

## **CAVEMOVE: Συλλογή ακουστικών δεδομένων για τη μελέτη τεχνολογιών φωνής μέσα σε κινούμενα οχήματα**

Νικόλαος Στεφανάκης<sup>1,2,\*</sup>, Μαρίνος Καλαϊτζάκης<sup>2</sup>, Δέσποινα Παυλίδη<sup>2</sup> και Αντρέας Συμιακάκης<sup>2</sup>

<sup>1</sup>Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ελληνικό Μεσογειακό Πανεπιστήμιο, Ρέθυμνο

<sup>2</sup> Ινστιτούτο Πληροφορικής, Ίδρυμα Τεχνολογίας και Έρευνας, Ηράκλειο

\* [nstefana@hmu.gr](mailto:nstefana@hmu.gr)

### **ΠΕΡΙΛΗΨΗ**

Το CAVEMOVE είναι ένα έργο το οποίο αποσκοπεί στην καταγραφή πραγματικών ακουστικών δεδομένων για τη μελέτη τεχνολογιών φωνής μέσα σε οχήματα. Η διαδικασία ηχογράφησης περιλαμβάνει (i) ηχογραφήσεις ακουστικών κρουστικών αποκρίσεων, οι οποίες αξιοποιούνται για την παραγωγή των συστατικών φωνής και (ii) ηχογραφήσεις ακουστικού θορύβου σε διαφορετικές συνθήκες, τόσο εν στάσει όσο και εν κινήσει. Τα δεδομένα καταγράφονται με δύο διαφορετικές διατάξεις μικροφώνων μέσα στην καμπίνα και ειδικότερα (i) με χρήση συστοιχίας μικροφώνων ή (ii) με καταναμημένα μικρόφωνα. Περιγράφουμε εν συντομία τις συνθήκες κάτω από τις οποίες πραγματοποιήθηκαν οι καταγραφές και εξηγούμε τις διαδικασίες πίσω από ένα API σε γλώσσα προγραμματισμού python που σχεδιάστηκε με στόχο να διευκολύνει το χρήστη στην αξιοποίηση των ηχητικών δεδομένων για το σχεδιασμό πειραμάτων. Μια πρώτη έκδοση του API και μέρος του περιγραφόμενου συνόλου δεδομένων διατίθενται για δωρεάν λήψη.

## **CAVEMOVE: An acoustic database for the study of voice-enabled technologies inside moving vehicles**

### **ABSTRACT**

CAVEMOVE is a research project dedicated to the collection of audio data for the study of voice-enabled technologies inside moving vehicles. The recording process involves (i) recordings of acoustic impulse responses, which provide the means for modelling the speech components and (ii) recordings of acoustic noise at a wide range of both static and in-motion conditions. Data is recorded with two different microphone configurations and particularly (i) a compact microphone array or (ii) a distributed microphone setup. We briefly describe the conditions under which the recordings were acquired and we give insights about a python API that we designed with the goal to support the research and development of voice enabled technologies inside moving vehicles. The first version of this python API and part of the described dataset are available for free download.

## Εισαγωγή

Την τελευταία δεκαετία, το ενδιαφέρον για εφαρμογές ήχου σε περιβάλλοντα κινούμενων οχημάτων έχει αυξηθεί σημαντικά. Αυτές οι εφαρμογές περιλαμβάνουν την αναγνώριση ομιλίας [1,2] την υποκειμενική αντίληψη του ήχου που παράγεται από τα αυτοκίνητα [3,4], τον ενεργό έλεγχο θορύβου [5] κ.α. Μέχρι το 2028, το Navigant Research προβλέπει ότι η διείσδυση των τεχνολογιών φωνής μέσα στα καινούρια αυτοκίνητα αναμένεται να αγγίξει το 90%. Ενδεικτικά, τέτοιες εφαρμογές είναι η ανίχνευση ομιλίας (voice activity detection), ο διαχωρισμός των ομιλητών, η καταπίεση του θορύβου στο σήμα ομιλίας, ο εντοπισμός της θέσης του ομιλητή και βέβαια, η αναγνώριση ομιλίας και η αναγνώριση ομιλητή. Ως εκ τούτου, χρειάζονται εργαλεία που να μπορούν να υποστηρίξουν την έρευνα που απαιτείται ώστε οι παραπάνω εφαρμογές να μπορούν να λειτουργήσουν επιτυχώς στα θορυβώδη περιβάλλοντα των αυτοκινήτων.

Σε αυτή την εργασία, περιγράφουμε τη διαδικασία που ακολουθήσαμε για την κατασκευή μιας ακουστικής βάσης δεδομένων και μιας αντίστοιχης διεπαφής προγραμματισμού εφαρμογών (Application Programming Interface - API) σε γλώσσα python με στόχο τη διευκόλυνση της έρευνας σε τεχνολογίες φωνής μέσα σε κινούμενα οχήματα. Σε αντίθεση με προηγούμενες εργασίες, το σύνολο δεδομένων μας δεν περιλαμβάνει μόνο διαφορετικά αυτοκίνητα και διατάξεις μικροφώνου, αλλά περιλαμβάνει επίσης κρουστικές αποκρίσεις μετρούμενες από διαφορετικές θέσεις επιβατών εντός της καμπίνας. Αυτό δίνει μέγιστη ευελιξία στην αξιοποίηση των δεδομένων, επιτρέποντας στους χρήστες να συνθέτουν σήματα μικροφώνου σε οποιαδήποτε επιθυμητή γλώσσα, παρέχοντας τις δικές τους καθαρές (ιδανικά ανηχοϊκές) εγγραφές ομιλίας. Μέρους του συνόλου δεδομένων που περιγράφεται, μαζί με το API διατίθεται δωρεάν για λήψη από το διαδίκτυο.

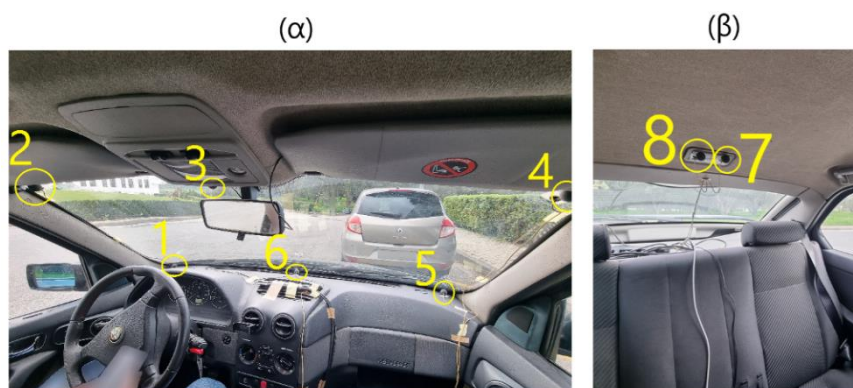
### 1. Καταγραφή των ηχητικών δεδομένων

#### 1.1 Μικροφωνικές διατάξεις

Όλες οι ηχογραφήσεις έγιναν με τη χρήση της USB κάρτας ήχου M-audio M-Track Eight audio, σε ρυθμό δειγματοληψίας 48kHz και μήκος δείγματος 24-bits. Οκτώ πανκατευθυντικά μικρόφωνα (Shure SM93) τοποθετημένα σε διαφορετικές θέσεις χρησιμοποιούνται για την καταγραφή του ακουστικού θορύβου και των κρουστικών αποκρίσεων (Impulse Responses – IRs). Αυτά είναι ελαφριά μικρόφωνα πέτου με επίπεδη απόκριση συχνότητας από τα 80 έως τα 20000 Hz. Δύο διαφορετικές διατάξεις μικροφώνων χρησιμοποιούνται μέχρι σήμερα, όπως περιγράφονται παρακάτω.:

- **Συστοιχία μικροφώνων:** τα οκτώ μικρόφωνα είναι ομοιόμορφα κατανεμημένα πάνω ένα κύκλο ακτίνας 5 cm. Η κυκλική συστοιχία τοποθετείται στην κορυφή του μπροστινού πίνακα, μεταξύ του οδηγού και του συνοδηγού (βλέπε Σχήμα 1(γ)). Τα μικρόφωνα τοποθετήθηκαν στη συστοιχία χωρίς αντιανέμιο.

- **Κατανεμημένα μικρόφωνα:** Τα μικρόφωνα είναι κατανεμημένα ως εξής: έξι από τα μικρόφωνα βρίσκονται στο μπροστινό μέρος του αυτοκινήτου, ενώ τα άλλα δύο είναι τοποθετημένα στην οροφή στο πίσω μέρος του αυτοκινήτου (βλέπε π.χ. Σχήμα 1(α) και (β)). Αυτή η διάταξη εξασφαλίζει ότι υπάρχει τουλάχιστον ένα μικρόφωνο



Σχήμα 1 Κατανομή των μικροφώνων μέσα στο Alfa Romeo. Η κατανομή είναι ενδεικτική και για την τοποθέτηση στα υπόλοιπα αυτοκίνητα που είχαν καταναμημένα μικρόφωνα.

κοντά σε κάθε επιβάτη. Όλα τα μικρόφωνα σταθεροποιήθηκαν με κατάλληλη κολλητική ταινία και καλύφθηκαν με αντιανέμιο.

### 1.2 Ηχογράφηση των κρουστικών αποκρίσεων

Οι ακουστικές κρουστικές αποκρίσεις μας επιτρέπουν να καταγράψουμε τις ακουστικές υπογραφές από κάθε ομιλητή προς τα μικρόφωνα. Η διαδικασία επαναλαμβάνεται σε κάθε αυτοκίνητο και διάταξη μικροφώνου, καλύπτοντας τουλάχιστον τη θέση του οδηγού, του συνοδηγού, του πίσω αριστερού, του πίσω μεσαίου και του πίσω δεξιού επιβάτη. Όλες οι κρουστικές αποκρίσεις καταγράφονται με τη χρήση του Room Eq Wizard, κάνοντας χρήση ημίτονου μέτρησης ως σήμα διέγερσης. Το μεγάφωνο που χρησιμοποιήθηκε για αυτές τις μετρήσεις είναι το Talkbox της NTi, το οποίο έχει σχεδιαστεί ειδικά για να προσομοιάζει στην απόκριση συχνότητας και την κατευθυντικότητα της ανθρώπινης φωνής. Το μεγάφωνο είναι βαθμονομημένο και διαθέτει ενσωματωμένα σήματα διέγερσης τα οποία μας επιτρέπουν να διεγείρουμε το εσωτερικό του αυτοκινήτου με ήχους που αντιστοιχούν σε συγκεκριμένη φωνητική προσπάθεια και συγκεκριμένα κανονική φωνητική προσπάθεια (60 dBA σε 1 m απόσταση) και υψηλή φωνητική προσπάθεια (70 dBA σε 1 m απόσταση) [6]. Αυτή η διαδικασία είναι πολύ σημαντική, καθώς θα μας επιτρέψει να υπολογίσουμε τις στάθμες σήματος που αντιστοιχούν σε συγκεκριμένες φωνητικές προσπάθειες, επιτρέποντάς μας έτσι πετύχουμε μια ρεαλιστική ισορροπία μεταξύ των συστατικών φωνής και των συστατικών θορύβου.

Θεωρήσαμε ότι ο ήχος από το ενσωματωμένο ηχοσύστημα του αυτοκινήτου αποτελεί μια ακουστική πηγή παρεμβολής που αξίζει να ληφθεί υπόψιν. Για αυτό το λόγο, κρουστικές αποκρίσεις μετρήθηκαν και από το ενσωματωμένο σύστημα ήχου στο αυτοκίνητο, όταν αυτό ήταν διαθέσιμο. Η διαδικασία αυτή περιλαμβάνει την ταυτόχρονη καταγραφή της απόκρισης από όλα τα ηχεία του αυτοκινήτου. Πραγματοποιήθηκε με την απευθείας τροφοδοσία του σήματος διέγερσης στη

βοηθητική είσοδο του ηχοσυστήματος ή με την αναπαραγωγή του από το CD player.

*Πίνακας 1. Αυτοκίνητα και μικροφωνικές διατάξεις που έχουν καλυφθεί έως σήμερα*

Κατασκευαστής	Μοντέλο	Κατανεμημένα	Συστοιχία	Ηχοσύστημα
Honda	CR-V (2009)	NAI	NAI	NAI
Alfa Romeo	146 TS (2000)	NAI	OXI	OXI
Smart	Forfour (2019)	OXI	NAI	NAI
Volkswagen	Golf (2011)	NAI	NAI	NAI

### 1.3 Αυτοκίνητα και ακουστικές συνθήκες

Μέχρι στιγμής, η διαδικασία καταγραφής έχει επαναληφθεί πέντε φορές, καλύπτοντας τέσσερα διαφορετικά αυτοκίνητα, όπως φαίνεται στον Πίνακα 1. Ως ελεγχόμενοι παράγοντες που επηρεάζουν τη σύνθεση και το επίπεδο του ακουστικού θορύβου σε κάθε μικρόφωνο θεωρούνται οι ακόλουθοι: η ταχύτητα του αυτοκινήτου, η ταχύτητα περιστροφής του κινητήρα, η θέση των παραθύρων και η ισχύς εξαερισμού/κλιματισμού. Πρόσθετοι παράγοντες που δεν είναι τόσο εύκολο να ελεγχθούν αλλά επηρεάζουν επίσης τον ακουστικό θόρυβο είναι οι καιρικές συνθήκες, η κυκλοφορία, η επιφάνεια του οδοστρώματος (π.χ. τραχιά, λεία κ.λπ.) και οι εξωτερικές ακουστικές πηγές (π.χ. οχήματα, άνθρωποι, ζώα). Όσον αφορά τους ελεγχόμενους παράγοντες, επιχειρήθηκε κατά την ηχογράφιση να έχουν σταθερές τιμές για συνεχή χρονικά τμήματα, ώστε να επιτευχθούν στάσιμες συνθήκες θορύβου. Οι ελεγχόμενες συνθήκες καλύπτουν ένα συγκεκριμένο εύρος ταχυτήτων οδήγησης και θέσης παραθύρων. Ειδικότερα όσον αφορά τη θέση των παραθύρων, εξετάζουμε τέσσερις καταστάσεις που χαρακτηρίζονται ως 0, 1, 2 και 3. Η κατάσταση 0 αντιστοιχεί σε τελείως κλειστά παράθυρα. Η κατάσταση 1 αντιστοιχεί σε παράθυρα του οδηγού και του συνοδηγού ελαφρώς ανοιχτά (περίπου 10 cm). Η κατάσταση 2 αντιστοιχεί σε εντελώς ανοιχτά παράθυρα μπροστά και, τέλος, η κατάσταση 3 αντιστοιχεί σε εντελώς ανοιχτά παράθυρα τόσο μπροστά όσο και πίσω. Όσον αφορά τις ταχύτητες οδήγησης, σε κάθε αυτοκίνητο καλύπτουμε τουλάχιστον το εύρος από 50 km/h έως 110 km/h με βήματα των 10 km/h. Σε κάθε περίπτωση καλύπτονται όλοι οι συνδυασμοί των ανοιγμάτων των παραθύρων και των ταχυτήτων οδήγησης, εκτός από τις περιπτώσεις που αυτό δεν είναι εφικτό. Επιπροσθέτως αυτών των εν κινήσει ηχογραφήσεων, ελήφθησαν στατικές ηχογραφήσεις για την καταγραφή του θορύβου που παράγεται από το ενσωματωμένο σύστημα εξαερισμού ή κλιματισμού. Σε κάθε αυτοκίνητο εξετάστηκαν δύο ή τρία διαφορετικά επίπεδα ισχύος εξαερισμού και η διαδικασία καταγραφής επαναλήφθηκε ώστε να καλυφθούν όλοι οι συνδυασμοί μεταξύ των διαφορετικών επιπέδων εξαερισμού και των διαφορετικών θέσεων παραθύρων.

### 1.4 Πρόσθετες σημειώσεις ως προς τη διαδικασία

Γενικά, η όλη διαδικασία της ηχογράφησης διαρκούσε περισσότερο από μία ημέρα κάθε φορά. Κατά τη διάρκεια αυτής της περιόδου, τα μικρόφωνα ήταν

σταθερά στερεωμένα στις αρχικές τους θέσεις και δόθηκε προσοχή ώστε να αποφευχθούν τυχόν ακούσιες μετατοπίσεις. Τα ποτενσιόμετρα της κάρτας ήχου, τα οποία επηρεάζουν το κέρδος προ-ενίσχυσης που σχετίζεται με κάθε κανάλι μικροφώνου, ήταν επίσης σε σταθερή θέση. Επιπλέον, πριν και μετά από κάθε ηχογράφηση, το καθένα κανάλι μικροφώνου υποβαλλόταν σε μια εμπειρική διαδικασία βαθμονόμησης που μας επέτρεπε να εξάγουμε μια κατά προσέγγιση εκτίμηση της ευαισθησίας ελεύθερου πεδίου. Η βαθμονόμηση έλαβε χώρα μέσα στον ημι-ανηχοϊκό θάλαμο του Ινστιτούτου Πληροφορικής του ΙΤΕ. Κάθε μικρόφωνο, ενώ ήταν συνδεδεμένο πάνω στην κάρτα ήχου, τοποθετήθηκε σε συγκεκριμένο σημείο εντός του θαλάμου και ηχοβολήθηκε με ροζ θόρυβο συγκεκριμένης στάθμης. Στη συνέχεια, ένα πανκατευθυντικό μετρητικό μικρόφωνο που ήταν συνδεδεμένο σε ένα ηχώμετρο NTi, τοποθετήθηκε στην ίδια ακριβώς θέση και ηχοβολήθηκε με τον ίδιο ακριβώς ροζ θόρυβο. Από το ηχώμετρο, καταγράφηκε η ισοδύναμη στάθμη θορύβου με φίλτρο A. Στη συνέχεια, το σήμα ροζ θορύβου που ηχογραφήθηκε με κάθε ένα μικρόφωνο πέτυε από κατάλληλο φίλτρο που σχεδιάστηκε ώστε να προσεγγίζει την χαρακτηριστική απόκριση συχνότητας του φίλτρου με στάθμιση A. Μέσω αυτής της διαδικασίας, μπορέσαμε να συσχετίσουμε την στάθμη σήματος του κάθε καναλιού με τη στάθμη ηχητικής πίεσης σε dBA, εξάγοντας έτσι μια κατά προσέγγιση εκτίμηση της ευαισθησίας ελεύθερου πεδίου κάθε μικροφώνου.

## 2 Συνδυασμός των ηχογραφήσεων

Ο απώτερος στόχος του API που σχεδιάστηκε στο πλαίσιο του CAVEMOVE είναι να επιτρέψει στους μηχανικούς και τους ερευνητές να αξιοποιούν εύκολα τα ηχητικά δεδομένα για να συνθέτουν τα μικροφωνικά σήματα που αντιστοιχούν σε καθορισμένα σενάρια. Για συγκεκριμένο αυτοκίνητο και μικροφωνική διάταξη η διαδικασία της σύνθεσης μπορεί να περιγραφεί

$$\mathbf{Y} = \mathbf{S}(p, Ls, w, \mathbf{x}) + \mathbf{A}(La, w, \mathbf{z}) + \mathbf{N}(s, w) + \mathbf{V}(l, w). \quad (2.1)$$

Εν συντομία, κάθε κεφαλαίο σύμβολο είναι ένας πίνακας σήματος  $N \times M$ , όπου  $M$  είναι ο επιθυμητός αριθμός καναλιών μικροφώνου και  $N$  είναι η διάρκεια του συντιθέμενου ηχητικού αποσπάσματος σε δείγματα. Το  $\mathbf{S}$  αντιπροσωπεύει το φιλτραρισμένο σήμα ομιλίας, το  $\mathbf{A}$  αντιπροσωπεύει τα ηχητικά σήματα παρεμβολής που προέρχονται από το ενσωματωμένο ηχοσύστημα (όταν είναι διαθέσιμο),  $\mathbf{N}$  είναι τα συστατικά θορύβου που αντιστοιχούν στη συγκεκριμένη κατάσταση οδήγησης και  $\mathbf{V}$  είναι τα συστατικά θορύβου που σχετίζονται με τον εξωτερικό/κλιματισμό. Υποθέτοντας ότι όλα αυτά τα ηχητικά συστατικά είναι στατιστικά ανεξάρτητα το ένα από το άλλο, το τελικό σήμα μικροφώνου μπορεί να συντεθεί όπως στην Εξ. (2.1), μέσω απλής υπέρθεσης. Όπως φαίνεται, κάθε συστατικό παράγεται ως συνάρτηση καθορισμένων παραμέτρων. Μια σύντομη εξήγηση αυτών των παραμέτρων είναι η ακόλουθη.

$w$ : είναι ένας ακέραιος αριθμός που παίρνει τιμές 0, 1, 2 ή 3, με κάθε ακέραια τιμή που αντιπροσωπεύει μια διαφορετική συνθήκη ως προς το άνοιγμα των παραθύρων

$x$ : είναι ένα μονοδιάστατο διάνυσμα με το σήμα της ομιλίας (ιδανικά ανηχοϊκό) που έχει επιλεγεί για το συγκεκριμένο σενάριο. Ο χρήστης είναι υπεύθυνος για την παροχή αυτού του σήματος σε μορφή PCM. Η ηχογράφηση μπορεί να έχει οποιοδήποτε μήκος και ρυθμό δειγματοληψίας (μετατρέπεται αυτόματα στο ρυθμό δειγματοληψίας που δουλεύει το πρόγραμμα).

$p$ : μπορεί να επιλεγεί από ένα δεδομένο σύνολο τιμών συμβολοσειρών, όπου η κάθε συμβολοσειρά αντιστοιχεί σε μια συγκεκριμένη θέση επιβάτη εντός της καμπίνας του αυτοκινήτου. Για κάθε συνδυασμό  $p$  και  $w$ , το API θα ανακαλέσει αυτόματα τις κρουστικές αποκρίσεις που απαιτούνται για την παραγωγή των συστατικών ομιλίας μέσω της συνέλιξης τους με το σήμα  $x$ .

$L_s$ : είναι μια τιμή σε dBA, που αντιστοιχεί στην επιθυμητή φωνητική προσπάθεια. Προτεινόμενο εύρος για ρεαλιστικής έντασης φωνή είναι μεταξύ 60 και 70, με 60 να αντιστοιχεί σε κανονική φωνητική προσπάθεια και 70 σε υψηλή φωνητική προσπάθεια [6].

$L_a$ : είναι μια τιμή σε dBA, που αντιστοιχεί στη μέση ακουστική στάθμη του ηχητικού προγράμματος που αναπαράγεται από το ενσωματωμένο ηχοσύστημα. Τιμές εντός του εύρους 50 με 90 θεωρούνται ρεαλιστικές.

$z$ : είναι ένα μονοδιάστατο διάνυσμα που αντιπροσωπεύει το ηχητικό πρόγραμμα που αναπαράγεται από το ηχοσύστημα. Ο χρήστης είναι υπεύθυνος για την παροχή του αρχείου ήχου σε μορφή PCM. Το αρχείο ήχου μπορεί να έχει οποιοδήποτε μήκος και ρυθμό δειγματοληψίας. Αρχεία ήχου με περισσότερα από ένα κανάλια είναι αποδεκτά αλλά αποδίδονται ως μονοφωνικά σήματα.

$s$ : είναι μια ακέραια τιμή από ένα σύνολο δεδομένων τιμών, όπου κάθε τιμή αντιστοιχεί σε διαφορετική ταχύτητα σε km/h. Τα συστατικά θορύβου που αντιστοιχούν σε συγκεκριμένη ταχύτητα και άνοιγμα παραθύρου φορτώνονται αυτόματα από το πρόγραμμα.

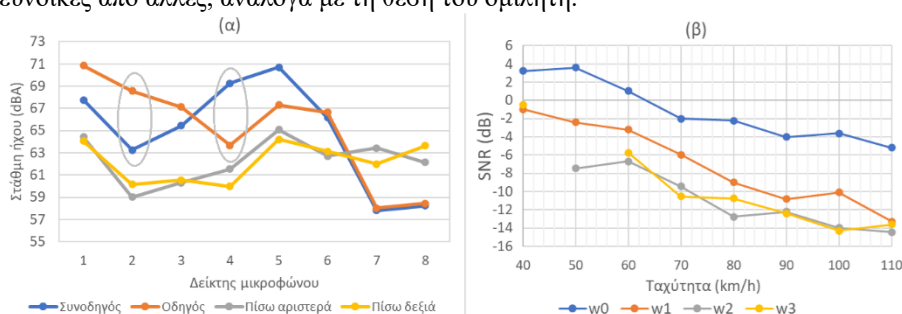
$l$ : είναι ακέραιος με τιμές στην περιοχή 1, 2 ή 3, όπου η κάθε τιμή αντιστοιχεί σε διαφορετική ισχύ λειτουργίας του συστήματος εξαερισμού/κλιματισμού.

Η προσέγγιση έχει σχεδιαστεί με τέτοιο τρόπο ώστε το προκύπτον μήκος  $N$  του σήματος εξόδου  $Y$ , καθώς και όλων των πινάκων στην Εξ.(2.1) να ταιριάζει με το μήκος του σήματος ομιλίας  $x$  που παρέχει ο χρήστης. Για να το πετύχουμε αυτό έχει ενσωματωθεί μηχανισμός που ανακυκλώνει τα σήματα θορύβου που είναι εγγενείς στην κατασκευή των  $A$ ,  $N$  και  $V$  όσες φορές χρειαστεί. Η προσέγγιση σύνθεσης παρέχει έναν εύκολο τρόπο για την παραγωγή ηχητικών μίξεων που διατηρούν μια ρεαλιστική ισορροπία μεταξύ της φωνής και των συστατικών θορύβου, χωρίς να απαιτούνται επιπλέον ενέργειες από τον χρήστη. Επίσης, εκτός από τις παραπάνω λειτουργίες, το API περιλαμβάνει βοηθητικές λειτουργίες, όπως για παράδειγμα την ικανότητα εξαγωγής των κρουστικών αποκρίσεων που εμπλέκονται στις συγκεκριμένες συνθήκες μελέτης αλλά και τη δημιουργία διανυσμάτων στροφής (steering vector) που να είναι συμβατά με την εν λόγω κυκλική συστοιχία. Αυτές οι λειτουργίες μπορούν να βοηθήσουν τον τελικό χρήστη στον σχεδιασμό των σχηματιστών λοβών (beamformers) και μετα-φίλτρων (post filters) που απαιτούνται για την ανάπτυξη τυπικών αλγορίθμων βελτίωσης ομιλίας. Τα δεδομένα ήχου που μοιραζόμαστε ανοιχτά με αυτό το έγγραφο είναι σε μορφή PCM, σε ρυθμό δειγματοληψίας 16 kHz και μήκος δείγματος 24 bit. Ενδιαφερόμενοι χρήστες μπορούν να κατεβάσουν το API της python από το GitHub του CAVEMOVE [7], όπου θα βρουν επίσης και το σύνδεσμο για τη λήψη των ηχητικών δεδομένων.

### 3 Ανάλυση ηχογραφήσεων

Η ανάλυση που ακολουθεί περιέχει αποτελέσματα που πηγάζουν κατευθείαν από τις ηχογραφήσεις που έγιναν σε συγκεκριμένα αυτοκίνητα, ωστόσο τα συμπεράσματα που προκύπτουν είναι ενδεικτικά και για τα υπόλοιπα αυτοκίνητα που ηχογραφήθηκαν στο CAVEMOVE.

Σε πρώτη ανάλυση, επιλέχθηκε η διάταξη με τα κατανεμημένα μικρόφωνα εντός του Alfa Romeo για να μελετηθεί αν κάποιες θέσεις μικροφώνων είναι πιο ευνοϊκές από άλλες, ανάλογα με τη θέση του ομιλητή.



Σχήμα 2. (α) Στάθμη ήχου (dBA) ανά μικρόφωνο για διαφορετικές θέσεις ομιλητή με κλειστά παράθυρα (w0) μέσα στο Alfa Romeo και (β) SNR συναρτήσει της ταχύτητας για κανονική φωνητική προσπάθεια από τη θέση του οδηγού μέσα στο Honda-CRV.

Λαμβάνοντας υπόψιν την ευαισθησία ελεύθερου πεδίου κάθε μικροφώνου (βλ. Ενότητα 1.3) απεικονίζουμε στο Σχήμα 2(α) τη στάθμη ήχου σε κάθε θέση μικροφώνου (σε dBA) ανάλογα με τη θέση του ομιλητή μέσα στο Alfa Romeo, για κανονική φωνητική προσπάθεια, όπως αυτή προέκυψε αξιοποιώντας το Talkbox. Η ανάλυση αυτή απεικονίζεται για τέσσερις θέσεις ομιλητών, του οδηγού, του συνοδηγού και των πίσω αριστερά και πίσω δεξιά επιβατών. Από το διάγραμμα μπορεί να παρατηρηθεί κανείς ότι η στάθμη της φωνής του οδηγού μεγιστοποιείται στο μικρόφωνο 1 και του συνοδηγού στο μικρόφωνο 5, γεγονός που έχει να κάνει με το ότι η απόσταση μεταξύ πηγής και δέκτη γίνεται ελάχιστη. Ωστόσο, το μικρόφωνο 2 φαίνεται να εξασφαλίζει καλύτερο ακουστικό διαχωρισμό (περίπου 6 dB) συγκριτικά με το μικρόφωνο 1, εάν υποθέσουμε ότι το επιθυμητό σήμα είναι η φωνή του οδηγού. Αντίστοιχα, το μικρόφωνο 4 εξασφαλίζει καλύτερο διαχωρισμό αν υποθέσουμε ότι το επιθυμητό σήμα είναι η φωνή του συνοδηγού. Τέλος, τα μικρόφωνα 7 και 8 στο πίσω μέρος εξασφαλίζουν έναν διαχωρισμό της τάξης των 5 dB σε σχέση με τις φωνές του οδηγού ή του συνοδηγού.

Στο Σχήμα 2(β), τα αποτελέσματα που παρουσιάζονται αφορούν το Honda CRV, όπου απεικονίζεται η διακύμανση του SNR συναρτήσει της ταχύτητας του οχήματος για διαφορετικές καταστάσεις παραθύρων. Το SNR προκύπτει ως η διαφορά μεταξύ της στάθμης της φωνής και του θορύβου στο μικρόφωνο 3 (βλ. Σχήμα 1), μετά την εφαρμογή του φίλτρου A σε κάθε συστατικό. Η στάθμη της φωνής υπολογίστηκε πάλι για κανονική φωνητική προσπάθεια, αγνοώντας το γεγονός ότι οι ομιλητές ενστικτωδώς καταβάλουν μικρότερη ή μεγαλύτερη φωνητική προσπάθεια ανάλογα με τη στάθμη θορύβου [8]. Στο διάγραμμα επιβεβαιώνεται η διαίσθησή μας, ότι η

στάθμη του θορύβου αυξάνεται με την ταχύτητα αλλά και με το άνοιγμα του παραθύρου. Φαίνεται επίσης ότι η ελάττωση που επέρχεται στο SNR κατά τη μετάβαση από τελείως κλειστά ( $w_0$ ) σε λίγο ανοιχτά παράθυρα ( $w_1$ ) μπορεί να φτάσει μέχρι και τα 8 dB στις μεγάλες ταχύτητες. Τέλος οι πολύ χαμηλές τιμές του SNR του Σχήματος 2(β) αναδεικνύουν τις σημαντικές προκλήσεις που πρέπει να αντιμετωπιστούν για την επιτυχή εφαρμογή των διαδεδομένων τεχνολογιών φωνής μέσα στα εν λόγω περιβάλλοντα

#### 4 Συμπεράσματα

Η μελέτη των τεχνολογιών φωνής μέσα σε κινούμενα οχήματα είναι μια σημαντική προϋπόθεση για τη βελτίωση της εμπειρίας χρήστη στα αυτοκίνητα του μέλλοντος. Το CAVEMOVE είναι ένα ερευνητικό έργο αφιερωμένο στη συλλογή δεδομένων ήχου για τη διευκόλυνση αυτής της μελέτης. Μαζί με τα ηχητικά δεδομένα, παρέχεται ένα API σε γλώσσα python ώστε ο χρήστης να μπορεί να συνθέτει εύκολα τα ηχητικά σήματα που αντιστοιχούν σε ένα ευρύ φάσμα διαφορετικών συνθηκών οδήγησης. Μελλοντικά, τα ηχητικά δεδομένα θα εμπλουτιστούν με ηχογραφήσεις και από άλλα αυτοκίνητα, και το API θα σχεδιαστεί και στη γλώσσα Matlab.

#### 5 Αναφορές

- [1] H. Hoshino, R. Terashima, T. Shimizu, and T. Wakita, “Noise robust speech recognition in a car environment based on the acoustic features of car interior noise,” *The R&D Review of Toyota CRDL*, vol. **39**, no. 1 (2004).
- [2] T. Matheja, M. Buck, and T. Fingscheidt, “A dynamic multichannel speech enhancement system for distributed microphones in a car environment,” *EURASIP Journal on Advances in Signal Processing*, pp. 191 (2013).
- [3] J. Yin, S. Damiano, M. Verhelst, T. van Waterschoot, and A. Guntoro, “Real-time acoustic perception for automotive applications,” in *Design, Automation & Test in Europe Conference & Exhibition*, pp. 1–6 (2023).
- [4] N. Kaplanis, S. Bech, J. Tervo, S. Patynen, T. Lokki, T. van Waterschoot, and S. Jensen, “Perceptual aspects of reproduced sound in car cabin acoustics,” *J. Acoust. Soc. Am.* **141**(3), pp. 1459–1469, (2017).
- [5] Samarasinghe, P.N., Zhang, W. and Abhayapala, T.D., “Recent advances in active noise control inside automobile cabins: Toward quieter cars,” *IEEE Signal Processing Magazine*, **33**(6), pp.61-73.
- [6] Talkbox manual, <https://www.ntiaudio.com/Portals/0/data/en/TalkBox-Manual.pdf>
- [7] CAVEMOVE API, <https://github.com/SPL-FORTH-ICS/CAVEMOVE>
- [8] V. Summers, D. Pisoni, R. Bernacki, R. Pedlow, and M. Stokes, “Effects of noise on speech production: Acoustic and perceptual analyses,” *J. Acoust. Soc. Am.* **84**(3), pp. 917–928 (1988).