θα αποτελέσει η απαίτηση, η χρήση των εργαλείων α) να μην προϋποθέτει την διενέργεια περίπλοκων ρυθμίσεων και ενεργειών κατά την διαδικασία της αποθρονοποίησης, β) να μπορεί να γίνει από άτομα χωρίς ιδιαίτερες τεχνικές γνώσεις στην επεξεργασία ήχου, και γ) αυτά να αποτελούν προσιτές οικονομικά λύσεις.

Έτσι δεν θα συμπεριληφθούν λογισμικά που μπορεί μεν να ανήκουν στα κορυφαία του είδους, όπως ενδεικτικά αναφέρεται το **RX**, της εταιρίας **Izotope** (<https://www.izotope.com/en/products/rx.html>), στα οποία όμως η παραγωγή του τελικού αποτελέσματος, λόγω της ιδιαίτερης εμπλοκής του χρήστη στην διαδικασία, συνεισφέρει μεγάλο βαθμό υποκειμενικότητας, δημιουργώντας σημαντική πιθανότητα η αξιολόγηση του τελικού προϊόντος -και κατ' επέκταση του ίδιου του εργαλείου- να είναι επηρεασμένη από την παράμετρο αυτή.

Όπως αναφέρεται στην φάση του σχεδιασμού της όλης διαδικασίας, ο τελικός αριθμός των εξεταζόμενων εργαλείων λογισμικού θα οριστεί στα **πέντε (5)**. Το δείγμα κρίνεται επαρκές, καθώς μεριμνάται ώστε να υπάρχει διασπορά στα ιδιαίτερα χαρακτηριστικά τους. Για τα λογισμικά αυτά θα αποκτηθεί άδεια χρήσης, με αγορά, εγγραφή, δημιουργία συνδρομής ή δικαίωμα δωρεάν χρήσης κατά περίπτωση, για την διεξαγωγή της έρευνας.

Στον Πίνακα 5.5 που ακολουθεί, περιλαμβάνονται με αλφαβητική σειρά, τα λογισμικά που θα διερευνηθούν κατά την διαδικασία της τελικής επιλογής. (Ο τελευταίος έλεγχος ορθής λειτουργίας των συνδέσμων που περιέχονται, έγινε στις 28/05/2026).

Πίνακας 5.5. Πίνακας εργαλείων TN που διερευνήθηκαν κατά την διαδικασία επιλογής λογισμικών

A/A	Όνομα	Σύνδεσμος
1	AI Mastering	https://aimastering.com/
2	AudioCleaner	https://audiocleaner.ai/
3	Audiodenise	https://audiodenise.com/
4	AUDIOEnhancer.ai	https://audioenhancer.ai/
5	Audo Studio	https://www.audo.ai/
6	Auphonic	https://auphonic.com/
7	BandLab	https://www.bandlab.com/ai
8	Cedar Studio (Restore)	https://cedaraudio.com/products/cedarstudio/...
9	Clarity Vx (Waves)	https://www.waves.com/plugins/clarity-vx
10	Cleanvoice	https://cleanvoice.ai/remove-background-noise/
11	Clear (Supertone)	https://www.supertone.ai/en/clear
12	Crumplepop (BorisFX)	https://borisfx.com/products/audiodenise/
13	Denoiser (Tape It)	https://tape.it/denoiser
14	dxRevive (Accentize)	https://www.accentize.com/product/dxrevive/...
15	Enhance (Adobe Podcast)	https://podcast.adobe.com/en/enhance
16	krisp	https://krisp.ai/
17	LALAL.AI	https://www.lalal.ai/voice-cleaner/

18	Magic Dust (Podcastle)	https://async.com/tools/magic-dust
19	MyEdit	https://myedit.online/en/audio-editor/...
20	Noise Reducer (Media.io)	https://noisereducer.media.io/
21	Noiseremoval.net	https://noiseremoval.net/
22	OpenVINO (Audacity)	https://www.audacityteam.org/download/opencvino/
23	ReHance (LANDR)	https://www.landr.com/vocal-enhancer
24	VocalPrep (Antares AutoTune)	https://www.antarestech.com/products/...
25	voice.ai	https://voice.ai/tools/audio-enhancer
26	Voice Isolator (11ElevenLabs)	https://elevenlabs.io/voice-isolator
27	VoicePea (HitPaw)	https://www.hitpaw.net/sem/...

Η παράθεση των λογισμικών που θα επιλεγούν, γίνεται στην συνέχεια με αλφαβητική σειρά. Η αποτύπωση χαρακτηριστικών και άλλων στοιχείων των λογισμικών, όπως και η καταγραφή των όποιων διαδικασιών υλοποίησης, ανταποκρίνεται στην χρονική στιγμή υλοποίησης της παρούσας διπλωματικής εργασίας (Οκτώβριος 2025 - Μάιος 2026).

Επισημαίνεται ότι ενδέχεται να υπάρχουν διαφοροποιήσεις σε σχέση με την περιεχόμενη πληροφορία, σε μικρότερο ή μεγαλύτερο βαθμό, ανάλογα με το που τοποθετείται χρονικά στο μέλλον η ανάγνωση του περιεχομένου της εργασίας.

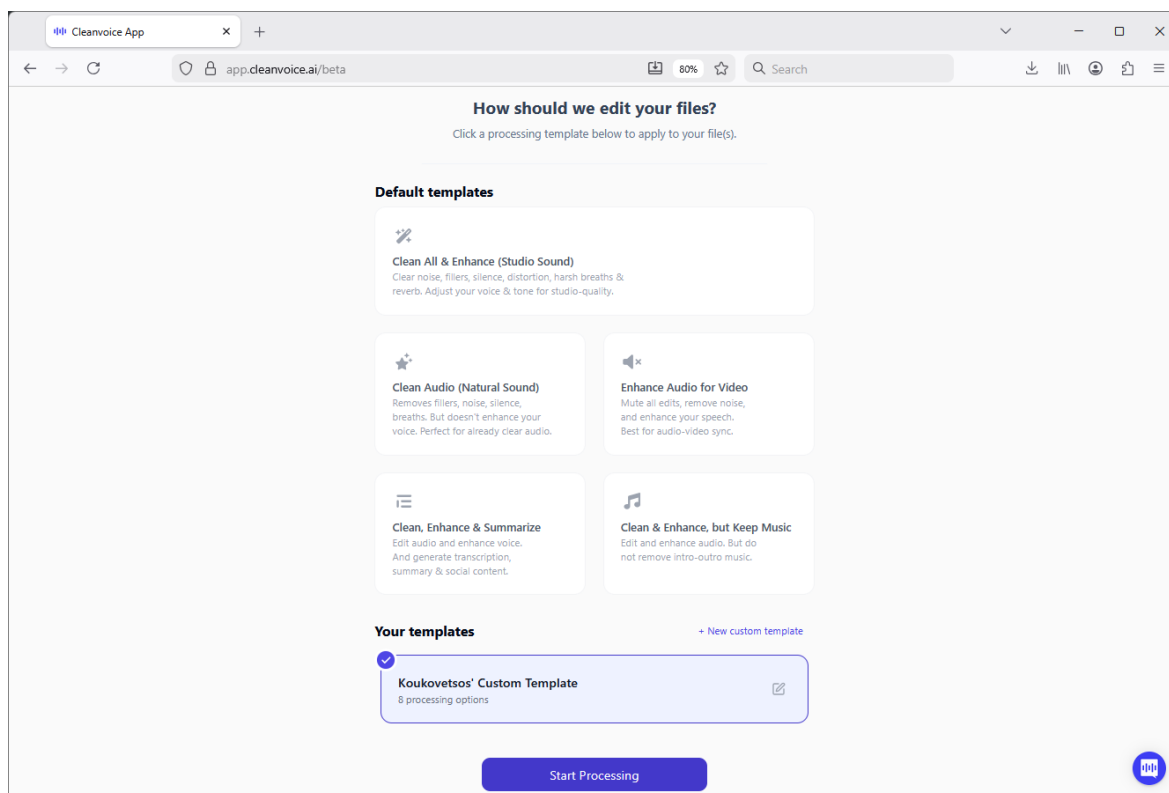
5.2.1 Λογισμικό 1 (S01) - Cleanvoice



Κύρια χρήση και λειτουργίες: Το **Cleanvoice** είναι ένα διαδεδωμένο διαδικτυακό εργαλείο επεξεργασίας ήχου με χρήση ΤΝ, το οποίο κυρίως προορίζεται για την αποθρομβοποίηση ηχογραφημένης ομιλίας, προερχόμενης από εργασίες όπως συνεντεύξεις ή εκπομπές διαδικτύου (podcasts). Στις βασικές του λειτουργίες, περιλαμβάνονται, εκτός από την διενέργεια αποθρομβοποιήσεων, η εξάλειψη παύσεων και στοματικών ήχων, καθώς

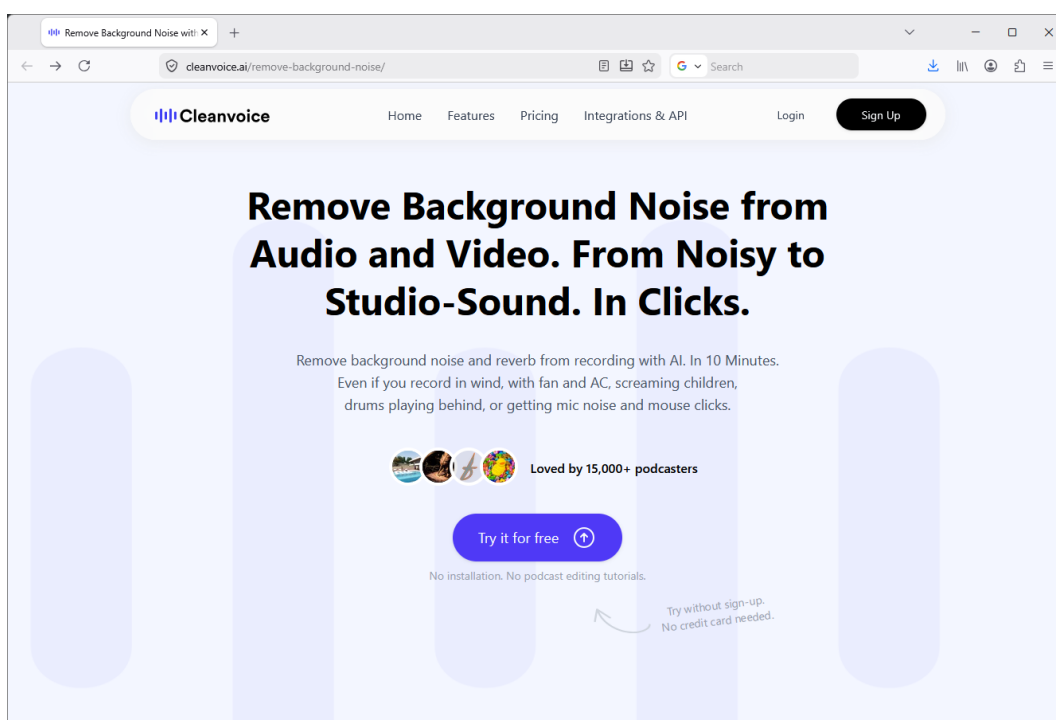
και η δημιουργία απομαγνητοφωνήσεων (transcriptions) με χρονικές σημάνσεις (timestamps). Μία ακόμα δυνατότητα του επιλεγμένου αυτού εργαλείου, είναι η υποστήριξη της επεξεργασίας πολυκάναλων ηχητικών σημάτων (stems), φροντίζοντας για την διατήρηση του συγχρονισμού μεταξύ των καναλιών. Παράλληλα, μπορεί να εφαρμοστεί και σε αρχεία βίντεο με ταυτόχρονη επεξεργασία του περιεχόμενου ήχου.

Βαθμός επέμβασης του λογισμικού: Το επιλεγμένο αυτό εργαλείο, ακολουθεί κυρίως τη λογική της αυτοματοποιημένης επεξεργασίας (one-click tool). Παρέχονται βασικές επιλογές ενεργοποίησης / απενεργοποίησης επιμέρους λειτουργιών, ενώ δεν προσφέρονται εκτεταμένες δυνατότητες λεπτομερούς ρύθμισης κατά την διάρκεια της επεξεργασίας. Έτσι η όποια διαφοροποίηση σε αυτή, πραγματοποιείται με την χρήση ήδη διαθέσιμων ή δημιουργημένων από τον χρήστη προτύπων (Templates), γεγονός που περιορίζει τον βαθμό χειροκίνητου ελέγχου της διαδικασίας, υπέρ της ευκολίας χρήσης της εφαρμογής. Στην Εικόνα 5.54 που ακολουθεί εμφανίζονται οι διάφορες επιλογές προτύπων επεξεργασίας.



Εικόνα 5.54. Cleanvoice - Επιλογή προτύπου επεξεργασίας (Template)

Διαθέσιμες εκδόσεις λογισμικού: Το Cleanvoice διατίθεται αποκλειστικά ως web-based εφαρμογή (online εργαλείο), και δεν παρέχεται ως plugin ή standalone desktop εφαρμογή. Στην ακόλουθη Εικόνα 5.55, αποτυπώνεται στιγμιότυπο από την αρχική οθόνη του ιστοτόπου, (<https://cleanvoice.ai/remove-background-noise/>), στον οποίο είναι διαθέσιμο το λογισμικό. Επισημαίνεται ότι το περιβάλλον εργασίας της πλατφόρμας, δεν είναι διαθέσιμο στα ελληνικά.



Εικόνα 5.55. Cleanvoice - Αρχική οθόνη εφαρμογής

Κόστος κτήσης - χρήσης: Το Cleanvoice παρέχει την δυνατότητα χρέωσης με κλιμακωτή μηνιαία ή ετήσια συνδρομή, ανάλογα τις επιθυμητές ώρες επεξεργασίας, ή μέσω επιλογής του μοντέλου χρέωσης τύπου «**Pay as You Go**», όπου προαγοράζεις συγκεκριμένο χρόνο εργασίας. Ενδεικτικά αναφέρεται, ότι κατά την στιγμή της υλοποίησης της παρούσας διπλωματικής εργασίας, προσφέρεται δωρεάν δοκιμαστική χρήση επεξεργασίας, διάρκειας 30 min, μετά από δημιουργία λογαριασμού στην πλατφόρμα, ενώ με το πλάνο Pay as You Go υπάρχει η επιλογή να προαγοραστούν 5, 10, ή 30 ώρες επεξεργασίας κόστους 10€, 18€ και 40€ αντίστοιχα, διαθέσιμες για κατανάλωση μέσα στα επόμενα δύο έτη. Περισσότερες πληροφορίες, υπάρχουν διαθέσιμες στην ηλεκτρονική διεύθυνση: <https://cleanvoice.ai/pricing/>.

Online λειτουργία: Η επεξεργασία πραγματοποιείται εξ ολοκλήρου μέσα από την διαδικτυακή πλατφόρμα της εφαρμογής, ενώ δεν παρέχεται δυνατότητα offline χρήσης. Τα αρχεία μεταφορτώνονται στους διακομιστές της υπηρεσίας, όπου γίνεται η επεξεργασία τους, μέσω αλγορίθμων Τεχνητής Νοημοσύνης.

Υποστηριζόμενα λειτουργικά συστήματα (OS): Η πρόσβαση στην εφαρμογή γίνεται μέσω φυλλομετρητή (browser), επομένως αυτή είναι διαθέσιμη για χρήση σε όλα τα λειτουργικά συστήματα (**Windows, macOS, Linux**), καθώς και σε φορητές συσκευές (**Android, iOS**), χωρίς να απαιτείται εγκατάσταση.

Υποστηριζόμενοι τύποι αρχείων ήχου: Υποστηρίζονται οι τύποι **WAV, MP3, FLAC** και **M4A** τόσο την είσοδο, όσο και στην έξοδο για την αποθήκευση του αποτελέσματος της επεξεργασίας.

Το Cleanvoice, αποτελεί ενδεικτικό παράδειγμα διαδικτυακού εργαλείου αποθορυβοποίησης με χρήση TN, το οποίο δίνει έμφαση στην αυτοματοποίηση και ευχρηστία, εις βάρος της λεπτομερούς παραμετροποίησης, γεγονός που το καθιστά ιδανικό για μη εξειδικευμένους χρήστες, αλλά ταυτόχρονα λιγότερο ευέλικτο για επαγγελματική, επεξεργασία, όπου απαιτούνται βέλτιστα αποτελέσματα.

5.2.2 Λογισμικό 2 (S02) - Enhance (Adobe Podcast)

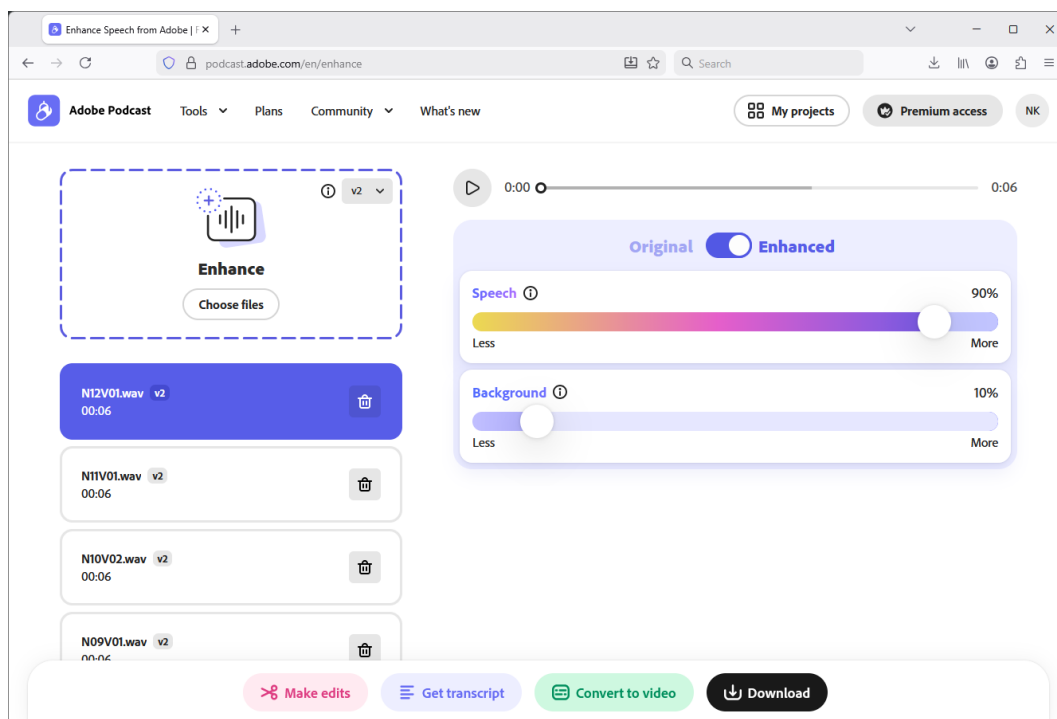


Adobe Podcast

Κύρια χρήση και λειτουργίες: Το **Adobe Podcast Enhance** (γνωστό και ως **Enhance Speech**) αποτελεί, όπως και το Cleanvoice, διαδεδομένο διαδικτυακό εργαλείο επεξεργασίας ήχου. Αναπτύχθηκε από την εταιρία **Adobe** και στοχεύει κυρίως στη βελτίωση της ποιότητας ομιλίας. Η βασική του λειτουργία είναι η μετατροπή ηχογραφήσεων χαμηλής ποιότητας σε ήχο που προσεγγίζει αποτελέσματα που επιτυγχάνονται σε επαγγελματικά studio, μέσω αποθορυβοποίησης, εξισορρόπησης φωνής και μείωσης της αντήχησης (reverb). Επικεντρώνεται κυρίως σε αρχεία ομιλίας, ενώ

υποστηρίζει (Jiroudek, 2026), εξαγωγή πολυκάναλων εγγραφών (stems), από τον Μάρτιο του 2026.

Βαθμός επέμβασης του λογισμικού: Υπάρχει δυνατότητα ρύθμισης του βαθμού ενίσχυσης ομιλίας ή της καταστολής του θορύβου, με την χρήση ράβδων κύλισης (sliders), οι οποίοι όμως στην δωρεάν έκδοση του λογισμικού -η οποία δεν απαιτεί την ύπαρξη δημιουργίας λογαριασμού- είναι απενεργοποιημένη. Στην περίπτωση αυτή, η λειτουργία του επιλεγμένου εργαλείου, βασίζεται σχεδόν αποκλειστικά σε one-click επεξεργασία, χωρίς δυνατότητα λεπτομερούς παραμετροποίησης. Ο χρήστης ανεβάζει το αρχείο και το σύστημα εφαρμόζει αυτόματα τις βέλτιστες ρυθμίσεις. Αυτό καθιστά το εργαλείο ιδιαίτερα εύχρηστο, αλλά περιορίζει τον έλεγχο του τελικού αποτελέσματος. Στην περίπτωση βέβαια της χρήσης με συνδρομή, οι ρυθμίσεις αυτές είναι διαθέσιμες στην χρήση (ενεργοποιημένες). Στην Εικόνα 5.56 που ακολουθεί, εμφανίζεται το περιβάλλον εργασίας της εφαρμογής Enhance, το οποίο είναι διαθέσιμο στην ηλεκτρονική διεύθυνση: <https://podcast.adobe.com/en/enhance>.



Εικόνα 5.56. Enhance - Περιβάλλον εργασίας της εφαρμογής

Διαθέσιμες εκδόσεις λογισμικού: Το Enhance, όπως και το Cleanvoice που επιλέχθηκε ήδη, διατίθεται αποκλειστικά ως web-based εφαρμογή (online εργαλείο), και

δεν παρέχεται ως plugin ή standalone desktop εφαρμογή. Το περιβάλλον εργασίας της πλατφόρμας, δεν είναι διαθέσιμο στα ελληνικά.

Κόστος κτήσης - χρήσης: Το λογισμικό διατίθεται τόσο για δωρεάν χρήση όσο και για χρήση με συνδρομή. Η κύρια διαφορά ανάμεσα στις δύο επιλογές, έγκειται στο ότι η συνδρομητική μορφή έχει ενεργοποιημένες περισσότερες δυνατότητες, όπως είναι η υποστήριξη video, η ταυτόχρονη μεταφόρτωση πολλών αρχείων για επεξεργασία, η ενεργοποίηση της ρύθμισης βελτίωσης του ήχου και η διάθεση περισσότερου χώρου και χρόνου επεξεργασίας ανά ημέρα. Κατά την διάρκεια της υλοποίησης της παρούσας διπλωματικής εργασίας, θα αξιοποιηθεί η δυνατότητα πλήρους χρήσης όλων των δυνατοτήτων της εφαρμογής, που παρέχεται δωρεάν για 30 ημέρες (trial version), με δημιουργία λογαριασμού στην πλατφόρμα. Περισσότερες πληροφορίες για τις επιλογές χρέωσης, υπάρχουν στην ηλεκτρονική διεύθυνση: <https://podcast.adobe.com/en/plans>.

Online λειτουργία: Όπως και στην περίπτωση του Cleanvoice, Η επεξεργασία πραγματοποιείται εξ ολοκλήρου μέσα από την διαδικτυακή πλατφόρμα της εφαρμογής, ενώ δεν παρέχεται δυνατότητα offline χρήσης. Τα αρχεία μεταφορτώνονται στους διακομιστές της υπηρεσίας, όπου γίνεται η επεξεργασία μέσω αλγορίθμων Τεχνητής Νοημοσύνης.

Υποστηριζόμενα λειτουργικά συστήματα (OS): Η πρόσβαση στην εφαρμογή γίνεται μέσω φυλλομετρητή (browser), επομένως είναι διαθέσιμο σε όλα τα λειτουργικά συστήματα (**Windows, macOS, Linux**), καθώς και σε φορητές συσκευές (**Android, iOS**), χωρίς να απαιτείται εγκατάσταση.

Υποστηριζόμενοι τύποι αρχείων ήχου: Υποστηρίζονται κοινοί τύποι αρχείων, όπως **MP3** και **WAV**, τόσο την είσοδο, όσο και στην έξοδο της επεξεργασίας.

Το **Enhance** αποτελεί χαρακτηριστικό και ιδιαίτερα δημοφιλές παράδειγμα εργαλείου ενίσχυσης της ανθρώπινης ομιλίας, το οποίο δίνει προτεραιότητα στην ποιότητα του αποτελέσματος, διαμέσου αυξημένης αυτοματοποίησης, προσφέροντας εντυπωσιακή βελτίωση ήχου, απαιτώντας ελάχιστη αλληλεπίδραση από τον χρήστη. Ωστόσο, η απουσία εκτεταμένης παραμετροποίησης, όπως και η έλλειψη δυνατότητας offline λειτουργίας, το καθιστά λιγότερο ευέλικτο για ορισμένες εξειδικευμένες επαγγελματικές εφαρμογές.

5.2.3 Λογισμικό 3 (S03) - LALAL.AI



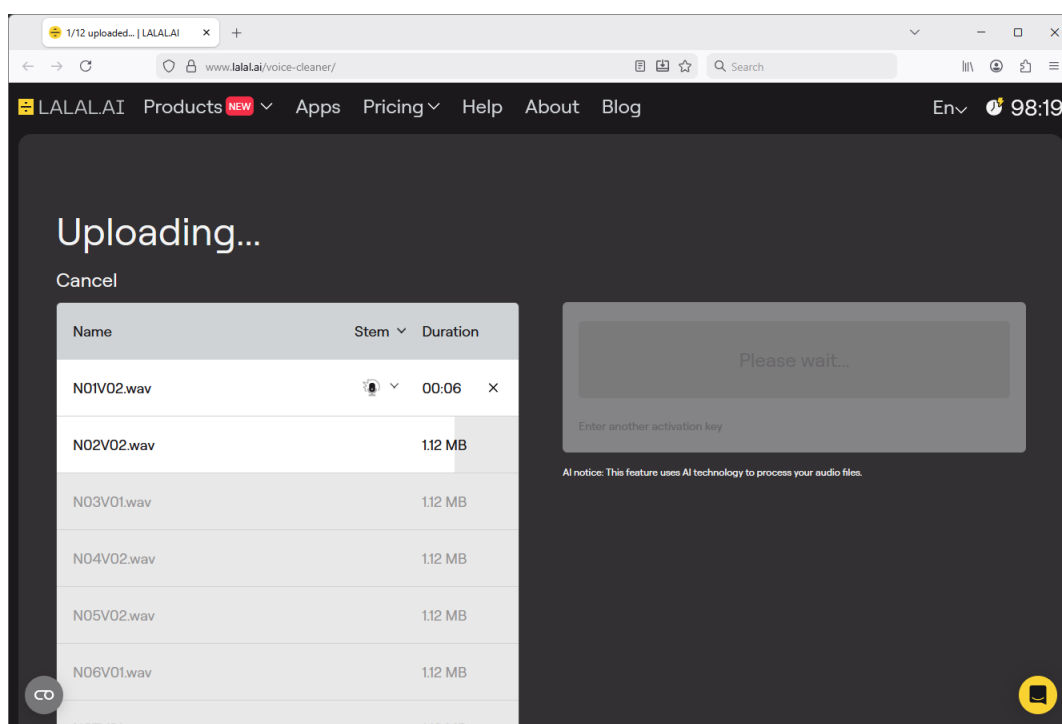
Κύρια χρήση και λειτουργίες: Το **LALAL.AI**, αποτελεί διαδικτυακό εργαλείο επεξεργασίας ήχου, που βασίζεται σε τεχνικές Τεχνητής Νοημοσύνης. Σε αντίθεση με τα δύο προηγούμενα λογισμικά, εξειδικεύεται κυρίως στον διαχωρισμό πηγών ήχου (source separation), σε μουσικά ηχητικά σήματα. Έχει σαν βασική του λειτουργία την απομόνωση φωνής, μουσικών οργάνων ή επιμέρους στοιχείων ενός ηχητικού ή μουσικού κομματιού (όπως φωνή, τύμπανα, μπάσο, ή πιάνο), δημιουργώντας ξεχωριστά κανάλια ηχητικών σημάτων (stems). Παράλληλα, προσφέρει δυνατότητες αποθορυβοποίησης και καθαρισμού του ήχου, καθώς και επεξεργασία αρχείων βίντεο με εξαγωγή και επεξεργασία του ηχητικού τους μέρους. Κατά την διεξοδική μελέτη, η οποία πραγματοποιείται στα πλαίσια της παρούσας διπλωματικής εργασίας, διαπιστώνεται ότι η δυνατότητα διαχωρισμού και αποθήκευσης των φωνών που ενυπάρχουν σε ένα ηχητικό σήμα, το καθιστά ιδιαίτερα αποτελεσματικό για αποθορυβοποιήσεις όπου στο σήμα υπάρχουν δευτερεύουσες ομιλίες, και αυτός ήταν και ο κύριος λόγος που συμπεριλήφθηκε στα επιλεγμένα λογισμικά.

Βαθμός επέμβασης του λογισμικού: Το επιλεγμένο αυτό εργαλείο, παρέχει βασικές δυνατότητες επιμέρους λειτουργιών, όπως είναι α) η επιλογή ανάμεσα στις διαφορετικές εκδόσεις των διαθέσιμων Νευρωνικών Δικτύων της εφαρμογής, (με τις ονομασίες «v4 Orion», «v5 Perseus» και «v6 Andromeda»), και β) η επιλογή επιπέδου ακύρωσης θορύβου («Mild», «Normal» η «Aggressive»). Άλλες επιλογές περιλαμβάνουν την δυνατότητα αφαίρεσης φαινομένων όπως αυτά της ηχούς και της αντήχησης.

Διαθέσιμες εκδόσεις λογισμικού: Το LALAL.AI διατίθεται κυρίως ως web-based εφαρμογή, ενώ προσφέρονται και desktop εκδόσεις (standalone εφαρμογές) για τοπική χρήση σε ορισμένα λειτουργικά συστήματα. Στην πράξη η desktop έκδοση δεν λειτουργεί ανεξάρτητα από την πλατφόρμα, αλλά φροντίζει περισσότερο για την καλύτερη τοπική οργάνωση και διαχείριση των αρχείων από και προς επεξεργασία. Στην συνδρομητική

έκδοση «**Pro**» το λογισμικό είναι διαθέσιμο και με την μορφή VST plugin, ενώ υπάρχει και η δυνατότητα ενσωμάτωσης σε τρίτες εφαρμογές, μέσω διαθέσιμου API.

Το περιβάλλον εργασίας της πλατφόρμας, όπως και στις προηγούμενες εφαρμογές, δεν είναι διαθέσιμο στα ελληνικά. Στην Εικόνα 5.57 που ακολουθεί, εμφανίζεται το διαδικτυακό περιβάλλον εργασίας της εφαρμογής LALAL.AI, το οποίο είναι διαθέσιμο στην ηλεκτρονική διεύθυνση: <https://www.lalal.ai/voice-cleaner/>.



Εικόνα 5.57. LALAL.AI - Διαδικτυακό περιβάλλον εργασίας της εφαρμογής

Κόστος κτήσης - χρήσης: Η χρήση του LALAL.AI διατίθεται σε τρεις εκδόσεις. Η δωρεάν «**Starter**» έκδοση δεν παρέχει δυνατότητα αποθήκευσης των προϊόντων της επεξεργασίας. Υπάρχουν δύο ακόμα εκδόσεις, η «**Lite**» και η «**Pro**» με μηνιαία συνδρομή, που παρέχουν μεταξύ άλλων την δυνατότητα αποθήκευσης των αποτελεσμάτων, ενώ προσφέρεται και η επιπλέον δυνατότητα αγοράς συγκεκριμένου χρόνου επεξεργασίας, χωρίς συνδρομή. Για τις ανάγκες της υλοποίησης της παρούσας διπλωματικής εργασίας, θα αγορασθεί από τον γράφοντα μηνιαία συνδρομή έκδοσης «**Lite**», με δημιουργία λογαριασμού στην πλατφόρμα. Περισσότερες πληροφορίες για τις επιλογές χρέωσης, υπάρχουν στην ηλεκτρονική διεύθυνση: <https://www.lalal.ai/pricing/>.

Online λειτουργία: Η επεξεργασία πραγματοποιείται μέσα από την διαδικτυακή πλατφόρμα της εφαρμογής, ή από την τοπικά εγκατεστημένη εφαρμογή. Και στις δύο περιπτώσεις, η χρήση είναι online, και τα αρχεία μεταφορτώνονται στους διακομιστές της υπηρεσίας. (Στην συνδρομητική έκδοση «**Pro**» παρέχεται δυνατότητα offline χρήσης με την διάθεση για τοπική εγκατάσταση του μοντέλου TN «**Lyra**» της εφαρμογής).

Υποστηριζόμενα λειτουργικά συστήματα (OS): Η πρόσβαση στην εφαρμογή γίνεται τόσο μέσω φυλλομετρητή (browser), όσο και με τοπική εγκατάσταση (desktop), σε όλα τα λειτουργικά συστήματα (**Windows, macOS,**). Η εφαρμογή είναι διαθέσιμη και για συσκευές Android: (<https://play.google.com/store/apps/details?id=com.omnisale.lalal.ai>).

Υποστηριζόμενοι τύποι αρχείων ήχου: Υποστηρίζονται οι τύποι **MP3, OGG, WAV, FLAC,** και **AAC,** τόσο την είσοδο, όσο και στην έξοδο για την αποθήκευση του αποτελέσματος της επεξεργασίας.

Το LALAL.AI αποτελεί χαρακτηριστικό παράδειγμα εργαλείου που αξιοποιεί την Τεχνητή Νοημοσύνη για προηγμένη ανάλυση και διαχωρισμό ηχητικών πηγών, επεκτείνοντας τη λειτουργικότητα πέρα από την απλή αποθορυβοποίηση. Η έμφαση στην αυτοματοποίηση και την ευκολία χρήσης το καθιστά ιδιαίτερα εύχρηστο, αν και η περιορισμένη δυνατότητα λεπτομερούς ελέγχου μπορεί να αποτελεί περιορισμό σε πιο απαιτητικά επαγγελματικά σενάρια.

5.2.4 Λογισμικό 4 (S04) - OpenVINO (Audacity Plugin)

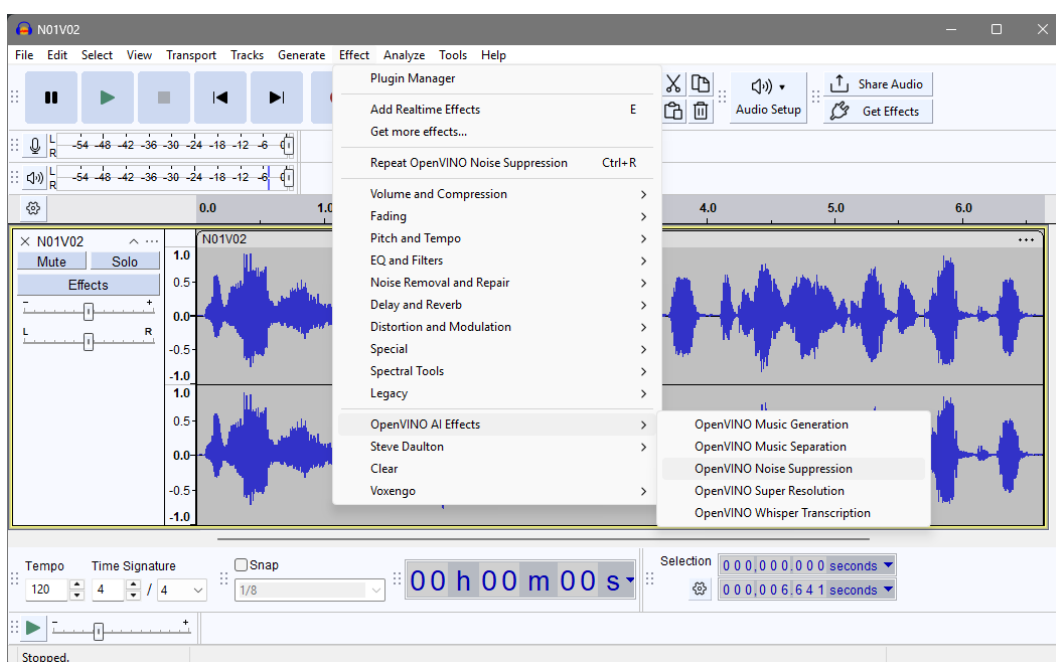


Κύρια χρήση και λειτουργίες: Το πρόσθετο (plugin) **OpenVINO**, είναι το τέταρτο κατά σειρά λογισμικό, που θα επιλεγεί για την υλοποίηση του πειραματικού σκέλους της παρούσας διπλωματικής εργασίας. Αποτελεί επέκταση που ενσωματώνει δυνατότητες Τεχνητής Νοημοσύνης στο περιβάλλον του σταθμού επεξεργασίας ψηφιακού ήχου (DAW) **Audacity**. Η βασική του χρήση αφορά την εφαρμογή μοντέλων επεξεργασίας ήχου, όπως είναι η αποθορυβοποίηση (noise reduction), η βελτίωση ομιλίας (speech enhancement) και

ενδεχομένως ο διαχωρισμός πηγών ήχου, ανάλογα με τα τοπικά εγκατεστημένα μοντέλα. Το plugin αυτό επιδρά σε ηχητικά αρχεία, τα οποία έχουν ήδη εισαχθεί στο Audacity, με εφαρμογή μεθόδων Τεχνητής Νοημοσύνης.

Βαθμός επέμβασης του λογισμικού: Υπάρχει η επιλογή της εγκατάστασης μίας ή περισσότερων, ανάμεσα σε 3 διαφορετικές εκδόσεις των διαθέσιμων μοντέλων καταστολής θορύβου του προσθέτου, (με τις ονομασίες «**deepfiltenet2**», «**deepfiltenet3**» και «**noise-suppression-denseunet-II-0001**»). Ο βαθμός επέμβασης του χρήστη είναι μέτριος, καθώς συνδυάζει αυτοματοποιημένες λειτουργίες με βασικές επιλογές ρύθμισης. Ανάλογα με την επιλογή μοντέλου, μπορεί να προσφέρει είτε ημι-αυτοματοποιημένη επεξεργασία (one-click) είτε περιορισμένες παραμέτρους επιλογής επίδρασης στην επεξεργασία.

Διαθέσιμες εκδόσεις λογισμικού: Το OpenVINO διατίθεται αποκλειστικά ως πρόσθετο, για χρήση μέσα από το περιβάλλον εργασίας του σταθμού επεξεργασίας ψηφιακού ήχου Audacity. Στην ακόλουθη Εικόνα 5.58, αποτυπώνεται στιγμιότυπο από την χρήση του προσθέτου OpenVINO στο Audacity.



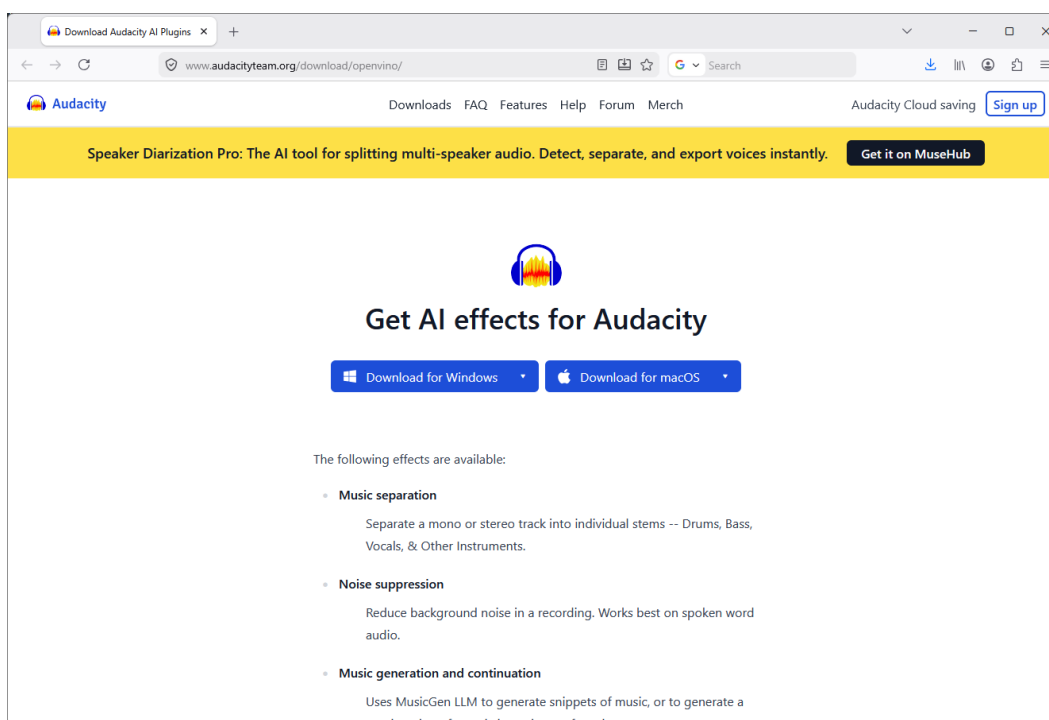
Εικόνα 5.58. OpenVINO - Στιγμιότυπο οθόνης κατά την χρήση του προσθέτου στο Audacity

Κόστος κτήσης - χρήσης: Το OpenVINO αποτελεί ανοιχτό λογισμικό (open source), το οποίο διατίθεται ελεύθερα για χρήση χωρίς χρέωση. Αυτός είναι και ένας

βασικός λόγος για τον οποίο το συγκεκριμένο λογισμικό θα επιλεγεί για να συμμετάσχει στην υλοποίηση της πειραματικής διαδικασίας της παρούσας διπλωματικής εργασίας.

Online λειτουργία: Για την λειτουργία του προσθέτου δεν είναι απαραίτητη η σύνδεση με το διαδίκτυο, καθώς τα μοντέλα επεξεργασίας έχουν αποθηκευτεί τοπικά κατά την εγκατάσταση. Αυτό αποτέλεσε έναν ακόμα λόγο της συμπερίληψης του συγκεκριμένου λογισμικού ανάμεσα στα επιλεγθέντα, καθώς αποτελεί και παράγοντα διαφοροποίησής του από τα υπόλοιπα.

Υποστηριζόμενα λειτουργικά συστήματα (OS): Το OpenVINO είναι διαθέσιμο για εγκατάσταση στα λειτουργικά συστήματα **Windows** και **macOS**, στην ηλεκτρονική διεύθυνση: <https://www.audacityteam.org/download/opencvino/>. Στην Εικόνα 5.59 που ακολουθεί, αποτυπώνεται στιγμιότυπο από την σχετική οθόνη του ιστοτόπου του Audacity, στον οποίο είναι διαθέσιμο το λογισμικό για μεταφόρτωση.



Εικόνα 5.59. OpenVINO - Οθόνη ιστοτόπου μεταφόρτωσης της εφαρμογής

Υποστηριζόμενοι τύποι αρχείων ήχου: Υποστηρίζονται οι τύποι **WAV**, **MP3**, **AIFF**, **FLAC**, **OGG** και γενικότερα όλοι οι τύποι που υποστηρίζονται από το Audacity,

τόσο την είσοδο, όσο και στην έξοδο για την αποθήκευση του αποτελέσματος της επεξεργασίας.

Η υλοποίηση και διάθεση του OpenVINO ως πρόσθετο (plugin), αποτελεί παράγοντα διαφοροποίησης του συγκεκριμένου λογισμικού τα προηγούμενα, καθιστώντας την επιλογή του κατάλληλη, ώστε να υπάρχει ευρύτερη και αντιπροσωπευτικότερη κάλυψη όλων των δυνατών περιπτώσεων, και πληρότητα κατά την διεξαγωγή της όλης έρευνας.

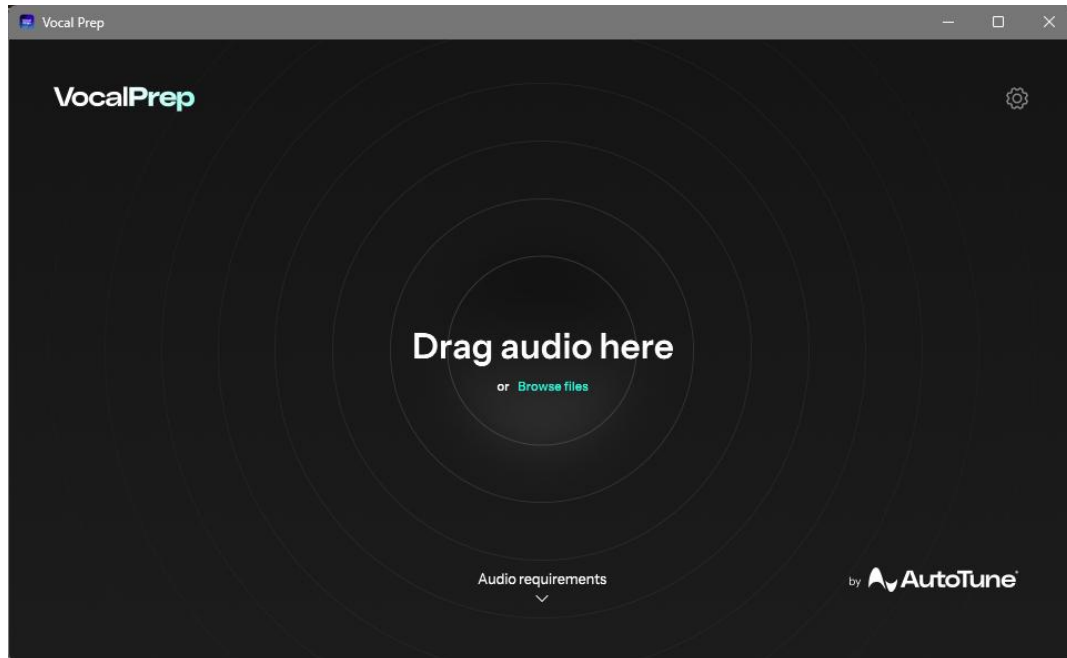
5.2.5 Λογισμικό 5 (S05) - VocalPrep (Antares AutoTune)



Κύρια χρήση και λειτουργίες: Το VocalPrep (AutoTune) της εταιρίας Antares, αποτελεί αυτόνομη (standalone) εφαρμογή επεξεργασίας φωνής που αξιοποιεί τεχνικές Τεχνητής Νοημοσύνης για την αυτόματη αποθορυβοποίηση ηχογραφήσεων. Η βασική του λειτουργία είναι η απομάκρυνση ανεπιθύμητων στοιχείων όπως ο θόρυβος περιβάλλοντος, διατηρώντας παράλληλα τον φυσικό χαρακτήρα της φωνής. Το εργαλείο είναι προορίζεται κυρίως για αποθορυβοποίηση φωνητικού περιεχομένου (τραγούδι, ομιλία, podcast), και επιτρέπει γρήγορη επεξεργασία αρχείων μέσω drag-and-drop, χωρίς να βασίζεται ή να έχει την ανάγκη χρήσης σταθμού επεξεργασίας ψηφιακού ήχου (DAW). Δεν υποστηρίζει λειτουργίες όπως τον διαχωρισμό σε κανάλια με διαφορετικό μουσικό περιεχόμενο (stems).

Βαθμός επέμβασης του λογισμικού: Το VocalPrep λειτουργεί με πλήρως αυτοματοποιημένη λογική (one-click tool), χωρίς δυνατότητα παραμετροποίησης από τον χρήστη. Η επεξεργασία πραγματοποιείται με χρήση Τεχνητής Νοημοσύνης, η οποία αναλύει το ηχητικό σήμα και εφαρμόζει βέλτιστη δυνατή αποθορυβοποίηση, χωρίς χειροκίνητη παρέμβαση. Αυτό μειώνει δραστικά την πολυπλοκότητα χρήσης, αλλά περιορίζει τον έλεγχο του αποτελέσματος. Στην πράξη, η διαδικασία περιλαμβάνει απλά το «σύρσιμο και απόθεση» (drag and drop) του αρχείου πάνω στο αρχικό «παράθυρο» του λογισμικού, ή την αναζήτησή του τοπικά, και η αποθορυβοποίηση συντελείται αυτόματα.

Στην Εικόνα 5.60 που ακολουθεί, αποτυπώνεται στιγμιότυπο από το αρχικό παράθυρο της εφαρμογής.



Εικόνα 5.60. VocalPrep - Αρχική οθόνη εφαρμογής

Διαθέσιμες εκδόσεις λογισμικού: Το VocalPrep διατίθεται αποκλειστικά ως εφαρμογή για τοπική εγκατάσταση, και δεν απαιτεί την ύπαρξη προσθέτων ή άλλων λογισμικών για τη λειτουργία του. Το περιβάλλον εργασίας της εφαρμογής, δεν είναι διαθέσιμο στα ελληνικά.

Κόστος κτήσης - χρήσης: Το λογισμικό παρέχει την δυνατότητα χρέωσης με μηνιαία ή ετήσια συνδρομή, ενώ είναι από τα λίγα πλέον λογισμικά, τα οποία είναι διαθέσιμα και με την αγορά μόνιμης άδειας (perpetual license). Αναφέρεται στο σημείο αυτό, ότι το λογισμικό θα αγοραστεί με μόνιμη άδεια από τον γράφοντα, για χρήση στα πλαίσια της υλοποίησης της παρούσας διπλωματικής εργασίας. Η αγορά θα γίνει με έκπτωση, στην τιμή των 50€, ενώ η τιμή μόνιμης άδειας κτήσης του λογισμικού βρίσκεται στα 100€.

Online λειτουργία: Η λειτουργία του VocalPrep, είναι όπως ήδη έχει αναφερθεί, πλήρως τοπική (offline), καθώς η επεξεργασία πραγματοποιείται στο υπολογιστικό σύστημα του χρήστη, χωρίς ανάγκη σύνδεσης στο διαδίκτυο

Υποστηριζόμενα λειτουργικά συστήματα (OS): Η πρόσβαση στην εφαρμογή γίνεται μέσω τοπικής εγκατάστασης. Υποστηρίζονται τα λειτουργικά συστήματα **Windows** και **macOS**, ενώ το λογισμικό εγκαθίσταται τοπικά μέσω του Auto-Tune Central με δημιουργία λογαριασμού στην πλατφόρμα.

Υποστηριζόμενοι τύποι αρχείων ήχου: Υποστηρίζονται οι ακόλουθοι τύποι ήχου στην είσοδο: **WAV, AIFF, FLAC, MP3** σε περιβάλλον **Windows**, και (επιπλέον) **M4A** σε **macOS**. Στην έξοδο για αποθήκευση, υποστηρίζονται οι τύποι: **WAV** σε **Windows**, και επιπλέον **AIFF, FLAC, MP3** σε **macOS**.

Βασικοί λόγοι που συντελούν στην συμπερίληψη του VocalPrep στα εξεταζόμενα λογισμικά, είναι η δυνατότητα τοπικής του εγκατάστασης, με συνεπακόλουθη την δυνατότητα offline επεξεργασίας, καθώς και η διαθεσιμότητά του και με την μορφή μόνιμης άδειας χρήσης, για όποιον δεν επιθυμεί το συνδρομητικό μοντέλο. Περισσότερες πληροφορίες για την εφαρμογή, είναι διαθέσιμες στην ηλεκτρονική διεύθυνση: <https://www.antarestech.com/products/ai-powered-vocal-chain/vocal-prep>.

5.3 Αποθορυβοποίηση αρχείων από τα επιλεχθέντα λογισμικά

Μετά από την δημιουργία των ενθόρυβων δειγμάτων, και την επιλογή των λογισμικών, θα πραγματοποιηθεί η διαδικασία των αποθορυβοποιήσεων. Σημειώνεται ότι μετά από την διενέργεια των αποθορυβοποιήσεων, δεν θα γίνει καμία περεταίρω επεξεργασία βελτίωσης ή άλλης τροποποίησης στα προϊόντα αρχεία πριν από τις ακροάσεις. Έτσι, το τελικό ηχητικό τους περιεχόμενο θα είναι ακριβώς αυτό που θα παραχθεί από τα λογισμικά σε κάθε περίπτωση, ώστε να μην υπάρχει αλλοίωση στο ηχητικό υλικό προς αξιολόγηση.

Οι τελικές αποθορυβοποιήσεις, θα προέλθουν μετά από εκτεταμένες δοκιμές για κάθε ενθόρυβο δείγμα, με στόχο την επίτευξη του καλύτερου δυνατού αποτελέσματος. Αυτές θα συμπεριλάβουν την δοκιμή διαφορετικών μοντέλων TN, ώστε να χρησιμοποιηθεί το καταλληλότερο σε κάθε περίπτωση, και την εφαρμογή διαφορετικών ρυθμίσεων, όπου βέβαια είναι διαθέσιμες αυτές οι επιλογές.

Ακολουθεί η καταγραφή των βασικότερων σημείων της διαδικασίας των αποθρομβοποιήσεων, που θα πραγματοποιηθεί από τα επιλεγμένα εργαλεία. Η παράθεσή των λογισμικών γίνεται και εδώ με αλφαβητική σειρά.

5.1 Επεξεργασία με το Cleanvoice

Η επεξεργασία των ενθόρυβων αρχείων από το Cleanvoice, θα γίνει με την χρήση των επιθυμητών παραμέτρων από τις διαθέσιμες, οι οποίες θα αποθηκευτούν πρώτα από τον γράφοντα στην πλατφόρμα, με μορφή νέου προτύπου. Οι επιλογές με τις οποίες θα δημιουργηθεί το πρότυπο, αποτυπώνονται στο στιγμιότυπο της Εικόνας 5.61.

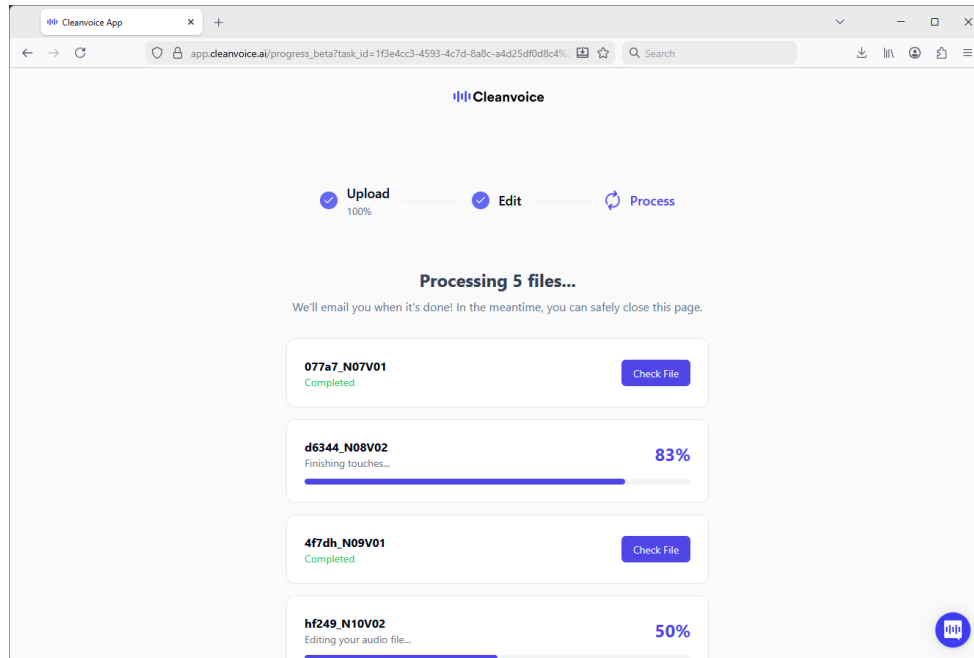
The screenshot shows the 'Koukovetsos' Custom Template' interface in Cleanvoice. It is divided into three main sections: 1. Edit, 2. Enhance, and 3. Export. Each section contains several options with checkboxes or radio buttons. In the 'Edit' section, 'Mute' and 'Mouth Sounds' are checked. In the 'Enhance' section, 'Remove Breath', 'Remove Noise', and 'Normalize' are checked. In the 'Export' section, 'WAV' is selected as the export format. A 'Create Template' button is at the bottom right.

Section	Option	Status
1 Edit	Mute	Checked
	Hesitations	Unchecked
	Long Silences	Unchecked
	Stutters	Checked
	Mouth Sounds	Checked
2 Enhance	Keep Music	Unchecked
	Remove Breath	Checked
	Remove Noise	Checked
	Normalize	Checked
3 Export	Transcription	Unchecked
	Summarize	Unchecked
	Social Content	Unchecked
	Export format	WAV (Selected)

Εικόνα 5.61. Cleanvoice - Οθόνη δημιουργίας προτύπου (Custom Template)

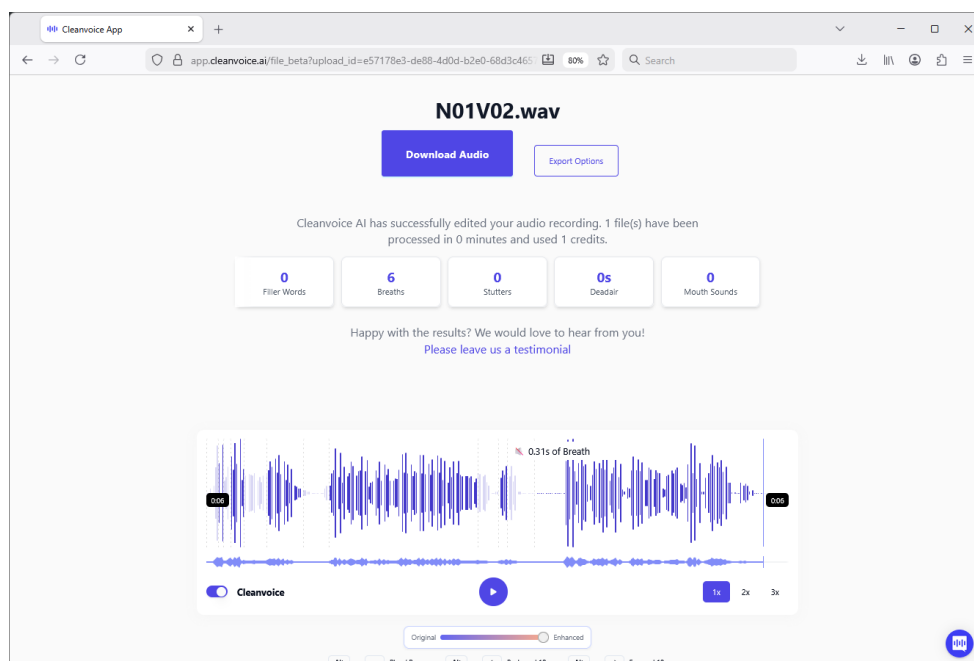
Όπως φαίνεται στην προηγούμενη εικόνα, θα επιλεγεί να γίνουν μικρές επεμβάσεις βελτίωσης του περιεχομένου, οι οποίες θα συμπεριλάβουν την αφαίρεση αναπνοών και άλλων στοματικών ήχων, και την κανονικοποίηση του αποτελέσματος. Για την εξαγωγή των αρχείων, θα επιλεγεί ο τύπος «WAV», χωρίς να διατίθενται περισσότερες επιλογές πάνω σε αυτό. Στο σημείο αυτό, αναφέρεται ότι το Cleanvoice εξάγει μονοφωνικά αρχεία.

Μετά την επιλογή του προτύπου, η επεξεργασία θα πραγματοποιηθεί αυτόματα από το λογισμικό, όπως φαίνεται στην Εικόνα 5.62 που ακολουθεί.



Εικόνα 5.62. Cleanvoice - Οθόνη από το στάδιο της επεξεργασίας

Στην επόμενη Εικόνα 5.63, αποτυπώνεται η οθόνη επισκόπησης, πριν την εξαγωγή.

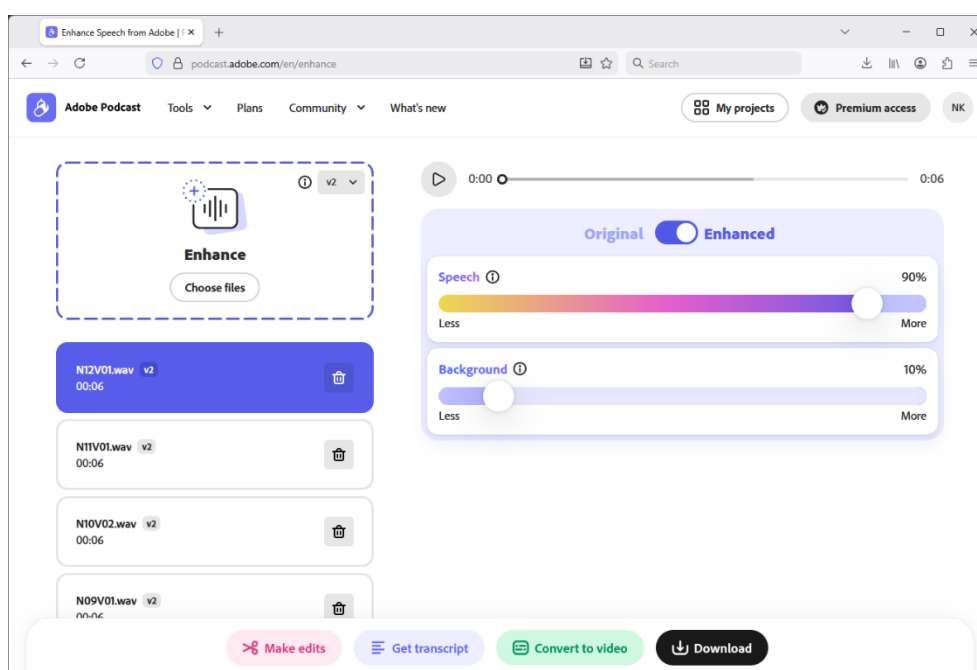


Εικόνα 5.63. Cleanvoice - Οθόνη επισκόπησης και εξαγωγής προϊόντων επεξεργασίας

5.2 Επεξεργασία με το Enhance

Σε αντίθεση με το Cleanvoice, η επεξεργασία των ενθόρυβων αρχείων από το Enhance, θα γίνει με περισσότερο αυτοματοποιημένο τρόπο. Στην περίπτωση αυτή, δεν διατίθεται η επιλογή ρυθμίσεων σε μορφή προτύπου, όπως στο προηγούμενο λογισμικό.

Μετά από το ανέβασμα των ενθόρυβων αρχείων στην πλατφόρμα, θα επιλεγεί το επιθυμητό ποσοστό της ομιλίας και του θορύβου παρασκηνίου, χρησιμοποιώντας δύο διαφορετικούς ράβδους κύλισης, όπως φαίνεται στην Εικόνα 5.64 που ακολουθεί, διαθέσιμους στην έκδοση του μοντέλου «v2», που είναι και το πιο πρόσφατο,.

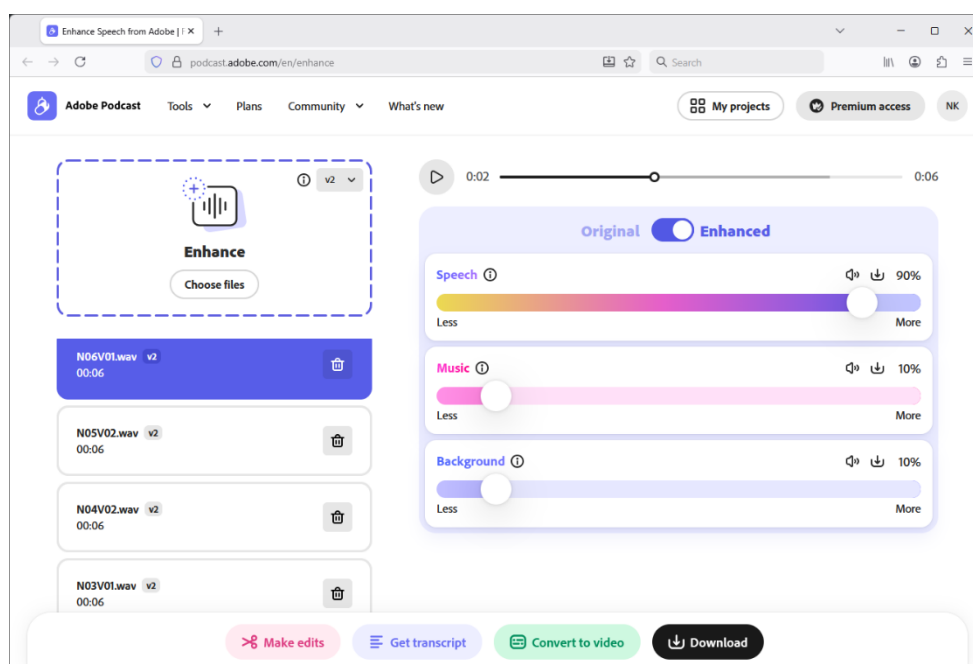


Εικόνα 5.64. Enhance - Οθόνη από το στάδιο της επεξεργασίας με το μοντέλο v2

Θα ακολουθήσει η διενέργεια εκτενών δοκιμών, όπου θα επιλεγεί -με υποκειμενικό τρόπο από τον γράφοντα- η «χρυσή τομή» ανάμεσα στην διατήρηση της ποιότητας του ήχου και της καταστολής του θορύβου. Διαπιστώνεται ότι οι βέλτιστες τιμές είναι διαφορετικές για κάθε περίπτωση. Για την πληρότητα της έρευνας, θα πραγματοποιηθούν δοκιμαστικές αποθρομβοποιήσεις και με το προηγούμενο μοντέλο της εφαρμογής, με το όνομα «v1», οι οποίες θα εμφανίσουν πολύ «φτωχά» αποτελέσματα σε σχέση με το νέο μοντέλο, ενώ για την επεξεργασία του είναι διαθέσιμη μόνο μία ράβδος κύλισης, αυτή της ρύθμισης

ποσοστού βελτίωσης της ομιλίας. Σημειώνεται ότι οι επιλογές ρύθμισης δεν είναι διαθέσιμες στην δωρεάν χρήση της εφαρμογής, και για τα δύο μοντέλα.

Η υλοποίηση της επεξεργασίας των ενθόρυβων αρχείων από το Enhance, θα ολοκληρωθεί τον Φεβρουάριο του 2026. Λίγο αργότερα, τον Μάρτιο του ίδιου έτους, η εφαρμογή θα ενημερωθεί, με την προσθήκη και τρίτης ράβδου κύλισης, στο μοντέλο «v2», όπως φαίνεται στην Εικόνα 5.65 που ακολουθεί.



Εικόνα 5.65. Enhance - Επιλογές μοντέλου v2 διαθέσιμες από τον Μάρτιο του 2026

Η αλλαγή αυτή, όπως και αρκετές άλλες, οι οποίες διαπιστώνονται στην διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας, αποδεικνύει ότι τα εργαλεία TN, εξελίσσονται καθημερινά, με ταχύτατο ρυθμό, καθιστώντας την οποιαδήποτε μελέτη δυνητικά ανεπικάιρη, προϊόντος του χρόνου.

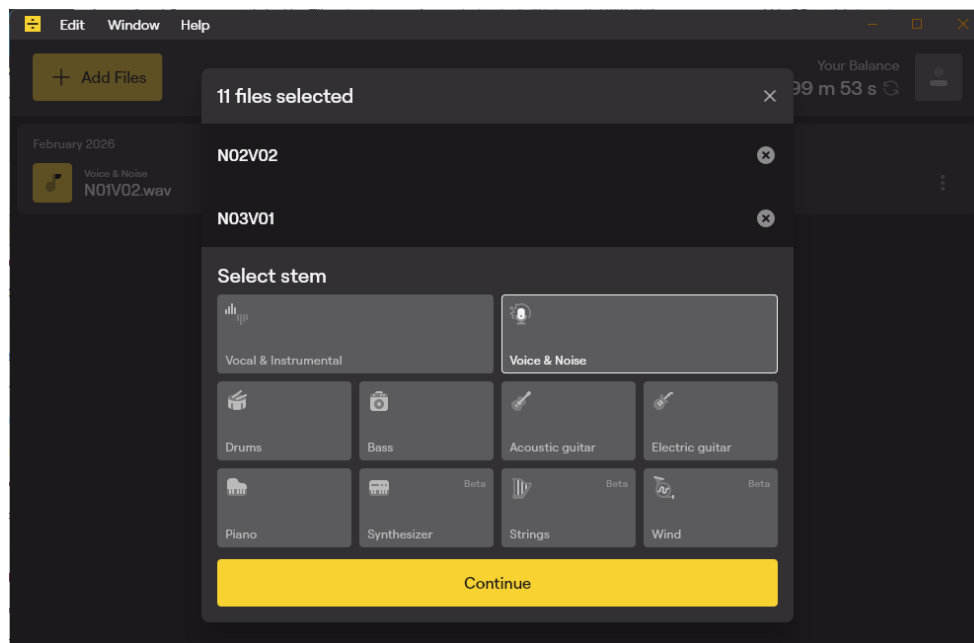
5.3 Επεξεργασία με το LALAL.AI

Για την αποθορυβοποίηση με το LALAL.AI, θα ακολουθηθούν δύο κύριες κατευθύνσεις, ανάλογα με το αν στο ηχητικό περιεχόμενο περιλαμβάνεται ή όχι θόρυβος από δευτερεύουσες φωνές ή μουσική στο παρασκήνιο. Ο λόγος για τον διαχωρισμό αυτό,

οφείλεται στο γεγονός ότι το λογισμικό είναι προσανατολισμένο στην επεξεργασία κυρίως μουσικού περιεχομένου.

Έτσι στην πρώτη περίπτωση, κρίνεται περισσότερο αποδοτική η επιλογή «**Vocal & Instrumental**», καθώς αυτή δίνει την δυνατότητα για διαχωρισμό σε κύριες και δευτερεύουσες φωνές, ενώ στην δεύτερη περίπτωση όπου στο σήμα δεν υπήρχαν δευτερεύουσες ομιλίες ή μουσική, δίνεται η επιλογή «**Voice & Noise**».

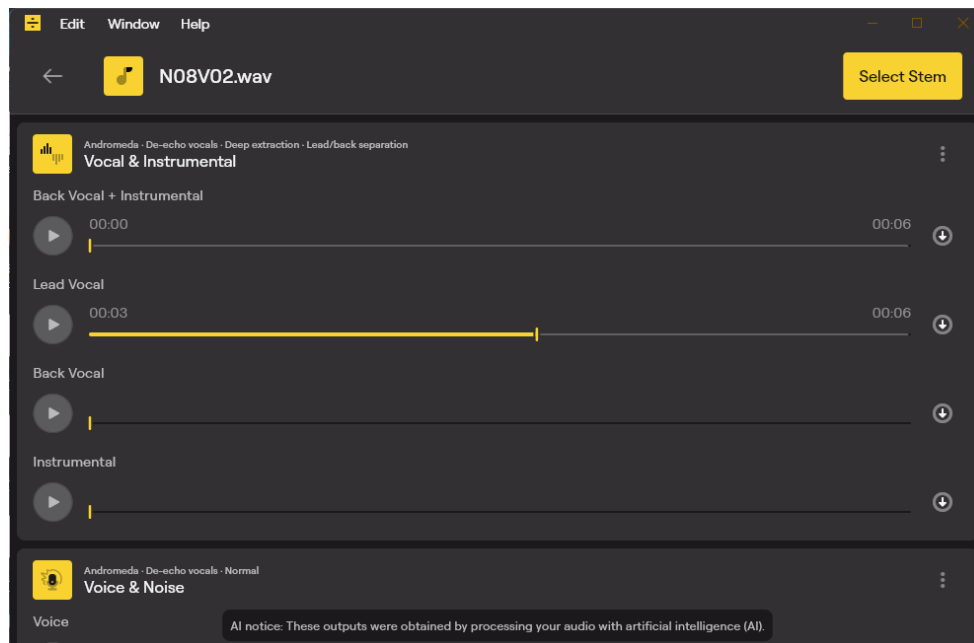
Στην Εικόνα 5.66 που ακολουθεί, περιλαμβάνεται η οθόνη επιλογής των διαφορετικών τρόπων διαχωρισμού σε κανάλια (stems).



Εικόνα 5.66. LALAL.AI - Οθόνη επιλογής καναλιών επεξεργασίας

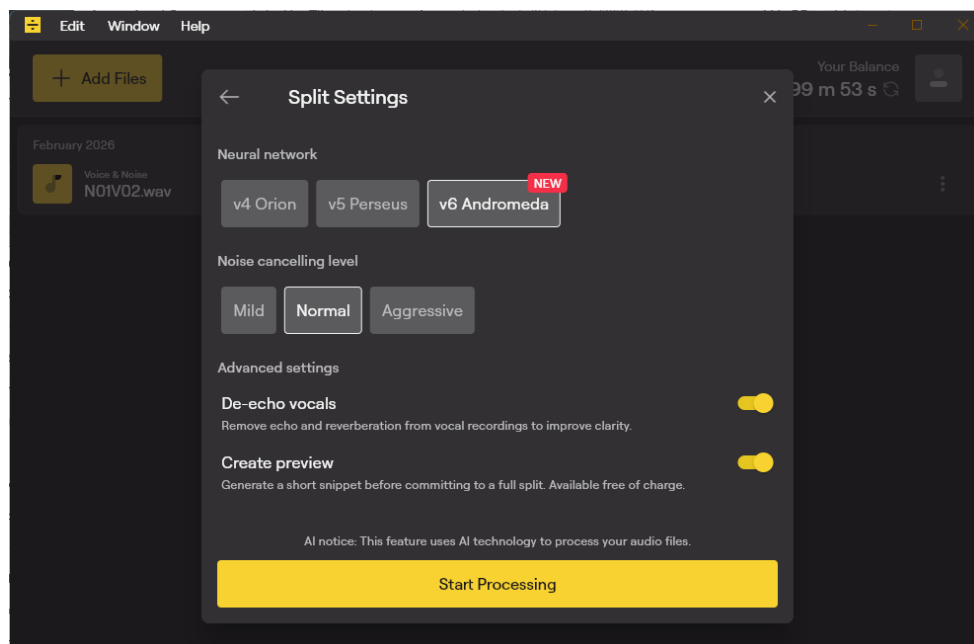
Σύμφωνα με την παραπάνω προσέγγιση, στην περίπτωση του θορύβου που περιλαμβάνει ομιλίες πλήθους (N06), με την επιλογή «Vocal & Instrumental», το τελικό αποθορυβοποιημένο αρχείο, θα προκύψει με εξαγωγή του καναλιού (stem) «**Lead Vocal**», ενώ οι ανεπιθύμητες ομιλίες παρασκήνιου, θα κατευθυνθούν από το λογισμικό στα κανάλια «**Back Vocal + Instrumental**» και «**Back Vocal**».

Η οθόνη που απεικονίζει τον διαχωρισμό αυτό, αποτυπώνεται στο στιγμιότυπο της Εικόνας 5.67 που ακολουθεί, στην επόμενη σελίδα.



Εικόνα 5.67. LALAL.AI - Οθόνη επιλογής καναλιών επεξεργασίας

Για όλες τις αποθορυβοποιήσεις, θα χρησιμοποιηθεί σαν περισσότερο αποτελεσματικό, το μοντέλο «v6 Andromeda», με επιλογή μέσης παρέμβασης (επίπεδο «Normal»), όπως φαίνεται στην οθόνη της ακόλουθης Εικόνας 5.68.



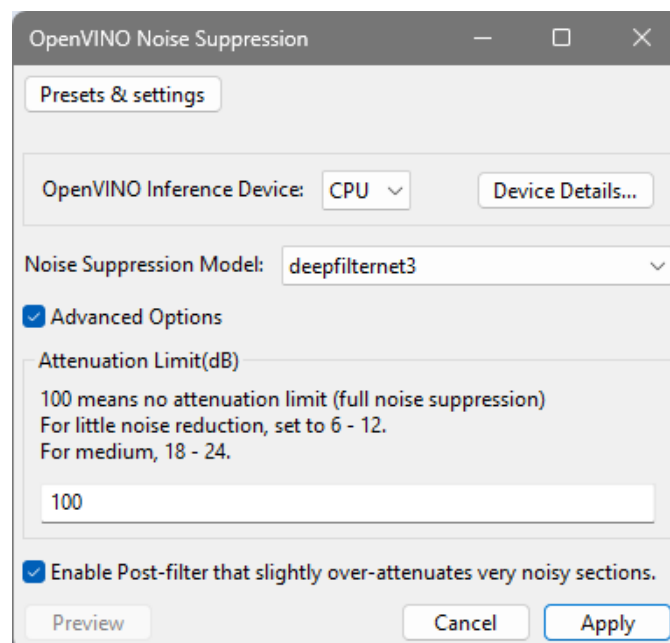
Εικόνα 5.68. LALAL.AI - Οθόνη επιλογής παραμέτρων επεξεργασίας

5.4 Επεξεργασία με το OpenVINO

Αφού εγκατασταθούν τοπικά και οι τρεις διαφορετικές εκδόσεις των διαθέσιμων μοντέλων καταστολής θορύβου του OpenVINO, («**deepfiltenet2**», «**deepfiltenet3**» και «**noise-suppression-denseunet-II-0001**»), η αποθορυβοποίηση με το OpenVINO, θα πραγματοποιηθεί μετά από διαδοχικές δοκιμές αποτελεσματικότητας, κυρίως με την χρήση του δεύτερου μοντέλου (**deepfiltenet3**), επειδή συνολικά δίνει πιο φυσικά αποτελέσματα, ενώ θα επιλεγεί η επεξεργασία να γίνει με χρήση της CPU του υπολογιστή (επιλογή: **OpenVINO Inference Device**).

Το μοντέλο **deepfiltenet3**, παρέχει όπως και το πρώτο μοντέλο (**deepfiltenet2**) την δυνατότητα ρύθμισης του επιθυμητού βαθμού περιορισμού του θορύβου (επιλογή: «**Attenuation Limit (dB)**»), με την επιπλέον δυνατότητα φιλτραρίσματος του αποτελέσματος μετά την διεργασία (επιλογή: «**Enable Post-filter that slightly over-attenuates very noisy sections**»), η οποία δεν είναι διαθέσιμη στο πρώτο μοντέλο.

Στην Εικόνα 5.69 που ακολουθεί, αποτυπώνεται στιγμιότυπο από το παράθυρο επιλογής των μοντέλων και των διαθέσιμων παραμέτρων τους

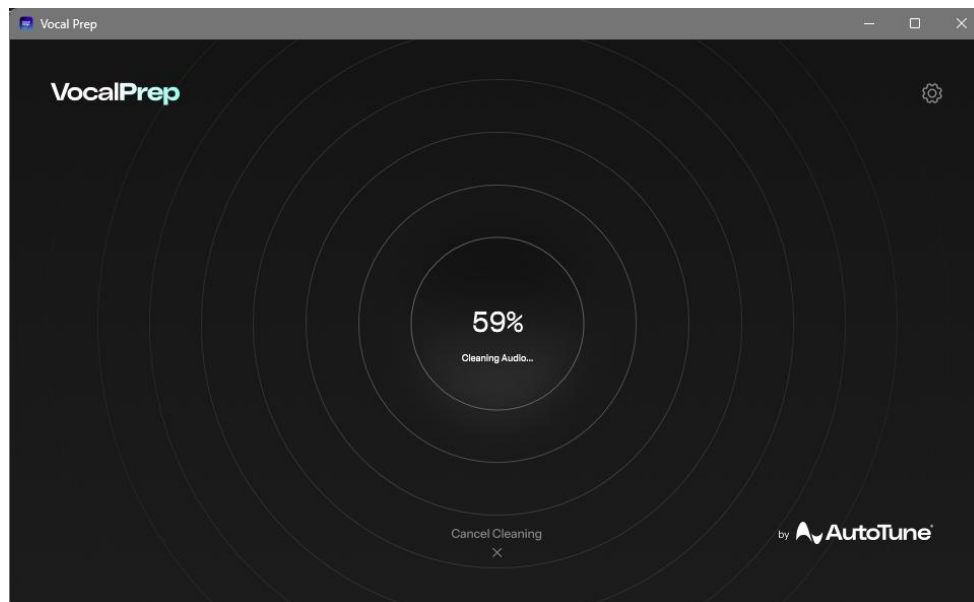


Εικόνα 5.69. OpenVINO - Οθόνη επιλογής μοντέλου αποθορυβοποίησης

Το μοντέλο noise-suppression-denseunet-II-0001, δεν διαθέτει κάποια από τις δύο αυτές δυνατότητες. Σε κάθε περίπτωση, η επεξεργασία με το OpenVINO, μπορεί να χαρακτηριστεί και αυτή ως αυτοματοποιημένη.

5.5 Επεξεργασία με το VocalPrep

Η αποθορυβοποίηση με την χρήση του VocalPrep, θα γίνει με εντελώς αυτοματοποιημένο τρόπο, με μηδενική δυνατότητα παρέμβασης, κατά την εκτέλεση της διαδικασίας. Στην Εικόνα 5.70 που ακολουθεί, εμφανίζεται στιγμιότυπο από την οθόνη εργασίας κατά την διάρκεια της αποθορυβοποίησης.



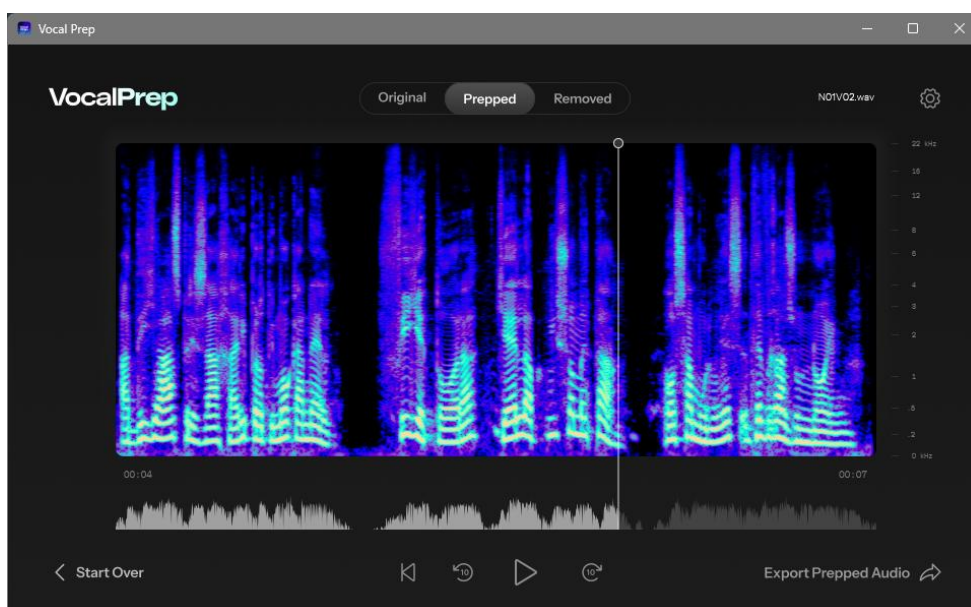
Εικόνα 5.70. VocalPrep - Οθόνη προόδου κατά την αποθορυβοποίηση

Η διάρκεια εκτέλεσης της αποθορυβοποίησης, όπως και στην περίπτωση του OpenVINO, θα εξαρτηθεί από την ισχύ του χρησιμοποιηθέντος υπολογιστικού συστήματος, επειδή το λογισμικό θα λειτουργεί τοπικά και όχι μέσω κάποιας πλατφόρμας. Για την πληρότητα αναφέρεται, ότι η διάρκεια αυτή είναι περίπου ισόχρονη με την διάρκεια των δειγμάτων, ενώ δεν θα εξετασθεί η χρονική επίδοση σε δείγματα σημαντικά μεγαλύτερης διάρκειας.

Οι τεχνικές προδιαγραφές των μέσων που θα χρησιμοποιηθούν, κατά την υλοποίηση της παρούσας διπλωματικής εργασίας, περιλαμβάνονται στο **Παράρτημα Β**.

Ενδιαφέρουσα προσθήκη στο λογισμικό αυτό, αποτελεί η γραφική απεικόνιση των φασματογραφημάτων, του ενθόρυβου (Original), του αποθορυβοποιημένου (Prepped) και του θορύβου που αφαιρείται (Removed), με δυνατότητα ακρόασης και των τριών σημάτων, για αξιολόγηση του αποτελέσματος πριν την αποθήκευση, χωρίς όμως να υπάρχει δυνατότητα άλλης παρέμβασης, ή αποθήκευσης του σήματος του θορύβου που αφαιρείται, σε περίπτωση που κάτι τέτοιο ήταν επιθυμητό.

Στην Εικόνα 5.71 που ακολουθεί, εμφανίζεται στιγμιότυπο από την σχετική οθόνη της εφαρμογής.

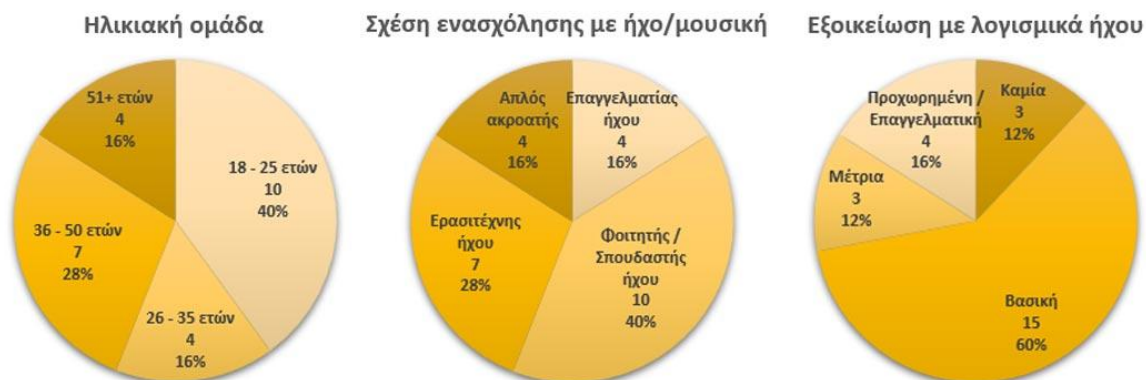


Εικόνα 5.71. VocalPrep - Οθόνη φασματογραφημάτων διαχωρισμένων ήχων

5.4 Αξιολόγηση αποτελεσματικότητας λογισμικών

Για την αξιολόγηση των αποτελεσμάτων της πειραματικής διαδικασίας, θα πραγματοποιηθεί διενέργεια «τυφλών» κριτικών ακροάσεων, από ομάδα **εικοσιπέντε (25)** ατόμων, που θα επιλεγούν για τον σκοπό αυτό. Οι ακροάσεις θα διενεργηθούν κατά την διάρκεια των μηνών Μάρτιου και Απριλίου 2026. Η διάρκεια των ακροάσεων, θα κυμανθεί από 20 έως 30 λεπτά της ώρας, για κάθε περίπτωση.

Στο Σχήμα 5.1 που ακολουθεί, αποτυπώνονται με γραφικό τρόπο τα δημογραφικά στοιχεία των ακροατών, όσον αφορά στην ηλικιακή τους κατανομή, την σχέση ενασχόλησής τους με τον ήχο ή / και την μουσική, και τον βαθμό εξοικείωσής τους με τη χρήση λογισμικών επεξεργασίας ήχου.



Σχήμα 5.1. Δημογραφικά στοιχεία της ομάδας ακροατών

Τα συνολικά τελικά αποθορυβοποιημένα δείγματα προς αξιολόγηση, θα είναι **εξήντα (60)**, σαν αποτέλεσμα του γινομένου των **δώδεκα (12)** διαφορετικών θορύβων, επί τα **πέντε (5)** λογισμικά, που εμπλέκονται στην διαδικασία. Για την αξιολόγησή τους, τα δείγματα θα χωριστούν σε ομάδες, οργανωμένες ανά διαφορετικό θόρυβο. Έτσι, θα δημιουργηθούν δώδεκα φάκελοι αντίστοιχα, με ονομασίες από «01» έως και «12».

Στην συνέχεια, θα γίνει ανακατανομή και μετονομασία των αρχείων ανά φάκελο, με τυχαίο καταγεγραμμένο τρόπο, ώστε η σειρά ακρόασης των ηχητικών δειγμάτων να αντιστοιχεί σε διαφορετικό λογισμικό κάθε φορά, για αντικειμενικότερα αποτελέσματα.

Με τον τρόπο αυτό, σε κάθε έναν από τους φακέλους θα συμπεριληφθούν **έξι (1+5)** αρχεία. Το ένα αντίστοιχο ενθόρυβο αρχείο, με την κωδική ονομασία «xx_(Noisy).wav» (όπου xx = 01 έως 12) και τα πέντε αρχεία των αποθορυβοποιήσεων από τα λογισμικά, με τις νέες κωδικές ονομασίες «xxA.wav» έως «xxE.wav», που θα προκύψουν από την τυχαία αντιστοίχιση με τα πέντε λογισμικά, (με κωδικές ονομασίες A, B, C, D και E).

Στον Πίνακα 5.6 που ακολουθεί, αποτυπώνεται η τυχαία διάταξη, από την οποία θα προκύψουν οι νέες κωδικές ονομασίες. Ο πίνακας αυτός, θα χρησιμοποιηθεί και μετά το τέλος των ακροάσεων, για την αντιστοίχιση τους με τα αρχικά αποθορυβοποιημένα αρχεία. Οι αριθμοί που εμφανίζονται, αποτυπώνουν την ταυτότητα των λογισμικών.

Πίνακας 5.6. Τυχαία διάταξη της σειράς ακρόασης των λογισμικών

Σειρά Ακρόασης	01	02	03	04	05	06	07	08	09	10	11	12
Ακρόαση Α	3	2	5	1	4	3	2	5	1	4	2	3
Ακρόαση Β	1	4	3	5	2	5	1	4	3	5	3	4
Ακρόαση C	5	1	2	4	3	1	4	3	5	2	1	5
Ακρόαση D	2	5	1	3	5	4	3	2	4	1	4	2
Ακρόαση E	4	3	4	2	1	2	5	1	2	3	5	1

Έτσι, για παράδειγμα, το αρχείο «**04B.wav**» αντιστοιχεί στην αποθορυβοποίηση του 4^{ου} θορύβου, που θα γίνει από το 5^ο σε αλφαβητική σειρά λογισμικό (ελικόπτερο από το VocalPrep), και είναι το 2^ο κατά σειρά αποθορυβοποιημένο στον φάκελο «**04**».

Όπως φαίνεται και στον προηγούμενο πίνακα, για την ευκολότερη εποπτεία των αποτελεσμάτων, θα εφαρμοστεί χρωματική κωδικοποίηση, με τυχαία απόδοση χρωμάτων στα εξεταζόμενα λογισμικά. Ο χρωματικός αυτός κώδικας, χρησιμοποιείται και στο επόμενο κεφάλαιο, και δίνεται στην Εικόνα 5.72, που ακολουθεί.

1 - Cleanvoice
2 - Enhance
3 - LALAL.AI
4 - OpenVINO
5 - VocalPrep

Εικόνα 5.72. Χρωματικός κώδικας περιγραφής λογισμικών

Για αντικειμενικότερη αξιολόγηση, δεν θα γίνουν γνωστές οι ταυτότητες των λογισμικών στους ακροατές πριν τις ακροάσεις, ενώ κατά την διάρκειά τους θα διασφαλιστούν οι απαραίτητες συνθήκες, με την διενέργειά τους σε ήσυχο περιβάλλον με την χρήση ακουστικών, όπως έχει προκαθοριστεί στην φάση του σχεδιασμού.

Στους ακροατές θα δοθούν οι απαραίτητες οδηγίες και στην συνέχεια θα κληθούν να βαθμολογήσουν την επίδοση των λογισμικών. Αξιολογούνται συνδυαστικά ο βαθμός της καταστολής του θορύβου και η ποιότητα το τελικού αποθορυβημένου ήχου. Για την

αξιολόγηση της επίδοσης, επιλέγεται η **5βαθμη** κλίμακα αξιολόγησης **MOS**, όπου: **1 = Πολύ κακή**, **2 = Κακή**, **3 = Μέτρια**, **4 = Καλή** και **5 = Πολύ Καλή** επίδοση.

Στην Εικόνα 5.73 που ακολουθεί, εμφανίζεται φωτογραφικό στιγμιότυπο από την διενέργεια των ακροάσεων. (Η φωτογράφιση έγινε με την συναίνεση της ακροάτριας).



Εικόνα 5.73. Στιγμιότυπο από την διαδικασία των ακροάσεων

Οι ακροάσεις θα πραγματοποιηθούν ανά θόρυβο, όπου οι ακροατές θα ακούνε το κάθε ενθόρυβο αρχείο και στην συνέχεια θα ακούνε τις πέντε (5) αποθρονοποιημένες εκδόσεις του, όπως εμφανίζονται στον Πίνακα 5.1. Τα ακούσματα θα γίνουν διαδοχικά, με μικρό κενό ανάμεσά τους, ώστε το αίσθημα από το κάθε δείγμα να είναι διακριτό.

Τα ποσοτικά αποτελέσματα των ακροάσεων, θα αποτυπωθούν μέσω έντυπου ερωτηματολογίου, που θα δημιουργηθεί για την συλλογή των δεδομένων που θα προκύψουν, όπως αποτυπώνεται και στην παραπάνω Εικόνα. Η επεξεργασία και η γραφική απεικόνιση των αποτελεσμάτων, θα γίνει με την χρήση του εργαλείου λογιστικών φύλων **Excel** της εταιρίας **Microsoft**, ενώ η ποιοτική τους αποτίμηση θα γίνει μέσα από την καταγραφή σχολίων και παρατηρήσεων. Τα αποτελέσματα θα καταγραφούν αναλυτικά στο επόμενο κεφάλαιο της εργασίας, ενώ το ερωτηματολόγιο μαζί με τις γραπτές οδηγίες προς τους αξιολογητές, θα περιλαμβάνεται για λόγους πληρότητας στο **Παράρτημα Γ**.

6. Αποτελέσματα Πειραματικής Διαδικασίας - Πιθανές Εφαρμογές

Μετά από την υλοποίηση των αποθορυβοποιήσεων, ακολούθησε η κριτική ακρόαση των προϊόντων αρχείων, και η συμπλήρωση ερωτηματολογίων από τους ακροατές, ώστε να αξιολογηθούν και να καταγραφούν τα αποτελέσματα της διαδικασίας.

Στο κεφάλαιο αυτό, περιλαμβάνεται η αποτίμηση της αποτελεσματικότητας των εργαλείων που εξετάστηκαν, με παράθεση των ευρημάτων που προέκυψαν κατά την διαδικασία των αποθορυβοποιήσεων από τον γράφοντα, και των ποσοτικών αποτελεσμάτων, τα οποία συγκεντρώθηκαν από τα ερωτηματολόγια που συμπληρώθηκαν κατά την διενέργεια των ακροάσεων.

Επίσης περιλαμβάνονται παρατηρήσεις και σχόλια των ακροατών, τα οποία προήλθαν μετά από την διενέργεια των ακροάσεων, ενώ γίνεται αναφορά στην υποκειμενικότητα της αξιολόγησης, όπως αυτή αποτυπώθηκε στα ερωτηματολόγια, μέσω της διασποράς που εμφανίστηκε στις απαντήσεις.

Το κεφάλαιο ολοκληρώνεται με την παράθεση των πιθανών εφαρμογών και χρήσεων, τις οποίες μπορούν να έχουν σε τομείς της καθημερινής πρακτικής, τα εργαλεία λογισμικού αποθορυβοποίησης ψηφιακού ήχου.

6.1 Αποτίμηση αποτελεσματικότητας επιλεγμένων εργαλείων

Η αποτίμηση της αποτελεσματικότητας των επιλεγμένων εργαλείων λογισμικού, πραγματοποιήθηκε τόσο από τον γράφοντα, κατά την διάρκεια της χρήσης των λογισμικών, όσο και από την ομάδα ατόμων που συμμετείχε στην διενέργεια των ακροάσεων. Τα ευρήματα ακολουθούν στην συνέχεια.

6.1.1 Ευρήματα κατά την διαδικασία αποθορυβοποίησης

Κατά την διενέργεια των αποθορυβοποιήσεων, δύο (2) από τα πέντε (5) εργαλεία λογισμικού που χρησιμοποιήθηκαν, (τα **Cleanvoice** και **Enhance**), παρήγαγαν μονοφωνικά αρχεία, παρότι είχαν τροφοδοτηθεί με στερεοφωνικά σήματα. Μάλιστα στην περίπτωση του Enhance παρατηρήθηκε ότι μεταβλήθηκε και ο ρυθμός δειγματοληψίας των προϊόντων αρχείων σε **48000 Hz**, από τα **44100 Hz** που είχαν αρχικά όλα τα δείγματα.

Μετά από έρευνα που πραγματοποιήθηκε στο διαδίκτυο, διαπιστώθηκε ότι η εξαγωγή μονοφωνικών αρχείων είναι κοινή πρακτική σε αρκετές περιπτώσεις λογισμικών, για τον περιορισμό του τελικού μεγέθους, που αποσκοπεί στην ευκολότερη (ταχύτερη) μετάδοση της πληροφορίας. Μάλιστα για τον σκοπό αυτό υπάρχουν περιπτώσεις, (Krisp, n.d.), όπου μειώνεται αντίστοιχα και ο ρυθμός δειγματοληψίας, με παράλληλη σχετική υποβάθμιση της ποιότητας του ήχου.

Από την υλοποίηση της διαδικασίας, διαπιστώθηκε και πειραματικά ότι οι επιδόσεις της TN εξαρτήθηκαν μεν σε κάποιο βαθμό, αλλά επηρεάστηκαν λιγότερο από τα SNR των ενθόρυβων σημάτων, και περισσότερο από το καθεαυτό περιεχόμενο του κάθε θορύβου. Αυτό πιθανότατα εξαρτάται από το κατά πόσο οι συγκεκριμένοι θόρυβοι συμπεριλαμβάνονται στα αποθηκευμένα δείγματα - μοντέλα του κάθε λογισμικού. Έτσι παρατηρήθηκε ότι ενθόρυβα σήματα με μικρότερο ή αρνητικό SNR, αλλά διαφορετικά χαρακτηριστικά από άλλα με μεγαλύτερο SNR, αποθορυβοποιήθηκαν με μεγαλύτερο βαθμό επιτυχίας από εκείνα.

Στις περιπτώσεις που υπήρχε η δυνατότητα επιλογής, προέκυψε ότι η απομείωση του θορύβου πάνω από κάποιο όριο προκάλεσε αλλοιώσεις στην ομιλία, οι οποίες υποβάθμισαν αισθητά την ποιότητα του ήχου. Έτσι έπρεπε να βρεθεί ή «χρυσή τομή» για κάθε περίπτωση, εισάγοντας το υποκειμενικό στοιχείο κατά την παραγωγή του τελικού αποτελέσματος προς ακρόαση, όπως έχει ήδη αναφερθεί σε προηγούμενο σημείο της παρούσας διπλωματικής εργασίας.

Ιδιαίτερο ενδιαφέρον παρουσίασαν τα ενθόρυβα σήματα στα οποία υπήρξαν ομιλίες παρασκηνίου (**Θόρυβος N06**) και μουσική (**Θόρυβος N08**). Στην πρώτη περίπτωση το λογισμικό **LALAL.AI** υπερείχε αισθητά από τα υπόλοιπα, λόγω του ιδιαίτερου προσανατολισμού του λογισμικού σε επεξεργασία μουσικού περιεχομένου, με δημιουργία

πολυκάναλων αποτελεσμάτων, παρέχοντας την επιλογή διαχωρισμού ανάμεσα σε κύριες και δευτερεύουσες φωνές (**lead** και **back vocals**).

Παρόλα αυτά, στην περίπτωση της μουσικής εμφάνισε αστοχία σε ένα πολύ συγκεκριμένο σημείο του δείγματος, όπου υπήρχε εφέ «σκρατς», και μάλιστα μόνο σε ένα από τα τρία παρόμοια εφέ που υπήρχαν στο δείγμα. Αυτό ήταν κάτι το αξιοσημείωτο καθώς ακριβώς την ίδια αστοχία εμφάνισαν και τα πέντε λογισμικά, κάτι που φανερώνει ότι οι αλγόριθμοι TN που εφαρμόστηκαν, θεώρησαν στο σύνολό τους τον θόρυβο αυτό σαν περιεχόμενο ομιλίας. Σημειώνεται ότι αυτός ήταν και ο κύριος λόγος για την χαμηλή βαθμολογία των λογισμικών στο συγκεκριμένο δείγμα, ενώ στην υπόλοιπη χρονική εξέλιξη του ήχου η επίδοση της αποθορυβοποίησης ήταν καλή.

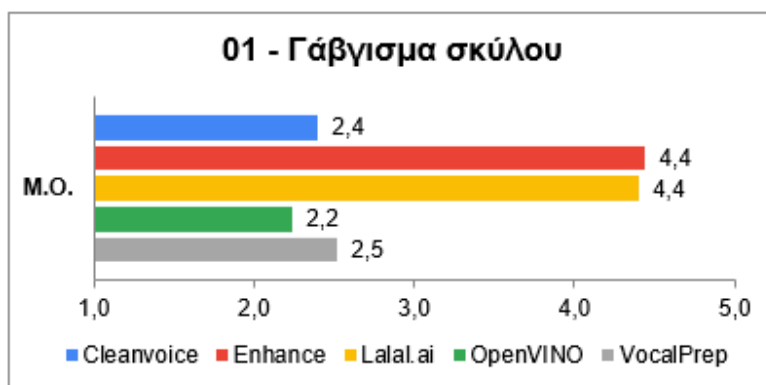
Διαπιστώθηκε επίσης ότι οι αλγόριθμοι TN που εφαρμόστηκαν σε κάποιες περιπτώσεις, επέτρεψαν θορύβους που βρίσκονταν στην αρχή των σημάτων να παραμείνουν, ενώ κατέστειλαν τους ίδιους θορύβους στην συνέχεια, κάτι που παρατηρήθηκε ιδιαίτερα στην περίπτωση του χτύπου του ρολογιού τοίχου (**θόρυβος N05**). Η παρατήρηση αυτή έρχεται σε συμφωνία με παρόμοιες διαπιστώσεις, οι οποίες έχουν κατά καιρούς δημοσιευθεί στο διαδίκτυο, αναφέροντας ότι μερικές φορές τα Νευρωνικά Δίκτυα μπορεί να επιτρέψουν τον θόρυβο στην αρχή των ηχητικών σημάτων που επεξεργάζονται (Richmond, 2022).

Τέλος ενδιαφέρουσα ήταν και η αστοχία του λογισμικού **VocalPrep** στην περίπτωση των ήχων σειρήνας με κορναρίσματα (**θόρυβος N11**), όπου ο αλγόριθμος ουσιαστικά σταμάτησε να λειτουργεί στο τελευταίο χρονικό σκέλος του δείγματος. Σημειώνεται ότι στην περίπτωση αυτή έγιναν αρκετές επαναληπτικές δοκιμές οι οποίες όμως οδήγησαν στο ίδιο αποτέλεσμα.

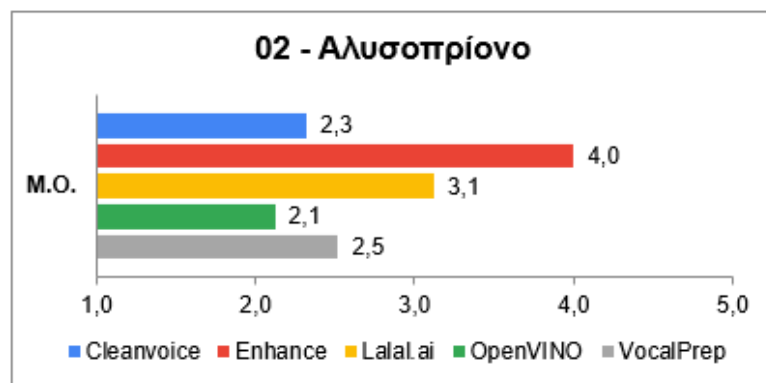
6.1.2 Αποτελέσματα από την διενέργεια των ακροάσεων

Στο σημείο αυτό, αποτυπώνονται με την χρήση κατάλληλων γραφημάτων, τα ποσοτικά αποτελέσματα των δεδομένων που συλλέχθηκαν και επεξεργάστηκαν από τα συμπληρωμένα ερωτηματολόγια, μετά από την διενέργεια των ακροάσεων, στα οποία εντοπίζονται ευρήματα που αναφέρθηκαν και στην προηγούμενη υποενότητα.

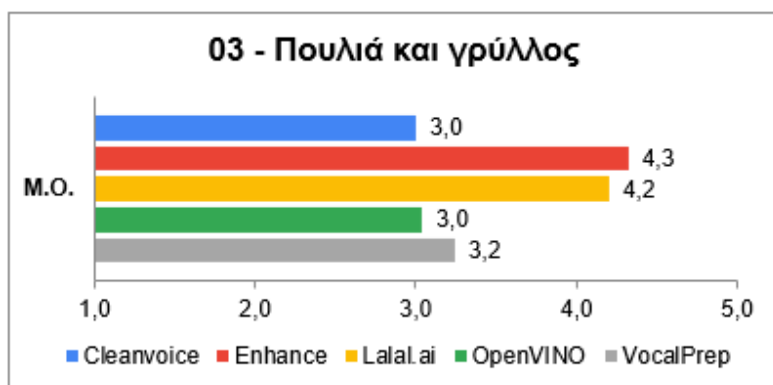
Στα Σχήματα 6.1 έως και 6.12 που ακολουθούν, εμφανίζονται με συγκριτικό τρόπο, οι μέσοι όροι της υποκειμενικής αξιολόγησης που διενεργήθηκε για την επίδοση των λογισμικών σε κάθε αποθορυβοποίηση, όπως βαθμολογήθηκαν από τους ακροατές. Οι απεικονίσεις είναι οργανωμένες ανά θόρυβο, ενώ για την οπτική διαφοροποίηση των λογισμικών ακολουθείται σταθερά ο ίδιος χρωματικός κώδικας, ο οποίος ήδη έχει αναφερθεί σε προηγούμενο σημείο της παρούσας εργασίας.



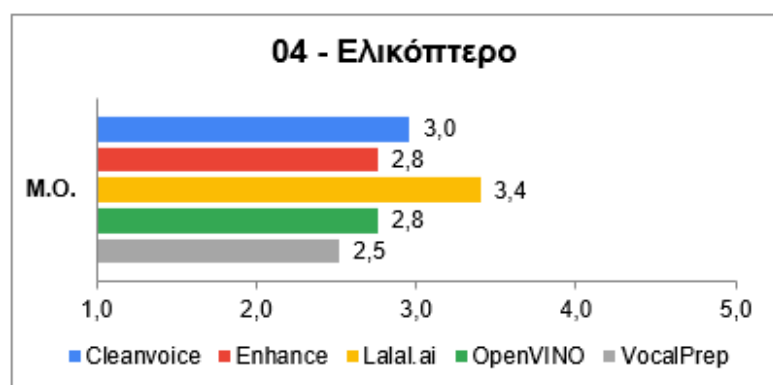
Σχήμα 6.1. Επίδοση λογισμικών στην αποθορυβοποίηση του 1^{ου} δείγματος



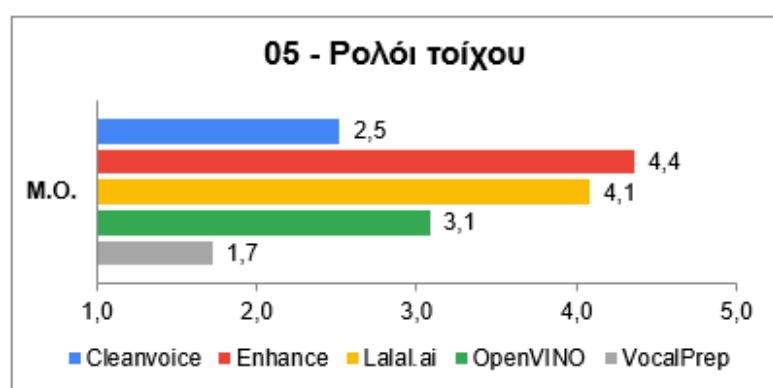
Σχήμα 6.2. Επίδοση λογισμικών στην αποθορυβοποίηση του 2^{ου} δείγματος



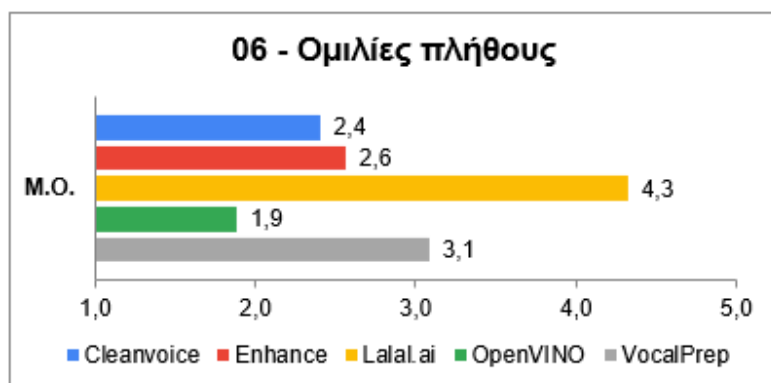
Σχήμα 6.3. Επίδοση λογισμικών στην αποθρομβοποίηση του 3^{ου} δείγματος



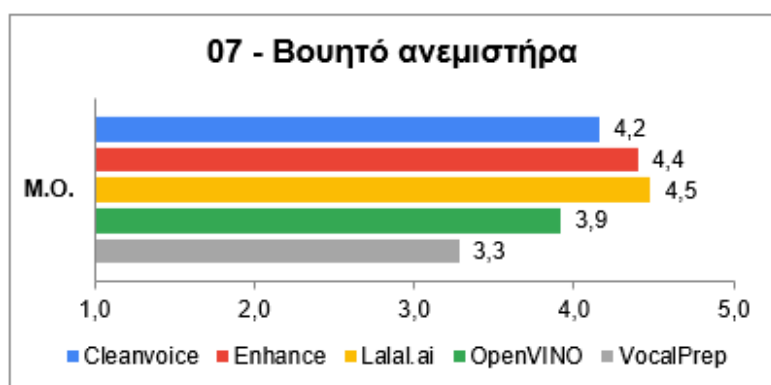
Σχήμα 6.4. Επίδοση λογισμικών στην αποθρομβοποίηση του 4^{ου} δείγματος



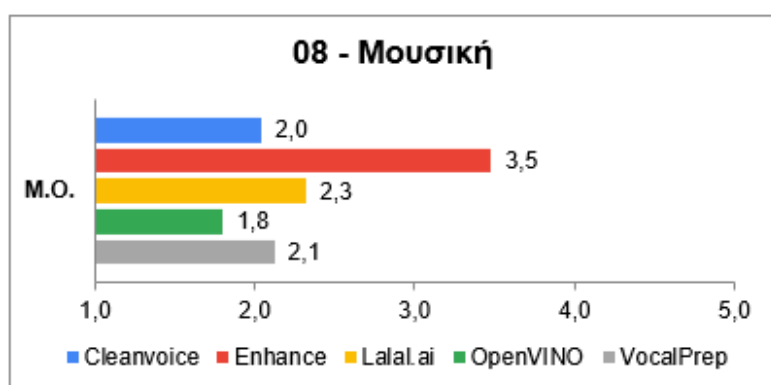
Σχήμα 6.5. Επίδοση λογισμικών στην αποθρομβοποίηση του 5^{ου} δείγματος



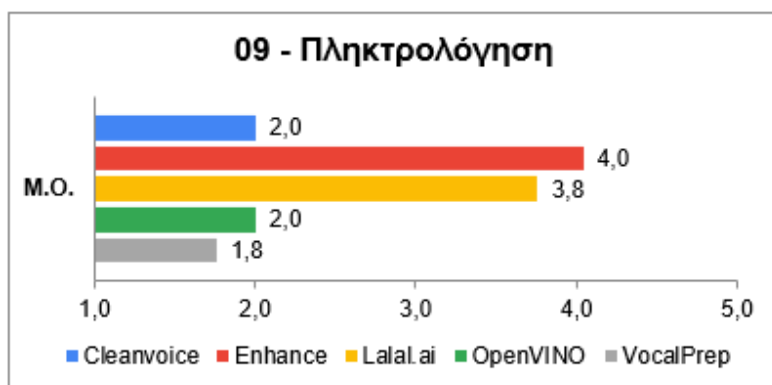
Σχήμα 6.5. Επίδοση λογισμικών στην αποθορυβοποίηση του 6^{ου} δείγματος



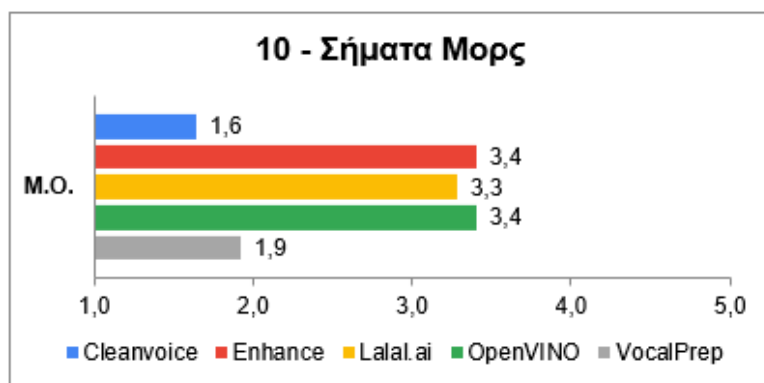
Σχήμα 6.7. Επίδοση λογισμικών στην αποθορυβοποίηση του 7^{ου} δείγματος



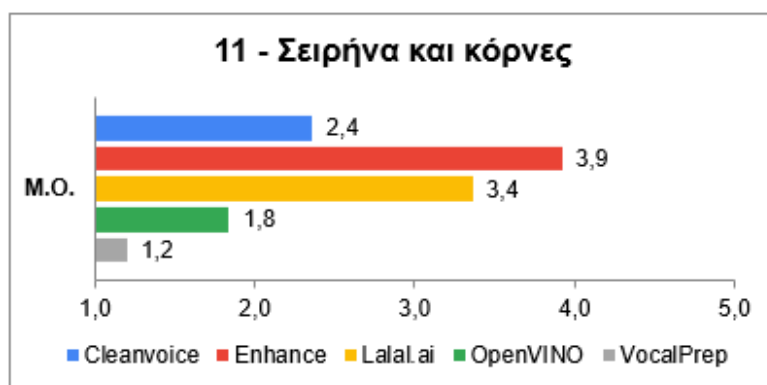
Σχήμα 6.8. Επίδοση λογισμικών στην αποθορυβοποίηση του 8^{ου} δείγματος



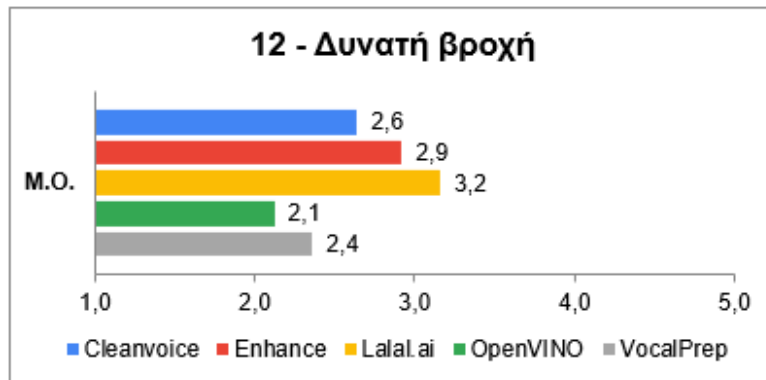
Σχήμα 6.9. Επίδοση λογισμικών στην αποθρομβοποίηση του 9^{ου} δείγματος



Σχήμα 6.10. Επίδοση λογισμικών στην αποθρομβοποίηση του 10^{ου} δείγματος



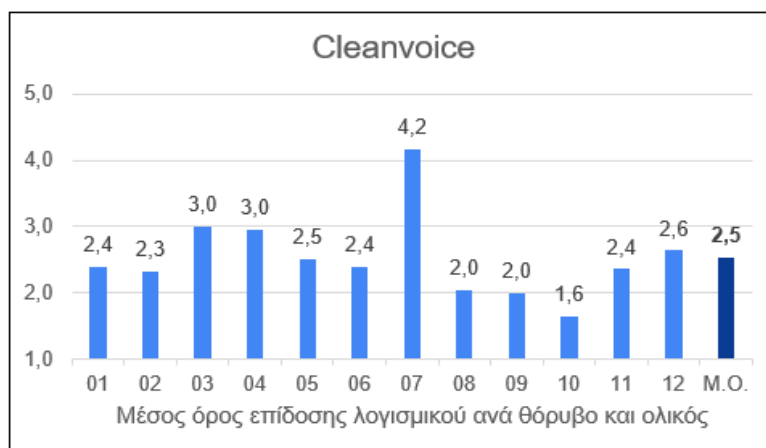
Σχήμα 6.11. Επίδοση λογισμικών στην αποθρομβοποίηση του 11^{ου} δείγματος



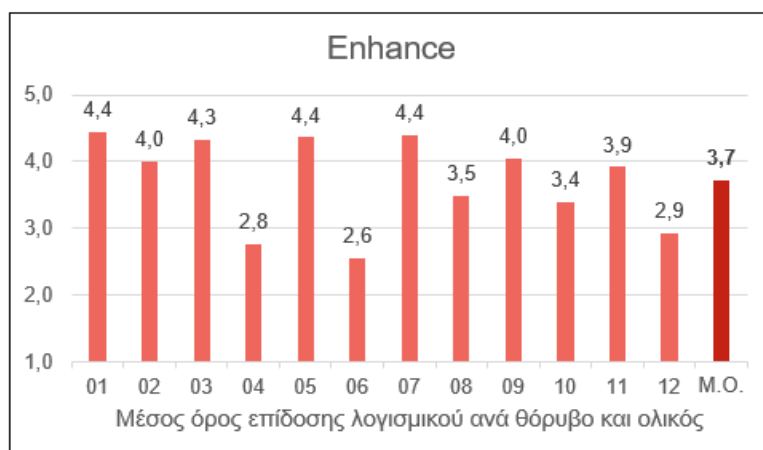
Σχήμα 6.12. Επίδοση λογισμικών στην αποθορυβοποίηση του 12^{ου} δείγματος

Για την καλύτερη εποπτεία, τα προηγούμενα αποτελέσματα οργανώθηκαν για απεικόνιση και ανά λογισμικό όπως παρατίθενται στα Σχήματα 6.13 έως και 6.17 που ακολουθούν. Εδώ εμφανίζονται οι μέσοι όροι της βαθμολογίας που συγκέντρωσαν τα λογισμικά για κάθε έναν από τους δώδεκα θορύβους, και στην δεξιότερη στήλη αποτυπώνεται ο συνολικός μέσος όρος επίδοσης. Οι αριθμοί 01 έως και 12 αντιστοιχούν στους δώδεκα διαφορετικούς θορύβους, με την ίδια σειρά που εκείνοι δημιουργήθηκαν.

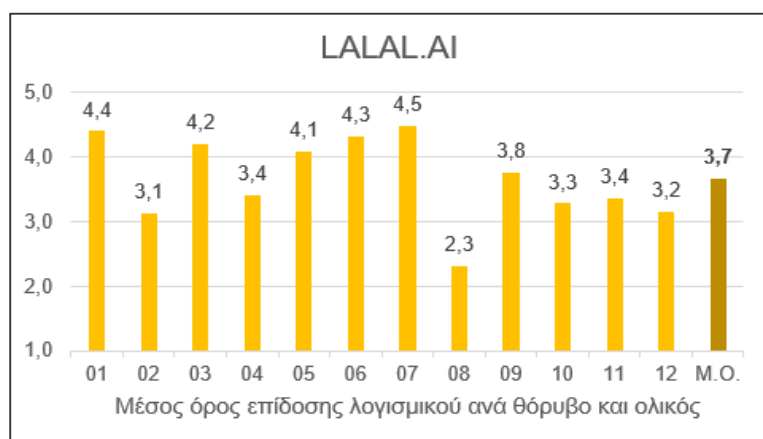
Επαναλαμβάνεται στο σημείο αυτό, ότι οι χαμηλές επιδόσεις σε ορισμένες περιπτώσεις, οφείλονται εν μέρει στο ότι τα ενθόρυβα δείγματα δημιουργήθηκαν εσκεμμένα με τρόπο ώστε το περιεχόμενό τους να είναι «απαιτητικό», με την επιθυμία δηλαδή να δοκιμαστούν τα όρια των λογισμικών.



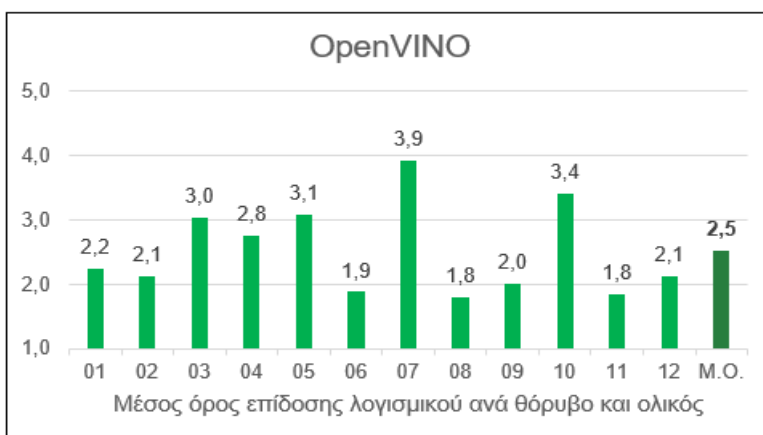
Σχήμα 6.13. Επίδοση λογισμικού Cleanvoice στην αποθορυβοποίηση των δειγμάτων



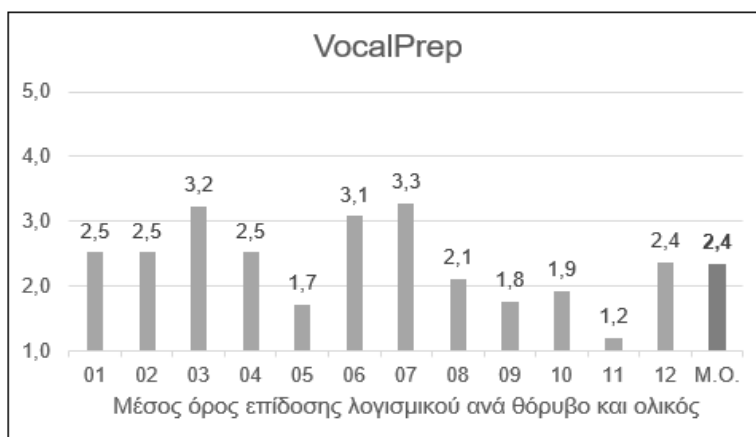
Σχήμα 6.14. Επίδοση λογισμικού Enhance στην αποθορυβοποίηση των δειγμάτων



Σχήμα 6.15. Επίδοση λογισμικού LALAL.AI στην αποθορυβοποίηση των δειγμάτων

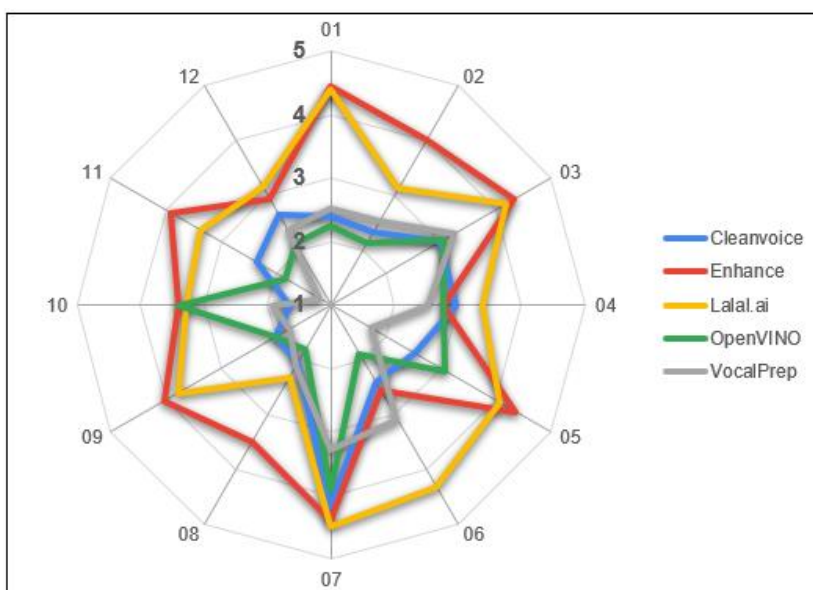


Σχήμα 6.15. Επίδοση λογισμικού OpenVINO στην αποθορυβοποίηση των δειγμάτων

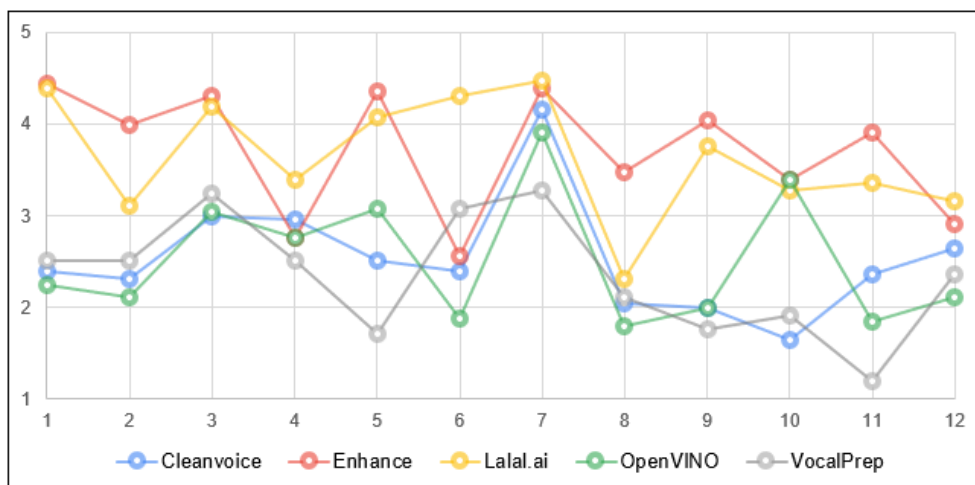


Σχήμα 6.17. Επίδοση λογισμικού VocalPrep στην αποθρομβοποίηση των δειγμάτων

Ακολουθούν στα Σχήματα 6.18 και 6.19, τα συγκεντρωτικά συγκριτικά γραφήματα απεικόνισης των αποτελεσμάτων, με την μορφή τύπου ραντάρ (Radar) και διασποράς (Scatter) αντίστοιχα. Οι αριθμοί 1 έως και 12 αντιστοιχούν με την ίδια σειρά στους δώδεκα διαφορετικούς θορύβους. Ισχύει και εδώ η ίδια χρωματική κωδικοποίηση.



Σχήμα 6.18. Συγκεντρωτικό γράφημα επίδοσης των λογισμικών τύπου ραντάρ



Σχήμα 6.19. Συγκεντρωτικό γράφημα επίδοσης των λογισμικών τύπου διασποράς

Από την μελέτη των παραπάνω αποτελεσμάτων, παρατηρεί κανείς ότι τα λογισμικά που εξετάστηκαν χωρίστηκαν σε δύο ζώνες επιδόσεων. Στην ζώνη με τις καλύτερες επιδόσεις, κατατάχθηκαν τα λογισμικά **Enhance** και **LALAL.AI**, με συνολικό μέσο όρο επίδοσης **3,7** (ο οποίος ταυτίζεται για τα δύο λογισμικά), ενώ στην επόμενη ζώνη με παρόμοιες κατώτερες επιδόσεις κατατάχθηκαν **Cleanvoice**, **OpenVINO** και **VocalPrep**, με συνολικούς μέσους όρους επίδοσης **2,5**, **2,5** και **2,4** αντίστοιχα.

Πιθανό παράγοντα για την διαφοροποίηση αυτή στις επιδόσεις, μπορεί να αποτελεί το διαφορετικό εύρος δειγμάτων και ο βαθμός εκπαίδευσης των αλγορίθμων του κάθε εργαλείου αποθρομβοποίησης, όπου ιδιαίτερα στην περίπτωση που τα γλωσσικά μοντέλα είναι εγκατεστημένα τοπικά, (**OpenVINO** και **VocalPrep**) τα λογισμικά φαίνονται να υστερούν, λόγω έλλειψης ενημερώσεων για την εκπαίδευση των μοντέλων.

Σημειώνεται πάντως ότι γενικά στην βαθμολογία των επιδόσεων με την κλίμακα αξιολόγησης MOS (Mean Opinion Score), η οποία και χρησιμοποιήθηκε εδώ, τα αποτελέσματα δύσκολα αγγίζουν το πάνω όριο, καθώς όπως έχει παρατηρηθεί υπάρχει διαδεδομένη η ανθρώπινη τάση να μην αποδίδονται εύκολα βαθμοί που αντιστοιχούν στην τέλεια επίδοση κατά την αξιολόγηση (Dialpad, n.d.).

6.1.3 Παρατηρήσεις - Σχόλια ακροατών

Κατά την διάρκεια των ακροάσεων, παρατηρήθηκε ότι σχεδόν σε όλες τις περιπτώσεις, οι ακροατές έκριναν αναγκαία την επανάληψη της ακρόασης των ήχων, πριν δώσουν την τελική βαθμολόγηση.

Παρατηρήθηκε επίσης ότι η βαθμολόγηση ήταν συγκριτική, με βάση τις επιδόσεις των υπολοίπων δειγμάτων για κάθε διαφορετικό θόρυβο, δηλαδή η επανάληψη δεν περιοριζόταν στην εκ νέου ακρόαση του ίδιου αρχείου, αλλά υπήρχαν παλινδρομήσεις σε προηγούμενα ή επόμενα δείγματα του ίδιου θορύβου. (Υπενθυμίζεται εδώ ότι τα αρχεία ήταν ομαδοποιημένα ανά θόρυβο).

Σε πολλές περιπτώσεις, αυτό συνδυάστηκε με αναθεώρηση της βαθμολογίας, ενώ συχνά υπήρξε αμφιταλάντευση ανάμεσα σε «διπλάνες» βαθμολογίες επίδοσης, όπως για παράδειγμα ανάμεσα στο «καλή» και «πολύ καλή».

Αμέσως μετά από την διενέργεια των ακροάσεων, ακολούθησε σύντομη συζήτηση με τους ακροατές, όπου τους ζητήθηκε να σχολιάσουν, εφόσον το επιθυμούσαν, όσα εκείνοι θεωρούσαν σημαντικά, αποτιμώντας τα αποτελέσματα της όλης διαδικασίας. Στα σχόλιά τους περιλαμβάνονται τα εξής:

«Παρατήρησα ότι σε κάποιες περιπτώσεις αλλοιώνονταν οι λέξεις, αντικαθιστώντας φωνήεντα με άλλα, ή δεν ήταν καθαρές οι καταλήξεις των λέξεων. Οι ηχογραφήσεις ήταν σε γενικές γραμμές ικανοποιητικής ποιότητας και το περιεχόμενο τους κατανοητό. Ωστόσο, σε ορισμένα σημεία παρατήρησα παρεμβολές, κάποια βουητά και ορισμένες αυξομειώσεις του ήχου».

«Σε κάποιες περιπτώσεις παρατήρησα ότι έγινε αντικατάσταση γραμμάτων με άλλα».

«Παρατήρησα ότι σε αρκετά δείγματα υπήρχε αλλαγή της ποιότητας όσο περνούσε ο χρόνος».

«Το ότι υπήρχε η αρχική μορφή με τον θόρυβο βοήθησε πολύ στο να γίνει καλύτερη αξιολόγηση των ήχων».

«Υπήρχαν θόρυβοι που τα λογισμικά είχαν καλές επιδόσεις στην απομόνωση και αφαίρεση των ήχων του παρασκήνιου. Ενώ σε άλλες περιπτώσεις τα αποτελέσματα δεν

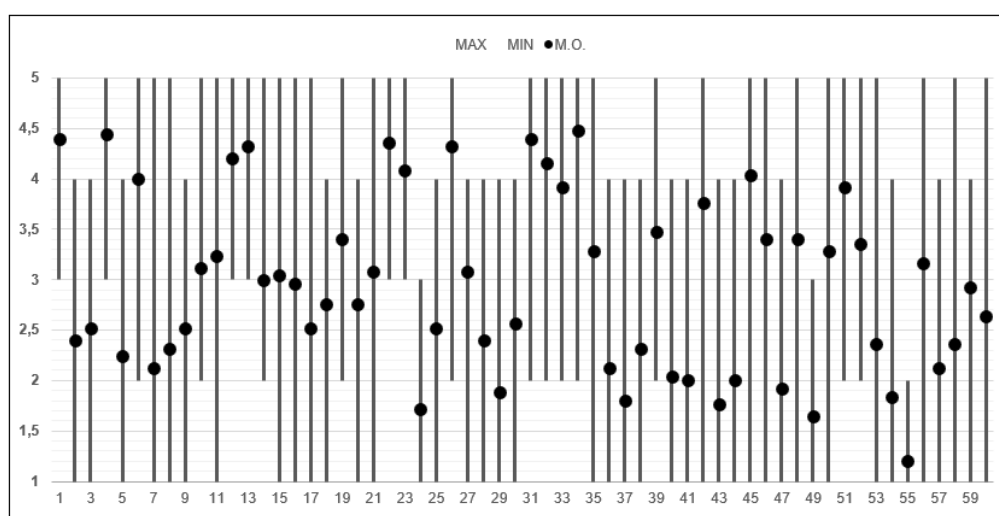
ήταν ικανοποιητικά. Όσο περισσότερο απομονωνόταν ο θόρυβος, τόσο υποβαθμιζόταν η ποιότητα του ήχου».

6.1.4 Αποτύπωση υποκειμενικότητας στις απαντήσεις

Από την επεξεργασία των δεδομένων των βαθμολογιών από τους ακροατές, διαπιστώθηκε ότι υπήρξε διασπορά στις απαντήσεις, η οποία σε ορισμένες περιπτώσεις ήταν σημαντική, καλύπτοντας σε κάποιες από αυτές ακόμη και τις πέντε δυνατές διαφορετικές απαντήσεις για το ίδιο δείγμα.

Η διασπορά στις βαθμολογίες επίδοσης -η οποία αποτελεί και παράμετρο / δείκτη υποκειμενικότητας στις αξιολογήσεις που πραγματοποιήθηκαν- είχε σαν σημαντικό αίτιό της την διαφορετική ιεράρχηση, όσον αφορά στην σημαντικότητα της απομάκρυνσης του θορύβου σε σχέση με την διατήρηση της ποιότητας του «καθαρού» ήχου. Έτσι, κάποιοι ακροατές «πριμοδότησαν» την αφαίρεση του θορύβου, θέτοντας σε δεύτερη μοίρα την παράμετρο της ποιότητας του αποτελέσματος, ενώ κάποιοι άλλοι έπραξαν το αντίθετο.

Αυτό παρατηρήθηκε και κατά την διάρκεια των ακροάσεων που πραγματοποιήθηκαν παρουσία του γράφοντα. Η διασπορά των απαντήσεων κατά την συμπλήρωση του ερωτηματολογίου, αποτυπώνεται γραφικά στο Σχήμα 6.20 που ακολουθεί.



Σχήμα 6.20. Διασπορά απαντήσεων κατά την συμπλήρωση του ερωτηματολογίου

Σημειώνεται ότι οι αριθμοί στον οριζόντιο άξονα αντιστοιχούν στα **εξήντα (60)** συνολικά αποθρομβοποιημένα δείγματα που βαθμολογήθηκαν, διατεταγμένα με την ακριβή σειρά που αυτά παρατέθηκαν στην φάση των ακροάσεων. Υπενθυμίζεται ότι η σειρά ακρόασης των δειγμάτων καθορίστηκε εσκεμμένα με τυχαίο τρόπο, για αντικειμενικότερα αποτελέσματα, όπως αποτυπώθηκε στον Πίνακα 5.6 του προηγούμενου κεφαλαίου.

6.2 Πιθανές εφαρμογές εργαλείων αποθρομβοποίησης

Από την διεξαγωγή της έρευνας που προηγήθηκε και την υλοποίηση της πειραματικής διαδικασίας, που πραγματοποιήθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας, διαπιστώθηκε ότι τόσο τα εργαλεία λογισμικού που συμπεριλήφθηκαν στην μελέτη όσο και τα υπόλοιπα λογισμικά που χρησιμοποιούν ΤΝ για την αφαίρεση του θορύβου από ένα ηχητικό σήμα, μπορούν να αξιοποιηθούν με την χρήση τους σε ένα ευρύ πεδίο εφαρμογών.

Το πρώτο λογισμικό που εξετάστηκε (Cleanvoice), μπορεί να χρησιμοποιηθεί κυρίως σε εφαρμογές οι οποίες αφορούν σε ηχητικά σήματα τα οποία περιλαμβάνουν ανθρώπινη ομιλία. Έτσι μπορεί να βρει εφαρμογή σε δημοσιογραφικές εκπομπές, σε απομαγνητοφωνήσεις συνεντεύξεων, σε μετάδοση ροής περιεχομένου (podcasting) και για την βελτίωση του ήχου σε παραγωγή βίντεο που αναρτώνται στο διαδίκτυο. Ακόμα μπορεί να αξιοποιηθεί από δημιουργούς online μαθημάτων, και επαγγελματίες οι οποίοι κάνουν χρήση τηλεδιασκέψεων, για την αυτόματη αφαίρεση του θορύβου που προέρχεται από τον περιβάλλοντα χώρο κατά την διάρκεια της συνομιλίας.

Η αυτοματοποίηση της διαδικασίας αποθρομβοποίησης με την χρήση ΤΝ, μειώνει τον απαιτούμενο χρόνο επεξεργασίας, συμβάλλοντας στην παραγωγή ηχητικού περιεχομένου, χωρίς μάλιστα την ανάγκη ύπαρξης εξειδικευμένων γνώσεων εκ μέρους του χρήστη, κάτι που καθιστά το Cleanvoice ιδιαίτερα χρήσιμο σε εφαρμογές όπου η καθαρότητα της ομιλίας είναι σημαντικότερη από την επεξεργασία του ήχου που στοχεύει στην παραγωγή καλλιτεχνικών αποτελεσμάτων. Επιπρόσθετα, η δυνατότητα του Cleanvoice να μετατρέπει μέσω μετεγγραφής (transcription) το περιεχόμενο της φωνής σε κείμενο –και μάλιστα σε διαφορετικές γλώσσες από την γλώσσα του ομιλητή- το καθιστά ιδιαίτερα αποδοτικό σε εργασίες υποτιτλισμού.

Το λογισμικό Enhance (Adobe Podcast), παρουσιάζει αξιόλογα αποτελέσματα στην βελτίωση ηχογραφήσεων φωνής, οι οποίες εμφανίζουν υποβαθμισμένο ηχητικό περιεχόμενο, που οφείλεται είτε στην χρήση εξοπλισμού χαμηλής ποιότητας κατά την εγγραφή ή μετάδοση του σήματος -όπως είναι τα ενσωματωμένα μικρόφωνα των κινητών συσκευών- είτε στο ενθόρυβο περιβάλλον στο οποίο βρίσκεται συνήθως ο ομιλητής.

Το Enhance μπορεί να αξιοποιηθεί από δημιουργούς περιεχομένου (content creators), ή άλλους επαγγελματίες οι οποίοι εργάζονται με τηλεεργασία, για την μετατροπή «οικιακών» ηχογραφήσεων σε ήχο που προσεγγίζει την ποιότητα που μπορεί να επιτευχθεί σε ένα στούντιο ηχογράφησης. Έτσι μπορεί για παράδειγμα να χρησιμοποιηθεί σε διαδικτυακά σεμινάρια, συνεντεύξεις, βιντεοσκοπημένες παρουσιάσεις και online μαθήματα, όπου απαιτείται καθαρή και ευχάριστη ομιλία. Επειδή το Enhance λειτουργεί αποκλειστικά με one-click λογική, αποτελεί ιδιαίτερα δημοφιλή επιλογή σε χρήστες οι οποίοι αναζητούν γρήγορες και εύκολες λύσεις, χωρίς να χρειάζεται να διαθέτουν προηγούμενη εμπειρία σε επαγγελματικά εργαλεία ήχου.

Το LALAL.AI, το οποίο αποτέλεσε το τρίτο εργαλείο λογισμικού TN που εξετάσθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας, χρησιμοποιείται κυρίως για διαχωρισμό ηχητικών πηγών (source separation) και δημιουργία ξεχωριστών καναλιών (stems) για κάθε μουσικό όργανο. Στην καθημερινή πρακτική εφαρμόζεται από μουσικούς παραγωγούς, DJs, και επαγγελματίες μουσικής για την απομόνωση φωνής, τυμπάνων, μπάσων ή άλλων μουσικών οργάνων, μέσα από μουσικές συνθέσεις.

Παρά το γεγονός ότι το LALAL.AI δεν αποτελεί αμιγές εργαλείο αποθορυβοποίησης, η δυνατότητα απομόνωσης της φωνής ή άλλων μουσικών πηγών, επιτρέπει την αφαίρεση του θορύβου, καθώς τον οδηγεί σε διαφορετικά ηχητικά κανάλια, επιτυγχάνοντας τον διαχωρισμό του από τα «ωφέλιμα» ηχητικά σήματα που θέλουμε να διατηρήσουμε. Έτσι, η δυνατότητα απομόνωσης της φωνής βρίσκει εφαρμογή σε εργασίες αποκατάστασης ήχου και καθαρισμού διαθέσιμου αρχειακού υλικού, όπου απαιτείται διαχωρισμός και αφαίρεση συγκεκριμένων στοιχείων από τον ήχο, όπως ο θόρυβος.

Το πρόσθετο OpenVINO του Audacity, χρησιμοποιείται κυρίως για τοπική επεξεργασία ήχου με την χρήση TN χωρίς να απαιτείται άλλη διαδικτυακή υποδομή για την λειτουργία του. Στην καθημερινή πρακτική μπορεί να αξιοποιηθεί από ερευνητές, φοιτητές, ανεξάρτητους δημιουργούς και ερασιτέχνες του ήχου, για αποθορυβοποίηση

ηχογραφήσεων, βελτίωση ομιλίας και πειραματισμό με μοντέλα τεχνητής νοημοσύνης, μέσα σε ένα δωρεάν περιβάλλον επεξεργασίας ανοιχτού κώδικα.

Η offline λειτουργία του OpenVINO, το καθιστά χρήσιμο σε περιπτώσεις όπου κατά την επεξεργασία του ήχου δεν υπάρχει σύνδεση στο διαδίκτυο, ή σε περιπτώσεις στις οποίες απαιτείται μεγαλύτερος έλεγχος στην διατήρηση και προφύλαξη των δεδομένων για την προστασία της ιδιωτικότητας. Επιπλέον, η ενσωμάτωσή του στο Audacity επιτρέπει την επεξεργασία με χρήση TN σε συνδυασμό με τις παραδοσιακές τεχνικές επεξεργασίας, κάτι ιδιαίτερα χρήσιμο στην διενέργεια εκπαιδευτικών ή ερευνητικών εργασιών.

Το VocalPrep μπορεί να χρησιμοποιηθεί κυρίως στην προετοιμασία φωνητικών ηχογραφήσεων από μουσικούς παραγωγούς και μηχανικούς ήχου, για αφαίρεση θορύβου ο οποίος σχετίζεται κυρίως από πρόκληση ανεπιθύμητων ήχων κατά την εκφορά λόγου. Όπως και στην περίπτωση του OpenVINO, η τοπική (offline) λειτουργία του VocalPrep επιτρέπει γρήγορη επεξεργασία χωρίς να χρειάζεται τα ηχητικά δεδομένα να μεταφορτωθούν στο διαδίκτυο, γεγονός σημαντικό σε επαγγελματικές ροές εργασίας μουσικών παραγωγών.

Παράλληλα, η απλότητα χρήσης του το καθιστά ιδιαίτερα κατάλληλο για γρήγορες παραγωγές περιεχομένου, και ηχογραφήσεις όπου απαιτείται ισορροπία μεταξύ ταχύτητας και ποιότητας στο τελικό αποτέλεσμα.

Οι παραπάνω πιθανές χρήσεις των εργαλείων αποθορυβοποίησης με χρήση TN, που εξετάστηκαν στην παρούσα διπλωματική εργασία, μπορούν σε γενικές γραμμές να πραγματοποιηθούν και από τα υπόλοιπα διαθέσιμα σύγχρονα εργαλεία Τεχνητής Νοημοσύνης. Έτσι τα εργαλεία TN μπορούν να χρησιμοποιηθούν σε ένα πλήθος εφαρμογών, όπου η αποθορυβοποίηση είναι επιβεβλημένη, όπως στην επεξεργασία βίντεο, την δημιουργία περιεχομένου, την υποστήριξη λογισμικών τηλεδιασκέψεων, την αντιστάθμιση θορύβων που οφείλονται στον χρησιμοποιούμενο εξοπλισμό και πολλές άλλες χρήσεις. Τέλος, το γεγονός ότι τα περισσότερα εργαλεία TN απαιτούν ελάχιστη παραμετροποίηση (one-click tools), τους δίνει το πλεονέκτημα να προσφέρουν γρήγορα αποτελέσματα με ελάχιστη προσπάθεια, καθιστώντας τα ιδανικά για χρήσεις όπου το κύριο ζητούμενο δεν είναι η τελική υψηλή ποιότητα του προϊόντος ηχητικού περιεχομένου.

7. Συμπεράσματα - Συνεισφορά - Μελλοντικές Επεκτάσεις

Σε αυτό το κεφάλαιο, περιλαμβάνονται τα συμπεράσματα που προέκυψαν, από την θεωρητική προσέγγιση του εξεταζόμενου θέματος και την εκτέλεση της αναλυτικής - πειραματικής διαδικασίας, στα πλαίσια της υλοποίησης της παρούσας διπλωματικής εργασίας. Παράλληλα καταγράφονται τα σημεία στα οποία η εργασία έχει συνεισφέρει και προτείνονται μελλοντικές επεκτάσεις για την διενέργεια περεταίρω έρευνας από την ακαδημαϊκή κοινότητα.

7.1 Συμπεράσματα

Τα παραδοσιακά λογισμικά επεξεργασίας ψηφιακού ήχου, αξιοποιούν την Τεχνητή Νοημοσύνη, ενσωματώνοντας πλέον υβριδικές υλοποιήσεις στα διαθέσιμα εργαλεία τους.

Υπάρχουν διαθέσιμες απομακρυσμένες και τοπικές επιλογές, αναφορικά με το πού βρίσκεται η βάση δεδομένων (δειγμάτων ήχου) της TN, ανάλογα με το αν γίνεται online χρήση εργαλείων μέσα από διαδικτυακές εφαρμογές, ή αν αυτά είναι εγκατεστημένα τοπικά με δυνατότητες offline χρήσης.

Η TN πλεονεκτεί σε απευθείας (realtime) μεταδόσεις, και γενικότερα στις περιπτώσεις όπου οι θόρυβοι μεταβάλλονται δυναμικά, έναντι των παραδοσιακών μεθόδων, οι οποίες διαθέτουν περιορισμένες ή μηδενικές δυνατότητες προσαρμογής στη μεταβολή του θορύβου, καθώς οι ρυθμίσεις τους έχουν προκαθορισμένες τιμές.

Υπάρχουν υλοποιήσεις TN, οι οποίες μπορούν να εξαιρούν κατ' επιλογή από την αποθορυβοποίηση, συγκεκριμένους «επιθυμητούς» θορύβους σε πραγματικό χρόνο, όπως είναι τα ηχητικά σήματα κρίσιμων ειδοποιήσεων παρασκηνίου, κάτι που τα συμβατικά εργαλεία αδυνατούν να κάνουν.

Οι σύγχρονες τεχνικές TN δεν προκαλούν πλέον σοβαρή καθυστέρηση στην μετάδοση του ήχου (latency), παρέχοντας βελτιωμένη επεξεργασία σε πραγματικό χρόνο, γεγονός που καθιστά την χρήση τους αποδοτικότερη στην περίπτωση των απευθείας μεταδόσεων. Σε αυτό συνέβαλε η απαίτηση για την κάλυψη της διαρκώς αυξανόμενης

ανάγκης για τηλεεργασία, η οποία πλέον περιλαμβάνει δραστηριότητα σε μη ιδανικές συνθήκες, σε ενθόρυβα περιβάλλοντα εκτός γραφείου.

Η αυτοματοποίηση της διαδικασίας μέσω TN, μειώνει δραστικά τον χρόνο επεξεργασίας, ιδιαίτερα όταν δεν υπάρχει απαίτηση για άριστη ποιότητα του τελικού αποτελέσματος. Έτσι η TN βρίσκει περισσότερο εφαρμογή σε περιπτώσεις δημιουργίας περιεχομένου όπου δίνεται προτεραιότητα στην παραγωγή γρήγορων αποτελεσμάτων.

Η ευρεία διάδοση της TN, συνέβαλε στην δημιουργία εργαλείων λογισμικού, τα οποία είναι προσιτά σε όλους, έχουν αισθητά μειωμένο κόστος σε σχέση με τα παραδοσιακά λογισμικά, ενώ δεν απαιτούν την ύπαρξη ιδιαίτερης εξειδίκευσης και εμπειρίας για την χρήση τους.

Αρκετές εφαρμογές TN παράγουν αυτόματα μονοφωνικό αποτέλεσμα, σαν προϊόν της διαδικασίας αποθορυβοποίησης αποθηκευμένου ήχου, παρόλο που μπορεί το αρχικό ενθόρυβο σήμα εισόδου να είναι στερεοφωνικό. Αυτό αποτελεί συνειδητή επιλογή στην υλοποίηση των εφαρμογών, με γνώμονα τον περιορισμό του μεγέθους των δεδομένων του επεξεργασμένου αρχείου, ώστε αυτό να μπορεί να μεταδοθεί ευκολότερα στην συνέχεια, μέσα από τα διαθέσιμα ψηφιακά κανάλια επικοινωνίας.

Οι αυτοματοποιημένες (One-Click) διαδικασίες αποθορυβοποίησης, και γενικότερα οι επιλογές λογισμικών που εμπλέκουν μικρότερο βαθμό παραμετροποίησης, είναι η γενικότερη τάση, για επεξεργασία αποθηκευμένου ηχητικού περιεχομένου. Για το λόγο αυτό παρατηρείται αύξηση στις διαθέσιμες επιλογές εργαλείων που υιοθετούν αυτόν τον τρόπο λειτουργίας.

Η αύξηση των διαθέσιμων επιλογών κατά την αποθορυβοποίηση, καθιστά το βέλτιστο αποτέλεσμα υποκειμενικό, καθώς ο χρήστης αναπόφευκτα πρέπει να επιλέξει το βαθμό ισορροπίας ανάμεσα στον περιορισμό του θορύβου και την συνεπακόλουθη εισαγωγή παραμόρφωσης, που προκαλείται στο σήμα της ομιλίας κατά την διαδικασία αποθορυβοποίησης. Σε κάθε περίπτωση, διαπιστώθηκε ότι η υπερβολή στην επεξεργασία εισάγει παραμορφώσεις και προκαλεί την εμφάνιση «αφύσικων» αποτελεσμάτων.

Υποκειμενικότητα υπήρξε και στην αξιολόγηση των αποτελεσμάτων της αποθορυβοποίησης, όπως παρατηρήθηκε και καταγράφηκε στην διάρκεια των ακροάσεων που πραγματοποιήθηκαν. Η υποκειμενικότητα αυτή, ενδεχομένως ενισχύθηκε από το

γεγονός ότι κατά την διαδικασία της βαθμολόγησης, χρησιμοποιήθηκε ένας μόνο δείκτης επίδοσης, ο οποίος ταυτόχρονα αξιολογούσε την καταστολή θορύβου και την ποιότητα της παραμένουσας ομιλίας, κάτι το οποίο συνέβαλε σε κάποιο βαθμό στην διαφοροποίηση των αποτελεσμάτων ανάμεσα στους ακροατές.

Σχετικά με την δυνατότητα επεξεργασίας της ελληνικής γλώσσας, διαπιστώθηκε ότι οι εφαρμογές TN επιτυγχάνουν αξιοσημείωτες επιδόσεις, καθώς τα μοντέλα τους έχουν πλέον εκπαιδευτεί με επαρκή αριθμό δειγμάτων ομιλίας στα ελληνικά. Δεν συμβαίνει όμως το ίδιο για την διαθεσιμότητα της ελληνικής γλώσσας στο περιβάλλον εργασίας του χρήστη, όπου στις περισσότερες εφαρμογές -και ιδιαίτερα σε αυτές που είναι διαθέσιμες διαδικτυακά- αυτό προσφέρεται κυρίως στα αγγλικά.

Οι επιδόσεις στην αποθορυβοποίηση ελληνικής ομιλίας ήταν πολύ καλές, ακόμα και στην μεμονωμένη εξεταζόμενη περίπτωση όπου στο παρασκήνιο υπήρχαν συζητήσεις στα αγγλικά.

Σε ορισμένες περιπτώσεις, υπήρχε αλλοίωση σε εκφερόμενες λέξεις, με αντικατάσταση ή / και αποκοπή φωνημάτων, τόσο στο εσωτερικό όσο και στην κατάληξη των λέξεων. Αυτό αποδεικνύει ότι γίνεται ανασύνθεση ομιλίας με χρήση δημιουργικής TN, ιδιαίτερα σε χρονικές στιγμές όπου η στάθμη θορύβου λαμβάνει υψηλές τιμές.

Παρατηρήθηκαν διαφοροποιήσεις στις επιδόσεις των εργαλείων TN, ανάλογα με τον ίδιο κατά περίπτωση θόρυβο. Εμφανίστηκαν αρκετές μεμονωμένες αστοχίες σε θορύβους, για τους οποίους άλλα λογισμικά δεν αντιμετώπισαν πρόβλημα. Αυτό πιθανόν σχετίζεται με το πώς και σε ποιο βαθμό είναι εκπαιδευμένο το μοντέλο του κάθε λογισμικού, και ειδικότερα με το κατά πόσο ο κάθε ιδιαίτερος θόρυβος περιλαμβάνεται επαρκώς στα δεδομένα με τα οποία έχει εκπαιδευτεί ο αντίστοιχος αλγόριθμος.

Σε ορισμένες περιπτώσεις, εμφανίστηκε αδυναμία επαρκούς καταστολής του θορύβου, όταν αυτός επαναλαμβανόταν μέσα στο ίδιο ηχητικό δείγμα, ανάλογα με το χρονικό σημείο στο οποίο αυτός ήταν τοποθετημένος. Η αδυναμία αυτή εντοπίστηκε ιδιαίτερα στην έναρξη των ήχων, γεγονός που υποδηλώνει ότι οι αλγόριθμοι αποθορυβοποίησης παρουσιάζουν διακυμάνσεις στον βαθμό επίδοσής τους, εξαρτώμενες από την χρονική παράμετρο, στην διάρκεια του ήχου. Σε μία μάλιστα περίπτωση, ο ίδιος επαναλαμβανόμενος θόρυβος, ενώ αφαιρέθηκε αποτελεσματικά στην μεγαλύτερη διάρκεια

του δείγματος, παρέμεινε ουσιαστικά αναλλοίωτος, στο τελευταίο τμήμα του αποθορυβημένου ήχου.

Διαπιστώθηκε πειραματικά, ότι οι όποιες αδυναμίες στην αποθορυβοποίηση με χρήση TN, είχαν σχεδόν αποκλειστική προέλευση το διαφορετικό περιεχόμενο θορύβου, ενώ ήταν σχεδόν ανεξάρτητες από τις διαφορετικές εμπλεκόμενες ομιλίες.

Σε περιπτώσεις όπου στο ηχητικό στο περιεχόμενο συνυπάρχουν περισσότερες από μία ομιλίες, όπως συμβαίνει σε συνθήκες όπου ο ομιλητής περιβάλλεται από ομιλούντες τρίτους, η ύπαρξη υψηλότερης στάθμης στον ήχο του ομιλητή, σε σχέση την ηχητική στάθμη των φωνών του παρασκηνίου, συμβάλει στην αύξηση της επίδοσης της αποθορυβοποίησης.

Στις περισσότερες περιπτώσεις, η TN πετυχαίνει ικανοποιητικά αποτελέσματα για πολύ μικρές ή ακόμα και αρνητικές τιμές SNR των ενθόρυβων σημάτων. Επίσης παρατηρήθηκε ότι τα συγκριτικά αποτελέσματα που προέκυψαν με την χρήση εργαλείων TN, αρκετές φορές δεν ήταν ανάλογα με το SNR των σημάτων. Διαπιστώθηκε έτσι, ότι ανάμεσα στα διαφορετικά σήματα υπήρξαν καλύτερες επιδόσεις, σε σήματα με μικρότερο SNR και το αντίστροφο.

Σε σχέση με τις παραδοσιακές μεθόδους τα εργαλεία TN παίρνουν περισσότερες αποφάσεις εκ μέρους του χρήστη για την επεξεργασία του ήχου, κάτι το οποίο συνεπάγεται λιγότερο έλεγχο στο τελικό αποτέλεσμα, καθώς αυτός δεν διαθέτει επιλογές λεπτομερέστερης επεξεργασίας και παρέμβασης, οι οποίες είναι απαραίτητες για επαγγελματικά αποτελέσματα σε ορισμένες περιπτώσεις.

Καταλήγοντας, οφείλεται να καταγραφεί στα συμπεράσματα, το γεγονός ότι οι εφαρμογές με χρήση TN εξελίσσονται μέρα με την μέρα, τόσο χάρη στην συνεχή «εκπαίδευση» τους, με την προσθήκη νέων μοντέλων, αλγορίθμων και δειγμάτων, όσο και στις διαρκώς νέες παρεχόμενες δυνατότητες και επιλογές επεξεργασίας, οι οποίες είναι διαθέσιμες στον χρήστη.

Αυτό διαπιστώθηκε πειραματικά, μέσα από τις αλλαγές που εντοπίστηκαν από τον γράφοντα, στην πολύμηνη διάρκεια της υλοποίησης της παρούσας διπλωματικής εργασίας, γεγονός που καθιστά την οποιαδήποτε αποτύπωση δυνητικά ανεπίκαιρη μετά από ένα εύλογο χρονικό διάστημα, καθώς οι εξελίξεις στον τομέα της TN είναι έντονες και διαρκείς.

7.2 Συνεισφορά της Εργασίας

Η ουσιαστικότερη συνεισφορά της παρούσας διπλωματικής εργασίας -κατά τον γράφοντα- συνίσταται στο γεγονός ότι αυτή συμπεριλαμβάνει συγκριτική αξιολόγηση σύγχρονων διαδεδομένων εργαλείων αποθρομβοποίησης, κάτω από πραγματικές πειραματικές συνθήκες, με ανθρώπινη ακουστική αξιολόγηση, η οποία έγινε από εκτενή ομάδα ακροατών.

Η εργασία αξιολογεί διαθέσιμα εργαλεία και όχι θεωρητικούς αλγορίθμους, το οποίο καθιστά τα αποτελέσματα άμεσα αξιοποιήσιμα από όσους καλούνται να επιλέξουν λογισμικό για σχετικές εφαρμογές αποθρομβοποίησης. Έτσι, το περιεχόμενό της αποτελεί πηγή χρήσιμων πληροφοριών για τους επαγγελματίες σχετικών με τον χώρο του ήχου, όπως ηχολήπτες και δημιουργούς περιεχομένου αλλά και ερευνητές ή φοιτητές / σπουδαστές του χώρου, οι οποίοι επιθυμούν να αξιοποιήσουν και ενσωματώσουν τα εργαλεία TN στην καθημερινή τους πρακτική, χωρίς να απαιτείται περεταίρω εξειδικευμένη γνώση.

Η εργασία αναδεικνύει τα πλεονεκτήματα αλλά και τους περιορισμούς των εργαλείων Τεχνητής Νοημοσύνης, αποφεύγοντας την άκριτη υιοθέτηση της αντίληψης ότι όλες οι λύσεις TN αποδίδουν εξίσου καλά σε κάθε περίπτωση. Παράλληλα χρησιμοποιεί εκτενές πειραματικό υλικό, κοινό για όλα τα λογισμικά, διασφαλίζοντας δίκαιες συνθήκες σύγκρισης, παρέχοντας έτσι μεγαλύτερη αξιοπιστία στα προϊόντα αποτελέσματα.

Το περιεχόμενο της εργασίας, περιλαμβάνει μία ενιαία μεθοδολογία αξιολόγησης, η οποία μπορεί να χρησιμοποιηθεί και σε μελλοντικές συγκριτικές μελέτες άλλων εργαλείων ή νεότερων εκδόσεων των ίδιων εφαρμογών, αναγνωρίζοντας από την υποκειμενικότητα στην αξιολόγηση μέσω των ακροάσεων, ότι η αντιληπτή ποιότητα του ήχου αποτελεί βασικό κριτήριο στην πράξη και δεν αποτυπώνεται πάντοτε πλήρως από τεχνικές μετρήσεις.

Με δεδομένη την διαρκή εξέλιξη που συντελείται στον τομέα της TN, η εργασία μπορεί να συνεισφέρει με την αξιοποίησή της σαν αποτύπωση ενός χρονικού στιγμιότυπου, ώστε ο μελλοντικός ερευνητής να μπορεί να έχει μία αναφορά ως προς τις επιδόσεις των λογισμικών TN στον τομέα της αποθρομβοποίησης του ψηφιακού ήχου, κατά την στιγμή της υλοποίησής της. Με τον τρόπο αυτό, η παρούσα διπλωματική εργασία δημιουργεί μία βάση για μελλοντική έρευνα, καθώς η ίδια μεθοδολογία μπορεί να εφαρμοστεί για την

αξιολόγηση νέων εργαλείων και μοντέλων TN, κάνοντας χρήση των ίδιων ή διαφορετικών κατηγοριών ηχητικών σημάτων, επιτρέποντας τη διαχρονική παρακολούθηση της εξέλιξης του συγκεκριμένου επιστημονικού πεδίου.

7.3 Μελλοντικές Επεκτάσεις

Μέσα από την έρευνα που πραγματοποιήθηκε στην διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας, επιβεβαιώθηκε ότι το αντικείμενό της είναι ευρύτατο, με αποτέλεσμα να υπάρχουν σχετικά αντικείμενα τα οποία αναγκαστικά τέθηκαν εκτός εξεταζόμενου πλαισίου και δεν συμπεριλαμβάνονται σε αυτή. Οι μελλοντικές επεκτάσεις, τις οποίες μπορεί να λάβει η εργασία ακολουθούν στην συνέχεια.

Η παρούσα εργασία επικεντρώθηκε στην αξιολόγηση επιλεγμένων εργαλείων αποθρομβοποίησης. Μία φυσική συνέχεια θα ήταν η επέκταση της μελέτης σε μεγαλύτερο αριθμό εμπορικών εφαρμογών αλλά και εφαρμογών ανοικτού κώδικα, ώστε να προκύψει μια ακόμη πληρέστερη εικόνα των δυνατοτήτων τους, αλλά και των διαθέσιμων λύσεων που μπορούν αυτά να παράσχουν.

Η παρούσα μελέτη επικεντρώθηκε κυρίως σε σήματα ομιλίας. Στο μέλλον θα μπορούσε να εξεταστεί η αποτελεσματικότητα των ίδιων εργαλείων σε μουσικές παραγωγές, ηχογραφήσεις συναυλιών, κινηματογραφικό ήχο ή αρχειακές ηχογραφήσεις, όπου οι απαιτήσεις αποθρομβοποίησης παρουσιάζουν σημαντικές διαφοροποιήσεις.

Έτσι η αξιολόγηση θα μπορούσε να επεκταθεί, με την χρήση περισσότερων ηχητικών δειγμάτων, τα οποία να συμπεριλαμβάνουν διαφορετικές γλώσσες ομιλίας και ηλικιακές ομάδες ομιλητών, ήχους από τραγούδι και μουσικά αποσπάσματα προς αποθρομβοποίηση, ώστε εξεταστεί η συμπεριφορά των εργαλείων TN σε ακόμη μεγαλύτερο εύρος συνθηκών.

Στην παρούσα εργασία, ιδιαίτερη βαρύτητα δόθηκε στην ακουστική αξιολόγηση από τους συμμετέχοντες. Στο μέλλον θα μπορούσε να πραγματοποιηθεί ανάλογη έρευνα με παράλληλη χρήση ποσοτικών δεικτών αξιολόγησης, όπως είναι οι PSNR, ESNR, SNR και CCI, ώστε να γίνει και ποσοτική αποτίμηση της ποιότητας των αποθρομβοποιημένων σημάτων, τα οποία προκύπτουν από τα εξεταζόμενα λογισμικά TN.

Ιδιαίτερο ενδιαφέρον παρουσιάζει η διερεύνηση της απόδοσης των εφαρμογών TN σε συνθήκες πραγματικού χρόνου (real-time processing), όπου εκτός από την ποιότητα του παραγόμενου ήχου, σημαντικό ρόλο διαδραματίζουν η καθυστέρηση επεξεργασίας (latency) και οι απαιτήσεις σε υπολογιστικούς πόρους, κάτι το οποίο ήταν έξω από τα πλαίσια της παρούσας διπλωματικής εργασίας και θα μπορούσε να εξετασθεί μελλοντικά.

Επίσης η συστηματική σύγκριση εφαρμογών που εκτελούνται τοπικά (offline) με αντίστοιχες υπηρεσίες που λειτουργούν μέσω υπολογιστικού νέφους (cloud), λαμβάνοντας υπόψη όχι μόνο την ποιότητα του αποτελέσματος αλλά και ζητήματα όπως η ταχύτητα, η ευχρηστία, το κόστος και η προστασία των προσωπικών δεδομένων, είναι κάτι το οποίο θα μπορούσε να διερευνηθεί σαν μια μελλοντική επέκταση της εργασίας.

Μία ακόμη ενδιαφέρουσα κατεύθυνση θα ήταν η δημιουργία και εκπαίδευση μοντέλων Τεχνητής Νοημοσύνης σε εξειδικευμένα σύνολα δεδομένων, ώστε να διερευνηθεί κατά πόσο η χρήση δεδομένων συγκεκριμένων εφαρμογών μπορεί να οδηγήσει σε περαιτέρω βελτίωση της απόδοσης της TN στην αποθορυβοποίηση του ψηφιακού ήχου.

Με βάση το γεγονός ότι αρκετά εργαλεία παρέχονται με συνδρομητικά μοντέλα ή διαφορετικές εκδόσεις χρήσης, θα είχε ιδιαίτερο ενδιαφέρον η μελέτη της σχέσης μεταξύ οικονομικού κόστους και ποιότητας των αποτελεσμάτων, ώστε να αξιολογηθεί μεταξύ άλλων το κατά πόσο το υψηλότερο κόστος συνεπάγεται και ουσιαστικά καλύτερη απόδοση.

Τέλος, επαληθεύτηκε και κατά την εκτέλεση της παρούσας διπλωματικής εργασίας, το γεγονός ότι οι εφαρμογές Τεχνητής Νοημοσύνης εξελίσσονται με ιδιαίτερα γρήγορους ρυθμούς. Με βάση το δεδομένο αυτό, προκύπτει ότι σαν επέκταση της εργασίας, η επανάληψη της ίδιας πειραματικής διαδικασίας σε μελλοντικό χρόνο, θα επέτρεπε τη διαχρονική παρακολούθηση της εξέλιξης των εργαλείων TN και την αξιολόγηση του βαθμού βελτίωσης που επιτυγχάνεται προϊόντος του χρόνου.

Βιβλιογραφικές Αναφορές

- Βρύσης, Λ. (2019). *Τεχνικές οπτικοακουστικής επεξεργασίας και ανάλυσης της διαμεσικής επικοινωνίας*. [Διδακτορική διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης]. IKEE <https://ikee.lib.auth.gr/record/305871>
- ΕΛΟΤ. (1996). *Ακουστική - Περιγραφή, μέτρηση και αξιολόγηση του περιβαλλοντικού θορύβου*. ISO 1996-2
- Κοντομίχος, Φ., Ζαρούχας, Θ., Χατζηαντωνίου, Π. (2013α). *Ακουστική Χώρου - Βασικές Αρχές και Σύγχρονες Εξελίξεις στην Κτιριακή Ακουστική*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Κοντομίχος, Φ., Ποτηράκης, Σ., Χατζηαντωνίου, Π. (2013β). *Ειδικά Θέματα Ακουστικού Σχεδιασμού και Πολυμέσων - Λογισμικό και Μέθοδοι για Ακουστικό Σχεδιασμό*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Ζαρούχας, Θ., Τσιλφίδης, Α., Χατζηαντωνίου, Π. (2013). *Ειδικά Θέματα Ακουστικού Σχεδιασμού και Πολυμέσων - Πολυμεσικές Εφαρμογές Ήχου*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Θωΐδης, Ι. (2024). *Αποκατάσταση της καταληπτότητας της ομιλίας σε σύνθετα ηχητικά περιβάλλοντα για άτομα με απώλεια ακοής με τη χρήση τεχνικών βαθιάς μάθησης*. [Διδακτορική διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης]. IKEE <https://ikee.lib.auth.gr/record/360339>
- Σκαρλάτος, Δ. (2013α). *Εισαγωγή στις Μεθόδους Ηχομόνωσης και Ηχοπροστασίας*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Σκαρλάτος, Δ. (2013β). *Εφαρμογές Ηχομόνωσης και Ηχοπροστασίας*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Σκαρλάτος, Δ. (2021). *Εφαρμοσμένη ακουστική*
- Τσιλφίδης, Α. (2013). *Ακουστική και Αρχιτεκτονική*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Φλώρος, Α. (2013). *Ήχος και ακουστική στη μουσική*. Ελληνικό Ανοικτό Πανεπιστήμιο
- Acon Digital. (n.d.). *Restoration Suite. Suite of four plug-ins for professional grade audio restoration*. <https://acondigital.com/products/restoration-suite>
- Albanese, A. (2026). *The 5 Best Online DAWs for Easy Music Creation*. <https://blog.landr.com/best-online-daw/>
- Alessi, M. S. (2024). *VST vs VST3: What's the Difference Between the Two?*. <https://borisfx.com/blog/vst-vs-vst3-whats-the-difference/>

- Brittain, A. M. (2024). *How AI Noise Reduction Works*. <https://www.hp.com/us-en/shop/tech-takes/ai-noise-reduction>
- Dhuriya, A., (2023). *Audio Enhancement and Denoising Methods*. <https://ankurdhuriya.medium.com/audio-enhancement-and-denoising-methods-3644f0cad85b>
- ASUS. (2024). *[Gaming NB] Two-Way AI Noise Cancelation Setup Guide*. <https://www.asus.com/gr/support/faq/1044576/>
- Audacity. (n.d.). *Noise reduction & removal*. <https://support.audacityteam.org/repairing-audio/noise-reduction-removal>
- Auphonic. (n.d.). *Audio Post Production Algorithms (Singletrack)*. <https://auphonic.com/help/algorithms/singletrack.html>
- Bachu, R. G., Kopparthi, S., Adapa, B., & Barkana, B. D. (2010). *Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy*. https://www.researchgate.net/publication/259823741_VoicedUnvoiced_Decision_for_Speech_Signals_Based_on_Zero-Crossing_Rate_and_Energy
- Boashash, B., Ali Khan, N., & Ben-Jabeur, T. (2015). *Time–frequency features for pattern recognition using high-resolution TFDs: A tutorial review*. *Digital Signal Processing*, Volume 40, 2015, Pages 1-30. <https://doi.org/10.1016/j.dsp.2014.12.015>
- BORISFX. (n.d.). *ClipRemover*. <https://borisfx.com/products/clipremover/>
- Dialpad. (n.d.). *Mean Opinion Score (MOS)*. <https://www.dialpad.com/glossary/mos-score/>
- Greatbatch, R. (2026). *Plugin Formats Explained (VST, AU, AAX, etc)*. <https://help.pluginboutique.com/hc/en-us/articles/6233336161428-Plugin-Formats-Explained-VST-AU-AAX-etc>
- Haario, M., Varis, K., & Ekblom, P. (2024). *Comparison of traditional perpetual licensing and SaaS licensing ICT models: A case study*. https://www.researchgate.net/profile/Keijo-Varis/publication/386567006_Comparison_of_Traditional_Perpetual_Licensing_and_SaaS_Licensing_ICT_Models_A_Case_Study/links/6756efc38eca147b25e57931/Comparison-of-Traditional-Perpetual-Licensing-and-SaaS-Licensing-ICT-Models-A-Case-Study.pdf

- Hitchins, R. (n.d.). Raggae's Delayed Adoption by Jamaican Radio Stations in the 60s and 70s. <https://jamaicatimeline.com/music/hitchins-riaa.html>
- Howard, K. (2009). *Cut and Thrust: RIAA LP Equalization*. <https://www.stereophile.com/content/cut-and-thrust-riaa-lp-equalization>
- Jiroudek, H. (2026). *What's new in Adobe Podcast? Updates and enhancements - March 2026*. <https://podcast.adobe.com/en/guides/whats-new-in-adobe-podcast-march-2026>
- Krisp. (n.d). *Do Krisp devices support stereo sound?*. <https://help.krisp.ai/hc/en-us/articles/7504823200540-Do-Krisp-devices-support-stereo-sound>
- Man, B., & Reiss, J. (2017). *The mix evaluation dataset*. Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17), Edinburgh, UK, September 5–9, 2017. <http://www.brechtdeinan.com/publications/pdf/DAFx-17.pdf>
- Mason, A. (2011). *Terminology for loudness and level dBTP, LU, and all that*. BBC Research & Development White Paper. <https://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP202.pdf>
- McGowan, P. (2023). *Splice and dice*. <https://www.psaudio.com/blogs/pauls-posts/splice-and-dice>
- Mendenhall, J. (2026). *Making Music on Chromebook: The DAWs That Actually Work in 2026*. <https://www.starryhope.com/chromebooks/music-production-on-chromebook/>
- Metlay, M. (2022). *5 Tips for Noise Removal with RX Elements*. <https://www.izotope.com/en/learn/tips-for-noise-removal-rx-elements>
- Petters, G. (2004). A large set of audio features for sound description. http://recherche.ircam.fr/equipes/analyse-synthese/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf
- QSC. (2024). *Audio Limiter Explained | Sound Advice*. [Video]. YouTube. <https://www.youtube.com/watch?v=Sihaz9XyekM>
- Richmond, J. (2022). *Does Waves Clarity Vx Beat iZotope RX?*. <https://www.bhphotovideo.com/explora/pro-audio/hands-on-review/does-waves-clarity-vx-beat-izotope-rx>

- Soren, H., Lund, N., Lund, T. (2000). *0dBFS+ Levels in Digital Mastering*.
https://www.researchgate.net/publication/228898071_0dBFS_Levels_in_Digital_Mastering
- Steinberg. (n.d.). *Our Technologies*. <https://www.steinberg.net/technology/>
- Streijl, R.C., Winkler, S. & Hands, D.S. (2016). *Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives*. *Multimedia Systems* 22, 213–227.
<https://www.stefan.winklerbros.net/Publications/mmsj2016.pdf>
- Taylor. (2025). *The Vinyl Records Revival and the Growth of Vinyl LPs: Why Is There a Resurgence of Vinyl Records?*. <https://www.taylor.com/blog/increased-vinyl-sales-and-the-vinyl-revival>
- Xu, R., Wu, R., Ishiwaka, Y., Vondrick, C., & Zheng, C. (2020). *Listening to sounds of silence for speech denoising*.
https://www.cs.columbia.edu/cg/listen_to_the_silence/paper.pdf
- Vanacoro, M. (2020). *The Top 4 Browser Based DAWs*.
<https://test.ask.video/article/audio/the-top-4-browser-based-daws>
- Zlatic, T. (2026). *Free VST Plugins*. <https://bedroomproducersblog.com/free-vst-plugins/>

Παράρτημα Α: Συνοδευτικά αρχεία

Στην διάρκεια της υλοποίησης της παρούσας διπλωματικής εργασίας, δημιουργήθηκαν αρχεία ήχου, τα οποία και χρησιμοποιήθηκαν στα διάφορα στάδια της διαδικασίας. Τα αρχεία αυτά είναι διαθέσιμα στον σύνδεσμο:

<https://drive.google.com/drive/folders/1YMj9sQZV4n74-QFpOv8voTyzG2lEowMH>

Για την καλύτερη οργάνωσή τους, τα αρχεία είναι κατανεμημένα σε υποφακέλους, με τα εξής ονόματα:

- «01 - Αρχεία ομιλίας και θορύβου πριν τη μίξη»,
- «02 - Ενθόρυβα αρχεία μετά την μίξη»,
- «03 - Αποθροvyβημένα αρχεία (οργανωμένα ανά λογισμικό)» και
- «04 - Αποθροvyβημένα αρχεία (οργανωμένα για ακρόαση ανά θόρυβο)».

Παράρτημα Β: Τεχνικά χαρακτηριστικά εξοπλισμού που χρησιμοποιήθηκε

Για την υλοποίηση του πειραματικού μέρους της παρούσας διπλωματικής εργασίας, χρησιμοποιήθηκε εξοπλισμός, ο οποίος συμπεριέλαβε σταθερό σταθμό εργασίας, φορητό σταθμό εργασίας για την διενέργεια των ακροάσεων, εξωτερική κάρτα ήχου και ακουστικά για τις ακροάσεις. Τα κύρια χαρακτηριστικά του εξοπλισμού, αναγράφονται στους Πίνακες Β.1 έως και Β.5 που ακολουθούν.

Πίνακας Β.1. Χαρακτηριστικά σταθερού σταθμού εργασίας

Μοντέλο - Τύπος	Dell OptiPlex 3080 (Desktop)
CPU	Intel i5-10500 @ 3.10 GHz
RAM	8 GB
GPU	Intel UHD Graphics 630
HD	256 GB SSD
OS	Windows 11 Pro

Πίνακας Β.2. Χαρακτηριστικά φορητού σταθμού εργασίας για διενέργεια ακροάσεων

Μοντέλο - Τύπος	Lenovo IdeaPad 3-15IIL05 (Laptop)
CPU	Intel Core i5-1035G4 @ 1.10 GHz
RAM	12 GB
GPU	Intel Integrated Graphics
HD	512 GB SSD
OS	Windows 11 Home

Πίνακας Β.3. Χαρακτηριστικά εξόδου κάρτας ήχου για διενέργεια ακροάσεων

Μοντέλο - Τύπος	Focusrite Clarett+ 4Pre (External)
Freq. Response 20-20kHz	< 0.06 dB
Freq. Response 20-35kHz	< 0.07 dB
Dynamic Range	118 dB
THD+N	-104 dB
Max Output level	16 dBu

Πίνακας Β.4. Χαρακτηριστικά ακουστικών Α, για διενέργεια ακροάσεων

Μοντέλο - Τύπος	AGK K44 (Over-ear, semi-closed design)
Audio Freq. Bandwidth	18-20000 Hz
Efficiency	115 dB SPL/V
Max. input power	200 mW
Rated impedance	32 Ω
Connection	Wired

Πίνακας Β.5. Χαρακτηριστικά ακουστικών Β, για διενέργεια ακροάσεων

Μοντέλο - Τύπος	Sennheiser HD 280 PRO (Over-ear)
Audio Freq. Response	8-25000 Hz
Sensitivity	113 dB SPL (1 kHz)
THD	<0,1 % (1 kHz, 100 dB SPL)
Impedance	64 Ω
Connection	Wired

Παράρτημα Γ: Ερωτηματολόγιο ακροάσεων - Οδηγίες

Η επίδοση αποθρομβοποίησης αξιολογείται σε κλίμακα από 1 έως 5. **Τσεκάρετε το πλαίσιο που αντιστοιχεί στον βαθμό επίδοσης του κάθε δείγματος.**

(Όπου: 1 = Πολύ κακή, 2 = Κακή, 3 = Μέτρια, 4 = Καλή και 5 = Πολύ Καλή).

Δείγματα	Βαθμός Επίδοσης				
01 έως 06	1	2	3	4	5
01A					
01B					
01C					
01D					
01E					
02A					
02B					
02C					
02D					
02E					
03A					
03B					
03C					
03D					
03E					
04A					
04B					
04C					
04D					
04E					
05A					
05B					
05C					
05D					
05E					
06A					
06B					
06C					
06D					
06E					

Δείγματα	Βαθμός Επίδοσης				
07 έως 12	1	2	3	4	5
07A					
07B					
07C					
07D					
07E					
08A					
08B					
08C					
08D					
08E					
09A					
09B					
09C					
09D					
09E					
10A					
10B					
10C					
10D					
10E					
11A					
11B					
11C					
11D					
11E					
12A					
12B					
12C					
12D					
12E					

Οδηγίες για την ακρόαση και συμπλήρωση του ερωτηματολογίου

Το ερωτηματολόγιο συμπληρώνεται για **συνολικό αριθμό εξήντα (60) δειγμάτων** αποθορυβοποιημένου ήχου, που προέκυψαν με βάση την επίδοση επιλεγμένων εργαλείων λογισμικού, τα οποία χρησιμοποιήθηκαν για την αποθορυβοποίησή τους.

Συνολικά αξιολογείται η αποθορυβοποίηση δώδεκα (12) διαφορετικών ενθόρυβων σημάτων, από πέντε (5) διαφορετικά λογισμικά το κάθε ένα. Έτσι **το ερωτηματολόγιο είναι οργανωμένο σε 12 πεντάδες**, μία για κάθε διαφορετικό θόρυβο.

Οι ακροάσεις πρέπει να γίνουν σε **όσο το δυνατόν πιο ήσυχο περιβάλλον και με χρήση ακουστικών**. Για αποτελεσματικότερη αξιολόγηση, υπάρχει η δυνατότητα επανάληψης κάθε ακρόασης πριν τη βαθμολόγηση.

Οι ακροατές καλούνται βαθμολογήσουν την επίδοση των λογισμικών, σαν απάντηση στο ερώτημα: «**Πώς αξιολογείτε την συνολική ποιότητα και καθαρότητα της ομιλίας, την οποία έχει επιδείξει το λογισμικό που έχει χρησιμοποιηθεί για την αποθορυβοποίηση του συγκεκριμένου δείγματος;**»

Για την απάντηση στο ερώτημα αυτό, πρέπει να ληφθούν υπόψη τα ακόλουθα αξιολογικά κριτήρια:

A). Βαθμός καταστολής θορύβου: Αξιολογείται το κατά πόσο και σε τι βαθμό, περιλαμβάνεται ή όχι διακριτός - ακουστός θόρυβος στο τελικό σήμα ομιλίας.

B). Ποιότητα αποθορυβημένου ήχου: Αξιολογείται το κατά πόσο εμφανίζονται ή όχι αισθητές αλλοιώσεις στο αποθορυβημένο σήμα ομιλίας. Σε αυτές συμπεριλαμβάνονται πιθανές ασυνέχειες στον λόγο, αποκοπή φωνημάτων, ή /και συριστικών συμφώνων, ή αντικατάσταση τους από άλλα (για φωνήεντα), φυσικότητα στην εκφορά λόγου, εμφάνιση "ρομποτικού", "μεταλλικού" ή γενικότερα παραμορφωμένου ήχου.

Η επίδοση αποθορυβοποίησης αξιολογείται με βαθμό σε κλίμακα 1 έως 5. (Όπου 1 = Πολύ κακή, 2 = Κακή, 3 = Μέτρια, 4 = Καλή και 5 = Πολύ Καλή).

Πληροφορίες για τα αρχεία (θέση και ονοματολογία):

Τα 60 αρχεία προς ακρόαση και αξιολόγηση **είναι οργανωμένα σε 12 υποφάκελους (με ονομασία από 01 έως 12)**, οι οποίοι αντιστοιχούν στους 12 διαφορετικούς θορύβους. **Ο κάθε ένας από αυτούς, περιλαμβάνει 6 (1+5) αρχεία. Το αντίστοιχο ενθόρυβο αρχείο με την ονομασία «xx_(Noisy).wav» (όπου xx = 01 έως 12) και τα αρχεία των αποθορυβοποιήσεων από τα 5 λογισμικά, με τις ονομασίες «xxA.wav» έως «xxE.wav».** (Σε κάθε υποφάκελο, τα A, B, C, D και E, αντιστοιχούν με τυχαίο τρόπο, σε διαφορετικό λογισμικό κάθε φορά, για αντικειμενικότερα αποτελέσματα).

Τα αρχεία είναι διαθέσιμα στον σύνδεσμο:

<https://drive.google.com/drive/folders/19Lt0roODYdSZ7EvsG57bah2b4lzdIBUr>

(Η συνολική διάρκεια της διαδικασίας είναι περίπου 30 λεπτά).

Σας ευχαριστώ πολύ για την συμμετοχή σας!

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.