

Προσομοίωση ακουστικής απόκρισης χώρου για αυθαίρετη επιλογή θέσης δέκτη-πηγής

Σωτήρης Λοής¹, Ρήγας Κωτσάκης², Χρήστος Σεβαστιάδης³, Νικόλαος Βρύζας³,
Λάζαρος Βρύσης³, Χαράλαμπος Δημούλας³, Γεώργιος Καλλιρής³

¹Imperial College London

²Διεθνές Πανεπιστήμιο της Ελλάδος

³Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

*rkotsakis@gmail.com

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία διερευνά τη δυνατότητα ηχητικής εικονικής πλοήγησης σε χώρο μέσω μιας ηχητικής προσομοίωσης της απόκρισης του χώρου για διαφορετικές θέσεις πηγής και δέκτη. Η χρησιμότητα αυτής της λειτουργικότητας είναι να παρέχει μια ενδεικτική πληροφορία σχετικά αφενός με την καταλληλότητα του χώρου για πραγματοποίηση ηχογραφήσεων, αφετέρου για τα ποιοτικά χαρακτηριστικά του για την υποβοήθηση του ηχητικού σχεδιασμού. Η ηχητική προσομοίωση βασίζεται στην πραγματοποίηση ηχογραφήσεων σε έναν χώρο με στόχο τη δημιουργία μεγάλων και δυναμικών αποθετηρίων με τη συνδρομή πολλών διαφορετικών χρηστών. Οι παραπάνω απαιτήσεις δικαιολογούν έναν συμβιβασμό στην πιστότητα και την ακρίβεια των μετρήσεων, οι οποίες πρέπει να μπορούν να εκτελεστούν από μη εξειδικευμένους χρήστες και με προσβάσιμο εξοπλισμό. Παρόλο, που πραγματοποιήθηκε πλήθος μετρήσεων από την ερευνητική ομάδα, στην παρούσα εργασία εξετάζεται σε αρχικό πειραματικό στάδιο η μεθοδολογία προσομοίωσης της απόκρισης σε συγκεκριμένη θέση πηγής-δέκτη με βάση τα δεδομένα αποκρίσεων από προϋπάρχουσες μετρήσεις χώρων, για να ελεγχεί πιλοτικά η εγκυρότητα του συνολικού εγχειρίματος.

Simulation of room acoustic response for arbitrary selection of receiver-source position

ABSTRACT

This study investigates the potentials of acoustic virtual navigation in a room/place through an acoustic simulation of the space's response for different source and receiver positions. The usefulness of this functionality is to provide indicative information regarding, on the one hand, the suitability of the place for conducting recordings, and on the other hand, its qualitative characteristics for aiding acoustic design. The acoustic simulation is based on conducting recordings in a place with the aim of creating large and dynamic repositories with the participation of many different users. These requirements justify a compromise in the fidelity and accuracy of the measurements, which must be executable by non-expert users with accessible equipment. Although numerous measurements were conducted by the research team, this study examines, at an initial experimental stage, the methodology for simulating the response at a specific source-receiver position based on response data from pre-existing space measurements, to pilot-test the validity of the overall project

Εισαγωγή

Η ηχητική κρουστική απόκριση (sound impulse response) είναι μια θεμελιώδης έννοια στην ακουστική, αναφερόμενη στον τρόπο με τον οποίο ένα σύστημα ή ένας χώρος αντιδρά σε μια σύντομη έκρηξη ηχητικής ενέργειας, που συχνά αναφέρεται ως κρουστικό σήμα (impulse). Όταν ένας ηχητικός κρουστικός ήχος εκπέμπεται σε ένα περιβάλλον, αλληλεπιδρά με τα διάφορα αντικείμενα, επιφάνειες και διαστάσεις του χώρου, προκαλώντας ανακλάσεις, απορροφήσεις και διαθλάσεις. Η απόκριση κρουστικού σήματος περιγράφει αυτήν την αντίδραση ως συνάρτηση του χρόνου, καταγράφοντας την απόκριση του συστήματος όταν διεγείρεται από μια σύντομη είσοδο. Ουσιαστικά, αποκαλύπτει τον τρόπο με τον οποίο ένας χώρος τροποποιεί τον ήχο που ταξιδεύει μέσα από αυτόν, προσφέροντας πληροφορίες για τις ακουστικές ιδιότητες του περιβάλλοντος [1]. Η μέτρηση της απόκρισης κρουστικού ήχου πραγματοποιείται με εξειδικευμένο εξοπλισμό, τεχνικές και σημάτων διέγερσης/ δοκιμής. [2]. Η απόκριση κρουστικού σήματος συχνά αναλύεται με την έννοια της απόκρισης χώρου (room impulse response - RIR), που αναφέρεται στον τρόπο με τον οποίο ο ήχος ταξιδεύει από μια πηγή σε έναν ακροατή μέσα σε ένα δωμάτιο. Η RIR μπορεί να αναλυθεί στον άμεσο ήχο (ο οποίος είναι ο αρχικός ήχος που ταξιδεύει απευθείας από την πηγή στον δέκτη), τις πρώτες ανακλάσεις (οι οποίες είναι οι ήχοι που ανακλώνται από επιφάνειες κοντά στην πηγή και φτάνουν αμέσως μετά τον άμεσο ήχο) και την καθαυτό αντήχηση (reverberation), που είναι το παρατεταμένο, διάχυτο πεδίο ανακλάσεων που συνεχίζουν να φτάνουν καθώς ο ήχος φθίνει. Αυτά τα στοιχεία βοηθούν στον καθορισμό της ευκρίνειας, της κατανοησιμότητας και της «ζεστασιάς» του ήχου στο περιβάλλον [3]. Τα δεδομένα απόκρισης κρουστικού ήχου χρησιμοποιούνται επίσης ευρέως στην επεξεργασία ψηφιακών σημάτων (DSP), όπου αποτελούν τη βάση για πολλούς αλγόριθμους που χρησιμοποιούνται στην παραγωγή ήχου. Για παράδειγμα, η συνελκτική αντήχηση (convolution reverb), ένα είδος ψηφιακής ηχητικής αντήχησης που χρησιμοποιείται στην παραγωγή μουσικής και στον ηχητικό σχεδιασμό, βασίζεται στην αρχή της αναδίπλωσης (convolution), που περιλαμβάνει την εφαρμογή μιας απόκρισης κρουστικού σήματος καταγεγραμμένης σε έναν πραγματικό ακουστικό χώρο σε ένα ψηφιακό ηχητικό σήμα. Αυτό επιτρέπει στους μηχανικούς να προσομοιώσουν τον ήχο διαφορετικών περιβαλλόντων, όπως αίθουσες συναυλιών, εκκλησίες ή μικρά δωμάτια, εφαρμόζοντας τις καταγεγραμμένες αποκρίσεις κρουστικού ήχου αυτών των χώρων στις ηχογραφήσεις τους. Το convolution reverb χρησιμοποιείται ευρέως στην παραγωγή μουσικής, την μεταπαραγωγή ταινιών και τις εφαρμογές εικονικής πραγματικότητας για τη δημιουργία ρεαλιστικών ηχητικών τοπίων [4], [5].

1. Μεθοδολογίες ανακατασκευής ηχητικού πεδίου

Η ανακατασκευή του ηχητικού πεδίου είναι κρίσιμη σε πολλές εφαρμογές, όπως ο χωρικός ήχος σε επαυξημένες, εικονικές και μικτές πραγματικότητες (AR/VR/XR), καθώς και ο έλεγχος του ηχητικού πεδίου για την ακουστική βελτιστοποίηση. Οι παραδοσιακές προσεγγίσεις στην ανακατασκευή ηχητικού πεδίου συχνά βασίζονται στη διεργασία παρεμβολής από περιορισμένες χωρικές και χρονικές μετρήσεις, αλλά νέες μέθοδοι αναπτύσσονται για να αυξήσουν την ακρίβεια και τον ρεαλισμό [6], [7]. Αυτές οι προσεγγίσεις έχουν ευρύ φάσμα εφαρμογών, συμπεριλαμβανομένης της ανάλυσης ηχητικού πεδίου, της βελτιστοποίησης των ακουστικών περιβαλλόντων και της ενίσχυσης εμπειριών εμπύθισης σε περιβάλλοντα AR/VR/XR [8], [9]. Άλλες χρήσεις περιλαμβάνουν τον έλεγχο του ηχητικού πεδίου, ο οποίος επιτρέπει τον χειρισμό του ήχου σε πραγματικό χρόνο για την επίτευξη συγκεκριμένων ακουστικών στόχων [10].

Ένας σημαντικός στόχος της ανακατασκευής ηχητικού πεδίου που είναι και το αντικείμενο της παρούσας εργασίας είναι η ακριβής πρόβλεψη των ηχητικών πεδίων σε περιοχές όπου δεν υπάρχουν διαθέσιμες μετρήσεις. Αυτό μπορεί να επιτευχθεί με τεχνικές παρεμβολής, οι οποίες εκτιμούν το ηχητικό πεδίο βάσει περιορισμένου συνόλου μετρήσεων [11]. Τεχνικές κανονικοποίησης καθώς και προσεγγίσεις μηχανικής μάθησης που προβλέπουν το ηχητικό πεδίο σε χαρτογράφητες τοποθεσίες, είναι μερικοί από τους τρόπους με τους οποίους έχει αντιμετωπιστεί αυτό το πρόβλημα [12]. Πιο πρόσφατα, ευρετικά μοντέλα που βασίζονται στην αποσύνθεση της απόκρισης κρουστικών κυμάτων (RIR) έχουν αποκτήσει εξέχουσα θέση λόγω της ικανότητάς τους να ερμηνεύουν τα φυσικά χαρακτηριστικά των ηχητικών πεδίων [13].

Ανάμεσα στις πιο καινοτόμες εξελίξεις στον τομέα αυτό είναι η άνοδος των μεθοδολογιών βαθιάς μάθησης, οι οποίες υπόσχονται να βελτιώσουν σημαντικά την ανακατασκευή του ηχητικού πεδίου [14]. Τεχνικές όπως τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs) έχουν δείξει εξαιρετικές δυνατότητες στην καταγραφή των περίπλοκων σχέσεων μεταξύ των δεδομένων εισόδου (π.χ. μετρήσεις ηχητικού πεδίου) και των αποτελεσμάτων εξόδου (ανακατασκευασμένα ηχητικά πεδία) [15], [16]. Αυτά τα μοντέλα λειτουργούν ιδιαίτερα αποτελεσματικά σε εφαρμογές που απαιτούν την επεξεργασία σύνθετων εργασιών, όπως η ανακατασκευή ηχητικού πεδίου, όπου τα χαρακτηριστικά των χώρων και η δυναμική διάδοσης του ήχου πρέπει να ληφθούν υπόψη [17], [18].

Μια ιδιαίτερα πολλά υποσχόμενη προσέγγιση που έχει πρόσφατα προσελκύσει το ενδιαφέρον είναι τα Φυσικά Ενημερωμένα Νευρωνικά Δίκτυα (PINNs). Τα PINNs είναι ένας τύπος νευρωνικών δικτύων που ενσωματώνουν τους φυσικούς νόμους που διέπουν τη διάδοση του ήχου, όπως η κυματική εξίσωση, απευθείας στη διαδικασία εκπαίδευσης [19]. Αυτό επιτρέπει στο δίκτυο να κατανοεί και να προβλέπει τη δυναμική του ηχητικού πεδίου με μεγαλύτερη ακρίβεια, ακολουθώντας θεμελιώδεις φυσικές αρχές [20]. Τα PINNs είναι ιδανικά για την επίλυση σύνθετων μερικών διαφορικών εξισώσεων (PDEs) που εμπλέκονται στη διάδοση ηχητικών κυμάτων, γεγονός που τα καθιστά ένα ισχυρό εργαλείο για εργασίες ανακατασκευής ηχητικού πεδίου. Επιπλέον, η συνεχής αναπαράσταση που προσφέρουν τα PINNs επιτρέπει χαρτογράφηση με τη χρήση πλέγματος, προσφέροντας ουσιαστικά υπολογιστικά οφέλη.

Η παρούσα εργασία εξετάζει το πλαίσιο στο οποίο μπορούν να εφαρμοστούν τα PINNs, χρησιμοποιώντας πειραματικές μετρήσεις και προσομοιωμένα δεδομένα. Τα πειραματικά δεδομένα σε πραγματικούς χώρους δείχνουν πώς τα PINNs μπορούν να προσαρμοστούν σε πολύπλοκα, θορυβώδη περιβάλλοντα, διατηρώντας παράλληλα υψηλά επίπεδα ακρίβειας [11].

2. Περιγραφή του Πειράματος

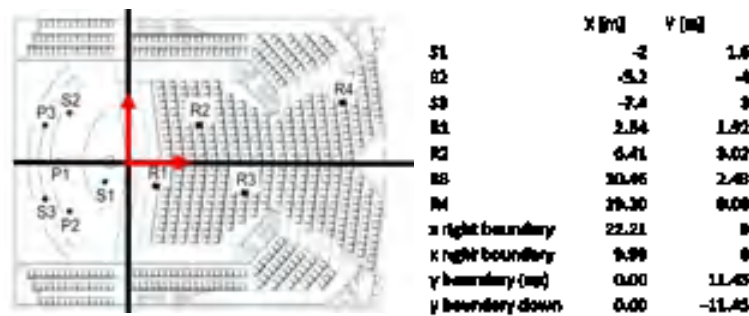
Στην παρούσα μελέτη, πραγματοποιήθηκε μια σειρά πειραμάτων για την προσομοίωση της ακουστικής απόκρισης της συναυλιακής αίθουσας Promenadikeskus στο Pori, Φινλανδίας, όπως περιγράφεται στο [21]. Το πείραμα επικεντρώθηκε στη μέτρηση αποκρίσεων κρουστικών σημάτων σε διάφορες θέσεις πηγής και δέκτη, χρησιμοποιώντας τόσο μια ομοιοκατευθυντική πηγή ήχου όσο και έναν υπογούφερ. Οι μετρηθείσες αποκρίσεις αποθρομβοποιήθηκαν και διορθώθηκαν για ατέλειες του συστήματος, ενώ τα δεδομένα περιλάμβαναν διάφορες διαμορφώσεις μικροφώνων (π.χ. ομοιοκατευθυντικά, δικαναλικά, και καρδιοειδή). Στη συνέχεια, εφαρμόσαμε ένα φυσικά ενημερωμένο νευρωνικό δίκτυο (PINN) για την αναπαραγωγή αυτών των μετρημένων αποκρίσεων από τα ομοιοκατευθυντικά μικρόφωνα στη θέση των θεατών, χρησιμοποιώντας τα συγκεντρωμένα δεδομένα για την εκπαίδευση του μοντέλου. Στη συνέχεια, χρησιμοποιήσαμε τα παρεχόμενα δεδομένα, όπως ο χρόνος αντήχησης (T30), για να

αναπαράγουμε την ακουστική συμπεριφορά του χώρου στο PINN. Ο στόχος ήταν να προσομοιώσει τη διάδοση του ήχου σε διαφορετικές συχνότητες και διαμορφώσεις πηγής-δέκτη, αξιοποιώντας τις προϋπάρχουσες μετρήσεις της αίθουσας για την επικύρωση της ακρίβειας του μοντέλου.

3. Μοντελοποίηση και Αποτελέσματα

Στην επεξεργασία των δεδομένων, χρησιμοποιήθηκαν μετρήσεις από την ομοιοκατευθυντική πηγή ήχου (omnidirectional), καθώς προσφέρει μια πιο ομοιόμορφη κατανομή της ηχητικής ενέργειας στον χώρο, επιτρέποντας την καταγραφή των ακουστικών χαρακτηριστικών του χώρου ανεξάρτητα από την κατεύθυνση της εκπομπής του ήχου. Αυτό μας εξασφάλισε μια πιο αντικειμενική και ακριβή καταγραφή της ακουστικής συμπεριφοράς του χώρου, χωρίς να επηρεάζεται το αποτέλεσμα από την κατευθυντικότητα της πηγής ήχου.

Για την απλοποίηση των δεδομένων και τη δημιουργία ενός συνεκτικού μοντέλου, θεωρήθηκε ότι η πηγή και οι δέκτες βρίσκονται σε ίδιο ύψος. Αυτή η υπόθεση επέτρεψε τη μείωση της πολυπλοκότητας των δεδομένων και εστίασαμε αποκλειστικά στις οριζόντιες διαστάσεις του χώρου (x, y). Οι συντεταγμένες των πηγών και των δεκτών τοποθετήθηκαν στις συντεταγμένες της Εικόνας 1. Το διάγραμμα αυτό βασίζεται στις ακριβείς γεωμετρικές πληροφορίες που παρείχαν τα δεδομένα μέτρησης και απεικονίζει τις θέσεις των πηγών (S1, S2, S3) και των δεκτών (R1, R2, R3) στον χώρο, όπως απεικονίζεται στην Εικόνα 1.



Εικόνα 1. Θέσεις Πηγών - Δεκτών

Στη συνέχεια, για κάθε συνδυασμό πηγής-δέκτη (S1-R1, S1-R2, κ.λπ.), επεξεργαστήκαμε τα ηχητικά σήματα που συλλέχθηκαν από τα αρχεία ήχου. Η χρονοσειρά που αντιστοιχεί στην ηχητική πίεση του αριστερού και του δεξιού καναλιού του κάθε δέκτη υπολογίστηκε και συνδυάστηκε για να παράγει την μέση τιμή της ηχητικής πίεσης. Τα δεδομένα εισόδου περιλάμβαναν τις συντεταγμένες της πηγής και του δέκτη, καθώς και την χρονοσειρά, ενώ τα δεδομένα εξόδου αναπαριστούσαν την ηχητική πίεση στο εκάστοτε σημείο δέκτη για το αντίστοιχο χρονικό διάστημα.

Αυτός ο τρόπος επεξεργασίας επέτρεψε την παραγωγή των συνόλων δεδομένων εισόδου και εξόδου σε μορφή numpy αρχείων, τα οποία ήταν έτοιμα για την είσοδο στο PINN ώστε να αναπαραχθούν τα αποτελέσματα της ακουστικής προσομοίωσης του χώρου.

Η κυματική εξίσωση που χρησιμοποιήθηκε για τη μοντελοποίηση του ηχητικού πεδίου είναι:

$$\frac{\partial^2 p}{\partial t^2} = \nabla^2 \left(\frac{\partial^2 p}{\partial x^2} - \frac{\partial^2 p}{\partial y^2} \right)$$

όπου p είναι η ηχητική πίεση, t είναι ο χρόνος, x και y οι χωρικές συντεταγμένες, και c η ταχύτητα του ήχου (θεωρούμενη ίση με 343 m/s).

3.1 Αρχιτεκτονική του PINN

Κατά την προεπεξεργασία των δεδομένων, αφαιρέθηκαν οι τιμές πολύ χαμηλού πλάτους (κάτω από 0.0002), οι οποίες θεωρήθηκαν ως μη ουσιώδεις ή πιθανώς θορυβώδεις. Η αφαίρεση αυτών των τιμών είχε ως στόχο τη μείωση του θορύβου στα δεδομένα εκπαίδευσης, που θα μπορούσε να επηρεάσει αρνητικά τη διαδικασία μάθησης του μοντέλου. Ειδικότερα, δεδομένα με σχεδόν μηδενικές τιμές ενδέχεται να περιέχουν θόρυβο ή πληροφορίες που δεν συνεισφέρουν ουσιαστικά στη μάθηση, καθώς δεν παρέχουν αξιόπιστη ένδειξη της ηχητικής πίεσης στο σύστημα. Η αφαίρεση τέτοιων τιμών είναι συχνά αποδεκτή πρακτική στη μηχανική μάθηση για τη βελτίωση της σταθερότητας του μοντέλου και την ταχύτερη σύγκλιση, καθώς επιτρέπει στο δίκτυο να επικεντρωθεί σε πιο σημαντικά μοτίβα [22]. Παρόλο που η απομάκρυνση αυτών των τιμών μπορεί να οδηγήσει σε απώλεια ορισμένων λεπτομερειών, η πρακτική αυτή αποσκοπεί στη βελτίωση της γενικής απόδοσης του μοντέλου, όπως έχει παρατηρηθεί σε άλλα προβλήματα εκπαίδευσης νευρωνικών δικτύων όπου χαμηλές ή θορυβώδεις τιμές εμποδίζουν τη σύγκλιση.

Το PINN αποτελείται από τρία εισαγωγικά χαρακτηριστικά (χρόνος, συντεταγμένες x και y) και μια έξοδο που αναπαριστά την ηχητική πίεση σε κάθε σημείο του χώρου. Η αρχιτεκτονική του δικτύου περιλαμβάνει ένα επίπεδο εισόδου και ένα αριθμό κρυφών επιπέδων. Στην περίπτωση αυτή, χρησιμοποιήθηκαν 2 κρυφά επίπεδα με 50 νευρώνες, και η λειτουργία ενεργοποίησης που εφαρμόστηκε είναι η \tanh . Η συγκεκριμένη λειτουργία επιλέχθηκε καθώς είναι κατάλληλη για μοντέλα που πρέπει να μάθουν σύνθετα μοτίβα μη γραμμικής φύσης. Η τελική έξοδος του μοντέλου παράγεται από ένα επίπεδο εξόδου χωρίς λειτουργία ενεργοποίησης (linear output).

Για την αρχική εκπαίδευση του δικτύου χρησιμοποιήθηκε ο αλγόριθμος Adam, καθώς είναι γρήγορος και σταθερός όταν ο αριθμός των δεδομένων είναι μεγάλος και η εκπαίδευση μπορεί να είναι ευαίσθητη στις αρχικές παραμέτρους. Στη συνέχεια, το μοντέλο βελτιστοποιήθηκε περαιτέρω με τον αλγόριθμο L-BFGS-B, ο οποίος επιτρέπει την πιο ακριβή ελαχιστοποίηση της συνάρτησης απώλειας και χρησιμοποιείται συχνά σε παραμετρικά προβλήματα με συνεχή δεδομένα.

Η συνάρτηση απώλειας του μοντέλου περιλαμβάνει δύο κύρια σκέλη: την απώλεια φυσικής (physics loss), που σχετίζεται με την κυματική εξίσωση, και την απώλεια δεδομένων (data loss), που αφορά τη σύγκριση μεταξύ των προσομοιωμένων και των πραγματικών δεδομένων. Ο συνδυασμός αυτός επιτυγχάνεται με τη χρήση προσαρμοζόμενων βαρών (adaptive weighting), με τις παραμέτρους ϵ_{data} και ϵ_{PDE} να ελέγχουν την αναλογία απώλειας μεταξύ των δύο.

Τα προσαρμοζόμενα βάρη χρησιμοποιούνται για να επιτρέπουν στο μοντέλο να ισορροπεί μεταξύ της μάθησης από τα δεδομένα και της ικανοποίησης της κυματικής εξίσωσης, δίνοντας μεγαλύτερη έμφαση στις φυσικές εξισώσεις (PDE) όταν απαιτείται ή στην απώλεια δεδομένων.

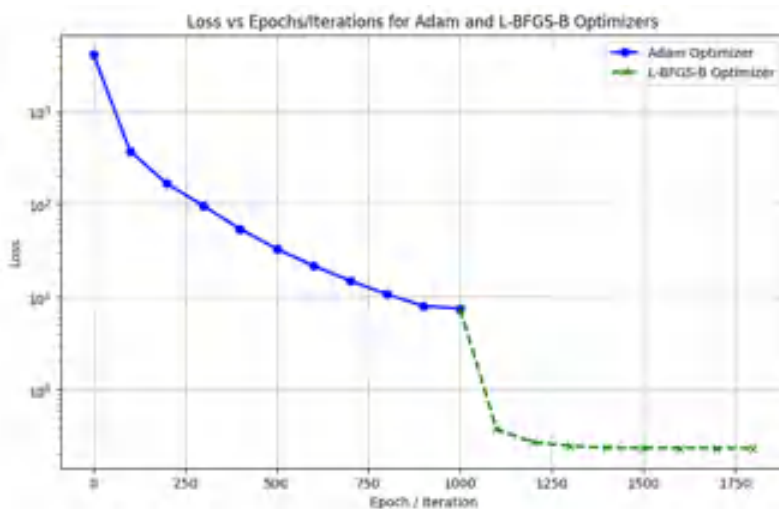
Σε αυτή την υλοποίηση, δεν χρησιμοποιήθηκαν ρητές συνοριακές συνθήκες (boundary conditions), κυρίως λόγω έλλειψης δεδομένων από τα όρια του χώρου. Επομένως, η προσομοίωση στηρίχθηκε κυρίως στα δεδομένα μέτρησης μεταξύ πηγής και δεκτών. Η εισαγωγή συνοριακών συνθηκών θα μπορούσε να βελτιώσει τη μοντελοποίηση, αλλά απαιτεί επιπλέον δεδομένα από τα σημεία αυτά τα οποία δεν ήταν διαθέσιμα από τις πηγές.

Η διαδικασία εκπαίδευσης ξεκίνησε με τον αλγόριθμο Adam για 1,000 εποχές. Η εκπαίδευση με τον Adam επιτρέπει γρήγορη σύγκλιση και αποφυγή τοπικών ελαχίστων σε πρώιμα στάδια της μάθησης. Αφού το μοντέλο συγκλίνει αρκετά, η βελτιστοποίηση

συνεχίζεται με τον αλγόριθμο L-BFGS-B, ο οποίος προσφέρει μεγαλύτερη ακρίβεια για την ελαχιστοποίηση της συνάρτησης απώλειας και μπορεί να εκμεταλλευτεί τη μαθηματική δομή του προβλήματος. Ο δεύτερος αλγόριθμος δεν έχει προκαθορισμένο αριθμό εποχών, αλλά εκτελείται έως ότου η τιμή της απώλειας σταθεροποιηθεί.

3.2 Αποτελέσματα

Η απώλεια του μοντέλου, όπως φαίνεται στην Εικόνα 2 «Loss vs Epochs/Iterations for Adam and L-BFGS-B Optimizers», παρουσιάζει σημαντική μείωση κατά την εκπαίδευση του PINN. Η απώλεια είναι σχεδιασμένη σε λογαριθμική κλίμακα, επιτρέποντας να φανεί η δραστική μείωση της αρχικής τιμής πάνω από 1000 σε τιμές κάτω του 1. Μετά από περίπου 1000 εποχές εκπαίδευσης με τον αλγόριθμο Adam, η βελτίωση με την προσθήκη του δεύτερου αλγόριθμου L-BFGS-B είναι εμφανής, καθώς η απώλεια μειώνεται ακόμα πιο γρήγορα, καταλήγοντας σε τιμές περίπου 0.2. Ο συνδυασμός αυτών των δύο αλγορίθμων ενισχύει την ακρίβεια του μοντέλου, καθώς ο Adam εξασφαλίζει γρήγορη αρχική σύγκλιση, ενώ ο L-BFGS-B προσφέρει μεγαλύτερη ακρίβεια στη βελτιστοποίηση, εκμεταλλευόμενος τη μαθηματική δομή του προβλήματος.



Εικόνα 2. Απώλειες Μοντέλου

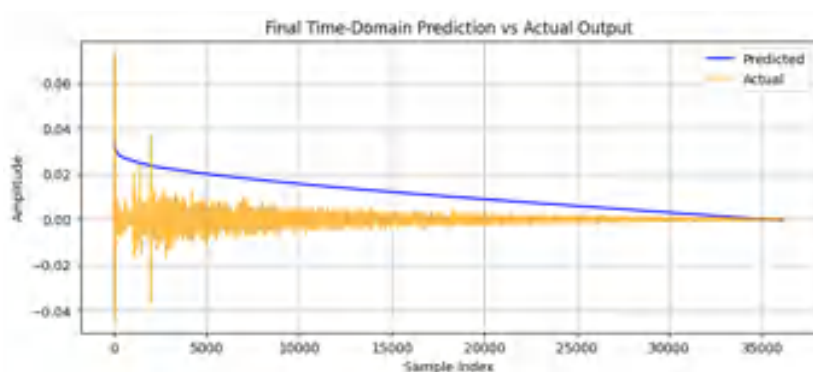
Στην Εικόνα 3 «Final Time-Domain Prediction vs Actual Output», απεικονίζεται η σύγκριση της προβλεπόμενης ηχητικής πίεσης με τα πραγματικά δεδομένα για το συνδυασμό R2 & S1. Στον άξονα των x βρίσκεται το δείγμα (sample index), το οποίο αντιστοιχεί στα χρονικά βήματα της ηχητικής απόκρισης, με κάθε τιμή να αναπαριστά μία χρονική στιγμή κατά τη διάρκεια της μέτρησης του σήματος. Ο άξονας των y αντιστοιχεί στο πλάτος της ηχητικής πίεσης. Παρά την οπτικά καλή σύγκλιση της γενικής μορφής της καμπύλης, η πρόβλεψη δεν καταφέρνει να αποδώσει πλήρως την πολυπλοκότητα των πραγματικών δεδομένων.

- **Έλλειψη συνοριακών δεδομένων:** Η έλλειψη λεπτομερών δεδομένων στα συνοριακά σημεία του χώρου περιορίζει τη δυνατότητα του μοντέλου να περιγράψει ακριβώς την ηχητική διάδοση στις άκρες του χώρου. Η προσθήκη συνοριακών συνθηκών θα μπορούσε να βελτιώσει την ακρίβεια της προσομοίωσης.

- **Απλότητα της κυματικής εξίσωσης:** Το μοντέλο χρησιμοποιεί μια βασική μορφή της κυματικής εξίσωσης, η οποία ίσως δεν επαρκεί για να περιγράψει τα πιο σύνθετα

φαινόμενα που παρατηρούνται στο πραγματικό ηχητικό πεδίο. Πιθανόν να απαιτείται η χρήση μιας πιο σύνθετης εκδοχής της κυματικής εξίσωσης, η οποία να περιλαμβάνει αποσβέσεις ή άλλους όρους για να περιγράψει την πραγματική διάδοση του ήχου στον χώρο με μεγαλύτερη ακρίβεια.

- **Απλή αρχιτεκτονική δικτύου:** Η αρχιτεκτονική του PINN που χρησιμοποιήθηκε είναι σχετικά απλή, με δύο κρυφά επίπεδα και 50 νευρώνες. Ενδεχομένως, ένα πιο περίπλοκο δίκτυο με περισσότερα επίπεδα ή περισσότερους νευρώνες θα μπορούσε να περιγράψει καλύτερα τα σύνθετα μοτίβα του ηχητικού σήματος. Ωστόσο, η εκπαίδευση πιο σύνθετων δικτύων απαιτεί σημαντικά μεγαλύτερη υπολογιστική ισχύ και χρόνο.



Εικόνα 3. Σύγκριση πραγματικών vs προβλεπόμενων τιμών

4. Συμπεράσματα και Μελλοντικές Κατευθύνσεις

Η παρούσα εργασία επιχειρεί να συγκρίνει και να προβλέψει τις ακουστικές ιδιότητες ενός χώρου μέσω φυσικά ενημερωμένων νευρωνικών δικτύων (PINN). Παρά τα αρχικά ενθαρρυντικά αποτελέσματα, σε μελλοντικές προσπάθειες θα μπορούσαν να ενσωματωθούν πιο εμπλουτισμένες συλλογές δεδομένων στα όρια του χώρου, καθώς και αξιολόγηση διαφορετικών αρχιτεκτονικών PINN για τη βελτίωση της ακρίβειας και της ταχύτητας σύγκλισης.

Επιπρόσθετα, χρήσιμη καθίσταται η διεξαγωγή υποκειμενικών/ ψυχοακουστικών πειραμάτων για την καταγραφή των αντιληπτικών συμπερασμάτων ενός ακροατηρίου αναφορικά με τις θέσεις πηγή-δέκτη και μετέπειτα σύγκριση με τα προσομοιωμένα αποτελέσματα.

5. Ευχαριστίες



Η παρούσα έρευνα είναι μέρος του ερευνητικού έργου SCENE που έχει λάβει χρηματοδότηση από το πρόγραμμα της EE Horizon Research and Innovation με αριθμό No 101095303.

6. Αναφορές

[1] Kuttruff, H., "Room Acoustics", (6th ed.). CRC Press, 2016. <https://doi.org/10.1201/9781315372150>

- [2] Farina, A., “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique”, Audio Engineering Society Convention 108, 2020;
- [3] Everest, F.A. and Pohlmann, K.C., “Master Handbook of Acoustics”, McGraw-Hill, New York, 2015.
- [4] Smith, J. O., “Physical Audio Signal Processing”, W3K Publishing, 2010.
- [5] Barron, M., “Auditorium Acoustics and Architectural Design”, E & FN Spon, 1993.
- [6] Ahrens, J. (2012). Analytic methods of sound field synthesis. Springer.
DOI: 10.1007/978-3-642-25743-8
- [7] Bogaert, Tim & Doclo, Simon & Wouters, Jan & Moonen, Marc. (2009). Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids. *The Journal of the Acoustical Society of America*. 125. 360-71. 10.1121/1.3023069.
- [8] Zhang X, Lou J, Zhu S, Lu J, Li R. Sound Field Reconstruction Using Prolate Spheroidal Wave Functions and Sparse Regularization. *Sensors*. 2023; 23(19):8312.
- [9] Bitzer, Joerg & Simmer, K.. (2001). Superdirective Microphone Arrays. *Microphone Arrays*. 10.1007/978-3-662-04619-7_2.
- [10] A. Politis, J. Vilkamo and V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852-866, Aug. 2015,
- [11] Liang, Z., Zhang, W. & Abhayapala, T.D. Sound field reconstruction using neural processes with dynamic kernels. *J AUDIO SPEECH MUSIC PROC*. 2024, 13 (2024). <https://doi.org/10.1186/s13636-024-00333-x>
- [12] Olivieri, M., Karakonstantis, X., Pezzoli, M. et al. Physics-informed neural network for volumetric sound field reconstruction of speech signals. *J AUDIO SPEECH MUSIC PROC*. 2024, 42 (2024). <https://doi.org/10.1186/s13636-024-00366-2>
- [13] Yokota, Kazuya & Ogura, Masataka & Abe, Masajiro. (2024). Synthesis of voiced sounds using physics-informed neural networks. *Acoustical Science and Technology*. 10.1250/ast.e24.55.
- [14] Donoho, D. L. (2006). "Compressed sensing." *IEEE Transactions on Information Theory*, 52(4), 1289-1306. DOI: 10.1109/TIT.2006.871582
- [15] Fei Ma, Sipei Zhao, Ian S. Burnett; Sound field reconstruction using a compact acoustics-informed neural network. *J. Acoust. Soc. Am.* 1 September 2024; 156 (3): 2009–2021.
- [16] Samuel A. Verburg, Efrén Fernández-Grande; Reconstruction of the sound field in a room using compressive sensing. *J. Acoust. Soc. Am.* 1 June 2018; 143 (6): 3770–3779.
- [17] Habets, Emanuël. (2006). Room Impulse Response Generator. Internal Report. 1-17.
- [18] Thrun, S., & Pratt, L. (2012). *Learning to Learn*. Springer. DOI: 10.1007/978-1-4615-5529-2.
- [19] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational Physics*, 378, 686-707. DOI: 10.1016/j.jcp.2018.10.045
- [20] Koyama, S., Ribeiro, J. G., Nakamura, T., Ueno, N., & Pezzoli, M. (2024). Physics-Informed Machine Learning For Sound Field Estimation. arXiv preprint arXiv:2408.14731.
- [21] Merimaa, J., Peltonen, T., & Lokki, T. (2005). Concert Hall Impulse Responses – Pori, Finland: Analysis Results. Helsinki University of Technology, Akukon Oy Consulting Engineers.
- [22] Zhang, Y., Song, Y., Fang, L., & Li, S. (2018). Noise removal in neural networks for better learning stability: A survey. *Journal of Machine Learning Research*, 20(3), 1-34.

