Doctoral Dissertation
Doctoral Program in Bioengineering and Medical-Surgical Sciences (36th Cycle)

Assessment of speech and auditory performances in adult patients with hearing aids and cochlear implants: the role of ecological environments and challenging communication scenarios

Andrea Albera

* * * * * *

Supervisors
Prof. Elio Berutti, Supervisor
Prof. Arianna Astolfi, Co-Supervisor

Doctoral Examination Committee:
Prof. Pasquale Bottalico, Referee, University of Illinois, Urbana-Champaign
Prof. Leone Giordano, Referee, Università Vita-Salute San Raffaele, Milano

Politecnico di Torino
April 15, 2024

I hereby declare that the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

…………………………………......

Andrea Albera

Turin, April 15, 2024

Politecnico di Torino
Doctoral course in *Bioengineering and Medical-Surgical Sciences*
XXXVI cycle October 2020 - October 2023

Doctoral course coordinator:

Gianluca Ciardelli, Full Professor
Department of Mechanical and Aerospace Engineering (DIMEAS)
Politecnico di Torino

Ph.D. supervisors:

Elio Berutti , Full Professor
Department of Surgical Sciences
Università degli Studi di Torino

Arianna Astolfi, Full Professor
Department of Energy (DENERG)
Politecnico di Torino

This PhD dissertation was completed in cooperation with:

Speech & Audiology Research Group
Department of Medical Physics and Acoustics
School of Medicine and Health Sciences
Oldenburg Universität
Ammerländer Heerstr. 114-118
26129, Oldenburg
Germany

S.C. ORL U.
Dipartimento di Chirurgia Generale e Specialistica
Città della Salute e della Scienza di Torino
Corso Bramante 88/90
10129, Torino
Italy

Summary

One of the main consequences of the evidence that population in western countries is aging is the progressive increase in diseases, such as the hearing loss, which are directly linked to aging. When comparing the 2020 EuroTrak results on self-reported hearing loss in Europe countries with the EuroStat data on the population 65 years or older, we observe indeed a very strong correlation. The higher the percentage of the population that is 65 years of age and older, the higher the percentage of the population that experiences hearing difficulties.

Hearing aids compensate, to a certain extent, for hearing impairment and, furthermore, ensure some individual rehabilitation: in 2018, 2 million Italian people owned hearing aids out of 7 million eligible people (29.5%). There is evidence that about 80% of adults who would benefit from hearing devices do not use them. Furthermore, more than 24% of hearing impaired given a hearing device do not really wear it in daily life, since they do not provide enough intelligibility improvement and comfort.

Differently, people with severe to profound hearing loss necessitate of hearing devices of greater power and complexity, such as the cochlear implants. A cochlear implant is a surgically implanted electronic medical device which can partially restore hearing in case of deafness.

All patients presenting hearing loss demonstrate outstanding auditory results in the use of hearing aids or cochlear implants in quite environments, but rather a significant performance deterioration and quite poor outcomes in the speech recognition for complex but typical everyday acoustic listening conditions.

In this context, there is also an important limitation of diagnostic tests that are currently performed in everyday clinical practice, namely a poor correspondence between audiometric measurements and the impairment reported by patients. A possible reason to explain the limitations of current audiometric tests is the use of simplistic diagnostic tools and sound systems that cannot adequately represent the spatial complexity of the real listening environments in which our patients live and work every day. Furthermore, although voice and phonation have been widely studied in the past with regards to prelingual deafness in childhood, very few studies in the literature report significant data regarding speech modifications in adults with hearing loss and consequent hearing remediation, by means of hearing aids or cochlear implants.

With this premise, the present doctoral thesis focuses on the increasing need to develop new diagnostic protocols to better assess speech and hearing loss to optimize hearing aid or cochlear implant fitting.

In clinical practice indeed, audiologists always face the critical phase to decide whether to recommend a hearing aid and, in the case of prescription, they do not have tools for a proper fitting on the different daily challenging listening conditions in the life environments (e.g., a noisy and reverberant office, day-care

center, restaurant or public place). The same limitations are evident in subjects with cochlear implants.
Through a multidisciplinary approach, the thesis aims to develop a new diagnostic tool for hearing loss (the SiIMax test, a simplified version of an adaptive audiometric test in noise), to conduct a comprehensive evaluation of the clinical consequences induced by the use of hearing aids or cochlear implants in adults (particularly regarding speech modifications), and finally, to assess the perceived quality and speech intelligibility in the use of modern flat-screen televisions and in complex acoustic scenarios such as daily environments with reverberant noise.

Acknowledgments

The Ph.D. years have been extremely demanding, sometimes difficult but most of the times totally illuminating. I am grateful to many people who made this experience possible and who have supported me in all the challenges I had to face.

I would like to express my sincere gratitude to my supervisor, Arianna Astolfi, for her invaluable guidance, support, and encouragement throughout the entirety of this research. The passion you imparted me allowed me to go in depth of everything and to be constantly curious.

A big thank you goes to Ania Warzybok for her extreme availability in hosting me within her research group in Oldenburg, allowing me to learn a lot about audiology and to participate in some of their interesting research projects.

I am deeply indebted to my family for their constant love, understanding and encouragement throughout this Ph.D, which represents three years of a multi-year journey of continuous study, hard work in the hospital, waking up early in the morning and returning home late in the evening. Particularly to my wife Giulia, whose patience, love, support and trust in me have never wavered and have provided me with the basis on which I have been able to pursue my (not yet finished) academic commitments.

I am also grateful to my parents for their unconditional love and unwavering belief in my abilities. Your unity has inspired me, your guidance has supported me, your presence has given me strength. I am not scared of what is going to happen in my future because my feet stand on very solid roots.

Dad, your infinite support and encouragement, along with all the knowledge and experience you've passed on to me, have been a guiding light in every aspect of my so far short career. I hope to achieve even just half of what you've accomplished over all these years.

Finally, I address my heartfelt appreciation to my friends and colleagues, to Giusy and Angela, who provided me with valuable insights and with whom I shared much of the work in this thesis.

Contents

List of Tables

List of Figures

List of Publications

As a result of the research that has been carried out within these three years of Ph.D. (2020-2023) , several publications were produced.
Some of them, all published in international scientific journals, constitute the core of this dissertation and are listed below.
Papers from I to IV have already been published in international scientific journals, while paper V, at the time of writing this thesis, is still under review for publication in an international scientific journal but has been already published and presented at the 10th Convention of the European Acoustics Association.

Paper I: Puglisi GE, Di Berardino F, Montuschi C, Sellami F, **Albera A**, Zanetti D, Albera R, Astolfi A, Kollmeier B, Warzybok A. *Evaluation of Italian Simplified Matrix Test for Speech-Recognition Measurements in Noise.* Audiology Research 2021;11:73-88.

Paper II: Cardella A, Ottaviani F, Luzi L, **Albera A**, Schindler A, Mozzanica F. *Daily speaking time and voice intensity before and after hearing aids rehabilitation in adult patients with hearing loss.* Folia Phoniatrica et Logopaedica 2023. [Epub Ahead of Print].

Paper III: **Albera A**, Puglisi GE, Astolfi A, Riva G, Cassandro C, Mozzanica F, Canale A. *Ambulatory Phonation Monitoring in Prelingual and Postlingual Deaf Patients after Cochlear Implantation*. Audiol Neurootol. 2022 Oct 4:1-11. DOI: 10.1159/000526936.

Paper IV: Astolfi A, Riente F, **Albera A,** Shtrepi L, Scopece L, Albera R, Masoero M. *Speech quality improvement of TV-sets for hearing-impaired older adult*s. IEEE Transactions on Broadcasting. 2023. DOI: 10.1109/TBC.2023.3254150

Paper V: Guastamacchia A, Puglisi GE, **Albera A**, Shtrepi L, Riente F, Masoero MC, Astolfi A. *Audiovisual recording and reproduction of ecological acoustical scenes for hearing research: a case study with high reverberation*. European Acoustics Association 2023:1765-1772. DOI:10.61782/fa.2023.0666

Moreover, throughout these three years I had the opportunity of contributing to several other works in the field of medical and surgical sciences, both with my research group and with multiple other researchers all over Italy and Europe, that allowed for the following publications:

- Urbanelli A, Nitro L, Pipolo C, Maccari A, **Albera A**, Fadda GL, Felisati G, Albera A, Pecorari G, Fuccillo E, Saibene AM. *Therapeutic approaches to sinonasal NUT carcinoma: a systematic review*. Eur Arch Otorhinolaryngol. 2024 Feb 8. doi: 10.1007/s00405-024-08489-0. [Epub Ahead of Print].

- Casani AP, Albera R, Piras C, **Albera A**, Noto A, Ducci, N, Atzori L, Lucisano S, Mussap M, Fanos V. *Clinical Efficacy and Metabolomics Modifications Induced by Polyphenol Compound Supplementation in the Treatment of Residual Dizziness following Semont Maneuver in Benign Paroxysmal Positional Vertigo (BPPV) of the Posterior Semicircular Canal (PSC): Preliminary Results*. Metabolites 2024, *14*,86. https://doi.org/10.3390/ metabo14020086. [Epub Ahead of Print].
- Canale A, Urbanelli A, Albera R, Gragnano M, Bordino V, Riva G, Sportoletti Baduel E, **Albera A**. *Binaural hearing in monaural conductive or mixed hearing loss fitted with unilateral Bonebridge*. Acta Otorhinolaryngol Ital. 2023. [Epub Ahead of Print].
- Gaffuri M, Di Lullo AM, Trecca EMC, Russo G, Molinari G, Russo FY, **Albera A**, Mannelli G, Ralli M, Turri-Zanoni M. *High-definition 3D exoscope in pediatric otorhinolaryngology: a systematic literature review*. J. Clin. Med. [Epub Ahead of Print].
- Riva G, Pecorari G, Motatto GM, Rivero M, Canale A, Albera R, **Albera A.** *Validation and reliability of the Italian version of the Self-reported Mini Olfactory Questionnaire (Self-MOQ)*. Acta Otorhinolaryngol Ital. 2023. [Epub Ahead of Print].
- Cardella A, Ottaviani F, Luzi L, **Albera A**, Schindler A, Mozzanica F. *Daily speaking time and voice intensity before and after hearing aids rehabilitation in adult patients with hearing loss.* Folia Phoniatrica et Logopaedica 2023. [Epub Ahead of Print].
- Canzi P, Berrettini S, **Albera A**, Barbara M, Bruschini L, Canale A, Carlotto E, Covelli E, Cuda D, Dispenza F, Falcioni M, Forli F, Franchella S, Gaini L, Gallina S, Laborai A, Lapenna R, Lazzerini F, Malpede S, Mandalà M, Minervini D, Pasanisi E, Ricci G, Viberti F, Zanetti D, Zanoletti E, Benazzo M. *Current trends on subtotal petrosectomy with cochlear implantation in recalcitrant chronic middle ear disorders*. Acta Otorhinolaryngol Ital. 2023 Apr;43(Suppl. 1):S67-S75. DOI: 10.14639/0392-100X-suppl.1-43-2023-09.
- Canale A, Urbanelli A, Gragnano M, Bordino V, **Albera A**. *Comparison of active bone conduction hearing implant systems in unilateral and bilateral conductive or mixed hearing loss*. Brain Sci. 2023;13:1150. DOI: 10.3390/brainsci13081150
- Giunta A, Candelori F, Zambonini G, Catalano S, Cassandro C, **Albera A**, Scarpa A, Viola P, Cortese A, Ricciardiello F, Cassandro E. *Spontaneous sphenoid mucocele in 6-year-old child.* Otorhinolaryngology 2023;73(1):18-21. DOI: 10.23736/S2724-6302.21.02399-9
- Trecca EMC, Gaffuri M, Molinari G, Russo FY, Turri-Zanoni M, **Albera A**, Di Lullo AM, Russo G, Mannelli G, Ralli M. *Impact of COVID-19 pandemic on pediatric otolaryngology: a nationwide study*. Acta Otorhinolaryngol Ital. 2023 Jul 28. doi: 10.14639/0392-100X-N2452 [Epub Ahead of Print].
- Astolfi A, Riente F, **Albera A**, Shtrepi L, Scopece L, Albera R, Masoero M. *Speech quality improvement of TV-sets for hearing-impaired older adults*. IEEE Transactions on Broadcasting. 2023. DOI: 10.1109/TBC.2023.3254150
- Cardella A, Riva G, Preti A, **Albera A**, Luzi L, Albera R, Cadei D, Motatto GM, Omenetti F, Pecorari G, Ottaviani F, Mozzanica F. *Italian version of the brief Questionnaire of Olfactory Disorders (brief-IT-QOD).* Acta Otorhinolaryngol Ital 2023;43:252-261. DOI: 10.14639/0392-100X-N2212.

- Canale A, Ndrev D, Macocco F, Albera R, Aschero G, Lovallo S, Gragnano M, Scozzari G, **Albera A**. *Binaural hearing using the ADHEAR bone conduction system in the monaurally occluded ear*. Audiol Neurootol. 2023 Feb 7:1-9. DOI: 10.1159/000528765.

- Ayache S, **Albera A**. *Transcanal Endoscopic Ear Surgery for Cholesteatoma*. B-ENT. 2022;229462022. DOI:10.5152/B-ENT.2022.22946.

- Burdo S, **Albera A**, Cantore I, Chiarella G, Di Nardo W, Girotto G, Maiolino L, Mancini P, Manfrin M, Martini A, Quaranta N, Ricci G, Vincenzi V. *Stato attuale delle Politiche Sanitarie Italiane sulla Sordità - Tavolo Epidemiologia.* Argomenti di Acta Otorhinolaryngologica Italica 2022;16(1):77-92.

- **Albera A**, Puglisi GE, Astolfi A, Riva G, Cassandro C, Mozzanica F, Canale A. *Ambulatory Phonation Monitoring in Prelingual and Postlingual Deaf Patients after Cochlear Implantation*. Audiol Neurootol. 2022 Oct 4:1-11. DOI: 10.1159/000526936.

- Canale A, Ndrev D, Sapino S, Bianchi C, Bordino V, **Albera A.** *Speech in Noise With Bilateral Active Bone Conduction Implant for Conductive and Mixed Hearing Loss*. Otol Neurotol. 2022 Oct 1;43(9):1000-1004. DOI: 10.1097/MAO.0000000000003671.

- Riva G, Lorenzi A, Borello A, **Albera A,** Canale A, Pecorari G. *Transoral approach to parotid tumors: a review of the literature*. Curr Oncol. 2022;29:9416-27. DOI: 10.3390/curroncol29120740.

- **Albera A**, Lucisano S, Spadola Bisetti M, Albera R. *Effectiveness of Bacopa Monnieri in the Therapy of Vertigo in association with Citicoline, Ginger, Vitamin B6 and Passionflower*. Sch J Oto. 2022;9(1):968-76. DOI: 10.32474/SJO.2022.09.000303.

- Fornaro G, Armeni P, **Albera A**, Barbara M. *The value of hearing aids for the Italian NHS: a cost utility analysis.* Otology & Neurotology Open 2022;2(4):e018. DOI: 10.1097/ONO.0000000000000018.

- Beatrice F, **Albera A**, Rossi Mason J. *Can you do without risk reduction in the fight against smoking?* J Community Med Public Health Care 2022;9:119. DOI: 10.24966/CMPH-1978/1000119.

- Perottino F, Di Furia D, Elia G, **Albera A**, Dalmasso G, Dumas G. *Otoplasty in a prominent pediatric series with Kaye's modified technique.* Otorhinolaryngology 2022;72(4):166-71. DOI: 10.23736/S2724-6302.22.02455-0

- Beatrice F, **Albera A**. *No smoke centers in Italy: Critical Issues & Perspectives*. J Community Med Public Health Care 2022;9:117 DOI: 10.24966/CMPH-1978/10000.

- Noto A, Piras C, Atzori L, Mussap M, **Albera A**, Albera R, Casani AP, Capobianco S, Fanos V. *Metabolomics in Otorhinolaryngology.* Front. Mol. Biosci. 2022;9:934311. DOI: 10.3389/fmolb.2022.934311.

- Salvago P, Immordino A, Plescia F, Mucia M, **Albera A**, Martines F. *Risk factors for sensorineural hearing loss and auditory maturation in children admitted to neonatal intensive care units: who recovered?* Children 2022;9,1375. DOI: 10.3390/ children9091375

- **Albera A**, Parandero E, Andriani R, Albera R, Riva G, Canale A. *Prognostic factors influencing postoperative air-bone gap in stapes surgery*. Acta Otorhinolaryngol Ital 2022;42:1-8. DOI: 10.14639/0392-100X-N0612.
- Pecorari G, Riva G, **Albera A**, Cravero E, Fassone E, Canale A, Albera R. *Post-operative infections in head and neck cancer surgery: risk factors for different infection sites*. J. Clin. Med. 2022;11:4696. DOI: 10.3390/ jcm11174969.
- Canale A, Dalmasso G, Albera R, Lucisano S, Dumas G, Perottino F, **Albera A**. *Control of disabling vertigo in Menière's disease following cochlear implantation without labyrinthectomy*. Audiol Res. 2022;12:393-403. DOI: 10.3390/audiolres12040040.
- Troisi D, De Luca P, Cassandro C, Ralli M, Viola P, **Albera A**, Gioacchini FM, De Campora L, Tassone D, Scarpa A. *Usefulness of auditory brainstem response in Cornelia de lange Syndrome.* Otorhinolaryngology 2022;72(2):93-5. DOI: 10.23736/S2724-6302.21.02396-3.
- Cassandro C, Ralli M, De Luca P, **Albera A**, Aschero G, Lovallo S, Landi V, Vernero I, Sammarco D, Manassero A, Scarpa A. *Minimal hearing loss in children: effect on speech in noise perception.* Otorhinolaryngology 2022;72(2):59-63. DOI: 10.23736/S2724-6302.21.02397-5.
- Inguscio BMS, Mancini P, Greco A, Nicastri M, Giallini I, Leone CA, Grassia R, Di Nardo W, Di Cesare T, Rossi F, Canale A, **Albera A**, Giorgi A, Malerba P, Babiloni F, Cartocci G. *Musical effort and musical pleasantness: a pilot study on the neurophysiological correlates of classical music listening in adults normal hearing and unilateral cochlear implant users.* Hear. Balance Commun. 2022. DOI: 10.1080/21695717.2022.2079325.
- Dumas G, Fabre C, Perottino F, Perrin P, **Albera A**, Schmerber S. *The skull-vibration-induced nystagmus test in 10 points: our experience and a review of the literature*. Otorhinolaryngology 2022;72(1):31-9.
- Zucca M, **Albera A**, Albera R, Montuschi C, Della Gatta B, Canale A, Rainero I. *Cochlear Implant results in older adults with post-lingual deafness: the role of "top-down" neurocognitive mechanisms.* Int J Environ Res Public Health 2022;19:1343. DOI: 10.3390/ijerph19031343.
- Riva G, Tavassoli M, Cravero E, Moresco M, **Albera A**, Canale A, Pecorari G. *Long-term evaluation of nasal polyposis recurrence: A focus on multiple relapses and nasal cytology.* Am J Otolaryngol. 2022 Mar-Apr;43(2):103325. DOI: 10.1016/j.amjoto.2021.103325.
- De Luca P, Cassandro C, Cavaliere M, Gioacchini FM, **Albera A**, Ralli M, De Vincentis M, Cassandro E, Scarpa A. *Pediatric adenoidectomy: where we are? A comparative review between cold curettage and emerging techniques.* Otorhinolaryngology 2021; 71(4):273-7.
- **Albera A,** Canale A, Boldreghini M, Lucisano S, Riva G, Albera R. *Contralateral delayed endolymphatic hydrops: clinical features and long term outcome.* J Otol 2021 Oct;16(4):205-9. DOI: 10.1016/j.joto.2021.02.003.
- Capaccio P, Cammarota R, Riva G, **Albera A**, Albera R, Pecorari G. *Transoral robotic surgery for bilateral parenchymal submandibular stones: the Flex Robotic System*. B-ENT 2021;17(1):45-8. DOI: 10.5152/B-ENT.2021.20190.

- Cassandro C, Manassero A, Landi V, Aschero G, Lovallo S, **Albera A**, Genovese E, Canale A. *Auditory processing disorders: diagnostic and therapeutic challenge*. Otorhinolaryngology 2021;71(3):120-4.
- Canale A, Macocco F, Ndrev D, Gabella G, Scozzari G, Albera R, Pecorari G, **Albera A**. *Cochlear Implant Outcomes in Prelingually Deafened Adults with and without Sound Deprivation: Are There Differences in Quality of Life?* Med Sci Monit. 2021 May 18;27:e93. DOI: 10.12659/MSM.930232.
- Bombaci A, Lazzaro C, Bertoli CA, Lacilla M, Ndrev D, Chiò A, **Albera A**, Calvo A, Canale A. *Stapedial Reflex: A Possible Novel Biomarker of Early Bulbar Involvement in Amyotrophic Lateral Sclerosis Patients.* Audiol Neurootol. 2021 Apr;13:1-8.
- Cavagnetto D, Carossa M, Deregibus A, Lacilla M, **Albera A**, Ceruti P, Carossa S. *Efficacy of transcutaneous electrical nervous stimulation in patients with somatosensory tinnitus and cervicofacial myalgia*. J Biol Regul Homeost Agents. 2021 Mar-Apr;35(2):751-6.
- Bruno G, Della Gatta B, Canale A, Albera R, **Albera A**. *Two challenging surgical cases of cochlear implantation under local anesthesia.* Otorhinolaryngology. 2021;71(2):112-25.
- Puglisi GE, Di Berardino F, Montuschi C, Sellami F, **Albera A**, Zanetti D, Albera R, Astolfi A, Kollmeier B, Warzybok A. *Evaluation of Italian Simplified Matrix Test for Speech-Recognition Measurements in Noise.* Audiology Research 2021;11:73-88.
- **Albera A**, Lucisano S, Cassandro C, Fantino C, Albera R, Canale A. *Hearing loss in Ménière disease.* Otorhinolaryngology 2020;70:128-31.
- **Albera A**, Lucisano S, Albera R, Canale A. *Evolution of symptoms in Ménière's disease*. Otorhinolaryngology 2020;70:117-20.
- Canale A, **Albera A**, Macocco F, Caranzano F, Albera R. *Microdrill stapedotomy for otosclerosis with small and large preoperative air-bone gap: a retrospective comparison of results.* Acta Otolaryngol. 2020;140(9):745-8.

Introduction

1.1 Brief overview of the Ph.D. research

In an era where the elderly population is rapidly expanding globally, understanding and addressing the multifaceted consequences of age-related health conditions is paramount. Among these conditions, hearing loss stands out as a prevalent and often underrecognized ailment affecting a substantial portion of the elderly population. The consequences of hearing loss in older adults extend far beyond the only impairment of auditory perception; they encompass various aspects of physical, cognitive, emotional, and social well-being. Despite its significance, hearing loss in the elderly remains an area deserving of further comprehensive investigation and attention.

This thesis aims to delve into the intricate web of consequences associated with hearing loss and to explore diverse perspectives of the nuanced interplay between deafness and its repercussions on various domains of individuals' lives.

Through this thesis we will undertake a multi-stage journey to understand the relationship between auditory perception, speech production, and communicative effectiveness, particularly in challenging acoustic scenarios.

The first part of the research activity within these three years was addressed to the investigation of speech intelligibility. At this aim, the validation of a new speech intelligibility test for pediatric population and the elderly that is accurate for repeated measurements and that was optimized in several languages based on the same algorithm, was performed. This test was validated within a joint project with the Oldenburg Universität (Germany) and Università degli Studi di Milano and, afterwards, was implemented to finalize a listening test aimed at understanding the effect of reverberation and noise with informational content on speech intelligibility.

The second part of the research was oriented to the investigation into speech modifications in cases of hearing loss aims to provide a holistic understanding of the complex factors influencing speech adaptation in individuals with varying degrees and of hearing impairment and the consequent application of hearing aid (HA) or cochlear implant (CI) for enhancing communication outcomes.

Finally, the last part of the research project aimed to explore the implications of the hearing loss in the elderly during common daily activities that pose acoustic challenges, such as listening the television, and to investigate the ability to communicate effectively in noisy, reverberant, and frequently attended acoustic environments.

The in-field audio-video recordings necessary for the development of our speech intelligibility test were conducted in the conference room of the Egyptian museum of Turin, and the experiments were then performed in a semi-anechoic room of the Audio Space Lab at Politecnico di Torino. Within this framework, the calibration procedures for the accurate estimation and measurement of speech in noise in case of deafness were defined, allowing us to develop and validate an innovative laboratory set-up and an ecological protocol for hearing device users which in the future we hope can also be replicated in the clinic for diagnostic purposes.

1.2 Social impact of the hearing loss

Deafness is a very frequent disabling condition that has a strong impact on the individual's quality of life: in terms of frequency, it represents the third most disabling pathology after rheumatic diseases and headaches.
Knowledge of the epidemiological data that characterize hearing loss has not only a scientific value, but also a socio-economic one considering the effects it causes on individuals, not only from a communicative point of view, but also cognitive and relational, without forgetting the financial implications, given the cost of the technology needed to treat the problem in most patients.
The social impact of hearing loss transcends mere audiological impairment, permeating various facets of an individual's personal and social spheres. Beyond the physical challenges of diminished auditory perception, hearing loss can profoundly shape an individual's social interactions, relationships, and overall well-being. At its core, hearing loss poses significant barriers to effective communication, undermining the fundamental human need for connection and belonging. In social settings, the inability to fully participate in conversations due to difficulties in understanding speech can lead to feelings of isolation, alienation, and exclusion. As a result, individuals with hearing loss may withdraw from social gatherings, avoid public spaces, and experience heightened levels of stress and anxiety in interpersonal interactions. Moreover, the impact of hearing loss extends beyond the individual, affecting familial dynamics, friendships, and professional relationships. Communication breakdowns may strain personal connections, leading to misunderstandings, frustration, and emotional distance between loved ones. In educational and occupational settings, untreated hearing loss can impede academic performance, limit career opportunities, and hinder professional advancement, perpetuating cycles of socioeconomic disadvantage. Furthermore, societal attitudes and misconceptions surrounding hearing loss can exacerbate the social challenges faced by affected individuals, fostering stigma, discrimination, and marginalization. Despite advancements in assistive technology and rehabilitation services, many individuals with hearing loss continue to encounter barriers to full societal participation, including limited access to communication support, inadequate accommodations, and insufficient awareness of their rights and needs.
From an epidemiological point of view, the method of collection and the choice of numerical cut-off data between normality and hearing loss play a decisive role for the correct discussion and the practical consequences of what is disclosed.
Data collection can be performed in three ways:

- on subjective declarations collected through questionnaires,
- with audiometric tests,
- with both subjective declarations and audiometric tests.

The World Health Organization (WHO) prefers the use of audiometric tests since subjective evaluations may be often affected by personal variables, leading many people to declare themselves as hearing impaired and many hearing impaired people to declare themselves as normal hearing.
The reason for this apparent contradiction is to be found in the absence of well-defined reference levels of normality in everyday life, for which everyone creates their own opinions on the concept of hearing normality.
Thus, the hearing impaired person who would not declare his or her hearing loss (leaving aside those who voluntarily hide it just for economic of working purposes) can be justified by the fact that, for example, he or she does not frequently encounter communication situations that are demanding for his or her lifestyle habits or uses recruitment as compensation for threshold drift.
When talking about audiometric tests, the choice of the decibel cut-off between normality and hearing loss is crucial.

Hearing loss according to the WHO, can be divided into 4 different degrees as highlighted in Table 1.

Grade of impairment	Corresponding audiometric ISO value[a,b]	Performance	Recommendations	Comments added to the previous classification
0: no impairment	25 dB or better	No or very slight hearing problems. Able to hear whispers	None	20 dB also recommended. People with 15 – 20 dB levels may experience hearing problems. People with unilateral hearing losses may experience hearing problems even if better ear normal
1: slight impairment	26–40 dB	Able to hear and repeat words spoken in normal voice at 1 m	Counselling. Hearing aids may be needed	Some difficulty in hearing but can usually hear normal level of conversation
2: moderate impairment	41–60 dB	Able to hear and repeat words using raised voice at 1 m	Hearing aids usually recommended	None
3: severe impairment	61–80 dB	Able to hear some words when shouted into better ear	Hearing aids needed. If no hearing aids available, lip-reading should be taught	Discrepancies between pure-tone thresholds and speech discrimination score should be noted
4: profound impairment including deafness	81 dB or greater	Unable to hear and understand even a shouted voice	Hearing aids may help in understanding words. Additional rehabilitation needed. Lip-reading and sometimes signing essential	Spoken speech distorted, the degree depending on the age at which hearing was lost

Table 1: Grades of hearing impairment and corresponding audiometric thresholds according to the World Health Organization (WHO). Pure-tone audiometric thresholds are expressed in decibels (dB).

In addition to the previously mentioned degrees of hearing loss, according to the WHO, communicative disability caused by deafness has traditionally been described for losses greater than 35 dBHL. However, unexpectedly, hearing losses between 20 and 34 dB have been considered to cause "great difficulty" in the latest “2021 World Report on Hearing”. This statement actually contradicts what was stated in previous reports and what emerge in clinical practice since all experts consider a PTA4 (average dB evaluated at 0.5, 1. 2 and 4 KHz) of 35 dB as the start of disabling hearing: this is also justified by the fact that presbycusis is almost always a cochlear-based hearing loss and, with recruitment, a slight threshold shift in the presence of supraliminal stimuli such as conversational voice can be compensated for.

In the present thesis, the value of 35 dB will be maintained as the cut off that separates hearing people from those who have difficulty comfortably audible speech.

In 2019 The Lancet published numerical data on the percentage of various degrees of deafness in Europe, reporting a prevalence of hearing loss of 3.5% among the entire population, which in Italy is reflected in approximately 2,100,000 hearing impaired people.
Since the HA is unfortunately a rehabilitation tool still not provided free of charge by the Italian National Health System, it is infrequent for companies that market HA to publish reports on their sales: the only original data reported by a "Centro Studi Investimenti Sociali – CENSIS" survey on HAs indicates that around 2,000,000 of HAs are used in Italy, which, in reality, corresponds more or less to the same number calculated using the percentages of 3.5% of real hearing disability.
According to the prevalence of the hearing loss, most hearing impaired people experience losses below 65 dB. It should be also underlined that about 70% of the total hearing impaired is represented by people with a mild deficit and among them that the HA adoption rate is very low.
As far as age is concerned, it is common knowledge that most hearing impaired people are over 65 years of age: Italy has the highest percentages of the 65+ population (23.1% of the population) and the second highest prevalence of self-reported hearing loss (12.2% of the population) all over Europe.
The percentage of individuals with hearing loss naturally increases progressively with advancing age, and for almost all age groups, it is slightly higher in males than females. The prevalence of hearing impairment approximately doubles with each decade of age. Consequently, in Italy, bilateral mild hearing loss is found in around 14.1% of individuals, moderate hearing loss in 3.6%, severe hearing loss in 0.4%, and profound hearing loss in 0.4 per thousand individuals.

The exploration of all the consequences of the hearing loss is not merely academic but holds profound implications for healthcare professionals, policymakers, caregivers, and society at large. By gaining a deeper understanding of hearing loss in the elderly, stakeholders can develop more effective strategies for early detection, intervention, rehabilitation, and support. A research study like the one addressed in this thesis seeks to foster greater awareness, recognition, and advocacy for the importance of optimizing acoustic conditions to facilitate effective communication and promote social inclusion in our increasingly noise-filled world.
With this knowledge we can advocate for greater awareness, inform the design of person-centered approaches to elderly care that prioritize the preservation of auditory health and we endeavor to pave the way for the design of more inclusive and accessible spaces, where individuals can communicate with clarity and confidence, regardless of the noise and distractions that surround them.

1.3 Objectives

The need of identifying practical strategies to maximize the diagnosis of hearing loss and its repercussions on speech and intelligibility in different everyday hearing contexts has been pointed out due to its tangible impact on society.
To contribute to the design of these strategies, this Ph.D. work has been addressed to five main aspects:

1. To validate an effective test that guarantees an accurate assessment of speech intelligibility in noise both in pediatric settings and in the evaluation of elderly subjects, which can be applied for both diagnostic and research purposes and which can be used in common clinical practice.
2. To investigate on the efficacy of HAs in the treatment of hearing loss and its impacts on daily communication and voice production.
3. To evaluate the impact of the CI in modifying phonatory parameters in adult patients with severe to profound deafness, both prelingual and postlingual, and with different environmental noise conditions.
4. To improve the listening experience of hearing-impaired older adults when using common flat screen televisions with built-in loudspeakers. In particular, to define a transfer function that dynamically amplifies certain frequencies in real time in order to increase intelligibility and improve the perceived quality of listening.
5. To investigate speech intelligibility of both normal hearing listeners and hearing-impaired elderly people in noisy reverberant environments. In particular, to develop an innovative laboratory setup where to present an intelligibility test with in-field audio-video recordings, which could subsequently be replicated and used routinely in the clinical activity of our hospitals and audiology centers.

1.4 Approach

All the research questions that were introduced throughout the work were of course in some ways connected. They were all oriented to define tools and procedures for the enhancement of listening and speech communication in the elderly with hearing loss.
To deepen all the aspect defined in previous paragraph, several experiments both in laboratory and in clinic were carried out in these three years. Moreover, in-field audio-video recordings were also conducted: for this reason, we extend our heartfelt gratitude to the Egyptian Museum of Turin for their generous provision of the conference room for conducting our recordings. We deeply appreciate the museum's kindness and support, which played a pivotal role in the success of our project. Their commitment to fostering academic and scientific pursuits is commendable and has made a meaningful impact on our work.
The long-standing collaboration with the Oldenburg University (Germany), and with the Physical Medicine Department of the Carl von Ossietzky Universität, has allowed us to work on the previous Matrix Test project and therefore to develop

and provide to the scientific community a new similar tool for evaluating intelligibility, specifically adapted for the pediatric and elderly population.
Thanks to the kind availability of the researchers of the Acoustic Division of the Department of Energy of the Politecnico di Torino, I had the opportunity of performing experiments in the semi-anechoic chamber called “Audio Space Lab” and to investigate on the effect of very different acoustic conditions on intelligibility and on speech production. Moreover, the generous cooperation of RAI - Radio Televisione Italiana, and the RAI - Centre for Research and Technological Innovation, has allowed us to carry on the part of the project dedicated to the implementation of the audio listening quality of TV programs on flat screen televisions in case of hearing loss. Finally, all the experiments regarding speech production and its alterations in hearing-impaired subjects who use HAs or CIs were conducted at the Audiology clinics of the Città della Salute e della Scienza di Torino Hospital - Department of Surgical Sciences of the University of Turin.
After acquiring all the experimental data and acoustic measurements, the statistical analysis was performed by means of traditional as well as advanced methods. This approach, which was supervised by experts in the statistical field, allowed for the identification of trends and relationships that were aimed at significantly improving the existing knowledge.

Focus on paper I

Title: Evaluation of Italian Simplified Matrix Test for Speech-Recognition Measurements in Noise

Authors: Puglisi GE, Di Berardino F, Montuschi C, Sellami F, **Albera A**, Zanetti D, Albera R, Astolfi A, Kollmeier B, Warzybok A.

Published on: Audiology Research – 11(1): 73-88 (2021)

Objective: To investigate speech perception in noise and the development of a simplified version of the Italian Matrix Sentence Test (SiIMax) for speech-recognition measurements in noise for both adults (elderly) and children.

Materials and methods: *Ad-hoc* designed experiments in laboratory. Speech tests with adults and children were conducted to examine the training effect and to establish reference speech-recognition thresholds of 50% (SRT50) and 80% (SRT80) correct responses.

Results: Test equivalence of the test lists and test-specific slope of intelligibility were assessed confirming the efficacy of the test. Mean SRT50 values for adults and children were obtained to set a reference for different ages. High test-retest reliability was also demonstrated.

Conclusion: The simplified Matrix Test is suitable for accurate and reliable speech-recognition measurements.

2.1 Speech perception in noise

Daily listening rarely occurs in silence. In recreational and work environments, as well as in educational settings, we are increasingly immersed in environments with background noise. The auditory process, specifically the journey from hearing a word or phrase to understanding it, is fascinating and complex, involving the auditory system.

Verbal language originates from the combination of different phonemes, which consist of sounds and noises. Sounds correspond to vowels, while consonants also include a portion of noise generated by articulation at the level of the vocal tract, composed of the pharyngeal, oral, and nasal-sinus cavities. The combination of phonemes to structure words follows the phonological rules of the language under consideration. The auditory system constitutes the sensory-perceptual apparatus of the communication loop and includes two components: a peripheral one consisting of the outer-middle-inner ear and the auditory nerve, and a central component consisting of the dorsal and ventral cochlear nucleus, the superior olivary complex, the lateral lemniscus nucleus, the inferior colliculus, the medial geniculate body, auditory cortical areas, associative areas, and nerve pathways connecting the various nuclei of the central auditory pathway. The vibratory energy contained in phonemes is transferred to the hair cells of the organ of Corti through the mechanical vibratory energy transmission system consisting of the pinna, the ear canal, the tympanic membrane, the ossicular chain, the oval and round windows, the endolymphatic fluids, and the membranes of the inner ear. The organ of Corti is responsible for transducing mechanical energy into nerve energy, which is then transferred, through the auditory nerve, to the central auditory pathway. It is the processing of the signal by all these central stations, associated with interaction with other associative areas and the auditory connectome, and with top-down modulation mechanisms, that allows the phenomenon of perception, i.e., the mechanism leading to conscious understanding of the input signal.

Speech perception in noise is one of the most challenging tasks for the auditory system. From the above, it can be easily understood how the central processing of the signal transduced by the organ of Corti is a determining factor for the understanding of verbal language in difficult listening conditions. There is evidence that professional musicians, compared to non-musicians, present greater auditory processing of speech at the neural level, and therefore a better auditory ability in noise, greater ability to perceive the intonation and melodic pattern of speech (Krause & Chandrasekaran, 2010). Furthermore, understanding in noise is a task that certainly requires significant cognitive effort. It is instrumentally demonstrated that the cognitive effort required when exposed to degraded signals includes domains of working memory and attention; this confirms that it is not only an auditory function, but also cognitive operations involving tasks not exclusively linguistic. It follows that verbal perception in noise may vary depending on age, experience, and central perceptual abilities (Peelle, 2017). These observations justify the noticeable differences among individuals, with the same hearing threshold. Such variations, to which reference is made, are within the normal range.

However, in examining verbal perception in noise, the role of the auditory periphery should not be overlooked. Word comprehension requires fine spectral and temporal analysis capabilities of the inputs by the auditory system, and this operation begins at the level of the auditory periphery. The ear breaks down the complex sounds that make up verbal language into different sinusoidal frequency components and allows hearing one sound in the presence of another sound (Giraud, et al., 1997). The ability to distinguish frequency changes is called frequency discrimination and should not be confused with frequency resolution, which is instead the ability to extract a sound with a certain frequency from other background sounds.

We can affirm that frequency discrimination and resolution are based on the coexistence and interaction of both peripheral and central mechanisms. Already in the transfer of acoustic energy from the outer ear to the hair cell, there is a rough modulation of the signal, but the first station where the real spectral and temporal analysis occurs is the inner ear. Correct peripheral analysis stimulates the central nervous system in an orderly manner, which, based on this discriminative analysis, carries out frequency resolution, allowing for the understanding of language even in the presence of a signal lower than background noise.

It can be inferred that the ways in which the human auditory system perceives language are complex and involve the intervention of interacting physiological and psychological mechanisms.

First, at the level of the middle ear, there occurs an initial active high-pass frequency filter that contributes minimally to the comprehension of complex sounds in noise. Then, the first spectral analysis occurs at the cochlear level and is due to the structural and mechanical properties of the basilar membrane. In the cochlea indeed, the pressure wave generated by a sound propagates along the basilar membrane from the base to the apex, and the vibration reaches its maximum amplitude at a precise point, closer to the apex as the frequency becomes lower. However, the discrimination for frequencies around 1,000 Hz is on the order of 3 per thousand at a level of 40 dB above the threshold. This excellent discriminatory capacity is crucial for comprehension in noise, allowing the distinction between a sound at 1,000 Hz and one at 1,003 Hz or 997 Hz. This frequency difference should correspond to a spatial difference on the basilar membrane of a few tens of microns.

Such first frequency analysis obtained is transferred to first-order neurons, whose response selectivity is therefore determined by cochlear mechanisms.

At the level of the auditory nerve and auditory centers, processed acoustic messages are distributed based on their frequency content. Spatial-based tonotopy alone (place locking) could not justify the maintenance of such selectivity. It has been demonstrated that the fine structure of action potentials, during a sound period, carries frequency information, and that there is synchronization between the input sound's carrier frequency and the firing frequency of the nerve action potential (Shamma, 1985). This frequency coding is called *phase locking* and is present for sounds with a medium-low frequency. For frequencies higher than

3,000 Hz, there cannot be synchronization of nerve action potentials with the fine temporal structure of the pressure wave to be encoded because the neurons' refractory periods are too long (Peterson & Heil, 2020). For low frequencies, however, the presence of the action potential tends to exhibit a sinusoidal oscillation that reproduces that of the sound. This synchrony provides complementary frequency information to the principle of spatial tonotopy, first defined as *place locking*. These two modalities coexist up to 3,000 Hz, although it is not clear which code is predominantly used by the auditory centers. However, in noisy conditions, it seems that the temporal aspect may be the only one preserved due to the impossibility of spatial coding.

One certainty is that frequency discrimination is much finer at low frequencies, where the two mechanisms coexist, while it significantly reduces at high frequencies.

Regarding frequency resolution, two sounds very close in frequency are perceived as a single sound; increasing the distance, there is a fluctuation in loudness called beating. This phenomenon has a physical basis, as the sum of the two phases alternately results in reinforcement and cancellation. When the frequency separation further increases, the beats are no longer perceived, but an unpleasant sound is experienced until the separation of the two tones is achieved. The frequency range for perceiving two tones, presented simultaneously, as distinct varies depending on the frequency range. Just like frequency discrimination, in frequency resolution, performance is better at lower frequencies and worse at higher frequencies. At 500 Hz, the separation is about 35 Hz, while at 5,000 Hz, it is about 700 Hz (McDermott, Keebler, Micheyl, & Oxenham, 2010). Even in the case of frequency resolution, the best abilities at low frequencies are due to frequency coding, not only place locked but also phase locked.

In conclusion, what happens when there are two sounds entering at different frequencies, for example, a signal and noise, that reach the cochlea simultaneously? Obviously, one component can mask the other.

One of the mechanisms is theorized by Fletcher and called "occupied line", according to which neurons involved in signal coding but connected to an area of the cochlea capable of responding to noise are, for this reason, found to respond to noise when the signal appears. A physiological overlap of critical bands, resulting in reduced frequency resolution, occurs in the perception of high-intensity acoustic stimuli. For stimuli with intensities greater than 60-70 dB, the greater oscillation amplitude of the basilar membrane leads to an increase in cochlear excitation patterns, in terms of discharge frequency, with an increase in the number of activated fibers (Fletcher, 2021). It is widely established that the sensation of sound power is encoded by the total number of action potentials conveyed by the auditory nerve. Recruiting fibers adjacent to those with the characteristic frequency under examination results, on one hand, in an increase in loudness and, on the other hand, in a widening of tuning curves and critical bands at the brainstem level.

Until now, we have analyzed auditory perception considering the auditory system as a monaural system. Physiological hearing is a binaural function, and this characteristic does not only have a summative function.

The previously described phase synchrony of action potentials at low frequencies, with the frequency of the incoming signal, plays a fundamental role in auditory perception in noise, not only for better frequency discrimination and resolution but also because it is a fundamental prerequisite for comparison by the central nervous system of signals from the two ears. Indeed, another significant factor for verbal comprehension in noise is *binaurality*.

There are three reasons underlying the better comprehension in noise due to hearing with two ears:

- Head diffraction (head shadow effect)
- The squelch effect.
- Binaural redundancy.

The above mechanisms, also important for sound source localization, are essential perceptually to effectively separate noise sources from useful signals.

Head diffraction is a purely physical phenomenon, independent of central nervous system processing, but it is the fundamental prerequisite for the second mechanism, namely the squelch effect, to occur. This physical phenomenon occurs because the position of the ears on either side of the head necessarily results in a different signal-to-noise ratio in the individual organs, depending on the angle of incidence of the signal and noise.

In summary, if the sound source and noise come from two different points, among the different signals reaching the two ears, there will be differences in intensity, time, and therefore phase. At low frequencies, the interaural intensity difference due to head diffraction is negligible, especially when the wavelength is greater than the diameter of the head (a wavelength of 20 centimeters corresponds to a cutoff frequency of 1,500 Hz). For frequencies below 3,000 Hz, the phase synchrony of action potentials allows the detection, at the level of the middle part of the superior olivary complex, of subtle time differences. Conversely, at high frequencies, phase synchrony of action potentials is insufficient for the interaural time difference to be conveyed in the nerve. In these frequency regions, the interaural intensity difference becomes predominant, analyzed in the lateral part of the superior olivary complex. The frequency must be above 2,000 Hz for the interaural intensity difference to be at least 1 dB, the smallest intensity difference discriminable by the auditory system according to Weber's law (Olsen & Carhart, 1975).

The central processing of these differences optimizes the separation of the useful signal from the background noise and is called the *squelch effect*.

Without going into detail, during central processing the phases of the different signals are compared, and the noise from the ear with a lower signal-to-noise ratio is used to partially remove the noise from the ear with a better signal-to-noise ratio. There is a kind of imperfect subtraction of the waveforms that arrive bilaterally, achieving a suppression (squelch) of noise (Bernstein, Schuchman, & Rivera, 2017). Another test confirming binaural squelch can be performed by stimulating one ear with a signal while simultaneously sending noise from the same side with successive intensity increments; there will be a point at which the signal will no longer be perceived because it is masked. If noise of equal intensity

is sent to the contralateral ear, the subject will once again perceive the signal. This is the simplest condition, but different levels can be obtained depending on the various combinations with variations in incidence angles. The amount of noise suppression, which occurs when the signal has a phase/time and intensity difference compared to the noise, is called the binaural masking level difference (BMLD). It is quite intuitive that the extent of binaural masking varies depending on the different incidence angles, and that noise suppression increases as the angle of incidence between signal and noise increases from 0 to 90 degrees.
Lastly, the binaural contribution to understanding speech in noisy environments is partly due to *binaural redundancy*. This phenomenon refers to the increase in loudness resulting from hearing with two ears even in cases of identical combinations of signal and noise on both sides. It is binaural summation and results in a 1-2 dB improvement in the signal-to-noise ratio (Ellen Peng & Litovsky, 2021).
One last aspect to consider in this overview of the physiology of verbal perception in noise is the role of *cognitive abilities*. While auditory abilities, with sound analysis at both peripheral and central levels, remain the strongest predictor for language comprehension accuracy, there is growing evidence of how cognitive aspects play an important role. Very simple behavioral observations highlight the need for cognitive effort in perception in noisy contexts. For instance, if a subject is asked to listen to a series of words, they will be memorized to a lesser extent when heard in a noisy context rather than a silent one. The auditory memory engaged in the difficult listening context is subtracted from that needed for storing the words heard. Other physiological observations have found that during auditory effort, there is pupil dilation and an increase in stress hormones. Neuroimaging techniques document activation of areas not strictly connected to auditory and linguistic aspects during listening that requires effort for intelligibility in a noisy context. In conditions of auditory effort, increases in neuronal activity compared to baseline are observed at the level of the cingulate operculum, which has a general performance monitoring function, at the level of the premotor cortex for auditory memory, and at the fronto-parietal level for attention (Peelle, 2018).

2.2 Speech intelligibility tests

Tests related to speech intelligibility have significant meaning in audiology as they contribute to defining the functional profile of the patient. At the same time, they are useful for assessing the effectiveness of prosthetic-rehabilitative treatments implemented, although their use is desirable in the follow-up of patients treated pharmacologically or undergoing surgical intervention. It should be noted that no verbal perception test can be considered perfect and even the most sophisticated laboratory simulation cannot fully express the complexity of real listening situations. Additionally, verbal perception is strongly influenced by cognitive factors as well as the quality of auditory input. However, the choice of verbal material and the tasks required of the patient cannot be individually tailored. Furthermore, test construction, administration methodology, and

calculation of indicators significantly influence the outcome of the investigations, which can yield highly variable results even in the same patient population.
For these reasons, the selection of tests to be included in clinical and scientific protocols requires an awareness of these limitations and pragmatic approaches. The first step is the clear definition of the assessment objectives. If an indicator synthesizing the patient's perceptual difficulty is desired, also for counseling purposes, a methodology extracting the percentage of intelligibility for a specific signal-to-noise ratio will be chosen. Conversely, to assess the effectiveness of a modification in a particular prosthetic setting, more precise and adaptive methodologies are required, with the calculation of the perception threshold (SRT) in the presence of noise. For pediatric evaluations, appropriate material for this age group, congruent administration modalities, etc., must be selected. If the goal is to compare with other contexts, tests with proven replicability that minimize the language effect should be used.
Speech audiometry is an essential tool for assessing the social impairment of hearing-impaired individuals: it allows determining word intelligibility and its recognition as a meaningful auditory stimulus, a complex function linked to numerous individual factors such as intellectual ability, cultural level, deterioration of nerve pathways and central integration, etc. While the perception of a pure tone is a function solely related to the auditory sensory modality, hearing and understanding a word, as well as recognizing its meaning, are consequences of a much more complex process involving both the ear as an anatomical-physiological unit and the psychological sphere aimed at recognizing and classifying the word itself. Vocal audiometry thus becomes an indispensable investigation for interpreting the social disability of the hearing-impaired individual. At the same time, it can provide invaluable data in the differential diagnosis of various degrees of hearing impairment, in the evaluation and adaptation of HA, in "pre-post" cochlear implantation evaluations, in anticipation of otosurgical interventions, and in audio-phoniatric rehabilitation.
In Italy, current clinical audiological standards still mainly rely on detecting pure-tone auditory thresholds and on vocal examinations based on mono- and disyllabic words, optimized to examine intelligibility in quiet environments. In the 1950s, the first vocal material in the Italian language adaptable for presentation with noise was organically developed. This consists of disyllabic words proposed by Bocca and Pellegrini in 1950 (Bocca & Pellegrini, 1950) and the sentences by Pietrantoni, Bocca, and Agazzi (Pietrantoni, Bocca, & Agazzi, 1956). The disyllabic words are phonetically balanced, and both words and sentences are grouped into lists of 10 items each. Recordings, performed by professional male speakers, are available on various media for clinical application and include material for competition (white noise and cocktail party noise). This material has been the diagnostic standard for over half a century and is still used in some clinics. However, this approach ignores the evaluation of the patient's ability to listen and understand in more complex acoustic situations, such as in the presence of background noise, sound reverberation, or multiple speakers.
A substantial innovation in investigative methodologies occurred with the introduction of the speech material proposed by Cutugno, Prosser, and Turrini in

2000 (Cutugno, Prosser, & Turrini, 2000). This was the most extensive and flexible material for vocal audiometry in the Italian language since the development of the Matrix test. Predominantly, in consisted in a new set of disyllabic words and this material was prepared for adaptive testing, although there was no software support to automate the investigation and facilitate the spread of this methodology in our country. The need to structure new disyllabic words (it is worth noting that there are no monosyllabic words of complete sense in the Italian language) arose from at least two reasons. The first is the obsolescence of the Bocca-Pellegrini disyllabic words: almost 30% of the words in these lists, in fact, occupied a position higher than 10,000th in the order of frequency of modern Italian, and some words appeared completely unusual. The second reason is phonetic balance, based only on the orthographic transcriptions of phonemes and not on phonemic ones, resulting in inadequate valorization of differences such as those between affricates and fricatives, open and closed vowels, etc.

Furthermore, historically used speech audiometric tests are typically tied to a single language, a limitation that restricts their applicability and makes it impossible to coherently estimate the harmful effects of noise on intelligibility across different languages due to differing design criteria.

Finally, most speech audiometric tests are only available in an open response format, requiring the involvement of an examiner who shares the language with the examinee and can identify and mark the patient's correct responses.

Fortunately, cochlear implantation has made it urgent to have reliable tests for use in different linguistic contexts for comparison purposes.

Therefore, there was a need to develop a new diagnostic test that was objective and reliable in assessing the intelligibility of a vocal signal in the presence of background noise. In this context, the development of the German Oldenburg Sentence Test (OLSA) (Wagener, Kühnel, & Kollmeier, 1999), subsequently adapted and validated in various foreign languages including Italian, under the name of Italian Matrix Test, fits perfectly (Puglisi, et al., 2015). It consists of an adaptive speech audiometry examination with competitive noise and allows evaluating the average discrimination threshold in noise (SRT), intended as the signal-to-noise ratio (SNR) relative to 50% intelligibility.

2.3 Introduction of the study

The Matrix Test, an adaptive speech in noise audiometric test, was developed in 1999 by Wagener et al. (Wagener, Kühnel, & Kollmeier, 1999) but the "matrix" principle was first proposed in 1982 by Hagerman for the Swedish language (Hagerman, 1982).

The test consists of matrix of a 50-word base matrix (10 proper names, 10 verbs, 10 numbers, 10 adjectives, and 10 nouns). Starting from the "base matrix," ten five-word sentences are automatically formulated by the software using a random combination of one word from each group (e.g., “Luca mangia quattro tavole belle”).

Names	Verbs	Numerals	Nouns	Adjectives
Sofia	compra (buys)	due (two)	scatole (boxes)	azzure (light blue)
Marco	vuole (wants)	poche (few)	matite (pencils)	piccolo (small)
Anna	prende (takes)	quattro (four)	tazze (mugs)	normali (normal)
Sara	dipinge (paints)	cinque (five)	pietre (stones)	nuove (new)
Chiara	vede (sees)	molte (many)	tavole (desks)	belle (nice)
Maria	cerca (looks for)	sette (seven)	palle (balls)	bianche (white)
Luca	trascina (drags)	otto (eight)	machine (cars)	grandi (big)
Andrea	regala (donates)	nove (nine)	sedie (chairs)	utili (useful)
Matteo	possiede (owns)	dieci (ten)	bottiglie (bottles)	nere (black)
Simone	manda (sends)	venti (twenty)	porte (doors)	rosse (red)

Figure 1: 50-word matrix of the Italian matrix sentence test.

The resulting sentences are grammatically correct but the content and semantic meaning are unpredictable, so as not to provide subjects with clues for easy memorization of the sentences and to ensure the possibility of subjecting the same patient to the test multiple times without compromising its validity. The use of 5 words for each sentence and a list of 10 different sentences has therefore proven to be a good compromise between the need to obtain a high number of words to ensure high measurement accuracy and a good representation of underlying linguistic properties, and the need not to evaluate the patient's memory capacity. Indeed, sentences that are too long would be more indicative of an estimate of the patient's individual memory than a measure of speech intelligibility. Moreover, it aims to overcome linguistic restrictions as it is available in both closed set and open set response formats. The closed response version allows self-administration of the test, presenting the examinee with possible response alternatives on a keyboard, touch screen, or any other electronic device: this bypasses the need for the patient to repeat the vocal material to the examiner. This opportunity, together with the validation of the Matrix Test in various languages, makes it a reliable test to administer to the patient in their native language even if the examiner does not understand it. This type of test has been designed to provide high repeatability of results with the same subject, characteristics that make it particularly effective for standardizing the examination in an international context.

Although all the audiometric tests previously described for the adult population are also routinely offered to children in clinical practice, they have not been studied and adapted specifically for this population and are therefore unreliable. This concept also applies to the Matrix test: sometimes it can be challenging to administer it to children, especially if they are younger than 9-10 years old and hearing-impaired, not only due to an incomplete perception of speech but also due to the low level of attention that the subject would be able to maintain for the duration of the entire test. We can also find the same difficulty in completely carrying out a normal test for adults in the elderly population, especially in elderly hearing-impaired subjects with initial cognitive decline and short-term memory problems, who often only remember the last 2-3 words of the phrase they just heard.

The early detection of hearing loss in children, appropriate treatment and rehabilitation are crucial for the development of speech, language, reading and cognitive abilities and the speech test must be a sensitive measure that objectively specific aspects of children's speech perception abilities. This can be obtained only if the principles of psychometric theory are considered and applied by the development of speech materials. The cognitive and attentional demands should be age appropriate. Furthermore, performance should be independent of higher-level language abilities and vocabulary knowledge. The speech material must be suitable for measurements with different populations of children, e.g., with varying degrees of hearing impairment, and of different ages and developmental abilities (Mendel, 2008).

For this reason, a simplified version of the Matrix Test has been developed (SiIMax), consisting of a selection of 21 words taken from the same matrix of the 50 words of the normal test, which are used to form a randomized list of 14 sentences of three-word each (a numeral, a noun and an adjective).

Name	Verb	Numeral	Noun	Adjective	English Translation
Sofia	compra	***due***	scatole	***azzurre***	Sofia buys two light-blue boxes.
Marco	vuole	poche	***matite***	***piccole***	Marco wants a few small pencils.
Anna	prende	***quattro***	tazze	normali	Anna takes four normal cups.
Sara	dipinge	***cinque***	***pietre***	nuove	Sara paints five new stones.
Chiara	vede	molte	***tavole***	***belle***	Chiara sees many nice desks.
Maria	cerca	***sette***	***palle***	***bianche***	Maria looks for seven white balls.
Luca	trascina	***otto***	***macchine***	***grandi***	Luca drags eight big cars.
Andrea	regala	***nove***	***sedie***	utili	Andrea donates nine useful chairs.
Matteo	possiede	***dieci***	bottiglie	***nere***	Matteo owns ten black bottles.
Simone	manda	venti	***porte***	***rosse***	Simone sends twenty red doors.

Figure 2: Base matrix of ITAMatrix test and selected words (italics/bold with a gray background) for the simplified ITAMatrix.

The simplified version of the Matrix Test significantly shortens the duration of the test and greatly facilitates the child or the elder who must repeat the phrase by focusing attention only on three words instead of five. Moreover, the limited number of speech items combined with the closed-set response format makes the test also interesting for remote testing or even for self-test applications via tablets, computers, or mobile telephones where only a few response alternatives can be displayed.

The simplified Matrix test was initially developed for normal hearing children only for German (Rosbach, 2017) and Russian (Garbaruk, et al., 2020), but nobody had translated and validated the Italian version yet.

The objectives of this study were multiple:

- To investigate the training effect;
- To test the equivalence of the test lists for normal-hearing listeners;
- To establish reference data related to SRT at 50% and 80% recognition for normal-hearing adults and children;
- To assess the test-retest reliability for both groups

In the present study, which led to the publication of the aforementioned article in the literature, my contribution mainly consisted in the investigation process, particularly in the execution of the experiments and in the collection of data in the Turin hospital. Furthermore, I participated in the literature review and manuscript writing.

2.4 Materials and methods

The phoneme distribution of the 21 words of the matrix was compared with the phoneme distribution of the original version of the ITAMatrix test (Puglisi, et al., 2015) and with a reference phoneme distribution of the Italian language taken from Tonelli (Tonelli, Panzeri, & Fabbro, 1998). The three phoneme distributions are shown in Figure 3.

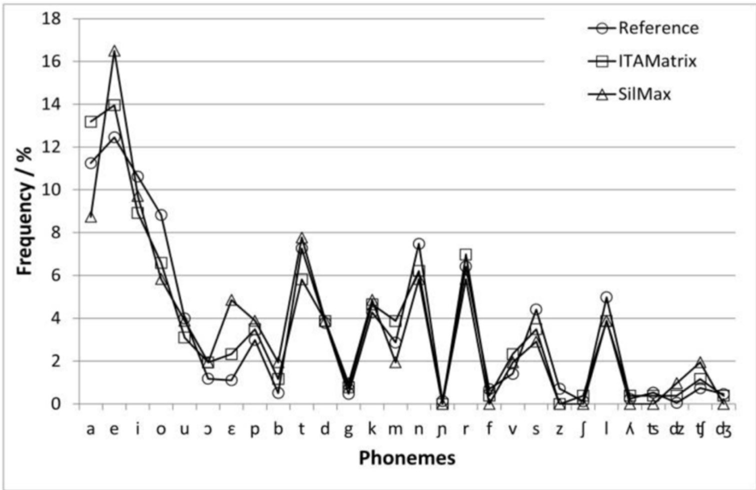


Figure 3: Phoneme distribution of the ITAMatrix (squares), SiIMax (triangles) and reference distribution for the Italian language (circles). The phonemes were transcribed using the symbols of the International Phonetic Alphabet.

Similarly to the study on the original ITAMatrix test (Puglisi, et al., 2015), singleton and geminate consonants were summarized as one phoneme class. The phoneme distribution resulted to follow the phoneme distribution of the ITAMatrix speech material and the language-specific distribution. The highest difference between the SiIMax and the reference distribution is observed for the phonemes /e/ (Δ = 4.1%) and /ɛ/ (Δ = 3.7%). These two phonemes are overrepresented in the SiIMax. The reason for this is related to the intrinsic structure of the speech phrases: since only female nouns in plural are used, all of them, as well as all adjectives (with the only exception of the adjective “grandi” - Italian term for big), end with the phoneme /e/. Comparing the phoneme distributions of the ITAMatrix and SiIMax, the highest difference of $\Delta = 4.4\%$ is observed for the phoneme /a/. In the speech material of the ITAMatrix, this phoneme is mainly included in the names and verbs. Since these word categories are not included in the speech material of the SiIMax, a drop in the occurrence of this phoneme can be noticed.

In choosing the appropriate words for the SiIMax, we considered only those familiar to school-aged children. In the category of numerals, only digits in the range between two to ten were selected, avoiding abstract words such as “pochi” (a few) or “molti” (many). Nouns were chosen based on the results obtained previously from adults during the development of the original version of the ITAMatrix. Many listeners indeed systematically confused certain nouns (e.g., “tazze” - cups, with “calze” - socks) and retaining such words might have been misleading for very young children. Out of the original list of 10 adjectives, three were omitted in the SiIMax, as they were the least commonly used among all. To validate the appropriateness of the final 3 × 7 matrix of words, experts in the field of pediatric audiology were engaged to perform a preliminary validation process. Additionally, the number of syllables within each word group was regulated to ensure uniformity, with all words having the same number of syllables (for example, all numerals were disyllabic) or a balanced distribution of syllables (for example, three nouns were trisyllabic and four were disyllabic).

From the 21-word matrix, ten test lists were generated randomly, each comprising 14 sentences such as "two new boxes” and in each test list, every word appeared exactly twice. The audio .wav files were identical to those utilized in evaluating the original ITAMatrix test (Puglisi, et al., 2015). For each word, seven realizations were employed to maintain coarticulation with the subsequent word within the speech phrase: these seven realizations captured the coarticulation between the given word and all possible subsequent words.

To assess all the characteristics of the SiIMax test, such as the Speech Reception Threshold (SRT), discrimination function slope, equivalence of test lists, size of the training effect, and test-retest reliability, measurements were conducted with two groups: normal-hearing adult listeners and normal-hearing children aged 5 to 10 years, divided into three age subgroups. The examination of list equivalence required lengthy evaluations with at least two different measurements per test list, and hence, was conducted solely with adult listeners. The literature indicates comparable results regarding test list equivalence between children and adults (Wagener & Kollmeier, 2005): therefore, outcomes of test list equivalence from measurements obtained with adults were assumed to be valid for children as well. A shorter protocol than adults was employed with children to maintain the evaluation time below one hour in total. Indeed, the duration of the test is crucial to obtain an objective result in children due to the ease with which they lose attention towards what they listen to, with the risk of obtaining much worse values in the tests carried out last.

All measurements were performed both with adults and with children:

Listeners - Adults: Twenty native Italian adult listeners (12 female, 8 male) with normal hearing participated in the experiments. Their ages ranged from 21 to 36 years, with a mean age of 24.5 years. The tests took approximately 2 hours per subject and were split into two sessions to allow for breaks and avoid cognitive fatigue or lose of attention.

Procedure - Adults: The tests were all conducted in a sound-attenuated booth in the Otolaryngology Unit of the Città della Salute e della Scienza Hospital in Turin. The setup included a notebook with an earbox "ear 3.0" sound card (Auritec, Germany) and Sennheiser HDA200 headphones. Speech recognition measurements were performed using the OMA - Oldenburg Measurement Applications software (HörTech GmbH, Germany), the same used for the regular test. Appropriate calibration was done to dB SPL using Brüel&Kjær instruments and monaural measurements were conducted at the listener's preferred ear.
Of course, speech intelligibility depends not only on the emission intensity of the speech material but also on the noise level. This relationship is normally expressed in decibels (dB) with a specific value defined by the signal-to-noise ratio (SNR). The diagnostic use of competitive noise allows the evaluation of the subject's ability to separate useful information (signal) from background noise and provides an accurate assessment of prosthetic results (traditional HAs, implantable devices, CIs, etc.) compared to reference values.
Speech-shaped stationary noise was used as a masker in this case, with an intensity of the noise fixed at 65 dB SPL and an adaptive speech signal level. The task assigned to the listeners was to repeat every word they understood, with the examiner marking all the correctly recognized words on a display. The responses were recorded using word scoring.
Training began with a fixed SNR of +5 dB (comfortable intensity, which allows to be correctly understood by almost all normal hearing listeners), followed by three adaptive test lists (T1, T2, and T3): the software, the same as the normal ITAMatrix test, presented the 3-words sentences with a specific loudness using a 1-up/1-down adaptive procedure with a varying step size, converging to an 80% speech recognition threshold (SRT80) (Brand & Kollmeier, 2002). The percentage of intelligibility achievable with masking noise that can be set in the software varies between 20% and 80% depending on the desired threshold, and as an adaptive test, it automatically sets the SNR level based on the correct/incorrect responses provided in the previous sentence by the patient. Another test list was further presented to establish a reference value for the 50% speech recognition rate (SRT50) and compare it with the original version of the ITAMatrix test for adults. The SRTs were estimated by fitting a logistic function to the measured scores. Subsequently, the equivalence of test lists was evaluated by measuring SRT20 and SRT80 for each listener with all ten lists.

Listeners - Children: 96 normal-hearing children (pure-tone thresholds not exceeding 20 dBHL at all octave-audiometric frequencies of 125 to 8000 Hz) were recruited from clinics (44 in Turin and 52 in Milan): among them, 51 were

female and 45 were male. In Turin, only children without articulation dysfunctions and normal scores in the Fisher checklist participated in the speech-recognition measurements. Articulation dysfunctions were assessed by a speech therapist using the Italian phonemic examination protocol (Vernero, Stefanin, Gambino, & Schindler, 1998), and the Fisher checklist was completed by parents to evaluate attention, auditory-visual integration, comprehension, figure-ground perception, and memory.
Contrarily, in Milan no specific test was conducted, but a careful conversation has been made with parents, who reported normal language development and no issues with comprehension, attention, or memory for their children.
The children's ages ranged from 5 to 10 years, and to consider the age effect on outcomes, they were divided into three groups as reported in Table 2. Each child attended the measurements once, with every session lasting about 30 minutes, including a break to avoid possible inattention and fatigue as well.

Group	Age Range	Number of Listeners in Turin	Number of Listeners in Milan	Sex (F—Female, M—Male)
1	5–6	15	15	21 F, 9 M
2	7–8	13	18	17 F, 14 M
3	9–10	16	19	13 F, 22 M
Total	5–10	44	52	51 F, 45 M

Table 2: Recruited listener sample for the measurements with children.

Procedure - Children: Similarly to adults, tests with children were conducted in sound-attenuated booths meeting specified requirements at two tertiary-care Italian University Hospitals: the Otorhinolaryngology Unit of the "A.O.U. Città della Salute e della Scienza," of Turin and the Audiology Unit of the "Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico," of Milan. As data were collected in two different clinics, a comparison between outcomes was previously performed to evaluate any effect of site where the test was conducted.
The measurement setup was identical to that used for adults. A compatible setup (notebook with an earbox "ear 3.0" sound card and Sennheiser HDA200 headphones) was prepared and calibrated. The OMA software was employed for speech-recognition measurements. Like measurements with adults, children were tested monaurally at their preferred ear, with the task of verbally repeating at loud voice understood words for the examiner to mark the correctly recognized words on a display.
Children were initially acquainted with the test through one test list presented at a fixed high SNR of 5 dB. The training effect was then evaluated through three subsequent measurements of the SRT80 (T1, T2, and T3) using an adaptive procedure. Two subsequent adaptive measurements, presented to children in a randomized order, converged to thresholds of 20 and 50% speech recognition to analyze the slope of the test-specific speech-recognition function. A logistic function was fitted to the SRT20 and SRT80 using a maximum likelihood procedure to obtain the test specific SRT50 and slope. In all conditions, the noise signal was fixed at 65 dB SPL and turned on and off 500 ms before and after the presentation of each speech sentence.

2.5 Results

Adults

Mean SRT80 values averaged across listeners and relative standard deviations measured within the first (T1, T2, and T3) and the second (T4 and T5) measurement session are showed in Figure 4. Analysis via repeated-measures ANOVA test on the SRT80s from T1, T2, and T3 revealed no statistically significant differences across adaptively measured thresholds ($F(1.5, 1.2) = 3.0$, $p = 0.76$). The mean SRT80 at T1, T2, and T3 was -4.5 ± 1.1 dB. On average, the SRT80s measured during the first measurement were 0.9 dB higher than those from the second session. This discrepancy proved statistically significant ($F(1, 1.19) = 10.2$, $p = 0.003$).

For assessing test-retest reliability, the SRT80 values from both the first and second sessions (T1, T2, T3, T4, and T5) were considered overall. Test-retest reliability, defined as the root mean square of the within-subject standard deviations of repeated SRT measures was found to be 1.0 dB for the SiIMax in adult listeners. Differently, the mean SRT50s from the first and second measurement sessions were -7.0 ± 0.6 and -7.7 ± 0.9 dB, respectively: the enhancement of speech perception in noise (0.7 dB) between the two session was like that observed for SRT80 and the test-retest reliability of SRT50 was 1.2 dB.

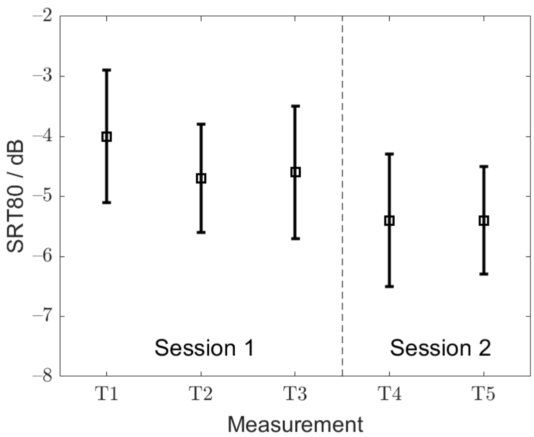


Figure 4: Mean SRT80s with corresponding standard deviations for training session within the 1st (T1, T2 and T3) and the 2nd measurement session (T4 and T5).

To quantitatively investigate the intelligibility test lists' equivalency, repeated-measures ANOVAs were carried out for the SRT20 and SRT80. Test-list equivalency was confirmed by the lack of statistically significant differences in the SRT80 ($F(9, 171) = 1.29$, $p = 0.25$) or SRT20 ($F(9, 171) = 1.35$, $p = 0.22$) across the test lists. The mean SRT20 and SRT80 values were −10.7 ± 1.3 and −5.3 ± 1.2 dB, respectively. A list-specific speech recognition function was fitted using the logistic function based on all the data obtained with the corresponding list (i.e., 40 data points for each list, that is, 20 for the SRT20 and 20 for the SRT80) to produce list-specific parameters, i.e., the SRT50 and slope. Figure 5 displays the list-specific functions as well as the average SRT and slope for each test list. With a standard deviation of 0.2 dB, the average SRT50 for all test lists was -8.0 dB. With a standard deviation of 0.6%/dB, the test-specific slope was 11.3%/dB.

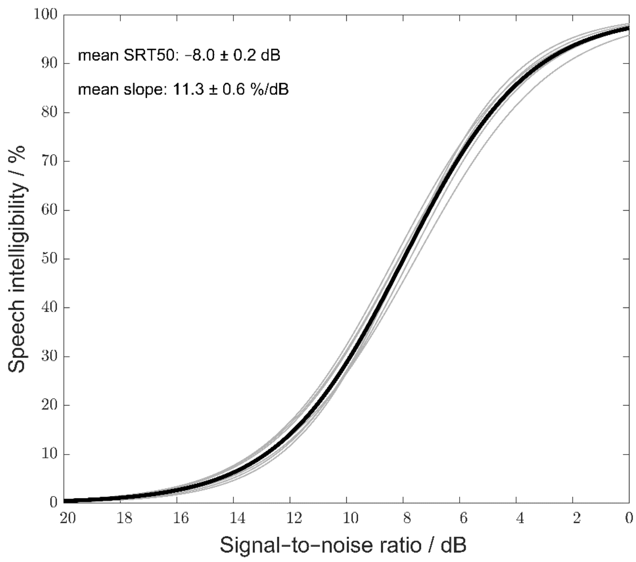


Figure 5: List-specific recognition functions (gray lines) and resulting average recognition function of the test (black line).

Children

For children measured in Milan and Turin, the initial training list delivered at a fixed SNR of 5 dB produced an average speech-recognition score of 100% and 95%, respectively. The median values (data aggregated across locations) for groups of age 1, 2, and 3 were, respectively, 93%, 98%, and 99%. First, the data from each age group at each measurement site were statistically combined. Concerning adaptive SRT80 measurements, nonparametric tests on medians were utilized and there were no statistically significant differences across the measurement sites within any of the age groups ($p = 0.88$, $p = 0.189$, and $p = 0.425$ for groups 1, 2, and 3, respectively). Consequently, the data collected in Turin and Milan were combined for further analysis. Then, a mixed-design ANOVA was conducted on the pooled data, with the training effect considered as a within-subject factor and age group as a between-subject factor. Significant differences in the SRT80s were observed between the age groups ($F(2, 93) = 14.3$, $p < 0.001$) but there was effect of the training list ($F(2, 186) = 0.24$, $p = 0.79$). Additionally, there was no interaction between age groups and training effect ($F(4, 186) = 0.92$, $p = 0.46$). Sidak's corrected post hoc test indicated no significant difference between group 2 (7–8 y.o.) and group 3 (9–10 y.o.) ($p = 0.26$), but significant differences were found between groups 1 (5–6 y.o.) and 2 ($p < 0.01$), and between groups 1 and 3 ($p < 0.001$). Figure 6 displays the mean SRT80s with corresponding standard deviations for each age group and training list. The youngest group of children (group 1) exhibited the highest mean SRT80 of -1.5 ± 2.7 dB, while groups 2 and 3 showed mean SRT80s of -3.0 ± 1.7 and -3.7 ± 1.4 dB, respectively.

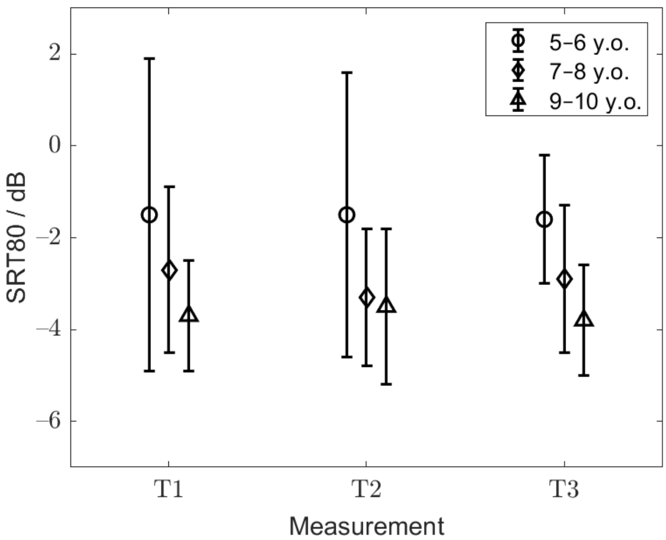


Figure 6: Mean SRT80 with standard deviation for the respective training list (T1–3) and age group.

No significant differences across the measurement sites were found for the SRT50s ($F(1, 94) = 0.46$, $p = 0.5$): therefore, data from both centers were pooled for further analysis. A one-way ANOVA revealed significant differences in the SRT50s across the age groups ($F(2, 94) = 6.6$, $p = 0.002$). Post hoc tests with Sidak's corrections showed significant differences between groups 1 and 3 ($p = 0.003$) as well as between groups 2 and 3 ($p = 0.03$). The mean SRT50s averaged across listeners from groups 1, 2, and 3 were -5.6 ± 1.2, -5.8 ± 1.2, and -6.6 ± 1.3 dB, respectively. Test-retest reliability was separately assessed for each age group from the adaptively measured SRT80 (T1, T2, and T3), resulting in 1.1 dB for groups 1 and 2, and 1.0 dB for group 3.

To characterize the test-specific speech perception function for each age group, a logistic function was fitted considering the SRT20s and SRT80s of all children within the respective age group. The test-specific function was defined by SRT50s and slopes of -5.0 dB and 9.1%/dB for group 1, -6.1 dB and 8.9%/dB for group 2, and -6.5 dB and 10.6%/dB for group 3, respectively.

For each independent age group, the average SRT50s and slopes were as follows: -5.0 ± 1.1 dB and 11.1 ± 3.5 dB for group 1, -6.2 ± 1.5 dB and 11.5 ± 3.3 dB for group 2, and -6.6 ± 1.0 dB and 14.0 ± 4.6 dB for group 3. Significant variations in SRT50 and slope between age groups were found using a two-way ANOVA analysis among the three children's group and the adult group ($F(3, 114) = 27.55$, $p < 0.001$ and $F(3, 114) = 4.16$, $p = 0.008$, respectively). The fitted SRT50 of group 1 significantly differed from those in groups 2 ($p = 0.001$) and 3 ($p < 0.001$), according to post hoc tests with Sidak correction. Between groups 2 and 3, there was no statistically significant change in SRT50 ($p = 0.49$).

Post hoc testing revealed that listeners in group 3 had a statistically steeper slope of the speech-perception function than listeners in groups 1 ($p = 0.01$) and 2 ($p = 0.033$) regarding the "Slope" factor. There was no discernible change in slope between groups 1 and 2 ($p = 0.96$).

2.6 Discussion

Concerning hearing loss in the pediatric population, its prevalence varies depending on factors such as age, geographic location, socioeconomic status, and the definition used for hearing loss. However, estimates suggest that hearing loss is a relatively common condition among children worldwide.

According to the WHO, it is estimated that around 32 million children globally have disabling hearing loss, with approximately 60% of cases being preventable. The prevalence of hearing loss in children varies by region, with higher rates observed in low- and middle-income countries compared to high-income countries (WHO, 2021).

It's important to note that hearing loss can occur at any age during childhood, from infancy to adolescence, and can be congenital (present at birth) or acquired later in childhood due to various factors such as infections, genetic conditions, ototoxic medications, trauma, or exposure to noise.

Early detection through newborn hearing screening programs and regular hearing evaluations during childhood are essential for identifying hearing loss promptly and implementing appropriate interventions to support language development, communication skills, and academic success in children with hearing loss.

In the present study we decided to translate, implement and validate the Italian version of the simplified Matrix Sentence test, a speech in noise test which is already considered the standard for hearing loss diagnosis in clinics.

The measurements conducted with adults demonstrated that the test lists of the SiIMax test exhibit equivalence concerning speech recognition. In other words, the outcomes are not influenced by the specific test list employed, suggesting that the test lists can be used interchangeably. Moreover, the standard deviation of the SRT50 across the test lists (0.2 dB) was notably smaller than across the listeners (0.6 dB), indicating a high level of homogeneity in intelligibility among the test lists. This variation in SRT across the test lists observed with the SiIMax is like findings from other speech tests like the standard version of the ITAMatrix and the German and Russian versions of the simplified Matrix test. Previous studies have shown that no differences in test-list equivalence were found between adults and children, suggesting that the small standard deviation of the SRT50 across the test lists of the SiIMax is valid not only for adults but also for children.

The mean SRT50 obtained with adults from the adaptive measurement, along with the variability across the listeners of the SiIMax (SRT = -7.0 ± 0.6 dB), closely resembled the reference data of the ITAMatrix (SRT = -6.8 ± 0.8 dB), which was also observed for the German language. The steepness of the test-specific slope of the SiIMax was 11.6 ± 0.6 dB%/dB, slightly lower than the test-specific slope of the ITAMatrix (13.3 ± 1.2%/dB). However, the SiIMax still exhibited a steeper slope compared to typical speech tests used in practice, leading to higher precision in SRT estimation for a given number of test items. In terms of test precision, the ITAMatrix is preferable when solely considering test efficiency, while the SiIMax is preferable if other test properties such as a short measurement time and low cognitive load are more important. This preference may be relevant for special patient groups such as children or adults with reduced

auditory memory span or cognitive decline. Similar to the ITAMatrix test, a training effect was found for the SiIMax. However, the training procedure is much shorter for the SiIMax than for the ITAMatrix, as only one training list with 14 speech phrases is needed for the SiIMax compared to two test training lists, each with 20 sentences, for the ITAMatrix.
Regarding test-retest reliability, both the SRT50 and SRT80 resulted in comparable reliabilities of 1.0 and 1.2 dB, respectively, which are slightly higher than the test-retest reliability of the ITAMatrix for SRT50. This slightly decreased reliability could be due to the lower number of speech items in the SiIMax compared to the ITAMatrix. However, the test-retest reliability of the SiIMax for both the SRT50 and SRT80 is low enough to ensure accurate speech-recognition measurements in noise.
In addition to be a valid and reliable tool for speech in noise assessment in the adults, the SiIMax test offers several advantages for audiological diagnostics in children, including a short training session requiring only one test list, and high accuracy in SRT measurement with a test-retest reliability of about 1 dB. This makes it suitable for use even with young children, as demonstrated by our experiments where a single test list was sufficient to familiarize even the youngest children of 5-6 years old with the speech material. The short duration of the test is crucial in clinical practice to minimize children fatigue and lack of concentration while maximizing the number of patients who can be diagnosed. Our multicenter experiments have shown the SiIMax to be highly reliable, a prerequisite for its diagnostic sensitivity in distinguishing between normal and abnormal cases, as well as in monitoring hearing rehabilitation processes. However, further studies are needed with hearing-impaired children to confirm the sensitivity of the SiIMax test.
The simplified Matrix test, such as the SiIMax, can be utilized not only for speech-recognition measurements in noise but also in quiet conditions. Studies have demonstrated its validity as an audiometric test for quantifying speech perception in quiet in children as young as 4 years old, including those with hearing devices (Neumann, et al., 2012).

2.7 Conclusion

The Simplified Italian Matrix test demonstrates highly consistent intelligibility results, with a narrow standard deviation of 0.2 dB across test lists at SRT50, allowing for their interchangeable use in repeated measurements.
For both adults and children, only one training list containing 14 speech items suffices to address the training effect.
With a high test-retest reliability of approximately 1.0 dB for SRT80, the test proves suitable for precise speech-recognition assessments in noisy environments, applicable across age groups. However, it's worth noting that while the Simplified Italian Matrix test offers reliability, the complete Matrix test may offer even higher accuracy per unit of measurement time.

Therefore, the SiIMax should be preferred over the ITAMatrix test in scenarios involving special patient groups, such as children or adults with reduced working memory capacity.

Finally, although the Matrix Test and its simplified version already represent an important added value in modern intelligibility diagnostics, current scientific research should focus on further enhancing the examination itself to recreate an acoustic masking situation, that is as faithful to real life as possible, rather than the stationary noise normally used.

Often the comprehension of speech messages is much disturbed by poor daily acoustic conditions, with background noises or chatter, such as that in a café, subway, theatres, restaurants and so on: a better assessment of a subject's social intelligibility in a more faithful and reliable manner could undoubtedly derive from immersing the subject in a real, ecological environment, in the presence of complex and everyday noises.

If we had the opportunity to subject a hearing-impaired school-aged child to the SiIMax with a real background noise recorded directly in the classroom, we would undoubtedly obtain reliable data regarding the comprehension difficulties encountered by the subject every day. Furthermore, the early identification of a comprehension problem would allow the school to implement measures aimed at protecting the child, such as seating him near the teacher's desk or using proper electronic voice amplification systems.

Focus on paper II

Title: Daily speaking time and voice intensity before and after hearing aid rehabilitation in adult patients with hearing loss.

Authors: Cardella A, Ottaviani F, c Livio Luzi L, **Albera A**, Schindler A, Mozzanica F.

Published on: Pholia Phoniatrica Logopedica – 2023 Nov 30 – Epub ahead of print. doi: 10.1159/000533371

Objective: To analyze trends in the use of voice in case of mild to moderate hearing loss, with remediation by hearing aids.

Materials and methods: All measurements were performed through an ambulatory phonation monitor (APM), which is a portable vocal dosimeter able to evaluate the variations in the vocal production. Subjective Quality of Life (QoL) data was collected through the Speech, Spatial, and Qualities questionnaire (SSQ) and the International Outcome Inventory for Hearing Aids (IOI-HA). Assessments were performed soon before and after rehabilitation with HAs.

Results: Trends in speech performance in a common daily condition were assessed and correlated to subjective data.

Conclusion: The APM has demonstrated its utility as a valuable tool for assessing the effectiveness of HA and can serve as an indicator of an individual's engagement in communication and social interactions.

3.1 The importance of the auditory feedback in voice production

Verbal language originates from the combination of different phonemes, which consist of sounds and noises. Sounds correspond to vowels, while consonants also include a portion of noise generated by articulation at the level of the vocal tract, composed of the pharyngeal, oral, and nasal-sinus cavities.

The concept of auditory feedback in voice production refers to the intricate interplay between the auditory system and the motor control processes involved in speaking. Auditory feedback encompasses the real-time monitoring and processing of acoustic signals generated by one's own vocalizations. It serves as a crucial mechanism for error detection, correction, and regulation of speech output. When individuals speak, they rely on auditory feedback to assess the accuracy and intelligibility of their vocalizations, making adjustments based on perceived discrepancies between intended and actual outcomes. This dynamic feedback loop is essential for maintaining vocal stability, pitch control, and speech fluency. Understanding the mechanisms underlying auditory feedback in voice production is fundamental not only for elucidating the complexities of speech motor control but also for informing clinical interventions aimed at addressing speech disorders and enhancing communication abilities.

Through this feedback loop, speakers continuously adjust their speech production in response to auditory cues, ensuring the accuracy and fluency of their utterances. Studies have shown that alterations or disruptions to auditory feedback, such as delayed auditory feedback or distorted auditory feedback, can significantly impact speech production, leading to disfluencies, alterations in pitch or loudness, and even speech errors. This underscores the importance of auditory feedback mechanisms in maintaining speech fluency and precision (Ubrig, et al., 2019). Moreover, research suggests that individuals with speech disorders, such as stuttering or dysarthria, may exhibit differences in their ability to utilize auditory feedback effectively, highlighting the clinical relevance of understanding the role of auditory feedback in speech production (Fiorin, et al., 2021).

Phonation and articulation are two essential physiological processes involved in speech production, each with distinct roles in shaping sounds of human speech.

<u>Phonation</u> refers to the process of sound production initiated by the vibration of the vocal cords within the larynx. It involves several key steps:

- *Vocal Cord Adduction*: the vocal cords come together, or adduct, closing the glottis, the opening between the vocal cords.
- *Airflow*: Air from the lungs is expelled through the closed glottis, causing the vocal cords to vibrate as the air pressure builds beneath them.
- *Vocal Cord Vibration*: the vibration of the vocal cords creates a buzzing sound, akin to a reed instrument.
- *Pitch Variation:* the frequency of vocal cord vibration determines the pitch of the sound produced. This frequency is controlled by adjusting the tension and length of the vocal cords.

- *Voice Quality*: the quality of the voice is influenced by factors such as the shape of the vocal tract and the resonance characteristics of the vocal cavities.

Articulation involves the precise shaping and movement of the articulators, including the tongue, lips, teeth, and palate, to modify the sound produced by phonation. Key aspects of articulation include:

- *Constriction and Release*: articulators come together or move apart to create specific sounds by constricting or releasing airflow.
- *Place of Articulation:* sounds are classified based on where and how the airflow is restricted in the vocal tract. For example, the /p/ sound is produced by closing the lips, while the /t/ sound is created by tapping the tongue against the alveolar ridge.
- *Manner of Articulation:* sounds are further categorized based on the degree and type of constriction in the vocal tract. For instance, /s/ is a fricative sound produced by creating a narrow channel for airflow, while /m/ is a nasal sound produced by lowering the velum to allow air to pass through the nasal cavity.
- *Coarticulation:* articulators often anticipate and adjust to neighboring sounds, leading to smooth transitions between speech sounds and contributing to the fluency of speech.

Together, phonation and articulation work in concert to produce the rich variety of sounds that form the basis of human speech. These processes involve intricate coordination of the respiratory, laryngeal, and articulatory systems, reflecting the complexity and precision of spoken language production (Kenney & Prather, 1986).

Hearing loss can have significant clinical implications for auditory feedback and voice production and indeed deaf people are more likely to suffer from voice and speech disorders than those with normal hearing. Auditory feedback plays a crucial role in regulating various aspects of speech production, including pitch control, articulatory precision, and speech fluency (Selleck & Sataloff, 2014). When individuals experience hearing loss, it can disrupt the normal functioning of the auditory feedback loop, leading to several potential consequences for voice production:

1. Pitch Control: individuals with hearing loss may have difficulty accurately perceiving the pitch of their own voice due to reduced auditory feedback. As a result, they may struggle to maintain consistent pitch levels during speech production, leading to variations in vocal pitch and intonation patterns. This can affect the naturalness and expressiveness of their speech.
2. Loudness control: normal hearing individuals commonly exhibit robust control of speech loudness and adapt their vocal production to compensate for complex acoustic scenarios, such as in presence of background noise. In such cases, the Lombard effect happens, allowing speakers to raise vocal loudness to be heard and intelligible. In case of hearing loss, the poor auditory feedback mechanisms may determine an increase of loudness

variability as well as problems in managing speech intensities, thus compromising social interactions.

3. Articulatory Precision: hearing loss can also impact articulatory precision, as individuals may have difficulty monitoring the clarity and accuracy of their speech sounds. Without clear auditory feedback, they may produce speech sounds with reduced clarity or precision, leading to difficulties in speech intelligibility.
4. Speech Fluency: auditory feedback plays a role in monitoring the rhythm and timing of speech production. Individuals with hearing loss may experience disruptions in speech fluency, such as hesitations, repetitions, or prolongations, as they struggle to maintain the natural rhythm and pacing of their speech without clear auditory cues.
5. Speech Monitoring and Self-awareness: hearing loss can affect individuals' ability to monitor their own speech and self-correct errors. Without accurate auditory feedback, individuals may have limited awareness of their speech production errors, making it challenging for them to make necessary adjustments to improve their speech clarity and fluency.
6. Psychosocial Impact: the clinical relevance of hearing loss in voice production extends beyond the physical aspects of speech production. It can also have psychosocial consequences, such as reduced self-confidence and social withdrawal, as individuals may feel self-conscious about their speech abilities and communication difficulties.

Given the clinical implications of hearing loss for auditory feedback and speech, it is essential for healthcare professionals, including speech-language pathologists and audiologists, to assess and address auditory feedback deficits in individuals with hearing loss. Rehabilitation strategies may include auditory training, speech therapy and the use of assistive listening devices, such as HAs or CIs, to optimize auditory feedback and support communication and voice production.

Auditory rehabilitation serves to mitigate the detrimental effects of hearing loss, enhancing functionality, activity, participation, and ultimately, improving the quality of life (QoL) for individuals with hearing impairment (WHO, 2021). Furthermore, it has the potential to reverse the adverse impacts of hearing loss on voice production (Selleck & Sataloff, 2014). Numerous studies have examined the influence on voice parameters in patients who have undergone CI procedures, revealing alterations in vocal parameters such as reduced F0 values, overall severity, strain, loudness, and voice instability following this form of treatment (Hamzavi, et al., 2000) (Mozzanica, Schindler, Iacona, & Ottaviani, 2019). Conversely, there was limited information available regarding the impact of HAs on voice production (Lee, Liu, & Lee, 2013), despite HAs being the most used rehabilitation method for mild to moderately severe sensorineural hearing loss. Moreover, most studies investigating the effects of hearing rehabilitation (either through HA or CI) on voice production focus on sustained phonation (Medved, et al., 2021), while there is scarce data available on daily speaking time and daily mean voice intensity.

3.2 Introduction of the study

Hearing loss is a prevalent condition, affecting more than 5% of the global population and over a third of adults over the age of 65. As mentioned before, this condition can significantly impair communication abilities, leading to feelings of loneliness, isolation, frustration, anxiety, depression, and hindered social interaction, especially among older individuals (Gao, Hu, & Yao, 2020). Deafness primarily results in considerable difficulties in comprehending spoken language, particularly in challenging environments such as noisy or multi-speaker settings. Additionally, hearing loss has a substantial impact on cognition and is the most significant potentially modifiable risk factor for age-related dementia (Kivimäki & Singh-Manoux, 2018). Furthermore, since hearing plays a crucial role in regulating voice production through the previously mentioned feedback and feedforward mechanisms, which provide information about vocal targets and allow for adjustments in pitch, volume, and other speech attributes, deafness may also affect various aspects of voice production (Mora, Crippa, Cervoni, Santomauro, & Guastini, 2012).

Regarding the study of speech production in adult subjects with hearing loss, there are indeed very few reports in the literature on this topic. Most of these studies focus solely on the technical aspects of vocal characteristics, without considering broader aspects such as phonation time and changes in voice usage throughout the day, or the impact on quality of life and individual subjectivity.

Recently, various methods for monitoring voice production have become available.

The necessity for accurately assessing the occurrence of voice disorders in clinical settings has led to the development of vocal dosimeters, which are designed to monitor an individual's voice production during their daily activities over extended periods.

A vocal dosimeter is a portable device consisting of a small transducer that is placed at the subject’s jugular notch, located at the anterior part of the neck between the cricoid cartilage and the sternum, to detect the skin acceleration provoked by vocal fold vibration. Typically, the transducer, which can be a contact microphone or an accelerometer, is connected to a data logger that records and/or analyzes the acquired signal.

Several requirements make all existing voice monitoring devices comparable:

- The device must be unobtrusive and wearable to ensure that normal activities or behaviors are not hindered;
- The sensor used to detect the voice signal should have a wide enough bandwidth to capture the entire vocal spectrum;
- The device should facilitate both short-term and long-term monitoring to meet various needs, including those of clinicians for diagnostic assessments (short-term monitoring) and those of voice experts for rehabilitation or continuous monitoring (long-term monitoring);
- The quantities measured or estimated from voice monitoring should be traceable, which necessitates proper calibration procedures. Additionally, their

uncertainty specifications should be declared to ensure accurate and reliable results.

Over the years, numerous authors have dedicated their research efforts to developing monitoring tools capable of detecting vocal activity without being intrusive (Cheyne, Hanson, Genereux, Stevens, & Hillman, 2003) (Švec, Titze, & Popolo, 2005). Notably, researchers from the Massachusetts General Hospital (Boston, U.S.A.) have developed and implemented a widely used vocal dosimeters: the Ambulatory Phonation Monitor (APM model 3,200 by Kay-PENTAX; Lincoln Park, NJ, USA). APM is worn in a waist pack by clients as they go about their normal daily routine and has been developed and commercialized to measure long-term phonation time, average and mode F0, and mean amplitude of voice production over an extended period. The transducer is a small accelerometer (contact microphone - vibrotactile unit) which is adhered to the base of the client's neck and can be hidden by the collar of a shirt or blouse. A cable runs from the accelerometer to the hardware module in the waist pack. The accelerometer senses the vibrations of the skin on the neck that are associated with phonation.

Figure 7: APM model 3,200 by Kay-PENTAX

In the clinic, the APM system is calibrated by clinicians prior to data collection. The patient then leaves the clinic and pursues his daily activities (Cantarella, et al., 2014). After wearing APM over a defined period, the unit is returned to the clinic and data is downloaded to a PC for analysis using APM software.

Data analysis includes both graphic and numeric displays of total phonation time, average fundamental frequency (Hz), and amplitude (dB SPL) values. s. Not only are values for phonation time and SPL reported, but the graphs also indicate when, during the data collection period, the vocalizations occurred. Additional graphical displays (e.g., histograms) reveal important characteristics of the client's phonatory behavior over many hours. It should be noted that APM only collects extracted voice parameters, not actual speech samples.

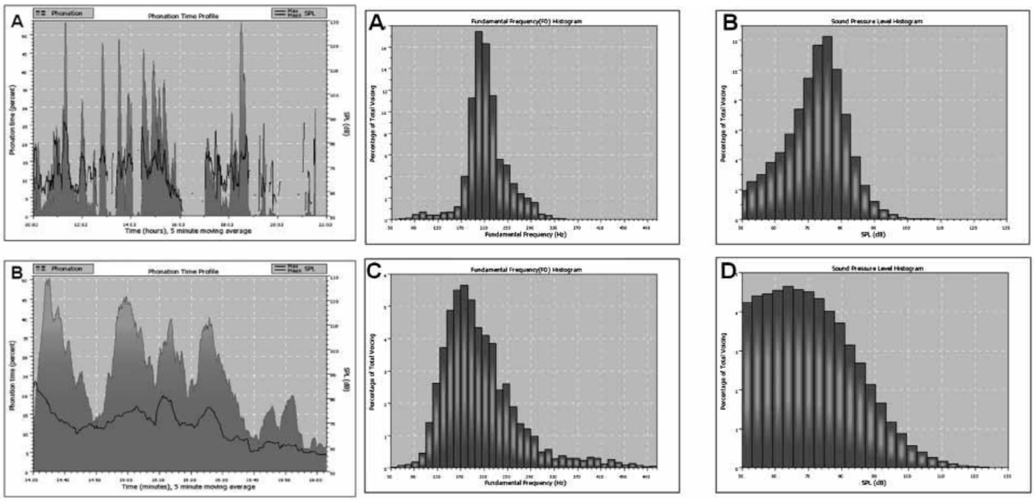


Figure 8: Histograms of phonation time (on the left), fundamental frequency (center) and sound pressure level (on the right) as reported by the software of the APM 3200.

This type of phonation recording has been demonstrated to be relatively insensitive to surrounding sounds and can distinguish between voluntary voice and other behaviors, such as throat clearing or coughing. Additionally, it records the amount of voicing produced but not the actual content of speech, enabling quantification while preserving privacy (Švec, Titze, & Popolo, 2005).

To ensure accurate assessment of vocal intensity, it's essential for each patient to calibrate the sensor properly. The accelerometer's signal can only provide reliable data on sound pressure levels if the calibration process is conducted correctly. During calibration, the subject stands in front of a microphone placed 15 cm away. They're instructed to take a deep breath and sustain the /a/ vowel, gradually increasing volume from soft to their loudest voice. If a patient can't sustain phonation for the entire range in one breath, they're guided to produce the /a/ vowel softly, medium, and loudly for shorter durations, with inhalation between each.

While the average speaking SPL range during calibration is around 35 dB, it can vary between individuals. Once phonation begins, calibration data points appear on the display, along with a straight red line representing the best linear correlation between microphone-recorded sound pressure levels and accelerometer-captured signal amplitude.

The software draws the red line after at least 7 data points, but this doesn't ensure successful calibration. An error message indicates if the drawn line isn't statistically valid based on available points, but it doesn't consider the patient's vocal range. Therefore, the examiner should prompt the patient to continue phonating until they reach their full amplitude range for accurate calibration.

Prior our studies, the APM had never been used in evaluating voice production in patients with hearing loss, treated either with HAs or CIs.

The aim of the first study, reported below, was to objectively evaluate alterations in daily speaking time and voice intensity among a group of adult patients with hearing loss treated with HA, utilizing the APM. Additionally, the study sought to explore the correlation between these changes and QoL measurements.

The underlying hypothesis posits that restoring hearing function through HA may lead to improvements in voice production. Furthermore, it suggests that enhancing the ability to produce voice could potentially influence changes in QoL following HA rehabilitation. Data on phonation time may indirectly reflect an individual's engagement in communication and social interaction, while alterations in voice amplitude post-HA rehabilitation could provide insights into shifts in the vocal behavior of deaf patients.

In the present study, which led to the publication of the aforementioned article in the literature, my contribution mainly consisted in the conceptualization of the project idea, in the recruitment of subjects with hearing loss and wearers of hearing aids, in the collection of part of the data and in the writing part of the manuscript.

3.3 Materials and methods

Performing a longitudinal observational study, we enrolled a group of 26 individuals with moderate sensorineural hearing loss who were treated with bilateral HAs.
The inclusion criteria were as follows: sensorineural postlingual hearing loss, being native Italian speakers, and having no prior history of HA use. Exclusion criteria included: age under 18, significant neurological or head and neck conditions, previous head and neck surgeries, reading limitations of any origin, speech disorders resulting from malformations or acquired damages to the speech organs, motor speech disorders, voice disorders of any origin, difficulties in HA rehabilitation, and associated disabilities.
Every enrolled subject underwent evaluation twice within a 4-month interval. The initial assessment occurred before the commencement of HA rehabilitation, while the second assessment took place after 4 months of HA usage.

The following parameters were recorded:

a) Phonatory Measurements: every subject identified a "typical" day for the measurements, comprising a standard workday with regular social interactions, while avoiding days with exceptional interaction levels such as holidays or celebrations. Before each recording, a sound pressure level (SPL) calibration was conducted using a microphone positioned 15 cm from the subject's mouth. The collected data included:
    - Phonation time (in minutes): duration when the vocal folds were in phonatory vibration;
    - Percentage of phonation time: proportion of time during APM usage when the vocal folds were in phonatory vibration;
    - Average F0 (in Hertz): mean frequency of vocal fold vibration.
    - Average amplitude (in SPL dB): mean energy level of the voice sound wave.

b) Auditory Measurements: Pure tone audiometry was performed before and after HA rehabilitation for each subject. The better-ear average (BEA) and worse-ear average (WEA) hearing thresholds across 500, 1,000, 2,000, and 4,000 kHz were recorded. The functional hearing gain provided by HAs was calculated from the differences between unaided and aided hearing thresholds tested, with assessments conducted in a sound-treated room. HA rehabilitation was deemed suitable when functional gains reached half of the unaided hearing levels or when the aided threshold at 1,000 Hz was less than 35 dB HL.

c) Subjective Evaluation: Baseline self-reported hearing ability upon entering the study (unaided) and after 4 months of HAs use was assessed using the *Speech, Spatial, and Qualities (SSQ) questionnaire*. The SSQ comprises 50 items measuring hearing ability/disability across various complex listening scenarios in everyday life. Scores for speech, spatial, and other auditory functions were obtained, along with a total score. Additionally, subjects completed the *International Outcome Inventory for Hearing Aids* (IOI-HA) independently during the second evaluation. The IOI-HA is a self-assessment questionnaire documenting the patient's perspective on the HA's impact on daily use, satisfaction, limitations, participation, impact on others, and quality of life. The questionnaire consists of seven items with five possible answers, scored from 1 to 5, with higher scores indicating greater satisfaction.

3.4 Results

Phonatory measurements

With a mean age of 71 ± 12.9 years (range: 53–86 years) and a mean duration of deafness of 7 ± 3.8 years (range: 4–10 years), the group of participants consisted of 8 females and 18 males. During the second evaluation, no significant changes in the patients' medical conditions - such as stroke, trauma, metabolic, or cardiologic diseases - that would have an impact on their quality of life or phonatory behavior were noted. It never took longer than five minutes to calibrate the APM and twenty minutes to complete the subjective evaluation questionnaires.

Regarding the assessment of voice production, every patient showed good tolerance to the APM. In the first assessment, the mean time of phonation monitoring was 467 ± 67 min (range: 329–544 min), and in the second assessment, it was 476 ± 69 min (range: 350–574 min). This represents the length of the complete APM evaluation. The Wilcoxon signed rank test revealed no difference in the mean length of data collection before and after 4 months of HAs use ($p = 0.715$).

Tables 3 and 4 present the APM results obtained both before and after 4 months of HA application. At the second evaluation, there was a noticeable increase in the phonation time and percentage of phonation time ($p = 0.002$ and 0.004, respectively, on the Wilcoxon signed rank test). These were the amount of time, expressed in minutes, that the vocal folds were vibrating and the percentage of the entire APM recording that exhibited this behavior. However, there was a discernible drop in the average amplitude ($p = 0.004$ on the Wilcoxon signed rank test).

	Before HA	After HA	p value
Phonation time, minutes	25.6±14.3 (9–62)	40.2±11.9 (15–64)	0.002
Percentage phonation time, %	5.7±3.3 (1.8–11.9)	8.7±2.9 (2.7–14.5)	0.004
Average amplitude, dB SPL	78.9±10.4 (59.1–101.6)	70.5±6.1 (59.6–81.1)	0.004

Table 3: Ambulatory phonation monitoring (APM) results in patients before and after 4 months of HA use.

Lastly, there was no discernible change in the average fundamental frequency (F0) for either gender before or after HA use.

	Before HA	After HA	p value
Average F_0, Hz			
Males	130.1±10.9 (110–154)	132.8±8.8 (112–144)	0.780
Female	212.4±20.1 (171–246)	197.1±21.1 (163–225)	0.773

Table 4: Ambulatory phonation monitoring (APM) results in patients before and after 4 months of HA use, according to the gender.

Auditory measurements

The BEA threshold was 41.1 ± 9.5 dB HL, and the WEA was 52.6 ± 13.4 dB HL prior to HA restoration. Aided average thresholds in the better and worse ears improved significantly to 26.3 ± 6.6 dB HL and 28.7 ± 8.8 dB HL, respectively, after 4 months of using HAs (p = 0.001 and p = 0.001 on Wilcoxon signed rank test). Furthermore, all patients exhibited assisted thresholds at 1 KHz of 35 dB HL or less during HA therapy. All enrolled participants were deemed suitable for HA rehabilitation based on their functional gain.

Subjective evaluation

Table 5 presents the average total SSQ score as well as the Speech, Spatial, and Qualities subscale scores before and after four months of HA use: there was a discernible improvement in the SSQ scores. The average score on the IOI-HA questionnaire, which was completed at the re-evaluation, was 28.8 ± 4.2 (range: 21–34).

SSQ	Before HA	After HA	p value
Speech	5.2±1.5 (2.9–7.3)	6.4±1.3 (3.7–8.7)	0.011
Spatial	5.7±2.1 (2.8–7.8)	6.9±2.1 (4.4–9.2)	0.049
Qualities	6.2±1.6 (3.6–8.6)	7.1±1.5 (5.1–9.1)	0.044
SSQ tot	5.7±1.5 (3.6–7.9)	6.8±1.5 (4.4–8.9)	0.021

Table 5: Speech, Spatial, and Qualities (SSQ) questionnaire scores before and after 4 months of HA use.

Tables 6-8 report the correlation analysis's findings. Regarding the relationship between phonatory measurements and SSQ scores (Table 6), Spearman test showed a weak but statistically significant relationship between phonation time and mean scores of the SSQ Speech subscale, both before and after HA usage. Furthermore, there was a significant correlation between the percentage of phonation time following four months of HA usage and mean scores on the SSQ Speech subscale at the second evaluation.

There were negative correlations ranging from low to very significant between the SSQ scores and auditory measures (Table 7). Specifically, before and after four months of HA use, there was a strong correlation between mean scores on the SSQ Speech and Qualities subscale and the BEA threshold. The aided BEA and WEA thresholds showed a substantial correlation with the mean scores of the SSQ Spatial subscale.
Lastly, a strong correlation between phonatory and auditory measures was discovered. The average amplitude and the BEA threshold prior to HA use showed the strongest association (Table 8).

		SSQ Speech		SSQ Spatial		SSQ Qualities		SSQ total	
		pre	post	pre	post	pre	post	pre	post
Phonation time	Pre	**0.354**	–	0.121	–	0.205	–	0.226	–
	Post	–	**0.393**	–	0.114	–	0.310	–	0.249
Percentage phonation time	Pre	0.259	–	0.196	–	0.155	–	0.190	–
	Post	–	**0.360**	–	0.147	–	0.190	–	0.254
Average amplitude	Pre	−0.141	–	−0.259	–	−0.094	–	−0.168	–
	Post	–	−0.145	–	0.133	–	−0.268	–	−0.189
Average F_0	Pre	−0.339	–	−0.319	–	−0.267		−0.273	–
	Post	–	−0.164	–	−0.224	–	−0.163	–	−0.248

Statistically significant correlations are reported in bold. Pre = before the HA use; Post = after 4 months of HA use.

Table 6: Correlation analysis between the Speech, Spatial, and Qualities (SSQ) questionnaire scores and phonatory measurements before and after HA use.

	Phonation time		Percentage of phonation time		Average amplitude		Average F_0	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
BEA								
Unaided	−0.228	–	0.121	–	**0.375**	–	0.146	–
Aided	–	−0.144	–	0.145	–	0.281	–	0.134
WEA								
Unaided	**−0.360**	–	−0.310	–	0.313	–	0.123	–
Aided	–	**−0.335**	–	−0.288	–	0.263	–	0.111

Table 7: Correlation analysis between Speech, Spatial, and Qualities (SSQ) questionnaire scores and auditory measurements before and after HA use.

	SSQ Speech		SSQ Spatial		SSQ Qualities		SSQ total	
	pre	post	pre	post	pre	post	pre	post
BEA								
Unaided	**−0.360**	–	−0.340	–	**−0.361**	–	−0.343	–
Aided	–	**−0.459**	–	**−0.569**	–	**−0.517**	–	**−0.603**
WEA								
Unaided	−0.318	–	**−0.580**	–	−0.412	–	**−0.488**	–
Aided	–	−0.155	–	**−0.412**	–	−0.197	–	−0.290

Table 8: Correlation analysis between phonatory and auditory measurements before and after HA use.

3.5 Discussion

According to the most recent estimates available from Global Burden of Disease study, age-related hearing loss accounted for 1,460 million cases worldwide in 2019 and was the fourth- ranked cause of years lived with disability in 2019 (WHO, 2021). The dominant cause is presbycusis, defined as the gradual loss of hearing with age. In Italy, it was estimated that almost 7 million people suffer from hearing loss and prevalence tends to increase with aging, with a prevalence of 25% in individuals between 61 and 80 years of age and of 50% in individuals

older than 80 years; as a consequence, prevalence is expected to increase over the next years due to population aging (CENSIS, 2019).
HAs are considered an effective rehabilitation instrument that can improve quality of life and mitigate other detrimental effects of age-related hearing loss, for example, limiting the excess risk for dementia. However, HA uptake is still relatively limited and frequently characterized by underuse or abandonment: potential barriers could be, among others, the persistence of social stigma around HA wearers, costs, inadequate customization of the device.
Concerning the benefits in the use of HAs, despite significant advancements in HA technology, a notable gap persists in our understanding of the daily implications of HA usage on speech production. While numerous studies have explored the efficacy of HAs in improving auditory perception and communication outcomes, little attention has been directed towards comprehensively assessing their impact on speech production dynamics in real-world settings. Moreover, the integration of vocal dosimetry within this context remains unexplored. This gap in research is particularly noteworthy given the fundamental role of speech production in everyday communication and social interactions. Understanding how the use of HAs influences speech production parameters is crucial for optimizing hearing rehabilitation strategies and enhancing overall communication outcomes for individuals with hearing impairment.
In the current study, a group of patients with mild to moderate hearing loss receiving HA treatment had their daily voice production analyzed using the APM for the first time, and its link with quality of life measures was examined. Notably, every recruited individual underwent two full days of phonation monitoring (before and after four months of HA usage), indicating that they all tolerated the monitoring well. Furthermore, the APM calibration process never took longer than five minutes, indicating that the process is quick and that the APM may be easily set up during standard ambulatory visits. There were notable variations between the APM parameters recorded before and after the use of HA. Following HA rehabilitation, mean amplitude fell but phonation time and percentage of phonation time increased. However, there were no appreciable variations in the research population's average F0 with respect to sex.
The amount of time that the vocal folds vibrate during a day with the intention of producing voice (phonation time) and percentage of phonation time during a typical full day of a HA user, to the best of our knowledge, have never been studied before and may be a sign of oral communication participation. An increase in speaking-intensive daily activities could account for a rise in these two characteristics as patients may have felt more comfortable speaking during conversations. These findings are in line with a previous study on recipients of CI, who had an average daily phonation time of 8.2% six months following surgery (Mozzanica, Schindler, Iacona, & Ottaviani, 2019). Interestingly, our results also align with those of Cantarella and colleagues, who examined the vocal behavior of 92 call center operators without hearing impairments and found a mean percentage phonation time of 7.1% (Cantarella, et al., 2014).

Voice amplitude considerably reduced following four months of HA use. This decrease in vocal intensity brought about by wearing HAs may be linked to an enhancement in auditory feedback, enabling patients to better manage their phonation by utilizing the sensory data they are gathering throughout the speaking task. As mentioned in a previous paragraph of this thesis, it makes sense that improved auditory input from hearing restoration could result in a decrease in voice intensity, as it is widely known that decreased auditory feedback increases voice amplitude (Brumm & Zollinger, 2011). Furthermore, our findings on voice amplitude during HA rehabilitation are in line with earlier research assessing normal-hearing people' speech production. In particular, the same 92 previously mentioned healthy call center operators had an average amplitude of 70.5 dB SPL (Cantarella, et al., 2014); additionally, an average speech amplitude of 69.6 dB SPL was observed by Franca et al., who examined the vocal demands of eight student singers (Franca & Wagner, 2015).

Interestingly, amplitude data before and after HA rehabilitation revealed that the interquartile range was wider before than after the use of HA; this could indicate that hearing restoration caused a decrease in amplitude variability in the patients.

The average F0 by gender did not change much in our investigation. This result defies logic and seems rather incongruent to us. Indeed, the normal speaking voice amplitude and pitch are somewhat correlated: normally, spontaneous pitch usually increases as vocal intensity decreases (Debruyne & Buekers, 1998). Yet, following 4 months of HA use in female subjects, our data indicate a nonsignificant tendency towards a decrease in average F0. The small sample size of eight females might have affected the findings, and it's likely that a larger sample would have produced different findings.

Following four months of HA administration, our patients' mean overall SSQ score as well as the scores for each of its subdomains showed a significant improvement. The SSQ is a self-report assessment tool used to gauge a listener's hearing comprehension in a variety of demanding and real-world contexts (Gatehouse & Noble, 2004). The idea that HA improves perceived difficulties connected to hearing is supported by the increase in its average score in the aided condition. Our findings align with earlier investigations. When compared to bilateral HA users, unaided patients' SSQ scores are much lower (Noble & Gatehouse, 2006). However, it is important to note that whereas Noble and Gatehouse examined various patient cohorts based on their HA status, our SSQ results pertain to the same group.

Regarding the IOI-HA questionnaire results, patients obtained an average score of 28.8 ± 4.2 after using HA for four months. This self-assessment questionnaire, with a range of 7 to 35, captures the patient's perspective on how their everyday usage of HA has changed over time by considering their level of pleasure, the constraints of their basic activities, their ability to participate, the impact they have on others, and their quality of life (Cox, Alexander, & Beyer, 2003). Our findings align with earlier research in the field. Cox evaluated the psychometric features of the IOI-HA and concluded that values above the middle of the scoring range could suggest satisfaction with the HA (Cox & Alexander, 2002). He found that the mean score for each of the seven questions that make up the index was

between 3.5 and 4.1. Later, after administering the IOI-HA to 108 patients, the average age of whom was 77 years, Kozlowski and colleagues investigated user satisfaction with HA and discovered a mean score of 27.3 (Kozlowski, Almeida, & Ribas, 2014). The authors interpreted this result as a high level of satisfaction with the HA and a positive change in QoL because the mark is larger than 50% of the score. Overall, our cohort's subjective evaluation results indicate that they were satisfied with their performance and saw a considerable gain from using HA.
Some of the APM parameters showed significant correlation with the speech subscale of the SSQ, indicating that changes in voice production were associated with changes in HA users' perceived disabilities, particularly in the communication-related domain. This raises the possibility that patients who spoke for longer periods of time during the day also self-reported being better able to follow speech, even in a setting where competition noise existed.

In these patients, HA appears to have served as a facilitating factor, lessening the perceived handicap in talks. To the best of our knowledge, no prior study has found a correlation between perceived hearing capacity or disability as indicated by SSQ scores and voice production as measured by the APM. However, the research that is currently available backs up the idea that HA has a positive impact on social activities, particularly among the older population (Holman, Drummond, & Naylor, 2021) (Sawyer, Armitage, Munro, Singh, & Dawes, 2019).

When comparing the intervention group to the controls, Holman et al. discovered significantly higher scores on the social activity and social participation questionnaires as well as lower marks on the social participation limitation questionnaire.

There was small to moderately significant negative correlation between the SSQ scores and the auditory data. These findings imply that patients' scores on the SSQ questionnaire and its subscales were lower among those who had greater hearing thresholds, or worse hearing, both before and after HA therapy. Similar findings were reported by Gatehouse and Noble who reported that the correlation between BEA and SSQ average was −0.51, while the correlation between WEA and SSQ was −0.52 (Gatehouse & Noble, 2004). Additionally, Most and colleagues discovered a correlation between poorer performance on the SSQ and higher unaided audiometric thresholds, which indicate a greater degree of hearing loss (Most, Adi-Bensaid, Shpak, Sharkiya, & Luntz, 2012).

Lastly, a strong link between the phonatory and auditory measures was discovered. In both the aided and unaided conditions, the worse ear's audiometric threshold showed a weak negative correlation with phonation time. This appears to support the idea that those who have higher hearing loss tend to speak less. Furthermore, there was a significant correlation between the average amplitude in the unaided setting and a greater BEA threshold prior to HA restoration. The idea on auditory input on vocal production is consistent with this outcome, since before to HA therapy, patients with more disability in their best ear tended to talk higher (Selleck & Sataloff, 2014).

3.6 Conclusion

In conclusion, this study has shed light on the intricate interplay between hearing loss, the utilization of hearing aids (HAs), and its impact on speech production dynamics and quality of life. The prevalence of age-related hearing loss continues to rise globally, posing significant challenges to individuals' communication abilities and overall well-being. Despite the acknowledged efficacy of HAs in improving auditory perception and communication outcomes, their uptake remains limited, often hindered by various barriers such as social stigma and cost concerns.

This research, utilizing ambulatory phonation monitoring (APM), has provided novel insights into the daily implications of HA usage on speech production parameters, an area that has been largely understudied. The findings indicate notable variations in phonation time and percentage of phonation time following both before and after a 4-month period of HA rehabilitation, suggesting a potential increase in oral communication participation among users. Moreover, the observed decrease in voice amplitude post-HA use aligns with the notion of enhanced auditory feedback facilitating better phonation control. The findings highlight the potential significance of increased speaking time and decreased amplitude as objective measures of acoustic rehabilitation effectiveness and surrogates for reduced social isolation and improved communication outcomes.

Importantly, improvements in quality of life measures, as evidenced by significant enhancements in SSQ scores and IOI-HA questionnaire results, underscore the positive impact of HA intervention on patients' perceived disabilities, particularly in communication-related domains. Furthermore, the correlation between changes in voice production and perceived hearing capacities highlights the interconnectedness of auditory and vocal functions in individuals with hearing impairment. The innovative application of daily measurements of speaking time offers a practical and informative approach to assessing the impact of HA intervention on daily communication and voice dynamics.

These findings not only contribute to a deeper understanding of the benefits of HA rehabilitation but also underscore the importance of personalized, comprehensive hearing rehabilitation strategies. By addressing barriers to HA uptake and optimizing rehabilitation interventions based on individual needs, healthcare professionals can enhance overall communication outcomes and quality of life for individuals with hearing loss. Moving forward, continued research in this area is essential to further refine rehabilitation approaches and improve outcomes for individuals living with hearing impairment.

Focus on paper III

Title: Ambulatory phonation monitoring in prelingual and postlingual deaf patients after cochlear implantation

Authors: **Albera A**, Puglisi GE, Astolfi A, Riva G, Cassandro C, Mozzanica F, Canale A.

Published on: Audiology and Neurotology – 2023 – 28 (1): 52–62. doi: 10.1159/000526936

Objective: Hearing loss is known to play a fundamental role in voice production due to a lack of auditory feedback. The purpose of the study was to evaluate both fundamental frequency (F0) and loudness of voice on adult deaf patients subjected to cochlear implantation, according to the prelingual or postlingual onset of the deafness.

Materials and methods: All measurements were performed through an ambulatory phonation monitor (APM), which is a portable vocal dosimeter able to evaluate the variations in the vocal production. 32 adults who had undergone cochlear implantation due to severe or profound bilateral hearing loss and their outcomes were compared with a control group of 32 normal hearing (NH) subjects. All subjects were asked to utter the sustained vowel /a/ for at least 5 s and then to read an Italian phonetically balanced text. Evaluations with cochlear implants (CIs) were made with both turn on and switch off conditions, as well as in quiet condition and with background noise.

Results: Both trends in speech performance in a common daily condition and phonatory characteristics of reading a text were assessed

Conclusion: The APM has demonstrated its utility as a valuable tool for assessing the effectiveness of CI and can serve as an indicator of an individual's engagement in communication. Our findings indicate comparable speech performances between individuals with prelingual and postlingual deafness, both in vowel phonation and reading tasks. These results underscore the CI's capacity to adjust certain aspects of everyday speech, such as fundamental frequency (F0) and loudness, by restoring auditory feedback.

4.1 Introduction of the study

Profound hearing loss represents a deep challenge to effective communication, as it fundamentally disrupts the auditory feedback loop crucial for speech production. The inability to perceive auditory cues leads individuals with profound hearing loss to adapt their speech patterns to compensate for this deficit. Consequently, speech modifications become not only a necessity but also a means of navigating the complexities of language expression. Severe hearing loss is known to disrupt the auditory feedback mechanisms, resulting in vocal alterations such as elevated pitch, increased loudness variability, and difficulties in managing speech intensities and clarity, thereby impacting social interactions adversely. However, recent advancements in auditory prosthetics have revolutionized the landscape of rehabilitative interventions for those with profound hearing loss. By re-establishing auditory feedback, cochlear implants offer a promising avenue for restoring speech intelligibility and quality.

Significant progress has been made in addressing these issues with the introduction of cochlear implants (CIs), which are electronic devices surgically inserted in the inner ear to direct stimulate auditory nerve fibers and restore sound feeling, by-passing death hair cells of the cochlea. Remarkably, studies have demonstrated how cochlear implants improve speech production by restoring auditory feedback (Wilson BS, Wolford, Eddington, & Rabinowitz, 1991). Research conducted on adult recipients of cochlear implantation has revealed declines in speech loudness (sound pressure level, SPL) and vocal pitch/fundamental frequency (F0) (Perkell, et al., 2007) (Schenk, Baumgartner, & Hamzavi, 2003), along with decreased variability in pitch and amplitude (Evans & Deliyski, 2007) (Gautam, Naples, & Eliades, 2019). Furthermore, shorter speech timing durations and enhanced phonatory control of vowels and consonants have been noted (Langereis, Bosman, van Olphen, & Smoorenburg, 1997). Still, there is a lack of study on prelingually deaf children, and most of the material that is now available focuses on postlingually deaf adults.

Furthermore, most studies evaluating speech production in CI recipients have assessed phonation under simplified conditions, neglecting real-world communication scenarios, such as noisy environments. Similarly, investigations into voice quality modifications have predominantly involved short vocal tasks, with limited exploration of sentence or text reading. Methodological limitations, including microphone placement variability and susceptibility to environmental interference, further complicate data interpretation in existing studies (Schenk, Baumgartner, & Hamzavi, 2003) (Gautam, Naples, & Eliades, 2019) (Ubrig, et al., 2019). To address these gaps, the study aimed to investigate phonatory parameter changes in adult CI recipients using a portable vocal dosimeter for ambulatory phonation monitoring (APM). Even though it wasn't created with this purpose in mind, the instrument has shown to be resistant to background noise and to yield accurate data on vocal parameters like F0 and SPL as opposed to the common unidirectional or multidirectional air microphones used in earlier studies (Mozzanica, Schindler, Iacona, & Ottaviani, 2019) (Svec, Titze, & Popolo, 2005). This study includes postlingually deaf individuals, in addition to prelingually deaf

adults, and evaluates full text reading alongside sustained vowel emissions. Even though numerous authors have examined changes in voice quality in individuals with severe hearing loss receiving CI treatment, all these studies only assessed a brief phonation that consists of repeating a single word or vowel for a few seconds at a comfortable pitch and consistent amplitude. Only two authors included reading sentences or short texts in their vocal assessments: Ubrig (but only for postlingually deaf adults) (Ubrig, et al., 2019) and Ruff (but only for adults and children), who assessed text reading but only for words recognition and reading difficulty following cochlear implantation (Ruff, et al., 2017).
Moreover, by assessing different listening conditions, including quiet and noisy environments, the goal was to elucidate the utility of phonation measurements in evaluating the success of cochlear implantation relative to speech production.

In the present study, which led to the publication in the literature of the aforementioned article in which I appear in the authorship as first author, my contribution mainly consisted in the supervision of the study, in the conceptualization of the project idea, in the design of the methodology, in the recruitment of subjects with hearing loss and cochlear implant users and assisting with data collection. Furthermore, I performed the statistical analysis of the results and participated in writing most of the manuscript.

4.2 Materials and methods

The study population comprised adults who had undergone cochlear implantation due to severe or profound bilateral hearing loss, meeting the institute's candidacy criteria, which included a pure-tone average hearing threshold exceeding 75 dB HL at 500 Hz, 1,000 Hz, and 2,000 Hz, along with a free-field speech perception threshold equal to or lower than 50% despite optimal amplification via hearing aids in the ear scheduled for implantation. Patients were categorized based on the onset of deafness, either prelingual or postlingual. Exclusion criteria encompassed any reading limitations, speech disorders resulting from malformation or acquired damages to the speech organs, motor speech disorders, voice disorders unrelated to deafness, and challenges with auditory rehabilitation or CI fitting, as well as associated disabilities.
The study included a cohort of 32 CI patients, comprising 16 males (8 prelingual and 8 postlingual) and 16 females (8 prelingual and 8 postlingual), with a mean age of 49.7 ± 6 years (19 to 81 years). The mean preoperative pure-tone average (PTA) at speech frequencies (0.5–1–2–4 kHz) was 78.5 ± 7 dB HL, while the mean post-implantation PTA was 27.3 ± 8 dB HL. Six patients (19%) underwent bilateral cochlear implantation, including four with congenital profound bilateral deafness. Among patients with prelingual deafness, all were implanted for the first time at a later age (mean age 30.8 years), retaining residual hearing at low frequencies (averaging 82 dB at 250 Hz, 86 dB at 500 Hz, and 110 dB at 1 kHz) at the time of surgery. Despite profound deafness, these patients had utilized hearing aids since childhood (average duration of use: 32 years), with some achieving minimal gains in speech recognition. Additionally, all had undergone extensive speech therapy rehabilitation, resulting in varying degrees of structured

oral language development. The same senior surgeon carried out every surgery. Four of the patients who underwent unilateral CI experienced a bimodal restoration of their hearing (CI plus contralateral HA). The companies that made the implanted CIs were Med-El (10 subjects, 31%), Cochlear (18 subjects, 56%), and Advanced Bionics (4 subjects, 13%). From the perspective of hearing rehabilitation, all the CI patients were stable as they had undergone auditory rehabilitation following cochlear implantation and had at least two years of consistent CI mapping following the activation of the prosthesis.

A control group of 32 normal hearing (NH) participants (mean age 29.7 ± 3 years), aged between 20 and 64 (sixteen males and sixteen females), was enlisted. Every NH individual showed a PTA of less than 15 dB HL (mean 9.18 ± 4 dB HL).

Initial room acoustic measurements were conducted to determine whether the chosen space's reverberation time (RT60), or the amount of time it takes for a signal to drop 60 dB from its peak, was appropriate for administering the test. According to the EN ISO 3382-1 standard [ISO, 2009], the assessments were carried out using the interrupted noise method using a pink noise generator (Minirator MR-1) and a sound level meter (Acoustilyzer AL1) that were attached to the primary speaker. The testing room was deemed acoustically acceptable for the investigation since it was acoustically treated, had a volume under 45 m^3, and had a measured RT60 of less than 0.5 s at medium frequencies.

NH subjects and patients with CI were asked to read a brief text in Italian called "Il ramarro della zia," which is a phonetically balanced content created by Vernero for speech therapy purposes (Vernero, Gambino, & Schindler, 1998), and to utter the sustained vowel /a/ for at least five seconds to evaluate the spectral and loudness modification of voice in terms of F0 and SPL, respectively, according to different hearing conditions. Both in a quiet environment and with a 50 dBA background energetic masking noise, NH subjects completed these tasks. In a similar vein, patients with CI completed these activities twice, under identical 50 dBA background noise and in calm conditions.

They were instructed to turn on their CI after turning it off the first time. To achieve the maximum amount of masking possible, three calibrated loudspeakers were used, one at 0° and the lateral ones at 110°, and placed at standard ear height (1 m from the floor) and at the same distance from the receiver (2 m).

Both CI patients and NH individuals were seated in a cozy manner. The most NH-like condition was selected among CI patients with processors that allowed them to choose the microphone's direction: a fixed orientation stimulating the pinna. Moreover, the ability of the CIs to reduce background noise by adaptive microphone adjustment has never been chosen to prevent any augmentation of the patient's voice intelligibility. Furthermore, the hearing aid was consistently taken out during the recordings for the four patients who received a bimodal hearing restoration.

1. Pronunciare a volume normale una lettera a prolungata

AAAAAAAAAAAAAAAAAAAAAAAAAAAAA

ATTENDERE 5 SECONDI

2. Leggere il seguente testo

Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso c'era una rosa rosso vivo e, lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana, che, saltando dalla strada, finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxì. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.

Figure 9: Speech evaluation protocol for NH subjects and deaf patients, to perform both with CI on and off.

The APM model 3200 (KayPENTAX, Lincoln Park, NJ, USA) was used in the study to ensure objective measurement of voice features. This device consists of an accelerometer that is attached to the front of the neck. It is designed to detect vibrations from the vocal folds that are conveyed through the tissues of the neck and translates them into speech pressure levels, or SPLs. At a rate of twenty samples per second, the APM records audio speech. The data is then sent to a microprocessor unit that is carried in a waist pack.

Among the other parameters that the APM yielded, the research concentrated on gathering the following data:

- Average fundamental frequency (F0) expressed in Hertz (Hz), representing the mean frequency of vocal fold vibration.
- Average loudness expressed in emitted SPL (in dB), indicating the mean energy level of the voice sound wave.

This technique for measuring phonation has shown to be somewhat insensitive to background noise and capable of distinguishing intentional vocalization from other sounds, such coughing or clearing the throat. The acquisition equipment needed to be individually calibrated before real-time voice monitoring could begin. This required calibrating an air microphone precisely 15 cm from the speaker's lips and a contact sensor at the jugular notch, which delivers reference SPL values, subject by subject.

Through the acquisition of voltage levels from the contact sensor resulting from skin acceleration produced by vocal fold vibration and referred SPL values from the air microphone, a calibration function comprising subject-specific constants could be created and used in further voice monitoring.

The initial calibration technique involves all 64 participants vocalizing a sustained vowel /a/ at varying loudness levels, from whispers to yells, to cover the whole range of expected loudness levels during later monitoring sessions. During the evaluations, all patients accepted the APM device well, and the calibration process usually took less than five minutes.

4.3 Results

Results at Mann-Whitney U test indicated no significant difference in postoperative PTA values between males and females ($p = 0.138$), between unilateral and bilateral cochlear implantation ($p = 0.524$), and between prelingual and postlingual deafness ($p = 0.491$). Given the similarity in postoperative auditory outcomes across these groups, all patients were considered comparable, validating the outcomes of phonatory tests. Furthermore, no differences were found between males and females in terms of age, as well as between unilateral and bilateral CI recipients ($p < 0.05$). However, patients with prelingual deafness were significantly younger (mean age 42.5 years) compared to postlingual deafness (mean age 62.5 years) ($p < 0.001$).

Speech F0 and loudness values obtained from both control NH subjects and CI recipients are detailed in Tables 9-11. There was no significant difference in speech characteristics according to different manufacturers (Advanced Bionics, n = 4; Cochlear, n = 10; Med-El, n = 18; $p > 0.05$), neither for speech F0 values nor for loudness.

NH subjects		
Male F_0 (*n* = 16)	Vowel	112.8±15 Hz
	Reading	122.9±12 Hz
	Reading + noise	129.3±13 Hz
Female F_0 (*n* = 16)	Vowel	202.6±27 Hz
	Reading	202.8±22 Hz
	Reading + noise	210.7±21 Hz
Loudness (*n* = 32)	Vowel	77.0±8 dB
	Reading	76.1±6 dB
	Reading + noise	79.9±7 dB

F_0, fundamental frequency; noise, background noise at 50 dBA.

Table 9: Phonatory outcomes of NH subjects.

Vowel task – deaf patients		
Male F_0 (*n* = 16)	CI off	156.5±40 Hz
	CI on	150.8±42 Hz
Female F_0 (*n* = 16)	CI off	251.2±54 Hz
	CI on	218.4±52 Hz
Loudness (*n* = 32)	CI off	82.5±11 dB
	CI on	80.9±13 dB

F_0, fundamental frequency; CI off, cochlear implant switched off; CI on, cochlear implant turned on.

Table 10: Phonatory outcomes of deaf patients with CI in the sustained vowel task.

Reading task – deaf patients		
Male F_0 (*n* = 16)	CI off	143.8±31 Hz
	CI on	136.8±34 Hz
	CI on + noise	143.0±31 Hz
Female F_0 (*n* = 16)	CI off	222.2±50 Hz
	CI on	218.4±52 Hz
	CI on + noise	226.7±55 Hz
Loudness (*n* = 32)	CI off	76.7±8.7 dB
	CI on	73.2±9.4 dB
	CI on + noise	76.2±8.9 dB

Table 11: Phonatory outcomes of patients with CI in the reading task. Background noise at 50 dB.

Sustained Vowel Task

A statistically significant difference in F0 values was found among males with NH (n = 16), deaf males without CI (n = 16), and deaf males with CI on (n = 16) (p = 0.001). Compared to the other two groups, deaf males who had their CI turned off showed greater F0 scores. Female patients with CI switched off had significantly higher F0 values than both NH women and women with CI on, indicating a similar difference between all three groups (p = 0.001).

Regarding the vowel /a/ loudness values, a significant difference was also seen between NH individuals (n = 32), patients with CI switched off (n = 32), and patients with CI turned on (n = 32) (p = 0.031). When compared to the other two groups, deaf patients who were not using CI showed greater loudness values. The Wilcoxon signed-rank test for deaf patients showed an evident but not statistically significant drop in F0 values in both males (p = 0.278) and females (p = 0.352) after the CI was activated. Similarly, following CI activation, there were no appreciable variations in loudness values in the vowel task (p = 0.286). Additionally, the F0 and loudness of the vowel task were compared between prelingual and postlingual deafness. Specifically, men with prelingual deafness had lower F0 values than men with postlingual deafness, both with CI off (p = 0.781) and with CI on (p = 0.486), however these differences were not statistically significant. On the other hand, both with CI off (p = 0.376) and with CI on (p = 0.133), females with prelingual deafness showed higher F0 values than females with postlingual deafness, but also these differences were not statistically significant. When it comes to loudness, prelingual patients reported higher but not statistically different values than postlingual patients, both with CI off (p = 0.174) and with CI on (p = 0.250). In the case of both prelingual and postlingual deafness, turning on and using the CI has not been demonstrated to significantly alter the values of F0 and loudness in the vowel task (p > 0.05) (Table 12).

Vowel task – deaf patients	CI off	CI on	*p* value
Prelingual deafness			
Females – F_0	263.6±61 Hz	266.6±59 Hz	0.821
Males – F_0	153.6±49 Hz	143.0±39 Hz	0.240
Loudness	85.2±12 dB	83.6±14 dB	0.360
Postlingual deafness			
Females – F_0	238.7±47 Hz	223.9±47 Hz	0.486
Males – F_0	159.5±31 Hz	158.5±48 Hz	0.927
Loudness	79.8±10 dB	78.2±12 dB	0.427

F_0, fundamental frequency; CI off, cochlear implant switched off; CI on, cochlear implant turned on.

Table 12: Phonatory differences between prelingual and postlingual deafness on deaf patients in the vowel task.

Reading Task

When reading the text "Il ramarro della zia," the NH individuals showed a significant increase in loudness after background noise was added at an intensity of 50 dBA (p < 0.001). Similarly, both NH males and females demonstrated a rise in F0 scores in the reading with background noise (p < 0.001). Similarly, the evaluation of deaf patients' speech using CI on showed that adding background noise significantly increased the F0 values in both male and female subjects (p =

0.007 and $p = 0.008$, respectively), and that loudness significantly increased the values in comparison to the assessment in quiet conditions ($p < 0.001$). When CI was activated, there was a significant decrease in F0 values ($p = 0.023$), with results that were now comparable to those of NH subjects ($p = 0.184$). The Mann-Whitney U test revealed that, in males and under quiet conditions, deaf patients with CI off had significantly higher F0 values than NH subjects ($p = 0.035$). On the other hand, $p = 0.402$ showed no significant difference between female NH participants and female deaf with the CI turned off, and $p = 0.717$ showed no significant change in the F0 in female patients upon turning on the CI further.
Regarding speech loudness in the quiet condition, there was a statistically significant decrease in the values after CI activation ($p < 0.001$), but no significant difference in values was found between NH subjects and deaf patients with CI turned off ($p = 0.989$). The sustained vowel test and the reading task yielded identical results for NH individuals in terms of loudness ($p = 0.640$) and F0 in females ($p = 0.717$). However, the average F0 value in NH men was considerably lower when it came to the phonation of the vowel /a/ ($p = 0.008$). On the other hand, in the vowel test as opposed to the reading task, deaf individuals with CI off demonstrated substantially higher F0 values ($p = 0.003$ for females and $p = 0.026$ for males) and loudness values ($p < 0.001$).
The Spearman correlation coefficient was used to examine the link between PTA levels and speech characteristics of deaf patients. The reading task, both with and without CI, did not show a significant correlation ($p > 0.05$) between mean post-implantation PTA thresholds and F0 values for males or females. The vowel task also did not yield significant results.
Conversely, a positive correlation was seen between the mean PTA thresholds and speech loudness (0.36 with CI off and 0.35 with CI on - $p < 0.05$). This means that higher speech loudness values were correlated to higher PTA thresholds.
Additionally, in the reading task, there was a negative connection seen in both genders and between the age of deaf patients and their mean F0 scores, with a confidence interval of ($r = -0.31$, $p < 0.05$) showing that younger patients had higher F0 scores. On the other hand, all other correlations between the patients' age and speech characteristics were determined to be non-significant ($p > 0.05$).
Additional comparisons between the postlingual and prelingual subgroups on the reading task revealed lower F0 values in all postlingually deaf patients, male and female, with and without CI. However, this difference was only statistically significant in deaf women who did not use CI ($p = 0.047$). In the case of postlingual deafness, lower but not statistically significant values were also shown for speech volume, both with CI off and CI on ($p > 0.05$). Furthermore, when speech was evaluated with background noise, we found no evidence of a significant change in speech features between prelingual and postlingual deafness ($p > 0.05$). After CI activation, there were no changes in prelingual deafness between males and females ($p > 0.05$), whereas in males with postlingual deafness the CI switching on significantly reduced the F0 values ($p = 0.011$). Conversely, in both prelingual and postlingual deafness condition, the use of the CI demonstrated a substantial decrease in the speech loudness values in all patients ($p < 0.05$) (Table 13).

Reading task – deaf patients	CI off	CI on	*p* value
Prelingual deafness			
Females – F_0	246.7±58 Hz	242.0±59 Hz	0.629
Males – F_0	147.2±40 Hz	141.3±44 Hz	0.455
Loudness	78.3±9 dB	75.2±8 dB	0.006*
Postlingual deafness			
Females – F_0	197.6±25 Hz	194.8±33 Hz	0.614
Males – F_0	140.6±21 Hz	132.5±24 Hz	0.011*
Loudness	75.1±8 dB	71.1±10 dB	0.001*

F_0, fundamental frequency; CI off, cochlear implant switched off; CI on, cochlear implant turned on. * Significant value <0.05.

Table 13: Phonatory differences between prelingual and postlingual deafness on deaf patients in the reading.

4.4 Discussion

This study's objective was to assess how cochlear implantation affected the voices of people with substantial hearing loss, with a special emphasis on the distinctions between prelingual and postlingual deafness. Thirty-two profoundly deaf adults, evenly divided between prelingual and postlingual deafness and gender, underwent cochlear implantation to make up our study group. There was also a control group that consisted of 16 NH males and 16 NH females. Both groups wore a contact-sensor-based voice monitoring equipment (i.e., the KayPENTAX APM device) and participated in voice recordings that involved them reading a phonetically balanced paragraph. Mean fundamental frequency and SPL were retrieved from the monitoring for every participant in both noisy and quiet environments.

From the results obtained, it is observed how the cochlear implant plays an important role in determining a change in voice management mode by implanted subjects, as evidenced by the variation in the values of the analyzed parameters. Despite undergoing cochlear implantation, patients with congenital deafness often exhibit pronunciation problems, vowel substitutions, and intonation difficulties, leading to highly understandable speech (Hocevar-Boltezar, et al., 2006) (Lenden & Flipsen, 2007). The restoration of auditory feedback through CI has also been demonstrated to cause adjustments in speech production, specifically in the reduction of fundamental frequency and speech loudness (Ubrig, et al., 2019) (Gautam, Naples, & Eliades, 2019) (Boisvert, Reis, Au, Cowan, & Dowell, 2020). Similarly, even subjects who experience the occurrence of deafness as adults demonstrate a degradation of the speech over time. However, as noted by Coelho in her comprehensive assessment of the literature, it is challenging to understand the true impact of the CI on the speech of deaf patients due to contentious results and the variability of the methods used in most of the research (Coelho, Brasolotto, & Bevilacqua, 2012). Only Ubrig examined a sizable case series that was like the one under consideration in this investigation, however he limited his analysis to adults who had postlingual deafness (Ubrig, et al., 2019). In line with the necessity of restoring auditory input earlier due to congenital deafness, individuals with postlingual deafness typically have a mean age of 62 years old, but the prelingual deaf group's mean age was significantly lower at 42 years old. All patients, however, attained a very excellent mean postoperative PTA threshold (27.3 dB HL in free-field evaluation), and there were no appreciable variations in hearing thresholds based on gender, unilateral versus bilateral implantation, or postlingual versus prelingual deafness. Recent research did not find any appreciable differences in the electrically evoked compound action potential of the auditory nerve in CI recipients between prelingual and postlingual deafness, despite the literature's suggestion that early cochlear implantation is crucial to patients' hearing outcomes (Harrison, Gordon, & Mount, 2005). Additionally, Canale reported no differences in perceived quality of life or in the benefit from CI between the postlingually and prelingually deafened groups, confirming our positive results on prelingually deaf patients. This suggests that the extent of rehabilitation with speech therapy and hearing aids performed during childhood also plays a role in the hearing outcomes achievable for subjects with congenital hearing loss implanted in adulthood (Canale, et al., 2019). Regretfully, unlike the positive hearing outcomes, there hasn't yet been any information published in the literature regarding the phonation differences among individuals who are prelingually deaf. In both control subjects and people with mild and severe dysphonia, Hillman demonstrated that a vocal accelerometer gives superimposable data of F0, voice loudness, and phonation duration to those captured by a standard microphone (Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006). Moreover, Švec showed that the APM is even more accurate than microphones in obtaining the average SPL value of gentle, pleasant, or powerful voices with an accuracy greater than ±2.8 dB in 95% of instances (Svec, Titze, & Popolo, 2005). This is consistent with the findings of Astolfi et al., who discovered that, despite its increased uncertainty, there is a significant benefit to using a contact microphone for other contact-sensor-based devices (Astolfi, Castellana, Carullo, & Puglisi, 2018). In fact, whereas a contact-sensor-based device can provide an uncertainty of up to 3 dB and a headworn air microphone up to 2 dB, the latter ignores background noise, even at high magnitudes, and allows for frequent, accurate, and long-term monitoring. Only Mozzanica has up to now incorporated the APM in the evaluation of voice production following cochlear implantation; however, this is limited to postlingual deafness and is associated with the registration of a 24-hour working day (Mozzanica, Schindler, Iacona, & Ottaviani, 2019). The vowel /a/ was chosen for its extended emission at habitual pitch and loudness in our voice recordings because it primarily relies on acoustic control rather than orosensitive control. However, with the aim of evaluating the speech in a condition as close as possible to everyday life, we also included the reading of a phonetically balanced text, both in quiet conditions and with a background noise of 50 dBA. But to assess the speech in an environment as like real life as feasible, we also had participants read a phonetically balanced text in both silent and 50 dBA background noise.

Except for a study by Lee et al., deaf people with CI have never had their speech characteristics assessed in a competitive acoustic setting (Lee, Liu, & Lee, 2013); instead, they have only ever been subjected to straightforward vocal tasks and quiet conditions (Hocevar-Boltezar, et al., 2006) (Evans & Deliyski, 2007) (Ubrig, et al., 2019) (Upadhyay, et al., 2019). As a result, little knowledge of speech production in actual communication conditions and noisy environments has been obtained. As expected, our findings demonstrated a substantial rise in

F0 and loudness during the reading task with background noise, which was noticeable in both deaf patients with CI on and NH participants. Lee et al. confirmed similar results, albeit restricted to postlingual deafness (Lee, Liu, & Lee, 2013). Patients with CI appear to react to background noise by modifying their speech production accordingly, which may be a perceptual benefit of the Lombard effect, which is regularly observed in NH subjects and appropriately restored when CI is activated. NH males demonstrated comparable loudness but much lower F0 values in the vowel task, while NH females were found to maintain both F0 and loudness reasonably steadily in the reading tasks. Regarding voice loudness stability, presuming that the vowel uttering, and the text reading are two consecutive voice production tasks, the results obtained support a study by Castellana et al. that discovered NH subjects have low intra-speaker variability, falling between 1 and 2 dB for mode SPL and within 1 dB for equivalent and mean SPLs (Castellana, Carullo, Astolfi, Puglisi, & Fugiglando, Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings., 2017). In contrast, all deaf patients showed higher loudness and F0 values during the vowel task than during the reading. Borden conducted a very helpful analysis of the evidence and suggests that motor control centers are not able to simultaneously govern speech production with very short auditory information (Borden, 1979). Otherwise, a one-minute reading gives the individual additional time to analyze his speech and potentially adjust its characteristics.

Comparable outcomes were observed about CI activation, emphasizing its function in bringing about a shift in the way patients manage their voices. Despite a small but not statistically significant decrease in F0 and loudness values after turning on the CI in the sustained vowel task, the entire sample of deaf patients did not exhibit the anticipated voice modifications, most likely because of the abrupt shift in auditory feedback. As Gautam pointed out, vocal control may in fact occasionally rely more on longer time scales than on instantaneous feedback, which would prevent adequate vocal adaptation in situations when the CI is turned on and off in a matter of minutes or where the job is too brief (Gautam, Naples, & Eliades, 2019). The literature emphasizes the following discordant outcomes in this regard: although adults and children were evaluated jointly, Monini found a considerably lower F0 in the voice samples of the Italian vowel /a/ at an early stage following cochlear implantation (Monini, Banci, Barbara, Argiro, & Filipo, 1997). In contrast, Kirk and Edgerton observed that only male patients had lower F0 values and less variability in loudness levels during the vowel /a/ assessment, while female patients displayed higher F0 values and more varied loudness with CI on (Kirk & Edgerton, 1983). Regarding text reading, turning on the CI appears to be able to considerably lower loudness and F0 in deaf men, reaching levels like those of NH subjects. This outcome is in line with findings from Hamzavi et al., who found that CI patients tended to have lower F0 gradually approaching the normal range of F0 (Hamzavi, et al., 2000). In this context, Leder showed that the F0 is the first acoustic characteristic to resemble normal values again following cochlear implantation, and that this is especially true for men (Leder, et al., 1987).

The examination of the patients' phonation did not reveal any notable variations in the phonatory outcomes among recipients of CIs made by various manufacturers. Of course, comparing the hearing results between two different cochlear implant companies is difficult because any type of device has different components, coding strategies, software, and so on. However, like our study, Withers' previous research in a case of bilateral cochlear implantation using different devices found no differences in PTA and speech perception, despite patients' opinions on perceived sound quality differing significantly (Withers, Gibson, Greenberg, & Bray, 2011). Any device that is correctly implanted and functioning can improve hearing and determine changes in the auditory feedback, even though each company's CI has unique technical features and heterogeneous hearing outcomes have been frequently described in literature depending on CI specific features. Thus, we may conclude that the F0 and loudness speech alterations that were previously mentioned have nothing to do with the specific model or brand of CI that was used: rather, they are solely related to the device's basic functioning.

Finally, concerning the role of prelingual or postlingual deafness, speech is known to be impacted by the time a deafness first appears since early auditory feedback deprivation impacts F0 control and articulation accuracy, and prelingual deafness makes it harder for a person to learn to speak clearly and concisely (Ruff, et al., 2017). The sustained vowel task and the reading task showed no significant differences in the speech characteristics of prelingual and postlingual deaf patients; additionally, postlingual deaf patients' speech quality deteriorates because of inadequate auditory feedback. Nevertheless, postlingual deafness was associated with lower values of both F0 and loudness. Female patients with postlingual deafness exhibited considerably lower F0 values than deaf females with prelingual deafness, which was the only exception documented. Following CI activation, comparable outcomes were also noted in the reading and vowel phonation, with no distinctions between prelingual and postlingual deafness. Therefore, we can confirm that, regardless of the type of deafness (prelingual or postlingual), almost all deaf patients behave similarly from a phonatory perspective, even though different postoperative auditory results are reported in the literature depending on the period of onset of the hearing loss. Furthermore, adding background noise to speech evaluations conducted on recipients of CIs did not show any appreciable variations in their phonatory traits for either postlingual or prelingual deafness.

Both patients with prelingual and postlingual deafness showed a significant decrease in loudness values when reading the sentence, according to an investigation of how the patients' speech characteristics altered after turning on the CI. In a similar vein, we discovered that - though this is limited to male patient - the application of the CI also significantly influences how the F0 is modified in postlingual deafness patients. In evaluating speech samples prior to and one to four years following cochlear implantation, Smoorenburg reported differing results. While focusing solely on postlingual deafness, he observed that after CI, abnormally high pitches of deaf people decreased in some implanted women but not in men (Smoorenburg, Huiskamp, Langereis, & Bosman, 1994).

The literature has demonstrated that subjects with better hearing outcomes following CI activation typically speak with a lower loudness, which translates into a reduced vocal effort and load (Bottalico & Astolfi, 2012). This was confirmed by the significant positive correlation that emerged between postoperative hearing thresholds and speech loudness. Moreover, older deaf patients - male or female - generally speak with a lower F0 while the CI is on, both in calm and in the presence of background noise. This is shown by the negative association between the overall age of the patients and their speech F0 values. This finding is consistent with previous research, albeit limited to NH listeners, as F0 tends to decline significantly with aging (Nishio & Niimi, 2008).
The slowdown of certain executive cognitive resources, like working memory, is known to affect multiple top-down mechanisms, including phonation (Zucca, et al., 2022). This slowdown of executive cognitive resources could also account for the correlation. The latter explanation could account for the correlation in addition to the simple application of the CI.

4.5 Conclusion

The study's objective was to assess the changes in voice in a cohort of patients undergoing cochlear implantation who were evenly distributed between prelingual and postlingual deafness. All patients wearing a contact-sensor-based voice monitoring device (KayPENTAX's APM device) were treated to speech recordings that included both vowels and reading a phonetically balanced paragraph. All things considered, we demonstrated how the CI can improve the vocal experience of most deaf patients in any acoustic setting by adjusting phonatory features like fundamental frequency and loudness merely by restoring auditory feedback. Specifically, we observed comparable speech patterns in the reading and vowel /a/ phonation between the prelingual and postlingual groups. Furthermore, our findings offer additional evidence that patients with congenital prelingual deafness may benefit greatly from cochlear implantation, even at a later age, even though adults with prelingual deafness have generally been shown to have worse auditory outcomes with CI because of their longer history of sound deprivation.

Focus on paper IV

Title: Speech quality improvement of TV-Sets for hearing-impaired older adults

Authors: Astolfi A, Riente F, **Albera A**, Shtrepi L, Scopece L, Albera R, Masoero M.

Published on: *IEEE Transactions on Broadcasting* – vol. 69, no. 2, pp. 495-504, June 2023, doi: 10.1109/TBC.2023.3254150

Objective: To develop a software for Rai, the national Italian TV broadcaster, that enhances the listening experience for older persons with hearing impairments, particularly for subject with hearing aids who claim to have trouble comprehending speech on television.

Materials and methods: Flat TV-sets sacrifice space for built-in loudspeakers and this implies a degradation of the speech quality. A Transfer Function (TF) that dynamically equalizes sound level in real time by a Digital Audio Optimizer prior to the broadcast to the broadcast tower was developed. The TF boosts the one-third octave band of 4 kHz, which is particularly crucial for speech comprehension, and amplifies the frequency range between 1 kHz and 4 kHz. Subjective testing of the proposed TF was conducted in a laboratory setting using a commercial flat-screen TV and 31 older individuals with hearing impairments in accordance with ITU-R standard BS.1116-3.

Results: The TF demonstrated an average improvement of 24.3% of audio quality felt for the group of elderly subjects.

Conclusion: The TF enhances the frequency spectrum of the audio signal from the Italian radio and TV broadcasting company RAI before the transmission and subjective tests improved perceived audio quality to a minimum 20% level of improvement as requested by RAI. Results with older people with age-related hearing loss and hearing aids agreed to a similar study run with normal hearing listeners.

5.1 Listening TV in the elderly with hearing loss

Comprehending television dialogue can pose a difficulty for many individuals, particularly those with hearing impairments. The diverse range of accents, dialects, backgrounds, sound effects, and music contributes to the complexity of this listening task, making it especially challenging for hearing aid users. Modern movies and TV shows often integrate impactful sound effects and music, which are integral to the narrative alongside character dialogue. While these elements enhance the viewing experience, they also present obstacles to understanding conversations. Despite these challenges, television remains a prevalent leisure activity among adults, with statistics showing significant daily and weekly viewing habits (Bureau of Labor Statistics, 2017).

Poor TV listening in the elderly with hearing loss presents significant challenges that affect their overall television viewing experience and quality of life. Age-related hearing impairment, a prevalent issue among older adults, exacerbates difficulties in understanding dialogue and discerning audio details, thereby diminishing the enjoyment derived from watching TV programs. The consequences of poor TV listening extend beyond mere inconvenience, impacting social engagement, cognitive stimulation, and emotional well-being.

One of the primary challenges faced by the elderly with hearing loss is reduced speech intelligibility. High-frequency hearing loss, a common manifestation of age-related hearing impairment, particularly affects the ability to discern consonant sounds essential for understanding speech. As a result, dialogue on television may appear muffled or indistinct, making it challenging for elderly viewers to follow plotlines, catch important details, or fully engage with the content.

Furthermore, the discrepancy between the volume required for optimal audibility and the comfort level for elderly individuals with hearing loss poses another significant challenge. While increasing the TV volume may enhance speech clarity, it can lead to discomfort for others in the household and strain relationships. Conversely, maintaining a lower volume to accommodate others may result in inadequate audibility for the elderly viewer, leading to frustration and a sense of isolation (Shirley & Oldfield, 2015).

Moreover, poor TV listening in the elderly with hearing loss can contribute to feelings of social exclusion and withdrawal. Television serves as a primary source of entertainment, information, and companionship, particularly for individuals with limited mobility or social interactions. However, the inability to fully participate in conversations about popular TV shows or share in the enjoyment of watching together with family and friends can exacerbate feelings of loneliness and isolation.

Addressing the issue of poor TV listening in the elderly with hearing loss requires a multifaceted approach that encompasses both technological and psychosocial interventions. Technological solutions such as assistive listening devices, sound amplifiers, or TV listening systems can help improve audibility and speech intelligibility for elderly viewers with hearing loss. Additionally, raising awareness about the impact of hearing loss on TV listening and promoting open communication within households can foster empathy and understanding, encouraging collaborative efforts to accommodate the needs of elderly family members. Surprisingly, research on TV listening by HA users has been limited, despite its importance in their daily lives.

Advancements in hearing aid technology aim to address the complexities of various listening environments, including TV watching. Some hearing aids offer specialized programs for speech clarity in noisy backgrounds or different listening scenarios, yet there's a lack of standardized programs optimized for TV viewing. Wireless accessories like Phonak's TV Link II, enabling audio streaming from TVs directly to hearing aids, were introduced with the promise of improving TV listening experiences. However, despite the evident need, these accessories are not widely adopted by users (Standaert, Rakita, & Strelcyk, 2017).

About half of adults in their seventh decade have hearing loss severe enough to interfere with communication, although the prevalence of presbycusis varies greatly depending on the pure-tone averaged frequencies and the classification system used (Rodríguez-Valiente, Álvarez-Montero, Górriz-Gil, & García-Berrocal, 2020). Similarly, speech intelligibility gradually declines with age-related hearing loss, both in quiet and in noisy environments (Int. Org. Stand., 2017).

Preminger and Van Tasell studied the relationship between speech quality and speech intelligibility and found that there is a close correlation between the two (Preminger & Van Tasell, 1995). They also found that a loss in speech intelligibility corresponds with a decline in speech quality. Measures of speech quality included overall impression, effort, volume, pleasantness, and intelligibility. When speech intelligibility is very high, close to 100%, individual differences in speech quality occur. Since music frequently contains lyrics, speech intelligibility may also be related to the quality of music heard by those with hearing impairments. Preminger and Van Tasell also demonstrated the importance of frequency distribution for speech intelligibility in the elderly using hearing aids, noting that decreasing low-frequency bands led to increased pleasantness (Preminger & Van Tasell, 1995). Similarly, French and Steinberg showed that reducing the cutoff frequency of a low-pass filter from 7 kHz to 2.85 kHz resulted in decreased correct identification of syllables presented in quiet (French & Steinberg, 1947).

The influence of frequency components above 3 kHz on speech intelligibility and sound quality in individuals with mild-to-moderate high-frequency sensorineural hearing loss (ARHL) has also been noted. Amplification of these frequencies could enhance speech understanding, particularly in noisy environments. High-frequency audibility, particularly between 0.5 kHz and 3 kHz, is emphasized in fitting hearing aids according to Scollie et al. (Scollie, et al., 2005)

5.2 Degradation of perceived sound quality with Flat TV

In the ever-evolving landscape of home entertainment technology, the advent of flat-screen televisions (Flat TVs) has revolutionized the way we experience visual media. From the bulky cathode-ray tube TVs of yesteryears to the sleek, high-definition displays adorning modern living rooms, the transition to flat-panel technology has brought about unparalleled advancements in visual fidelity, energy efficiency, and aesthetic appeal. However, amid the marvels of razor-sharp images and vibrant colors, a subtle yet significant compromise often lurks beneath the surface – the degradation of perceived sound quality.

The pursuit of slimmer profiles and minimalist designs inherent in flat-screen TVs has necessitated a reimagining of audio delivery systems. Gone are the days of robust built-in speakers housed within spacious TV enclosures; instead, consumers are presented with increasingly compact speaker arrays nestled within slender chassis. While this design evolution undoubtedly contributes to the seamless integration of televisions into modern living spaces, it also poses inherent challenges to the reproduction of high-fidelity audio.

The degradation of sound quality and intelligibility with built-in loudspeakers of flat-panel televisions (Flat TVs) is a multifaceted issue that arises from various factors inherent in their design and construction. Firstly, the slim form factor of flat-panel TVs imposes significant constraints on the size and placement of built-in speakers. Limited enclosure space restricts the size of speaker drivers and the volume of the speaker cabinet, compromising their ability to reproduce full-range audio with depth and clarity. As a result, bass frequencies are often underrepresented, leading to a thin and unbalanced sound profile that lacks the richness and impact of a dedicated audio system. Moreover, the proximity of the speakers to the display panel can introduce unwanted resonances and reflections, further degrading sound quality and muddying the overall listening experience (Fuchs & Oetting, 2014).

Additionally, the drive towards aesthetic minimalism in flat-panel TV design can lead to compromises in speaker engineering and placement. Manufacturers may prioritize sleek and slim designs over acoustic performance, resulting in suboptimal speaker configurations and inadequate acoustic isolation. Furthermore, the integration of multiple electronic components within the TV chassis can exacerbate thermal concerns, limiting the available space for speaker enclosures and ventilation. Consequently, built-in speakers may suffer from thermal compression and distortion, particularly during prolonged periods of high-volume playback.

Furthermore, the inherent trade-offs between audio performance and visual appeal pose significant challenges for flat-panel TV manufacturers. Achieving a balance between compact form factors and satisfactory sound reproduction requires careful engineering and acoustic tuning, often necessitating compromises in one area to optimize the other (Astolfi, et al., 2021).

5.3 Strategies adopted to improve intelligibility and sound quality

The room and monitoring environment have a major influence on speech intelligibility; different people in different monitoring conditions have very different listening preferences: the listening volume, the playback acoustics with possible interference levels (open windows, children's screams, other background

noises), the quality of equipment and the different playback formats used all have an impact.
Increasing the loudness of the TV or hearing aids, using closed captioning, and streaming TV audio directly to headphones or modern hearing aids with an available wireless connection are some compensating techniques (Strelcyk & Singh, 2018).
Speech recognition is further enhanced by using TV adapters, which are augmentative TV-listening devices that are connected to the television (Sjolander, Bergmann, & Hansen, 2009). TV adapters enable the digital transfer of audio data from the television to the hearing aids over Bluetooth. Systems that use frequency modulation (FM) yield an enhanced signal-to-noise ratio (SNR). In any case, these gadgets are unable to offer improved SNR in relation to the broadcast mix. In a noisy listening environment, which is not the case in the situation under discussion, they increase the signal-to-noise ratio (SNR) (at least in the usual living room of hearing-impaired listeners).
Numerous audio options on most televisions can enhance the listening experience. Indeed, elderly people with high frequency hearing loss may attempt to make up for it by using EQ presets that automatically amplify higher frequencies and the upper mid-range while lowering the bass and lower mid-range frequencies.
TV broadcasters employ various strategies to enhance the intelligibility and perception of sound on television, ensuring a more engaging and immersive audiovisual experience for viewers. One common approach is the implementation of audio normalization techniques, such as loudness management, to maintain consistent volume levels across different programs and commercials. By reducing abrupt volume fluctuations, broadcasters aim to prevent viewer discomfort and enhance overall listening comfort.
Furthermore, broadcasters may utilize dynamic range compression (DRC) algorithms to mitigate the disparity between quiet dialogue and loud action sequences, ensuring that speech remains intelligible even during scenes with high levels of background noise or music. DRC helps to preserve the clarity of dialogue while preventing excessive volume spikes, thereby improving the overall balance and coherence of audio playback. (Baumgartner, Van Everdingen, Schreiner, & Krämer, 2022)
Moreover, broadcasters may employ advanced audio processing technologies, such as Dolby Digital and DTS surround sound formats, to deliver spatially immersive audio experiences that enhance the perception of sound on TV. These technologies enable the creation of multidimensional audio environments, with discrete channels for dialogue, ambient effects, and directional sound cues, enriching the viewer's auditory engagement with on-screen content.
Broadcasters may invest in high-quality audio production and mastering techniques, including dialogue editing, sound mixing, and mastering for broadcast, to ensure optimal sound reproduction on TV. By prioritizing clear and articulate dialogue delivery, broadcasters can enhance the intelligibility and perception of sound, facilitating viewer comprehension and immersion in the narrative. A set of recommended practices for programmers was released by the British Broadcasting Corporation (BBC) to enhance TV voice clarity throughout

the production process. The goal of the paper is to make it easier for those who have hearing loss to access BBC services and content. One of the primary points of the guidelines is this: "Intelligibility can be affected by a variety of factors, including unclear speech, strong or unfamiliar accents, background noise, and background music." When multiple of these problems come together, audibility might be seriously impaired (BBC, 2011).

Recently, a method based on sound source separation and recurrent neural networks (RNN) has enhanced the speech augmentation for listeners with hearing impairments. Real TV broadcast content was assessed to have superior sound quality and required less listening effort due to the separation of voice from background signals and remixing at a higher signal-to-noise ratio (Westhausen, et al., 2021). These methods have the potential to be more effective than simple equalization since they can improve SNR. Many problems are still unresolved, though, including the difficulty to process the signal in real-time, the need for a substantial quantity of training data, and the challenges associated with noise detection in voice-over-voice situations.

5.4 Introduction of the study

The collective auditory changes that occur with aging are termed presbycusis or age-related hearing loss (ARHL). It represents the most common cause of hearing loss, clinically affecting, albeit to varying degrees, all individuals over 40-60 years of age. Currently in Italy there are 12 million people over sixty-five years, and the elderly population is steadily increasing; it is believed that by 2050, the over-sixty-five population will quadruple compared to 2000. This is particularly relevant considering that the prevalence of individuals affected by hearing loss increases from less than 5% in the pediatric age to 70% in sixty-year-olds and 90% over eighty years old. Also, hearing impairment attributable to presbycusis is already demonstrable in 10% of individuals aged 45 to 55 years and in 25% of individuals aged 55 to 65 years (Yamasoba, et al., 2013).

Several theories have been proposed to explain the origin of cochlear suffering secondary to presbycusis (Lin, Thorpe, Gordon-Salant, & Ferrucci, 2011), including:

- Vascular hypothesis: the inner ear damage would be secondary to chronic ischemia consequent to atherosclerosis;
- Hyperlipidemic hypothesis: an imbalance in circulating fats in the blood (dyslipidemia), with higher values of cholesterol and triglycerides, drastically increases the risk of atherosclerosis, especially when accompanied by diabetes, kidney failure, or metabolic syndrome. There are several risk factors that can contribute to hypercholesterolemia, including diet, overweight or obesity, lack of physical activity, and smoking;
- Socioacusis: hearing loss secondary to exposure to environmental noise to which all individuals are constantly subjected in their daily routine. It is certainly a contributing factor for presbycusis, that justifies the more

pronounced hearing loss observed in individuals living in industrialized countries;
- Genetic hypothesis: it may explain the diverse behavior of hearing loss in different individuals of the same age.

The main symptom of presbycusis is the presence of high-frequency hearing loss, which results in difficulty perceiving the harmonic frequencies of vocal signals, while initially maintaining a relative normality of the audiometric threshold at low frequencies, allowing for the perception of fundamental frequencies and their first formants. As the years go by, there is a reduced ability to perceive sounds of high pitch, such as the doorbell or the phone ringing, and there is reduced intelligibility, initially in noisy environments and then in all conditions (typically expressed by the individual as "I hear sounds, but I don't understand words"). The hearing loss in presbycusis reaches severe levels in a small number of cases, in which the application of a CI may be helpful. Additionally, patients often complain of tinnitus, and this symptom becomes more frequent with advancing age. Alongside the hearing issues, the patient may also experience deterioration of mood and other emotional responses or may experience setbacks in their professional (if still active working) or social life (church, theater, attending conferences, cinema, concert hall). Furthermore, presbycusis can interfere with personal relationships (couples, family, friendships) or annoy others due to excessively loud volume kept on the radio or television.
With advancing age, physiological changes within the auditory system often result in a gradual decline in hearing acuity, presenting formidable challenges for older adults in navigating various auditory environments. As such, understanding the physiological intricacies underlying age-related hearing loss is paramount in devising tailored interventions aimed at ameliorating its consequences and enhancing the overall quality of life for affected individuals. Among the myriad of social activities, television viewing holds a central position in the lives of many older adults, serving not only as a source of entertainment but also as a means of staying informed and connected with the world. However, for hearing-impaired older adults, the TV listening experience can be significantly compromised, exacerbating feelings of isolation and hindering full engagement with audiovisual content. This thesis embarks on a multifaceted exploration, delving into the physiological dimensions of age-related hearing loss and meticulously examining its implications on the TV listening experience of older adults with hearing impairments. By elucidating the intricate interplay between auditory physiology and television consumption habits among this demographic, this research endeavor seeks to offer valuable insights that can inform the development of targeted interventions and technological solutions aimed at optimizing auditory capabilities and enriching the television viewing experience for hearing-impaired older adults.

In the present study, which led to the publication in the literature of the aforementioned original article, my contribution consisted mainly in the recruitment of subjects with presbycusis to be subjected to evaluations and in the execution of audiometric tests to assess whether the hearing condition corresponded to what was initially established . Furthermore, I directly participated in the data collection and writing of the original draft, writing part of the introduction, the materials and methods section and the discussion.

Preliminary study

A way to enhance the quality of voice on flat commercial TVs has been proposed in a prior work by the same research team of this study (Astolfi, et al., 2021). In such research project it was decided to test a Digital Audio Optimizer (DAO) boosts the audio signals toward a flat frequency spectrum while maintaining the same loudness level by dynamically equalizing the sound levels in real-time. This previous study was conducted on normal-hearing subjects, unlike the clinical study that will be described later, part of this doctoral thesis, which deals with subjects with presbycusis.
The DAO applies a Transfer Function (TF), called Heavy, to the TV station's audio signal's frequency spectrum prior to the transmission to the broadcast tower. Figures 10 and 11 represent the schematic illustration of the implemented system and the audio processing system inside the DAO, respectively.

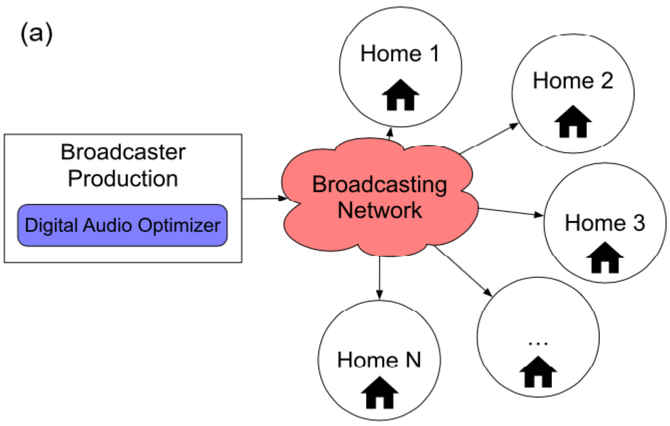


Figure 10: Schematic representation of the implemented system.

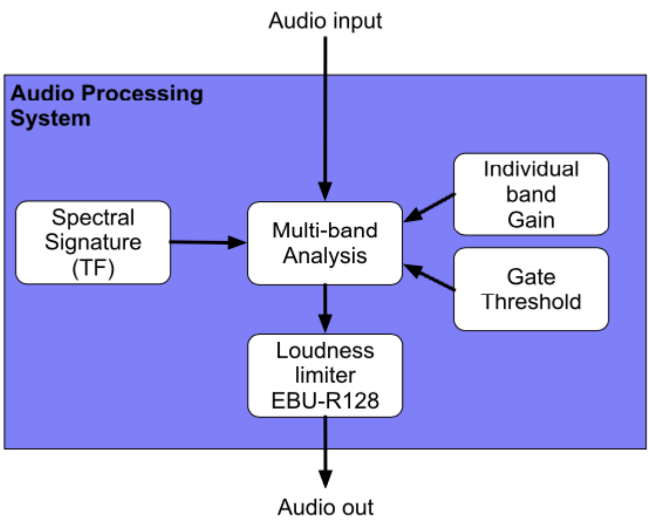


Figure 11: Representation of the audio processing system within the Digital Audio Optimizer (DAO).

The previous research aimed to improve the quality of speech when watching television.
It was conducted on three commercial TVs, whose frequency responses were recorded in both anechoic and ambient environments. It has been demonstrated that the sound pressure level for the three TVs drops with a slope from -7 dB/oct to -15 dB/oct above 2kHz.
Selections of audio clips from TV shows categorized into three categories: Speech, Singing and Music, and Sports. The tracks were ten-minute audio-visual snippets from television shows.
The considerations on the decline in sound pressure level for each of the three TVs and three genres in the frequency spectrum served as the foundation for the definition of the TF. Additionally, the TF was designed with the knowledge that the human ear is most sensitive in the range of 3 kHz to 4 kHz (Int. Org. Stand., 2003) and that the frequency range that is most crucial for speech intelligibility is 0.5 kHz - 4 kHz (Steeneken & Houtgast, 1999) (Int. Electrotech. Commission, 2020). Consequently, to provide a flat frequency spectrum in the 100Hz - 4kHz range, the TF boosts the energy beginning at the one-third octave band, 1kHz.
Using a "spectral signature," which stands for a reference curve, the TF is applied dynamically. The "spectral signature" dynamically increases and decreases the signal's spectral components as a multi-band filter (Junger, 2020). Figure 12 illustrates a "spectral signature" in action. The size of the white spheres indicates the highest gain, and Table 14 report the equivalent value. When the input signal's spectrum is compared to the reference curve, the equalization is activated if each band's signal level is higher than a predetermined gate threshold.

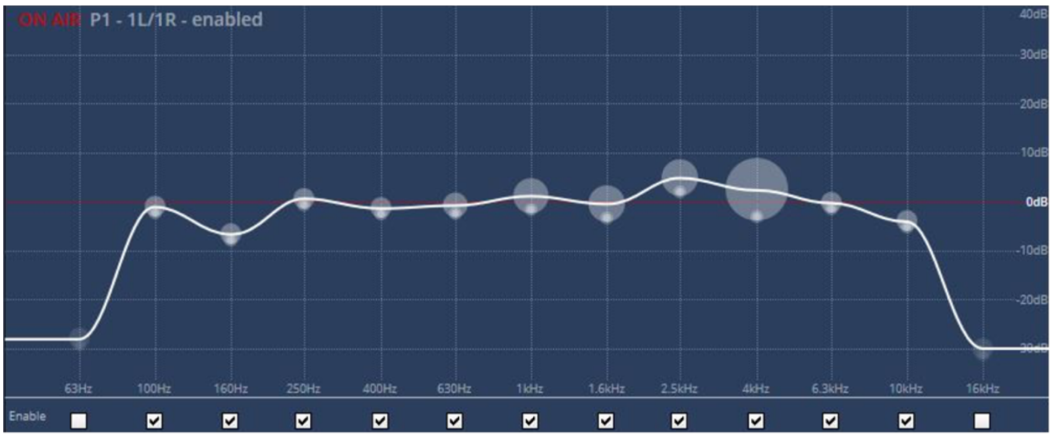


Figure 12: Frequency spectrum of the filter with the spectral signature represented by the solid white line and the maximum gain (white circles) for each one-third octave band center frequency for the Heavy filter.

Frequency $[Hz]$	Max Gain $[dB]$	Threshold level $[dB]$
100	2	-1.1
160	1.6	-6.7
250	1.5	0.6
400	2.2	-1.4
630	3.7	-0.8
1000	5.8	1.1
1600	6.1	-0.5
2500	6.1	4.8
4000	12.0	2.3
6300	1.1	-0.3
10000	0.7	-4.1

Table 14: Settings of the digital audio optimizer for each one-third octave band taken one every two of the Heavy filters.

A gate threshold is adjusted to stop noise in a band from getting amplified, notably buzz. Amplification won't occur if the energy in the band is less than the threshold. The distinction between the input and processed signals in the context of a voice sample is depicted in Figure 13. Under the frequency bands, a row of colored round circles - yellow or green, respectively - indicates whether the relative gate is engaged. The band is turned off, as indicated by the gray circle. There can be an individual gain for each band, which controls the amount of amplification or attenuation.

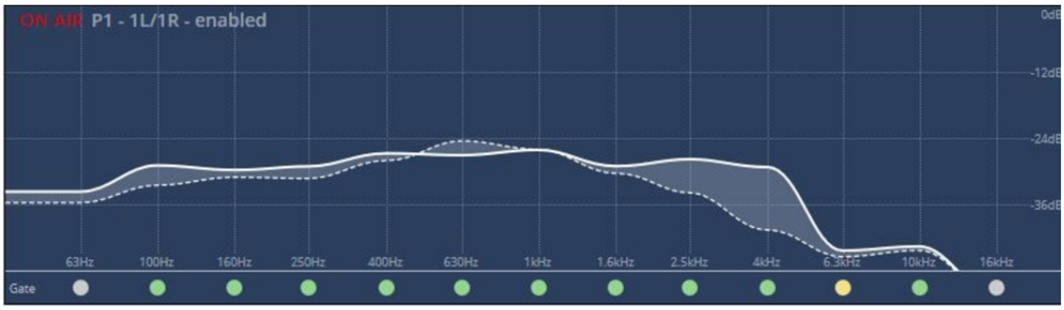


Figure 13: Difference between the input signal (dashed line) and the processed audio signal with the Heavy TF (solid line) in case of speech. The colored circles under frequency bands (yellow or green) indicate if the relative gate is activated or not.

The video stream was delayed by roughly 6 ms while the audio signal is being processed. This is because the audio processor needs time to equalize the signal. Ultimately, a loudness limiter block that complies with the EBU-R128 standard receives the processed audio.
Using the "double-blind triple-stimulus with hidden reference" test, a total of 72 normal-hearing participants between the ages of 21 and 53 participated in the subjective examination.
The tests are based on the “Subjective Difference Grade” (SDG), which is the difference between the evaluation of the filtered signal and the reference signal as shown in Equation:

$$SDG = Evaluation_{\text{signal fdt}} - Evaluation_{\text{hidden ref}}$$

The SDG ranged from −3 to +3, where +3 corresponds to a +100% speech improvement, while −3 to a −100% worsening. The next equation shows the SDG in percentage:

$$SDG_{\%} = SDG \cdot \frac{100}{3} [\%].$$

Results of the first study showed improvement in the audio quality of 25.3% on average, over the three TVs and the three genres.

5.5 Materials and methods

The study's objective was to evaluate the Heavy filter with patients with ARHL. The filter boosts high frequencies, maximizing the emphasis at 4 kHz. Research indeed indicates that frequencies higher than 3 kHz are crucial for hearing-impaired listeners to perceive speech, particularly in situations where background noise is present (Moore B. , 2016) (Hornsby & Ricketts, 2003). Based on the following considerations, hearing-impaired listeners can be subjected to the TF Heavy:

- The external auditory canal's (EAC) maximal resonance is reported to be between 2.5 and 4 kHz in human physiology research (Silva, Blasca, Lauris, & Oliveira, 2014);
- The most significant frequency range for speech intelligibility is between 0.5 and 4 kHz (Steeneken & Houtgast, 1999);
- The significance of medium frequencies for hearing is supported by Italian law, specifically by the National Institute for Insurance against Accidents at Work (INAIL), which applies a biological damage of 25% and 35% in the case of hearing loss at 1kHz and 2kHz, respectively, while only 5% of damage is attributed when hearing loss hits the 4kHz frequency (INAIL, 1994). Nevertheless, the literature emphasizes that the audiometric threshold at 4kHz and possibly 6kHz should be considered when assessing noise-induced hearing loss in a medico-legal context (Moore B. , 2016) (Gomez, Hwang, Sobotova, Stark, & May, 2001)
- Preferences for high-frequency amplification (up to 9kHz) are also revealed for wearers of hearing aids with mild-to-moderate hearing loss (Ricketts, Dittberner, & Johnson, 2008)

TV Selection

The “model A” commercial TV, which was the one that showed the highest average audio quality improvement in the subjective tests, was selected for this study out of the three that were employed with subjects with normal hearing. Its primary features are a 55-inch ultra-HD 4K display with 3840x2160 pixels, a Dolby Digital audio decoder, and two channels of downward-facing, 20-watt loudspeakers. The collected results from prior studies were considered in choosing the adopted TV. A convolution approach with an exponential sweep signal from 50Hz to 20kHz was used to do two observations. The first measure was carried out in an anechoic chamber to investigate the response free from reflection, and the second one was carried out in an ASL, which is a typical indoor listening environment, to assess how the TV would act in the room for the subjective test. Figure 5 illustrates the two responses that were received. The response in the anechoic chamber is flat between 100 Hz and 3 kHz, with a 40 dB dip at 1.5 kHz. The response decays at a rate of -14dB/oct above 3 kHz. The response in the ASL is often less flat and more jagged, decaying at a rate of -7 dB/oct above 3 kHz.

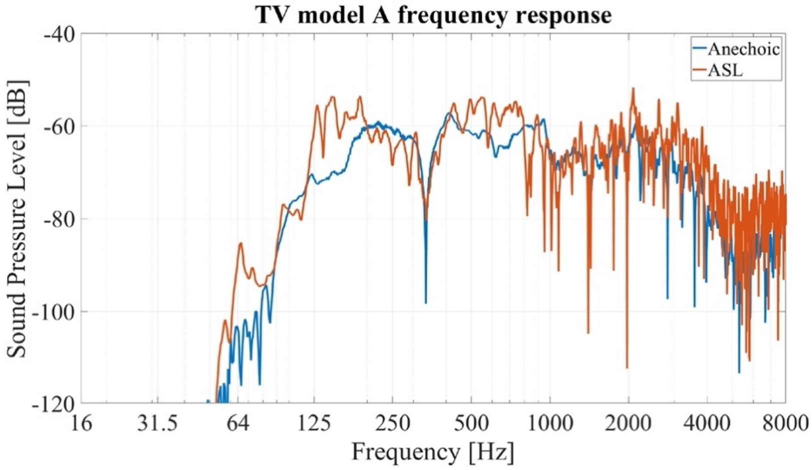


Figure 14: Frequency response of the model A TV-set in the anechoic chamber and in the ASL.

TV Tracks

The TV tracks that were chosen for the earlier trial with subjects who could hear normally were also used in this investigation (Astolfi, et al., 2021). The tracks are made up of thirty 10-second audio-video clips that were provided by Rai and taken from shows that were broadcast between 2017 and 2019. The three genres of "Speech," "Singing and Music," and "Sport" comprise the tracks. Speech include TV fiction, news, and movies. Commentaries on sporting events belong to the sports genre. Specifically, volley, cycling, soccer, and basket. A few vocal samples from TV music shows and the Sanremo Festival were used in "Singing and Music". Specifically, there were 18 tracks for "Speech" - 12 fiction, 6 news, and 12 movie scenes - 5 "Singing and Music," and 7 tracks for "Sport," which includes 2 soccer, 2 cycling, 2 volley, and 1 basket event commentary.

Figure 15: Examples of Sport, Music and Speech programs.

Every single one of the chosen fragments had singing or speech with almost little background noise. Each track's frequency content was calculated, and the average spectrum for each genre was retrieved. For "Speech" and "Sport," the average spectra show a similar tendency beginning at 600 Hz, with a slope of roughly -10 dB/oct. For "Singing and Music," the trend is a little less steep.
This was considered when designing the TF, as signal processing applied in this frequency range can improve speech intelligibility across all genres.

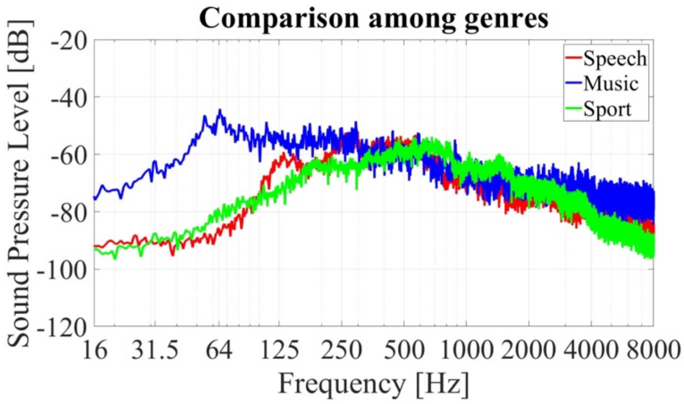


Figure 16: Overlapping of the average frequency spectra for each genre.

Selection of Subjects

In the study testing the Heavy TF, thirty-one elder listeners, ranging in age from 62 to 85 years (mean = 71.7 years; SD 5.9), took part. Five people took part in a preliminary pilot test, while twenty-six subjects took part in the main trial. Pure-tone thresholds were assessed in a sound-attenuated booth at octave and interoctave frequencies ranging from 250Hz to 8kHz using TDH-39 headphones and a clinical audiometer (Triangle, Inventis Srl, Padova, Italy). Every participant exhibited a high-frequency sensorineural hearing loss, albeit slight in certain cases, in line with ARHL.
At high frequencies, listeners' hearing loss was symmetrical mild to moderate on average. The mean hearing threshold of the 26 individuals with ARHL is displayed in Figure 17. All patients had normal otoscopy and none of them had a medical history of major ear illness, persistent exposure to excessive noise, or use of potentially ototoxic drugs. Individuals who had conductive hearing loss (10 dB or more in the air-bone gap) that was clinically relevant were not allowed to participate.

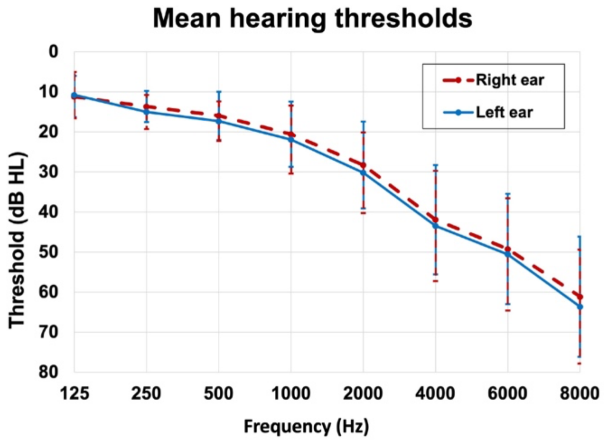


Figure 17: Mean hearing threshold of the 26 elderly subjects with ARHL

Subjective Tests in ASL

Five subjects with ARHL who were not engaged in the study and who were between the ages of 65 and 75 took a preliminary exam to test setup and instructions: the main test was started after the preliminary test produced satisfactory findings in terms of comprehension and tool usability for this group of older adults. The Audio Space Lab at Politecnico di Torino was used for conducting the research.

Every participant sat on a chair two meters away from the TV. During the test, the subjects were only shown the audio extract of the track; the only visual element was the TV user interface, which was used by the subjects to provide their subjective assessments (Fig. 18).

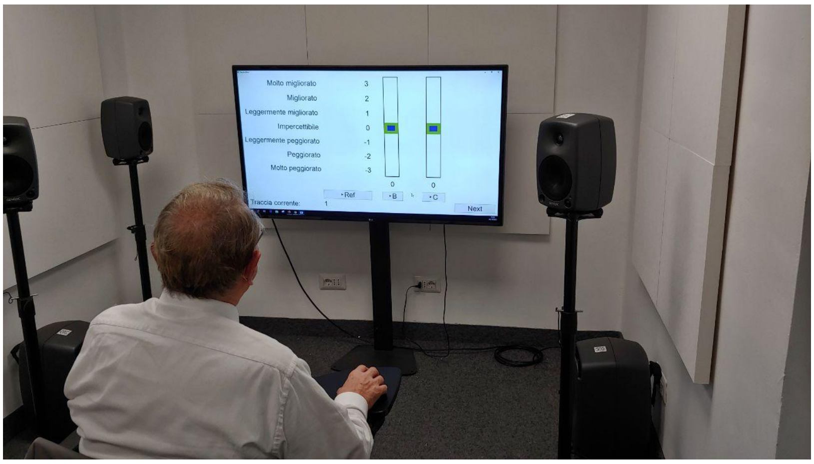


Figure 18: Subject conducting the test, judging the quality of the adjusted tracks compared to the reference.

The participants were asked to score each excerpt's speech quality in terms of voice clarity, noting any improvements or declines compared to the reference track.

The entire exam consisted of two sections:

- Training (15 minutes): During this time, the subject was shown the test's interface by the conductor who was present in the room. Every listener received three extracts. The TV volume was asked to be adjusted by the subject to his or her comfort level, which was maintained during the main test.

- Main test (30 minutes): The listener was requested to listen to and assess the 30 excerpts after being left alone in the room.

To prevent distractions from the visual portion of the sample, it was agreed to exclusively use the audio portion during the tests. The user interface was adjusted to full screen on the TV at startup to accommodate the older subjects' possible visual impairment. A single SDG was calculated for each subject by averaging the SDGs associated with each recording within each genre.

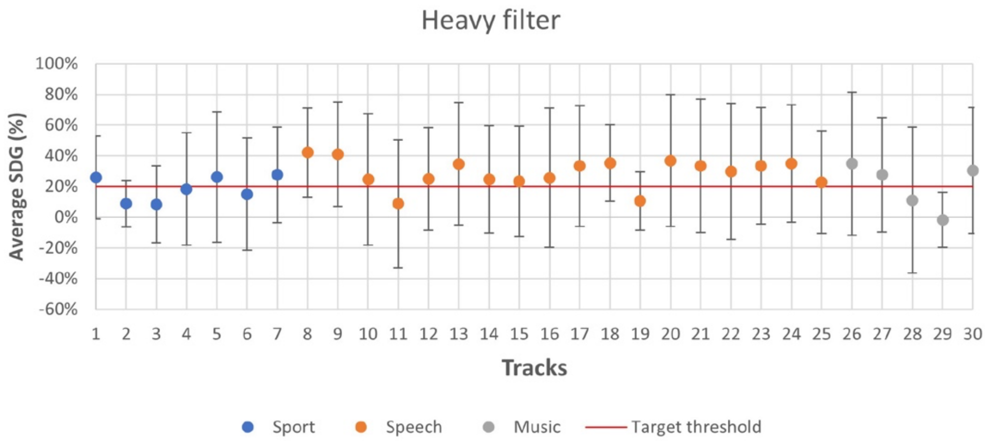


Figure 19: Average Subjective Difference Grade for each track as a result of the subjective tests with 26 hearing-impaired listeners. The genre Singing and Music has been shortened in Music.

5.5 Results and discussion

The average SDG for each track into the categories of "Speech," "Singing and Music," and "Sport" is shown in Figures 20-22. Most of the tracks, across all genres, had an average SDG higher than 20%, which was the lowest limit Rai asked us in terms of overall increase of the perceived audio quality. Speech, singing and music, and sports categories had average SDGs of 29% (SD 9.2%), 21% (SD 15.3%), and 19% (SD 8.2%), respectively. The values and standard deviations for speech, singing and music, and sports were comparable to those found with normal hearing listeners. The intrinsic variability of the subjective tests is the cause of the quite large standard deviations. With a SD of 10.9%, the overall average SDG across all genres was 24.3%, meeting the 20% target. This finding is consistent with the average SDG of 29.9% that was achieved in the previous investigation with normal hearing subjects on the same TV set. The slightly lower SNR in the audio sample of both sport and singing and music genres, which intrinsically involve noise or music overlapped with speech - could be a plausible reason for the lower mean SDG for these two categories as opposed to speech.

The study's goal was to enhance flat TV speech quality in speech samples with extremely low noise. For this reason, SNR enhancement-based techniques have not been considered. Regarding audio processing based on SNR enhancement in the broadcasting industry, there are still a lot of unresolved concerns. Specifically, one of the most promising and recently proposed methods for enhancing perceived voice quality and lowering listening effort in TV broadcasting is source separation and subsequent remixing at a higher SNR compared to the original mix (Westhausen, et al., 2021). This technique accomplishes the objective of making speech perception easier while retaining as much of the original acoustic environment as feasible. Based on the accuracy of the estimate of the ambient noise from the speech pauses, the method's results demonstrate that it can lower the listening effort by 2 points out of 13 for common background noise settings for "Music," "Sport," and "Environment." Because there is a backdrop sound during speech pauses and the algorithm does not recognize it as background noise, the method is ineffective in voice-over-voice situations. The primary challenge with this approach includes selecting the optimal architecture to accomplish speech separation in real-time without requiring the listener to do any particular action. The range of 0.5 kHz to 4 kHz is the most important for speech intelligibility, and it is well known that the human ear is most sensitive in the range from 3 kHz to 4 kHz (Int. Org. Stand., 2003). Furthermore, the Heavy TF obtained here for the hearing-impaired listeners, which boosts the amplification of 4kHz, is somewhat improved. These may be the causes of the Heavy filter's enhancement to the point where it is audible to listeners with normal hearing.

We performed statistical analysis on both subject categories to further explore the importance of the results acquired from both the normal hearing and the hearing impaired participants.

For the three genres "Speech," "Singing and Music," and "Sport," Fig. 9 displays the occurrences of the SDG scores ranging from -3 to +3. The most frequent

grade for both topic areas is 1 (Slightly improved) and this rating preference was consistent with findings from people with normal hearing.
Statistical findings indicate that when compared to tracks without the transfer function, listeners consistently rated the ones with the implemented transfer function as being better.

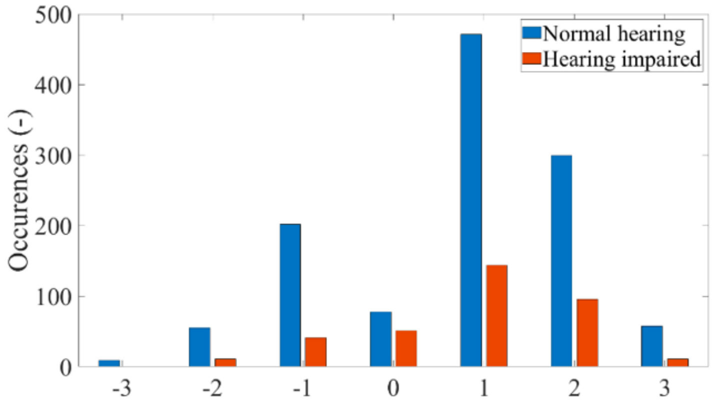


Figure 20: Occurrences of the SDG scores for Speech category.

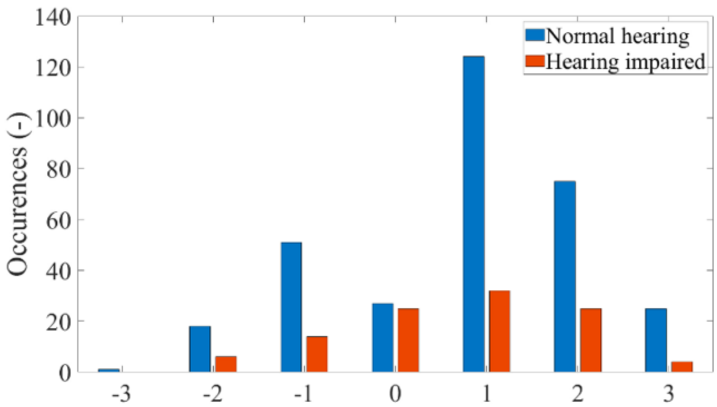


Figure 21: Occurrences of the SDG scores for Singing and Music category.

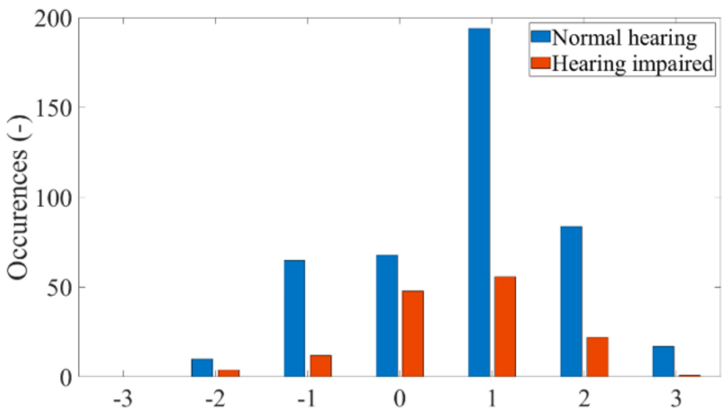


Figure 22: Occurrences of the SDG scores for Sport category.

5.5 Conclusion

In this study, we delve into the realm of enhancing the audio quality of commercial flat-screen TVs by implementing real-time modulation of the audio signal. The experiment focuses on assessing the perceived audio quality improvement of a commercial TV set, particularly for individuals with age-related hearing loss.

As the population ages, there's a growing recognition of the importance of accommodating the specific auditory needs of older viewers. With advancements in sound technology and the integration of various features aimed at enhancing clarity and accessibility, modern flat-screen TVs are poised to revolutionize the television viewing experience for individuals with age-related hearing impairments.

The crux of the enhancement proposed lies in the implementation of a Transfer Function (TF) that dynamically modifies the audio signal in real-time. This TF is strategically designed to augment the frequency spectrum of the audio signal sourced from the Italian radio and TV broadcasting company Rai, right before transmission to the broadcasting tower. Notably, the TF amplifies the frequency range crucial for speech intelligibility, predominantly targeting frequencies between 1 kHz to 4 kHz. A special emphasis is placed on boosting the one-third octave band centered around 4 kHz, known to significantly impact speech clarity.

Subjective testing involving individuals with age-related hearing loss showcased notable improvements in perceived audio quality. These enhancements not only met the stringent 20% improvement benchmark set by Rai but also aligned closely with results from a similar study conducted on listeners with normal hearing capabilities. Statistical analysis of the subjective assessments revealed that subjects unanimously judged the applied transfer function to yield a "Slightly improved" perceived audio quality across the audio tracks tested.

This empirical evidence underscores the efficacy of real-time modulation of the audio signal in elevating the auditory experience of flat-screen TVs, particularly catering to individuals with age-related hearing impairments. The strategic application of the Transfer Function not only enhances speech intelligibility but also ensures a more immersive and enjoyable audio experience for viewers across various content genres.

The proposed TF, thanks to these results, has been heavily promoted on social media channels by RAI and has been subsequently implemented in all broadcasts offered on RaiPlay and on live TV channels Rai Premium, Rai Movie, and Rai Storia.

Focus on paper V

Title: Audiovisual recording and reproduction of ecological acoustical scenes for hearing research**:** a case study with high reverberation

Authors: Guastamacchia A, Puglisi GE, **Albera A**, Shtrepi L, Riente F, Masoero MC, Astolfi A.

Published on: *European Acoustics Association 2023* – pp. 1765-1772 - 2023. DOI:10.61782/fa.2023.0666

Objective: To validate a proper laboratory set-up and procedures to perform ecological tests for the assessments of intelligibility in noise, under the complex acoustical conditions of everyday life environments.

Materials and methods: Speech intelligibility tests based on seven audio-visual scenes were administered inside an immersive virtual 3D environment reproduced through a spherical 16-speaker array synced with a head-mounted display. Audio-Visual scenes were collected in a medium-sized reverberant conference hall through in-field 3rd-order Ambisonics impulse response recordings and 360-degree stereoscopic video shootings. The visual scenes included cues on the spatial location of the sound sources without lip-sync-related cues.

Results: Both normal-hearing subjects and patients with hearing loss wearing hearing aids were submitted to assessments. The audio-only tests in the static condition resulted in the highest speech intelligibility scores, followed by a tie between audio-visual with self- motion and in the static condition, and the lowest scores were observed in the audio-only with self- motion test. Overall, speech intelligibility decreased as the target-to-listener distance increased in all noisy scenes. Additionally, speech intelligibility increased when the noise azimuth was at 120° compared to 180° and 0°.

Conclusion: The results suggest a significant spatial release from masking in the presence of reverberation and one-talker interfering noise. The laboratory set-up, consisting of AV scenes directly recorded in everyday life conditions such as a conference room through in-field 3rd-Order Ambisonics recordings, allows us to demonstrate the full feasibility of carrying out intelligibility tests in virtual ecological conditions. Furthermore, this is a highly reproducible system, which is suitable for replication in medical research environments and for use in clinical practice.

6.1 The pathophysiology of hearing in noise

About 90% of adult hearing impairments and 75% of childhood hearing impairments exhibit a loss manifesting from 500 Hz to 4,000 Hz, with a morphology of the audiometric threshold curve descending. This data is clinically relevant because it is known that speech components are weaker at high frequencies compared to low frequencies, and speech loudness depends mainly on low-frequency components. Consequently, individuals with sensorineural hearing loss with the morphology described above may perceive speech loudness well and may not be aware of losing signal comprehension due to hearing deficits. Hearing loss affecting frequencies starting from 500 Hz can lead to difficulties in discriminating, for example, the vowel "o" from the vowel "i". Indeed, the two vowels differ in the second formant, which in "o" is around 700-1,000 Hz, while in "i" it is at 3,000 Hz. Considering these findings, we can understand the not uncommon statements of subjects with such hearing impairments, who report that "speech is loud, if only people didn't mumble words ...".

However, the reduction in speech intelligibility in a noisy context cannot be explained solely by the decrease in auditory capacity. In 1986, Plomp described a model that considers two independent factors for the degradation of verbal recognition in the case of sensorineural hearing loss: the attenuation factor and the distortion factor (Plomp, A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired., 1986). The attenuation factor simply represents a worsening of the hearing threshold. If only this factor were present, as in purely conductive hearing loss forms, deterioration in comprehension would only occur at low noise levels, and conversely, comparable performance to normal-hearing individuals would be achieved at high noise levels (evaluating the signal-to-noise ratio as the outcome).

The distortion factor is due to the reduction of the following functions:

- Dynamic range of loudness;
- Discrimination and frequency resolution;
- Temporal resolution.

Although it is confirmed that sensorineural hearing loss is associated with the distortion factor (Hülsmeier, et al., 2022), which includes the aforementioned three parameters, the relationship between the latter and the elevation of the threshold is not proportional. In fact, even in cases of mild sensorineural deafness, difficulties in understanding speech in noise can be significant.

The factor we have defined as distortion can be exclusively related to peripheral deficit or can be secondary to a central deficit. In some cases, as in the elderly hearing-impaired patient, it is secondary to alterations of both peripheral and central functions simultaneously.

At the peripheral level, a simple and common hearing loss with alterations of the outer hair cells (OHCs) is responsible for the loss of cochlear analysis function with alterations in the three parameters mentioned above.

The reduction of the dynamic range of loudness occurs because, in the case of sensorineural hearing loss, there is a greater increase in the hearing threshold

compared to the discomfort threshold, both due to the loss of the cochlear amplification mechanism for low intensities and due to the loss of the protection mechanism operated by the efferent system through action on the OHCs.
In a damaged cochlea, a narrow-band sound produces broad stimulation that is not selective in a limited region of the basilar membrane. Therefore, if cochlear resonances are widened following OHC pathologies, there is a *loss in frequency discrimination* (Tyler, Wood, & Fernandes, 1982). If background noise contains energy at frequencies close to the narrow-band signal, the ear is unable to separate the two components and there will be an overlap. The excitement due to noise invades regions of the basilar membrane affecting neurons that are not normally responsive to the considered noise. This condition leads, at the brainstem level, to functionally "widened" critical bands with difficulty in "resolving" the fundamental frequency of speech signal, which is an important requirement for comprehension in difficult listening conditions.
The *reduction of temporal resolution* capacity of the hearing-impaired subject also determines difficulties in listening in noise. The first aspect to be altered is the ability of the basilar membrane to follow cycle-by-cycle the timing of the complex waveform of the incoming signal, which allows, at subcortical levels, to analyze the fine envelope of speech. The second aspect is that louder sounds can mask weaker sounds that immediately precede or follow (Zwicker & Schorn, 1982).
This alteration compromises two aspects of comprehension: the recognition of transients (even in silence) and the listening in the gaps. Transients are frequency variations of the vowel formants that precede or follow the consonant. A reduction in temporal analysis reduces the ability to correctly analyze transients, resulting in difficulties in verbal recognition. Moreover, noise in daily life often features rapid fluctuations. These fluctuations, in the case of an undamaged cochlea, allow for the extraction of fragments of information during moments when the background noise is weaker. This ability is called listening in the gaps and presupposes good temporal resolution.
The reduction of the temporal analysis capacity of the incoming signal is mainly responsible for the significant comprehension difficulties in noise observed in the elderly.
It is also certainly true that a part of hearing difficulties in noise in elderly patients may result from a reduction in cognitive processing in general; for example, the elderly are more vulnerable than young people to the distracting effect of semantic content in a noisy environment (Göthberg, et al., 2023). Indeed, comprehension in noise in the elderly heavily involves working memory and attention, which will be available for the semantic aspect only to a lesser extent (Arehart, Souza, Baca, & Kates, 2013). For this reason, to achieve adequate verbal recognition in elderly patients, it is necessary to significantly increase the signal-to-noise ratio.
In summary, the pathophysiology of hearing in noise is a complex field that requires a thorough understanding of the sensory and cognitive challenges involved in verbal comprehension in noisy environments.

6.2 Environmental noise and hearing loss

We all inhabit a world constantly enveloped in noise, whether it's the bustling streets of city centers, the hum of activity in workplaces, or the background buzz within our homes filled with the sounds of radios, TVs, smartphones, and various appliances. Even in seemingly quieter settings like restaurants, the noise of traffic outside can make conversations challenging to follow. Despite efforts to minimize occupational noise exposure and promote awareness of hearing protection, noise-induced hearing loss (NIHL) persists as a significant issue in developed nations.

Defining noise isn't straightforward, as it can carry dual meanings. On one hand, it can denote any bothersome sound event, but this perception varies among individuals. On the other hand, noise can be described as an acoustic event lacking a discernible mathematical relationship between its frequency components. In non-work settings, noise exposure can be both unavoidable and avoidable, depending on personal choices. A 2015 study by Eurisko, involving 8800 adults across 46 cities in 11 developed countries, found that car and public transportation noises were the most annoying, followed by background music, TV, radio, conversations, and household appliances (Eurisko, 2015). Environmental sounds of lower intensity, such as mobile phone rings, background music, and crowded room conversations, contribute to a continuous noisy backdrop. This background noise, along with reverberation, can sometimes overwhelm conversations, posing challenges, particularly for individuals with hearing impairments.

Background noise poses a particularly insidious threat to speech intelligibility—the ability to understand and interpret spoken words with clarity. As individuals vie to make themselves heard above the din, the clarity of their messages becomes compromised, leading to misunderstandings and misinterpretations.

The impact of noise on communication affects everyone, regardless of age, gender, or lifestyle, in various aspects of daily life. This interference can lead to hearing impairment and, in some cases, even jeopardize personal safety, for instance, causing accidents due to reduced awareness of surroundings. Additionally, noise can hinder children's education and learning, impact health by causing vocal issues or hearing loss, and impede social engagement among the elderly.

Older adults commonly exhibit degraded auditory processing, particularly in perceiving the amplitude envelope of sounds. The significance of such cues varies depending on the task at hand, whether involving speech or music, and the nature of age-related hearing loss, whether related to hair cell or neural damage. Simulated temporal aspects of auditory processing suggest inter-dependencies between periodicity coding and speech envelope cues (Thoidis, Vrysis, Markou, & Papanikolaou, 2020). Jittering, a simulation technique, has been shown to negatively affect memory, consequently reducing working memory span. The role of cognitive factors, particularly attention, is evident in self-reported age-related differences in Speech and Spatial tasks. Additionally, strong correlations have been identified between behavioral measures of hearing and cognition, as well as between self-reported listening abilities and socio-emotional status.

To enhance our comprehension of the interactions between auditory, cognitive, and socio-emotional factors in everyday life, longitudinal studies are imperative. It remains unclear whether self-appraisal precipitates or results from auditory declines in older adults, and whether auditory declines precede or follow cognitive declines. Addressing these uncertainties could facilitate the development of better services for treating older adults.

6.3 Introduction of the study

Speech quality, defined as spoken communications to one or more listeners, is measured in terms of intelligibility. Traditionally, this has been done by calculating the proportion of a speaker's words that the listener correctly understands relative to the total number of words spoken. Satisfactory intelligibility denotes understanding of sentences with a proportion of no less than 95%. Unfortunately, most audiological tests to evaluate speech intelligibility are very outdated and almost all of them are not adapted to be performed with background noise performance. Air conditioning system noise, background talk, bad acoustics, etc. can all interfere with understanding spoken words. It is incorrect to evaluate noise only in terms of the highest and lowest sound levels because noise in an ecological environment is typically erratic and impulsive rather than stationary. Instead, it is important to determine an appropriately averaged sound level that is representative of the duration of the event itself. Practical cases where a correct understanding of the voice is essential are numerous: ranging from theaters, conference halls, cinemas, school classrooms, etc. The list should also include other public places such as railway stations, airports, churches, or supermarkets, where understanding verbal messages may be important for safety implications (risk and/or danger warnings). In the above-mentioned cases, namely in ecological environments and situations that each of us experiences daily, it is important to understand the message conveyed through speech, which manifests itself through the voice.

If an intelligible vocal signal is radiated in a room or outdoor environment, it will inevitably undergo alteration due to the presence of background noise; similarly, a vocal signal radiated in an environment with a high reverberation time undergoes profound alterations due to the overlay of countless reflections produced by the walls of the room or the furnishings present. It is therefore fundamental to control the phenomenon of noise masking and its ability to influence verbal communication, especially in critical environments such as a conference room, where the perfect understanding is crucial for work purposes or science.

In general, most people placed in unfavorable acoustic conditions present a comprehension problem. Considering also that life expectancy is continuously increasing, as are age-related hearing impairments (presbycusis), in the coming years, we will be increasingly confronted with an elderly population with communicative difficulties in noise and therefore consequent limitations in social activities.

Imagine a subject with bilateral moderate hearing loss during his daily routine: he will probably still be able to recognize and discriminate correctly most individual sounds and noises encountered during the day, as well as maintaining vocal comprehension towards family and friends in relatively quiet situations. But if the same person is placed daily in an ecological environment, such as the noise of a chaotic city center, he will likely experience a reduction in intelligibility values. However, a peculiar aspect of this phenomenon is that the extent of this decrease depends on a series of subjective and environmental characteristics: there is indeed a great variability both in the responses of the same individual to the same noise in different situations, and among different individuals in the same situation. Current scientific research is focused on further enhancing the examination itself to recreate an acoustic masking situation that is as faithful to real life as possible. A test that is perfectly suited to be carried out in noisy conditions is the Matrix Sentence Test (Puglisi, et al., 2015), already explained extensively in the first part of this thesis. Having the Matrix Test to be included in routine clinical practice to cope with intelligibility evaluations with ecological masking noises would allow for improving the study of vocal comprehension in hearing-impaired patients, monitoring over time their "social discomfort," and comparing the potential auditory benefit of proper acoustical prostheses in real-life settings.

Speech intelligibility is the main goal of acoustics in both large and small classrooms, conference and court halls, restaurants, and other places where speech communication is the primary means of communication. To guarantee the ecological validity of the results, speech testing has been usually conducted in laboratories that faithfully replicate real-life acoustic settings (Van De Par, et al., 2022). Thus, the task of our project was to recreate Audio-Visual (AV) situations such that users behave as though they were physically present in the environment tested and feel fully immersed in the virtual area (i.e., recalling natural movements of the head, torso, and eyes that maximize voice recognition (Grimm, Hendrikse, & Hohmann, 2020).
This becomes even more pertinent when considering participants who use hearing aids, as the directional filtering integrated into these devices relies heavily on the listener's head orientation (Abdipour, Akbari, Rahmani, & Nasersharif, 2015). To enhance the authenticity of depicted scenarios, it's preferable to construct them using real-life audio and video recordings of communication situations rather than simulations. While simulations offer optimal flexibility for research purposes, allowing for quick modifications as needed, some studies indicate a preference for real videos over virtual renderings (Llorach, Hendrikse, Grimm, & Hohmann, 2020) (Hendrikse, Llorach, Grimm, & Hohmann, 2018). Despite being less adaptable, video recordings prove to be more effective in situations requiring high realism, especially in relatively simple scenes with few actors or vehicles (Llorach, Grimm, Hendrikse, & Hohmann, 2018).
However, striving for lifelike listener experiences by replicating realistic immersive AV environments might not suffice for genuinely ecological listening tests. While Grimm et al. emphasized the importance of Self-Motion for greater ecological validity, they also noted its relevance might vary based on factors like

environment, age, noise level, task, and instructions (Grimm, Hendrikse, & Hohmann, 2020). Therefore, future efforts to enhance the ecological validity of speech intelligibility tests should consider incorporating real-time social interactions between speakers and listeners. Additionally, studies have shown that head orientation significantly contributes to enhancing the Signal-to-Noise Ratio (SNR), especially in subjects with hearing loss, with participants often orienting themselves in ways that improve SNRs rather than facing the sound source. Visual cues in this context may aid in sound source localization but can also serve as potential distractors, as individuals commonly look away from the primary talker during multi-talker conversations (Hendrikse, Eichler, Hohmann, & Grimm, 2022).

Beyond self-motion, visual cues have varying impacts on ecological validity: contextual and source-related visual cues affect localization and acceptance of auditory illusions, while observing face and mouth movements significantly aids speech comprehension. These factors have spurred recent investigations into the role of visual cues in speech intelligibility tests. Virtual renderings of contextual and source-related cues were explored. For instance, reverberant scenarios were simulated, presenting a virtual ring of loudspeakers to indicate possible noise sources, while the target speech lacked visual correspondence. Results indicated worse speech intelligibility scores in reverberant conditions but no significant differences between Audio-Only (AO) and Audio-Visual (AV) tests (Fichna, Biberger, Seeber, & Ewert, 2021). Similarly, a reverberant AV scenario was examined, with a fixed interfering talker and a target talker changing positions around the listener. Significant differences were observed, particularly when the target talker was positioned at 90° or -90° azimuth: the enhancement brought by the AV condition suggests that participants likely utilized visual cues for spatial orientation, leveraging self-motion to improve speech intelligibility (Hládek & Seeber, 2023).

Recently, an open-source database of audiovisual (AV) environments was published with the aim of promoting ecological auditory research and facilitating collaboration among laboratories (Hládek, Van de Par, Ewert, & Seeber, 2021). The initial contributions, detailed in [1], involved multi-channel recordings of Room Impulse Responses (RIR) conducted in real-world settings to enable the auralization of speech intelligibility (SI) tests. These recordings were coupled with virtual renderings of visual scenes for three distinct environments. However, only a limited number of studies have attempted to address SI measurements using genuine recordings of visual scenes, while also considering the influence of lip-reading on the target source.

For instance, Seol conducted SI tests using a 360° video depicting a café scene, featuring a conversational partner in the foreground and chatting customers in the background (Seol, Kang, Lim, Hong, & Moon, 2021). While participants demonstrated improved speech recognition with visual cues, the study didn't fully replicate the acoustic conditions of the café environment, as it presented anechoic speech with unmatched background noise. Similarly, Moore incorporated a one-talker video recording into a 360° video to account for lip movements, but the masking noises lacked a visual counterpart. Nevertheless, comparing speech

intelligibility tests in Audio-Only (AO) and Audio-Visual (AV) conditions revealed a significant improvement in Speech Reception Thresholds (SRT50) of up to approximately 9 dB in the AV condition when the speaker and noise source were co-located (Moore, Green, Brookes, & Naylor, 2022).
Despite these efforts, further exploration into the role of visual cues, particularly in virtual reality contexts, is warranted, encompassing all types of visual cues for all sound sources.
In complex listening scenarios involving a target speaker and multiple speech sources at varying azimuths and distances, researchers often investigate the Spatial Release from Masking (SRM) effect. SRM refers to the enhancement of intelligibility in noise when speech and noise sources are spatially separated. While SRM in reverberant environments has been studied since the 1970s, its efficacy in immersive virtual reality settings requires deeper examination. Previous researches by Plomp and Kidd highlighted contrasting findings regarding SRM in reverberant conditions, suggesting that listeners utilize the "precedence effect" to distinguish the target from the masker (Plomp, 1976) (Kidd, Mason, Brughera, & Hartmann, 2005). Differently, other studies further investigated SRM under various reverberation conditions, shedding light on its impact on intelligibility, particularly on classroom acoustic scenario (Puglisi, Warzybok, Astolfi, & Kollmeier, 2021).
In our study, we collected AV scenes in a medium-sized conference room using in-field 3rd-Order Ambisonics (3OA) RIR recordings and 360° stereoscopic video footage. The scenes include spatial cues for sound source location but lack lip-sync cues, which will be included in further studies on the topic. Speech intelligibility tests were administered using a spherical 16-speaker array and a Head-Mounted Display (HMD) to normal-hearing participants. We examined the effect of high reverberation on SRM in the presence of one-talker noise, different listener-to-target talker distances, and the influence of video recordings and self-motion of the head.

In the present study, which led to the publication in the literature of the aforementioned original article, my contribution consisted mainly in the conceptualization of the project and the formulation of general research objectives and goals as well as in the development of the methodology and the idea of adapting the speech material of the Italian Matrix Sentence Test as a basis for the administration of our speech perception test in noise. Furthermore, I was personally involved in the writing and submission of the Institutional Review Board application, as well as I contributed to the writing of original draft of the study to be submitted to peer-review.

6.4 Materials and methods

Participants

Forty native Italian speakers with normal hearing, who were naïve to the study, participated voluntarily. Among them, there were 36 males and 14 females, with ages ranging from 22 to 49 years (average 28.5 years). Prior to participation, all individuals underwent a pure-tone audiometry test to ensure they did not have any hearing loss (HL) that could affect the test results. A maximum threshold of 16 dB for the average pure-tone hearing at each ear from 500 Hz to 4 kHz was chosen as inclusion criteria. Participants either had normal vision or wore corrective lenses, and they did not have any conditions that could affect their mobility. Prescription glasses were permitted during the experiments as they did not compromise the integrity of the test.

AV Scenes and their acquisition

The highly reverberant conference hall located within the Egyptian Museum of Turin was selected as the recording environment for the audiovisual (AV) scenes. This hall, known for its challenging acoustics, presents a typical setting where clear speech comprehension is crucial. With a volume of 1500 m^3, the hall is minimally furnished, featuring 100 light chairs and two wooden tables. One table is positioned on a 30 cm high wooden stage at the front, designated for the main speaker, while the other table serves as the control station for the two-loudspeaker amplification system at the back. Figures 23 and 24 depict a 3D model of the hall, indicating the positions of the loudspeakers and the main talker at the front table, and a photograph of the related space.

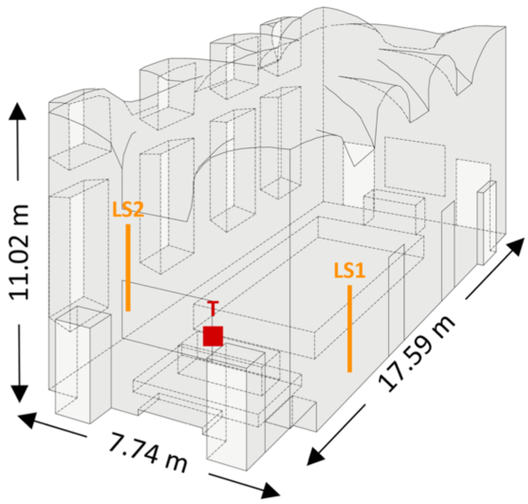


Figure 23: 3D model of the conference hall: LS are the room loudspeakers and T is the position of the target talker.

Figure 24: Picture of the conference hall taken with the same orientation as the 3D model above.

Seven distinct scenes representing common communication scenarios within the hall were defined and captured. Each scene portrayed a unique spatial arrangement for the listener, the main talker, and the interfering talker representing the source of background noise. To authentically replicate the hall's usage, the main speech was consistently projected through the two room loudspeakers positioned on the side walls. Figure 25 illustrates the floor plan of the conference hall, delineating the positions of the room loudspeakers (LS), as well as the designated locations for the listener (L), the main talker (T), and the interfering talkers (N) across all seven scenes.
Specifically, two listening positions within the audience area were selected, positioned at sitting positions approximately 1.2 meters above the floor. These positions included one closer and one farther from the main talker seated behind the table, positioned at a height of 1.5 meters from the floor. Furthermore, for each listening position, one-talker interfering noise was introduced from at least two different directions (always at 1.2 meters from the floor). This allowed for the assessment of speech recognition variations when the noise was presented alongside the main talker at 180° or 0° azimuth, compared to when the noise was spatially separated at 120° azimuth.

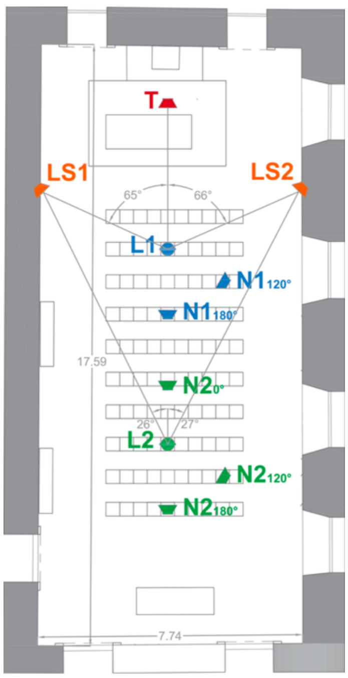


Figure 25: Conference hall floor plan with locations of loudspeakers (LS1, LS2), target speech (T) and competitive noise (N1120°, N1180°, N2120°, N2180°, N20°) sources for all listening positions (L1, L2).

To capture the audiovisual (AV) scenes, 4K 360° stereoscopic videos and 3rd-Order Ambisonics Room Impulse Responses (3OA RIRs) were obtained. The recording systems were positioned at the listening locations, with the sound source placed either at the target or noise positions, oriented towards the listening point. However, for the N20° position, the sound source was rotated 180°. The sound source utilized was the NTi Audio Talkbox acoustic signal generator, known for its flat frequency response from 100 Hz to 10 kHz and energy distribution resembling that of the human voice's polar diagram. Audio recordings were captured using the 19-capsule Spherical Microphone Array (SMA) Zylia

ZM-1, which offers a nominal flat frequency response from 28 Hz to 20 kHz, while video acquisition employed the Insta360 Pro 360° camera.
During recordings, a few actors were present in the hall to simulate a worst-case occupancy scenario, as confirmed by the conference hall management. A total of seven spatial RIRs were collected, with the Talkbox emitting a 5-second-long exponential sine sweep signal from 20 Hz to 20 kHz. For RIR measurements associated with the Talkbox in the target talker location, the room microphone connected to the 2-loudspeaker system was activated and placed in front of the Talkbox at a 20 cm distance to capture the overall effect of the room amplification system.

Figure 26: Picture of the RIR recording procedure in case of SMA placed in the farthest listening location (L2) and the Talkbox placed in the respective 120° azimuth noise location (N2120°).

Regarding the visual scenes, seven 2-minute 3D video recordings were captured, positioning the 360° camera at the listening locations. The Talkbox and a Brüel&Kjær 4128 dummy head were placed at the target and interfering talker positions, respectively, to provide visual reference for the spatial arrangement of the reproduced sounds during the AV speech test. The dummy head was only utilized during video recordings to visually distinguish between the target talker and the interferer. The same Talkbox was employed as the sound source in both target speech and noise positions for acoustic RIR acquisition.

Figure 27: Picture of the video recording procedure in case of the 360° camera placed in the listening location closest to the target speaker represented by the Talkbox, and one-talker noise at 180° azimuth represented by the dummy head.

Figure 28; Equirectangular preview of the visual scene with the listener closer to the target speaker (T) represented by the Talkbox in the front and the one-talker noise at 120° azimuth (N120°) represented by the dummy-head.

Acoustical characterization of the AV scenes

To characterize the acoustics of the conference hall in unoccupied conditions, the reverberation time (T30) was measured following the EN ISO 3382-2:2008 standard (EN ISO 3382-2, 2008). The measurements were conducted using the Brüel&Kjær 4292-L omnidirectional sound source and the NTi Audio XL2 omnidirectional class-1 Sound Level Meter (SLM), with analyses performed using the ITA Toolbox MATLAB library. The T30 was found to be 3.19 seconds ± 0.44 seconds, which exceeds the optimal value for good speech comprehension in small conference halls, as per recent Italian standards on schools.
Additionally, the A-weighted equivalent background noise level was measured using the SLM, resulting in a value of 39.1 dB based on a 3-minute integration time.
To estimate the Speech Intelligibility (SI) and target speech levels typically experienced in the two listening positions during a conference speech, the STIPA test signal was emitted with an "elevated vocal effort" by the Talkbox placed in the target speech position and amplified by the room loudspeakers. STIPA values of 0.62 ± 0.01 and 0.55 ± 0.01, along with LAeq values of 73.3 dB and 71.8 dB, respectively, were measured in the two listening positions. These STIPA values approach the optimal threshold for conference halls, indicating good speech comprehension in quiet conditions.
Furthermore, binaural parameters in the listening positions were derived from the 3OA RIRs to provide insight into the auditory scenes. Binaural Room Impulse Responses (BRIRs) were obtained from the output of the IEM plug-in suite Binaural Decoder, with parameters such as Interaural Level Difference (ILD), Interaural Time Difference (ITD), and Inter-Aural Cross Correlation (IACC) computed to assess sound localization and spatial impression. Speech Clarity (C50) and Direct-to-Reverberant energy Ratio (DRR) were also calculated for both ears and averaged based on the left and right ear RIRs.
The broadband ITD and ILD were estimated using appropriate methods, considering their roles in lower and higher frequency perception, respectively. Broadband DRR values were determined for each ear using a time window of 5 ms centered on the peak of the squared impulse response. Binaural speech clarity was computed across octave bands and provided as average values from 250 Hz

to 2 kHz. Differences in speech clarity between the target and noise sources were also evaluated to assess the gap in speech clarity between them.

Virtual reality system

The tests were conducted in the Audio Space Lab (ASL), a small sound-treated listening room at Politecnico di Torino, compliant with ITU-R BS.116-3 recommendation. The ASL is equipped with a 3rd-Order Ambisonics (3OA) audio reproduction system synchronized with the Meta Quest 2 Head-Mounted Display (HMD) to create an immersive virtual 3D AV environment.

The Ambisonics playback system comprises a spherical array of 16 Genelec 8030B 2-way active loudspeakers arranged in three rings: one horizontal ring of eight speakers at ear level, and two 4-speaker rings at +45° and -45° elevation angles. Additionally, two frontal Genelec 8351A 3-way active loudspeakers serve as subwoofers to cover the lower frequency range. All loudspeakers are connected to the Antelope Orion32 32-channel sound card driven by a high-end desktop PC.

To manage the AV reproduction, three software tools are used, which communicate via the OSC protocol to maintain AV synchronization. Bidule DAW handles real-time audio signal processing for Ambisonic decoding and sweet spot equalization, Unreal Engine by Epic Games manages playback of visual scenes via streaming 360° stereoscopic videos onto the HMD, and a MATLAB routine is implemented to trigger and synchronize the AV reproduction and collect outcomes of the speech test. Figure 29 depicts the ASL during the test session with a participant.

Figure 29: Audio Space Lab during the execution of the AV speech test.

Material and generation of the AV speech intelligibility test

The audio tracks for the ecological speech intelligibility tests were generated using a MATLAB routine based on the Room Impulse Responses (RIRs) obtained from the conference hall. The target speech corpus was derived from the validated, extended version of the Italian Matrix Sentence Test, featuring 5-word sentences spoken by a female speaker (Puglisi, et al., 2015). For the interfering noise, a standardized phonetically balanced speech, commonly used for speech recognition testing, spoken by a female talker, was employed (Castellana, Carullo, Astolfi, Puglisi, & U., 2017).

The auralized target signals were adjusted to achieve the same signal level in the center of the loudspeaker array, corresponding to the measured levels in the conference hall listening positions (73 dBA for the closest listening position to the target source and 72 dBA for the farthest one). In-noise scenes were generated by combining each auralized target sentence with a separate clip of auralized noise speech, maintaining a -5 dB Signal-to-Noise Ratio (SNR). This SNR value was chosen to present a moderately challenging acoustical condition, as SNR values around -5 dB correspond to SRT80 in anechoic conditions.

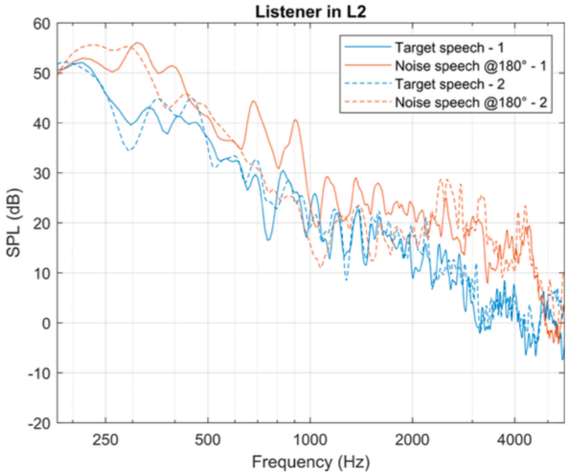


Figure 30: Examples of auralized target and noise speech spectra with 5 dB SNR for the scene with the listener in L2 location and noise at 180° azimuth.

The spectra demonstrate similarities between target and noise speeches, particularly in critical frequency ranges important for speech intelligibility.

To prepare the participant for the upcoming target sentence, the noise onset was presented a few seconds before the target speech, consistent with previous studies (Cubick & Dau, 2016). Each audio track initiated with 2 seconds of interfering noise (or silence for in-quiet scenes), followed by the presentation of the 5-word target sentence, and concluded with an additional 2 seconds of silence or interfering noise, resulting in an overall duration of 6-7 seconds for every scene.

Experimental procedure

The 40 participants were divided into equal-sized groups of 10 individuals, each characterized by different test administration configurations:

1. Audio-Only test with Self-Motion (AO-SM)
2. Audio-Only test in the Static condition (AO-S)
3. Audio-Visual test with Self-Motion (AV-SM)
4. Audio-Visual test in the Static condition (AV-S)

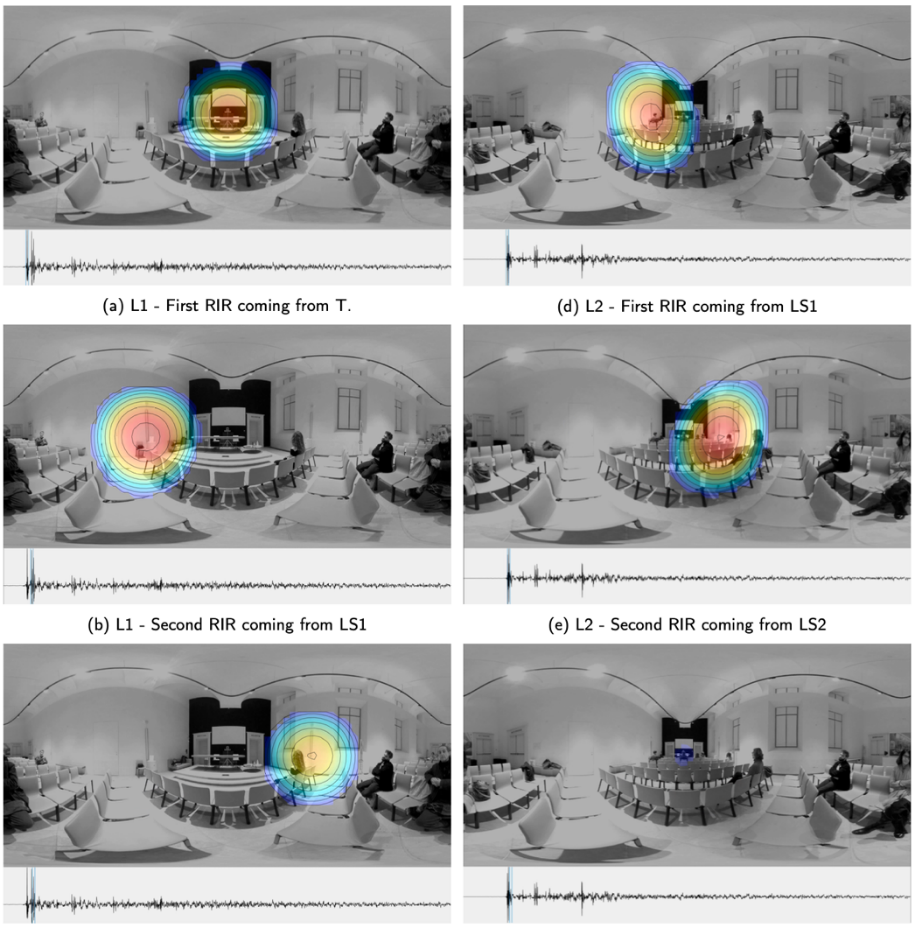
Before starting the experiment, participants underwent a training procedure to acquaint themselves with the system used for scene reproduction and the speech test. For the Static condition (S) tests, participants were instructed to refrain from turning their heads during the test to maintain the original spatial configuration of target speech and masking noise relative to the listening position. Conversely, in the Self-Motion (SM) tests, participants were informed they were free to move the head and the trunk in whatever direction they wanted to improve auditory perception.

In all test configurations, all seven scenes were presented, with each scene auralizing 20 sentences from different lists of the speech-in-noise test. The sequence of scenes was randomized and counterbalanced across participants.

The SI tests were conducted in an open format, with participants verbally repeating the words they understood, and the experimenter recording the correct responses. Overall, each test session lasted approximately 35 minutes per participant, with a 5-minute break after the initial 15 minutes of testing.

6.5 Results

The direction from which the target speech would be perceived during the speech test, emitted from three sound sources active simultaneously, was evaluated by measuring the time and level differences between all sound sources using the spatial RIRs. The view captured with the 360° camera from both listening positions is illustrated in figure 31 at the time instants when the RIR emitted from T, LS1, and LS2 sound source locations arrive, coupled with the SPL color map showing the direction of the incoming sound. As it can be seen from Figure 31, in the case of the listener in the L1 location, the first sound reaching the listener comes from the Talkbox in location T, while the second RIRs comes from LS1 and arrives after 2.4 ms from the first RIR, and, finally, the last RIR comes from LS2 after 1.1 ms from the LS1 one.



(a) L1 - First RIR coming from T.
(d) L2 - First RIR coming from LS1
(b) L1 - Second RIR coming from LS1
(e) L2 - Second RIR coming from LS2
(c) L1 - Third RIR coming from LS2
(f) L2 - Third RIR coming from T

Figure 31: SPL color map on the equirectangular view from the listening positions L1 and L2, showing the time history of the direction of arrival of the normalized RIR emitted from the Talkbox in position T and amplified by LS1 and LS2.

In a typical stereophonic configuration, when two identical signals are presented from left and right locations simultaneously, the auditory event is perceived as coming from a single "phantom source" in the front. Conversely, with a short delay and level differences between the two coherent signals reaching the listener’s ear, the sound is perceived at intermediate locations between the two sources, a phenomenon known as "summing localization." When the time difference (Δt) between two signals is from about 1 ms to 5 ms, the auditory event's location coincides with the position of the sound source whose signal arrives first, referred to as "localization dominance" (Litovsky, Colburn, Yost, & Guzman, 1999) Under this effect, in reverberant conditions, humans localize sounds based primarily on the direction of the preceding direct sound instead of the later-arriving reflections.

For the L1 listening position, where the first arriving signal comes from the Talkbox and is followed by signals from LS1 and LS2, the direction of the target speech is perceived as coming from the front, matching the T position. For the L2 position, where the first approaching signal comes from LS1, the auditory event's direction is perceived as approximately coming from the center between the two loudspeakers.

Analysis of the binaural parameters, including ITD, ILD, IACC, and DRR, further supported these findings (Tab. 15). Higher values of ITD and ILD were observed

for spatially separated sound sources, indicating differences in time delay and level between the two ears. The IACC values were lower for spatially separated sources compared to co-located ones. DRR analysis showed differences in energy distribution between direct and reverberated sound components, with reverberated components dominating in certain scenarios.
The binaural values for C50, related to speech clarity, followed a similar trend as DRR. Higher C50 values were observed for noise sources at 120° azimuth, indicating better clarity at the right ear. Conversely, lower values were observed for noise sources at 180° and 0° azimuth, as well as for the target speech. The differences in clarity between the target and noise sources varied depending on the listening position and the direction of the noise source.

	ACTIVE SOUND SOURCE POSITIONS						
	T@4.1 m,0° LS1@4m,-65° LS2@4m,66°	N@1.8m,120°	N@1.8m,180°	T@9.8m,0° LS1@8.2m,-26° LS1@8.3m,-27°	N@1.8m,120°	N@1.8m,180°	N@1.8m,0°
ITD (μs)	-20.8	540.0	0.0	-360.0	540.0	20.8	0.0
ILD (dB)	0.7	-6.0	-1.1	0.5	-6.3	1.3	0.7
IACC (-)	0.2	0.4	0.7	0.3	0.4	0.7	0.6
Left C_{50} (dB)	1.8	1.5	-0.1	-1.4	0.6	-0.2	-3.6
Right C_{50} (dB)	1.1	6.7	0.2	-1.1	6.0	0.0	-3.5
Mean C_{50} (dB)	1.5	4.9	0.0	-1.0	4.2	-0.1	-3.5
Left DRR (dB)	0.0	-0.4	-2.0	-5.0	0.1	-0.4	-7.7
Right DRR (dB)	-1.1	6.7	-0.8	-4.6	6.8	-1.6	-8.0
Mean DRR (dB)	-0.5	4.5	-1.3	-4.8	4.6	-0.9	-7.8

Table 15: ITD (low-pass filtered with a cut-off frequency of 1.3 kHz), ILD (High-pass filtered with a cut-off frequency of 1.3 kHz), broadband IACC, left, right and within-ear mean DRR (broadband) and C50 (mean from 250 Hz to 2 kHz) values from the 3OA.

Figure 32 displays the means for the speech intelligibility percentage scores achieved in each test condition (AO-SM, AO, AV-SM, AV-S) across each scene. Results revealed that the target distance and noise azimuth significantly predicted speech intelligibility scores, while visual cues and self-motion were not significant individually but showed significance in their interactions.

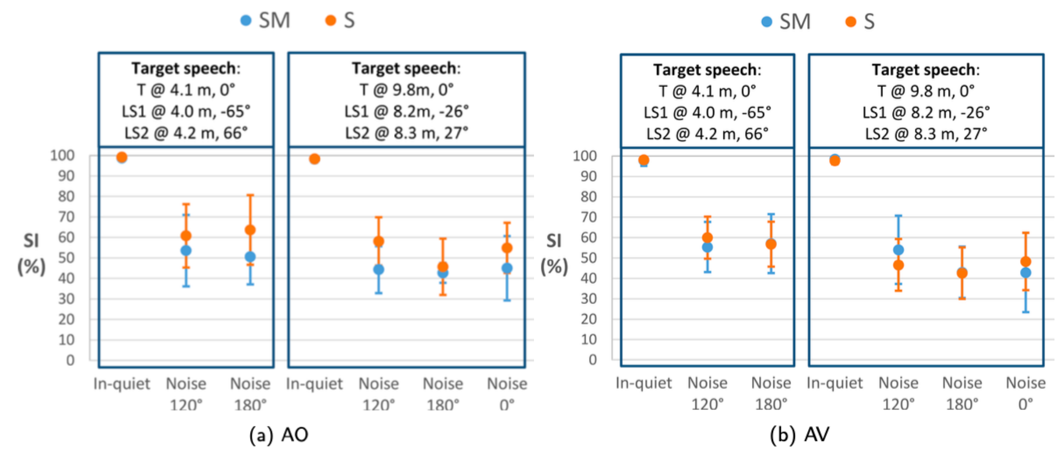


Figure 32: Mean and standard deviation values of the percentage speech intelligibility scores for each scene for the comparison between the Self-motion (SM) and the Static condition (S) in case of (a) Audio-Only and (b) Audio-Visual test.

The analysis indicated that the intra-subject variability (σα) was lower than the inter-subject variability (σε), suggesting consistency within subjects across different test configurations.

To account for the ceiling effect, the data were transformed into Rationalized Arcsin Units (RAU) before performing the statistical analysis (Studebaker, 1985). The mean SI scores expressed in RAU for the four test configurations (AV-SM, AV-S, AO-SM, AO-S) showed significant differences, with the AO-S condition achieving the highest mean SI score (tab. 16).

Test configuration	N	Mean	SD
AV-SM	1000	50.99	32.64
AV-S	1000	51.18	31.09
AO-SM	1000	47.81	29.88
AO-S	1000	**57.16**	30.59

Table 16: Mean and standard deviation (SD) of the speech intelligibility scores in RAU for each test configuration (AV-SM, AV-S, AO-SM, AO-S).

Further comparisons between test configurations and auditory scenes revealed nuanced effects. In particular, scenes with self-motion tended to score lower than static conditions for certain scenes, especially in the audio-only setup. However, in the audio-visual condition, self-motion led to better intelligibility scores in some scenes. Comparisons between AO-S and AV-SM configurations indicated that the AO-S condition performed better in most scenes.
Analyzing differences between auditory scenes, significant variations were observed based on target-to-listener distance and noise azimuth. For instance, intelligibility scores improved when the noise azimuth was 120° compared to 180°. Additionally, speech intelligibility increased when the target was farther away from the listener, especially with noise at 0° azimuth compared to 180° azimuth. This improvement was attributed to differences in speech clarity between target and noise sources, where the target was clearer than noise when noise was directed away from the listener.
Comparing scenes with different target-to-listener distances and noise azimuths also revealed expected trends. Speech intelligibility scores were worse for farther listening positions, with a more pronounced effect observed for noise at 180° azimuth. This aligns with findings from previous studies (Fichna, Biberger, Seeber, & Ewert, 2021) (Puglisi, Warzybok, Astolfi, & Kollmeier, 2021), highlighting the impact of distance and noise characteristics on speech intelligibility.

(a)	COMPARISON BETWEEN TEST CONFIGURATIONS (X_1 vs X_2) FOR EACH SCENE						
Scene number	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Active sound source position	T@4.1 m,0° LS1@4m,-65° LS2@4m,66°	T@4.1 m,0° LS1@4m,-65° LS2@4m,66° N@1.8m,120°	T@4.1 m,0° LS1@4m,-65° LS2@4m,66° N@1.8m,180°	T@9.8m,0° LS1@8.2m,-26° LS1@8.3m,-27°	T@9.8m,0° LS1@8.2m,-26° LS1@8.3m,-27° N@1.8m,120°	T@9.8m,0° LS1@8.2m,-26° LS1@8.3m,-27° N@1.8m,180°	T@9.8m,0° LS1@8.2m,-26° LS1@8.3m,-27° N@1.8m,0°
AV-SM vs AV-S					***0.011***		
AO-SM vs AO-S	**0.010**	**0.000**		**0.000**		**0.001**	
AV-SM vs AO-S	**0.015**	**0.042**	**0.029**				**0.000**

(b)	COMPARISON BETWEEN SCENES (X_1 vs X_2) FOR EACH TEST CONFIGURATION						
Scene number	*3* (T,LS~4m, ... N@180°) *vs* *2* (T,LS~4m, ... N@120°)	*6* (LS~8m, ... N@180°) *vs* *5* (LS~8m, ... N@120°)	*7* (LS~8m, ... N@0°) *vs* *5* (LS~8m, ... N@120°)	*7* (LS~8m, ... N@0°) *vs* *6* (LS~8m, ... N@180°)	*4* (LS~8m, ...) *vs* *1* (T,LS~4m, ...)	*5* (LS~8m, ... N@120°) *vs* *2* (T,LS~4m, ... N@120°)	*6* (LS~8m, ... N@180°) *vs* *3* (T,LS~4m, ... N@180°)
AV-SM		**0.000**	**0.000**				**0.000**
AV-S				***0.048***		**0.000**	**0.000**
AO-SM						**0.000**	**0.012**
AO-S		**0.000**		***0.001***			**0.000**

Table 17: U-Mann Whitney statistical analyses results for the comparison between: (a) the different test configurations (AV-SM vs AV-S, AO-SM vs AO-S, AV-SM vs AO-S) for each scene, and (b) the different acoustical conditions of the scenes for the same test

6.6 Discussion

The direction from which virtual sound environment, the effect of head self-motion (SM) on speech intelligibility when different spatial configurations for one-talker interfering noise are presented was investigated. In the audio-only (AO) condition, head SM led to a decrease in speech intelligibility compared to the static (S) condition, as observed in Table 16 and confirmed by the analysis in Table 17(a), where SM resulted in lower scores for 4 out of 7 scenes. Head movements can potentially optimize the signal-to-noise ratio or spatial unmasking (Brimijoin, McShefferty, & Akeroyd, 2012), but in this study, the negative impact of SM on speech intelligibility suggests that it might have interfered with these benefits, contrary to findings in other studies. Other studies did not find any significant effect of SM on speech intelligibility. For example, Hladek and Seeber did not show significant differences in SI when the speech noise and the target were at 0° for the three conditions AV-SM, AO- SM, and AO-S, while the AO-S condition determined higher SI when the target was at 90° and -90° and the speech noise at 0°, followed by the AV-SM condition (Hládek & Seeber, 2023). Frissen et al. studied the effect of speech-irrelevant head movements on speech intelligibility with multiple maskers in the acoustic scene and did not find any significant positive effect of the head movement (Frissen, Scherzer, & Yao, 2019).

Moreover, contrary to expectations, AO-S tests perform better than AV tests, leading to better speech intelligibility scores compared to AV-SM configurations in 4 out of 7 scenes. Furthermore, it seems that results from AV are equivalent in both SM and S conditions, as found in (Fichna, Biberger, Seeber, & Ewert, 2021), although one could expect that allowing SM during the AV tests should lead to better intelligibility scores than the S case, being the AV-SM the test configuration that gets closer to the real-life listening experience. Previous studies had indeed suggested that visual cues enhance speech intelligibility, especially when combined with SM (Hládek & Seeber, 2023) (Neidhardt, Schneiderwind, & Klein, 2022).

The lack of significant improvement in speech scores in the AV-SM condition in this study might be attributed to the absence of lip movements (not included in

this preliminary study) which are known to aid speech reading (MacLeod & Summerfield, 1987).
Additionally, the distance between the listener and the speaker in this study may have been too far for visual cues to substantially benefit speech intelligibility: indeed, the accuracy of speech reading decreases rapidly as a function of the distance from the speaker, and, in our cases, with about 4 m and 8 m, it is unlikely that it could have brought a significant improvement.
Finally, regarding the detectability of spatial release from masking (SRM) in the presence of one-talker interfering noise at different distances from the target source, with and without SM and visual cues, Table 17(b) provides insights. Notably, SRM was evident when the target was farthest from the listener, with noise azimuths at 120° compared to 180° in certain scenes. This improvement in intelligibility cannot be solely explained by differences in speech clarity between target and noise sources. Instead, it may be attributed to the precedence effect (PE), which facilitates the segregation of competing voices from the target stream in reverberant environments (Litovsky, Colburn, Yost, & Guzman, 1999) (Freyman, Helfer, McCall, & RK., 1999).

This study advances our understanding of speech intelligibility in reverberant environments, particularly regarding the influence of head SM and visual cues.

Final discussion

The necessity of a greater awareness regarding deafness cannot be overstated. Deafness, a condition that affects millions worldwide, encompasses a spectrum of experiences, challenges, and opportunities that often go unnoticed or misunderstood by the general population: although hearing loss permeates various facets of human existence, affecting not only auditory perception but also intertwining with broader aspects of communication, social interaction, and everyday living, is a condition often underestimated.

Raising awareness about deafness is crucial for fostering inclusivity and understanding within society. Deaf individuals face numerous barriers in communication, education, and employment due to misconceptions and lack of awareness. By educating the public about deaf culture and assistive technologies, we can create a more inclusive environment where deaf individuals can fully participate and thrive.

We have seen how deafness is often associated with a range of voice disorders, yet this correlation remains relatively obscured in mainstream discourse. Individuals with hearing impairments frequently encounter challenges in vocal production, ranging from alterations in pitch, volume, and timbre to difficulties in articulation and speech clarity. Moreover, prolonged reliance on compensatory mechanisms to navigate communication barriers can exacerbate vocal strain and lead to the development of secondary voice disorders. Despite the profound impact of voice disorders on the quality of life and social integration of individuals with deafness, there exists a pervasive lack of recognition and understanding of this symbiotic relationship.

Limitations faced especially by elderly individuals with deafness have been highlighted when it comes to television viewing: in case of deafness, even the seemingly simple act of watching TV can present significant challenges, mostly in programs incorporating important audiovisual cues, such as background music, sound effects, and non-verbal communication, to enhance the viewing experience. However, by promoting inclusive viewing environments through the implementation of targeted strategies, such as the real-time dynamic modulation of the audio signal proposed, we can enable older people with deafness to overcome these barriers and fully enjoy television programming.

Finally, the domain of speech perception in challenging listening conditions is one of the most sensitive areas in the assessment of hearing-impaired patients. It is closely related to quality of life, as everyday conversations, an implicit indicator of relational well-being, often occur in acoustically unfavorable environments or in the presence of multiple speakers and background noise. The ability to perceive sound spatially and discriminate between different auditory sources is fundamental to effective communication and situational awareness in everyday life. However, individuals with hearing loss encounter substantial impediments in binaural listening, especially amidst background noise and reverberant environments characteristic of public spaces.

On the one hand, clinical practice laboratory procedures lack ecological validity, relying on listening tests in quiet environments or with stationary noise using plain loudspeaker setups. On the other hand, in-field procedures lack reproducibility, making it difficult to detect slight changes in individuals' hearing abilities. Thus, aiming at retaining control over the tested acoustical stimuli and conducting listening tests matching real-life auditory demands, we embraced the use of spatial sound playback systems, which reproduce Virtual Sound Environments that mimic everyday listening payloads involving multiple sound sources and noise from different directions. Particularly, we investigated the impact of head self-motion and visual cues on speech intelligibility in reverberant environments. Contrary to expectations, head SM during audio-only conditions led to decreased speech intelligibility, suggesting interference with potential benefits like spatial unmasking. Furthermore, audio-only tests performed better than audio-visual tests, with no significant improvement observed when combining visual cues and head SM, probably due to the lack of lip movements and distance between speaker and listener.

It highlights the complex interplay between auditory and visual cues in shaping intelligibility and the importance of considering ecological factors in understanding auditory perception. These insights are valuable for designing technologies and interventions aimed at enhancing speech communication and auditory perception in real-world settings.

Further research is needed to elucidate the mechanisms underlying spatial release from masking in naturalistic contexts and its implications for auditory processing, as well as to understand the functioning of conventional hearing aids in binaural listening under ecological conditions.

References

Abdipour, R., Akbari, A., Rahmani, M., & Nasersharif, B. (2015). Binaural source separation based on spatial cues and maximum likelihood model adaptation. *Digital Signal Processing*, 36:174–183.

Arehart, K., Souza, P., Baca, R., & Kates, J. (2013). Working memory, age, and hearing loss: susceptibility to hearing aid distortion. *Ear Hear.*, 34(3):251-60.

Astolfi, A., Castellana, A., Carullo, A., & Puglisi, G. (2018). Uncertainty of speech level parameters mea- sured with a contact-sensor-based device and a headworn microphone. *J Acoust Soc Am.*, 143(6):EL496.

Astolfi, A., Riente, F., Shtrepi, L., Carullo, A., Scopece, L., & M., M. (2021). Speech quality improvement of commercial flat screen TV-sets. *IEEE Trans. Broadcast.*, 67(3):685–695.

Švec, J., Titze, I., & Popolo, P. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *J. Acoust. Soc. Am.*, 117(3), 1386-1394.

Baumgartner, H., Van Everdingen, R., Schreiner, B. K., & Krämer, U. (2022). *Speech Intelligibility in TV - Technical Review.* EBU - Operating Eurovision and Euroradio.

BBC. (2011). *Editorial policy guidance: Hearing impaired audiences.*

Bernstein, J., Schuchman, G., & Rivera, A. (2017). Head Shadow and Binaural Squelch for Unilaterally Deaf Cochlear Implantees. *Otology & Neurotology*, 38(7):p e195-e202.

Bocca, E., & Pellegrini, A. (1950). Studio statistico sulla composizione della fonetica della lingua italiana e sua applicazione pratica all'audiometria con la parola. *Arch. Ital. Otol.*, 116–141.

Boisvert, I., Reis, M., Au, A., Cowan, R., & Dowell, R. (2020). Cochlear implantation outcomes in adults: a scoping review. *PLoS One.*, 15(5): e0232421.

Borden, G. (1979). An interpretation of research on feedback interruption in speech. . *Brain Lang.*, 7(3):307–19.

Bottalico, P., & Astolfi, A. (2012). Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *J Acoust Soc Am*, 131(4):2817–27.

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am*, 111:2801–2810.

Brimijoin, W., McShefferty, D., & Akeroyd, M. (2012). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hearing research*, 283(1- 2):162–168.

Brumm, H., & Zollinger, S. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Beyond Behav.*, 148(11–13): 1173–98.

Bureau of Labor Statistics. (2017). *American Time Use Summary - 2016 results.*

Canale, A., Santagata, F., Massaia, M., Caranzano, F., Boggio, V., & Albera, A. (2019). Cochlear implant in elderly deaf patients with adverse predictors of audiological outcome. *Otorinolaringologia*, 69(1):21–5.

Cantarella, G., Iofrida, E., Boria, P., Giordano, S., Binatti, O., & Pignataro, L. (2014). Ambulatory phonation monitoring in a sample of 92 call center operators. *J Voice*, 28(3): 393.e1–393.e6.

Castellana, A., Carullo, A., Astolfi, A., Puglisi, G., & Fugiglando, U. (2017). Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings. *J Acoust Soc Am.*, 141(4):2353.

Castellana, A., Carullo, A., Astolfi, A., Puglisi, G., & U., F. (2017). Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings. *The Journal of the Acoustical Society of America*, 141(4):2353–2363.

CENSIS. (2019). *Sentirsi bene. Il valore sociale dell'audioprotesi.* Available at: https://www.censis.it/welfare-e-salute/sentirsi-bene-1.

Cheyne, H., Hanson, H., Genereux, R., Stevens, K., & Hillman, R. (2003). Development and testing of a portable vocal accumulator. *J. Speech Lang. Hear. Res.*, 46(6), 1457–1467.

Coelho, A., Brasolotto, A., & Bevilacqua, M. (2012). Sys- tematic analysis of the benefits of cochlear im- plants on voice production. *J Soc Bras Fono- audiol*, 24(4):395–402.

Cox, R., & Alexander, G. (2002). The international Outcome inventory for hearing aids (IOI- HA): psychometric properties of the English version. *Int J Audiol.*, 41(1):30–5.

Cox, R., Alexander, G., & Beyer, C. (2003). Norms for the international Outcome inventory for hearing aids. *J Am Acad Audiol.*, 14(08): 403–13.

Cubick, J., & Dau, T. (2016). Validation of a virtual sound environment system for testing hearing aids. *Acta Acustica united with Acustica*, 102(3):547–557.

Cutugno, F., Prosser, S., & Turrini, M. (2000). *Audiometria Vocale.* Padova: GN ReSound.

Debruyne, F., & Buekers, R. (1998). Interdependency between intensity and pitch in the normal speaking voice. *Acta Oto-Rhino-Laryngol Belg.*, 52(3):201–5.

Ellen Peng, P., & Litovsky, R. (2021). he Role of Interaural Differences, Head Shadow, and Binaural Redundancy in Binaural Intelligibility Benefits Among School-Aged Children. *Trends Hear*, 25:23312165211045313.

EN ISO 3382-2, .. (2008). *Acoustics - Measurement of Room Acoustic Param- eters – Part 2: Reverberation Time in Ordinary Rooms.*

Eurisko. (2015). *Indagine "Coping with noise.*

Evans, M., & Deliyski, D. (2007). Acoustic voice analysis of prelingually deaf adults before and after cochlear implantation. *J Voice*, 21(6):669– 82.

Fichna, S., Biberger, T., Seeber, B., & Ewert, S. (2021). Effect of acoustic scene complexity and visual scene representation on auditory perception in virtual audio-visual environments. In *Immersive and 3D Audio: from Architecture to Automotive (I3DA)* (p. 1–9). IEEE.

Fiorin, M., Marconato, E., Palharini, T., Picoloto, L., Frizzo, A., Cardoso, A., & Oliveira, C. (2021). Impact of auditory feedback alterations in individuals with stuttering. *Braz J Otorhinolaryngol.*, 87(3):247-254.

Fletcher, M. (2021). Can Haptic Stimulation Enhance Music Perception in Hearing-Impaired Listeners? . *Front Neurosci*, 15:723877.

Franca, M., & Wagner, J. (2015). Effects of vocal demands on voice performance of student singers. *J Voice*, 29(3):324–32.

French, N., & Steinberg, J. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Amer.*, 90–119.

Freyman, R., Helfer, K., McCall, D., & RK., C. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, 106(6): 3578–3588.

Frissen, I., Scherzer, J., & Yao, H. (2019). The impact of speech-irrelevant head movements on speech intelligibility in multi-talker environments. *Acta Acustica united with Acustica*, 105(6): 1286–1290.

Fuchs, H., & Oetting, D. (2014). Advanced clean audio solution: Dialogue enhancement. *SMPTE Motion Imag. J*, 123(5):23–27.

Göthberg, H., Skoog, I., Tengstrand, T., Magnusson, L., Hoff, M., Rosenhall, U., & Sadeghi, A. (2023). Pathophysiological and Clinical Aspects of Hearing Loss Among 85-Year-Olds. *Am J Audiol.*, 32(2):440-452.

Gao, J., Hu, H., & Yao, L. (2020). The role of social en- gagement in the association of self-reported hearing loss and health-related quality of life. *BMC Geriatr.*, 20(1):182.

Garbaruk, E., Goykhburg, M., Warzybok, A., Tavartkiladze, G., Pavlov, P., & Kollmeier, B. (2020). Application of the matrix sentence test Russian version in children. *Vestn. Otorinolaringol.*, 85:34–39.

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int J Audiol*, 43(2):85–99.

Gautam, A., Naples, J., & Eliades, S. (2019). Control of speech and voice in cochlear implant patients. *Laryngoscope*, 129(9):2158–63.

Giraud, A., Garnier, S., Micheyl, C., Lina, G., Chays, A., & Chery-Croze, S. (1997). Auditory efferents involved in speech-in-noise intelligibility. *Cognitive Neuroscience and Neuropsycology*, 8: 1779 – 1783.

Gomez, M., Hwang, S., Sobotova, L., Stark, A., & May, J. (2001). A comparison of self-reported hearing loss and audiometry in a cohort of New York farmers. *J. Speech Lang. Hearing Res.*, 44(6): 1201–1208.

Grimm, G., Hendrikse, M., & Hohmann, V. (2020). Review of self-motion in the context of hearing and hearing device research. *Ear and hearing*, 41:48S–55S.

Hülsmeier, D., Buhl, M., Wardenga, N., Warzybok, A., Schädler, M., & Kollmeier, B. (2022). Inference of the distortion component of hearing impairment from speech recognition by predicting the effect of the attenuation component. *Int J Audiol.*, 61(3):205-219. .

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian audiology*, 11(2), 79–87.

Hamzavi, J., Deutsch, W., Baumgartner, W., Denk, D., Adunka, O., & Gstoettner, W. (2000). Cochlear implantation and auditory feedback. *Wien Klin Wochenschr.* , 112(11):515–8.

Harrison, R., Gordon, K., & Mount, R. (2005). Is there a critical period for cochlear implantation in congenitally deaf children? Analyses of hearing and speech perception performance after implantation. *Dev Psychobiol.*, 46(3):252–61.

Hendrikse, M., Eichler, T., Hohmann, V., & Grimm, G. (2022). Self-motion with hearing impairment and (directional) hearing aids. *Trends in Hearing*, 26:23312165221078707,.

Hendrikse, M., Llorach, G., Grimm, G., & Hohmann, V. (2018). Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters. *Speech Communication*, 101:70–84.

Hillman, R., Heaton, J., Masaki, A., Zeitels, S., & Cheyne, H. (2006). Ambulatory monitoring of disordered voices. *Ann Otol Rhinol Laryngol.*, 115(11):795–801.

Hládek, L., & Seeber, B. (2023). Speech intelligibility in reverberation is reduced during self-rotation. *Trends in Hearing*, 27: 23312165231188619.

Hládek, L., Van de Par, S., Ewert, S., & Seeber, B. (2021). Audio-visual scenes repository: How to contribute. doi: 10.5281/zenodo. 5532673.

Hocevar-Boltezar, I., Radsel, Z., Vatovec, J., Geczy, B., Cernelc, S., & Gros, A. (2006). Change of phonation control after cochlear implantation. *Otol Neurotol.*, 27(4):499–503.

Holman, J., Drummond, A., & Naylor, G. (2021). Hearing aids reduce daily-life fatigue and increase social activity: a longitudinal study. *Trends Hear.*, 5:23312165211052786.

Hornsby, B., & Ricketts, T. (2003). The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding. II. Sloping hearing loss. *J. Acoust. Soc. Amer.*, 113(3):1706–1717.

INAIL. (1994). *Circolare n. 22 del 7 luglio 1994.*

Int. Electrotech. Commission, .. (2020). *Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index.* Standard IEC 60268-16, Geneva, Switzerland.

Int. Org. Stand., .. (2003). *Acoustic - Normal Equal-Loudness - Level Contours.* Standard ISO 226, Geneva, Switzerland.

Int. Org. Stand., .. (2017). *Acoustics - Statistical Distribution of Hearing Thresholds Related to Age and Gender.* Standard ISO 7029:2017, Geneva, Switzerland.

Junger. (2020). D*AP4—Digital Audio Processor. Berlin, Germany.

Kenney, K., & Prather, E. (1986). Articulation development in preschool children: consistency of productions. *J Speech Hear Res.*, 29(1):29-36.

Kidd, G., Mason, C., Brughera, A., & Hartmann, W. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta acustica united with acustica*, 91(3):526–536.

Kirk, K., & Edgerton, B. (1983). The effects of cochlear implant use on voice parameters. *Otolaryngol Clin North Am.*, 16(1):281–92.

Kivimäki, M., & Singh-Manoux, A. (2018). Prevention of dementia by targeting risk factors. *Lancet*, 391(10130):1574–5.

Kozlowski, L., Almeida, G., & Ribas, A. (2014). Level of user satisfaction with hearing aids and environment: the international Outcome inventory for hearing aids. *Int Arch Oto- rhinolaryngol.*, 18(3):229–34.

Krause, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci*, 11, 599 – 605.

Langereis, M., Bosman, A., van Olphen, A., & Smoorenburg, G. (1997). Changes in vowel quality in post-lingually deafened cochlear implant users. *Audiology*, 36(5):279–97.

Leder, S., Spitzer, J., Milner, P., Flevaris-Phillips, C., Kirchner, J., & Richardson, F. (1987). Voice intensity of prospective cochlear implant candidates and normal hearing adult males. *Laryngoscope*, 97(2):224–7.

Lee, G., Liu, C., & Lee, S. (2013). Effects of hearing aid amplification on voice F0 variability in speakers with prelingual hearing loss. *Hear Res.*, 302:1–8.

Lenden, J., & Flipsen, J. P. (2007). Prosody and voice char- acteristics of children with cochlear implants. *J Commun Disord.*, 40(1):66–81.

Lin, F., Thorpe, R., Gordon-Salant, S., & Ferrucci, L. (2011). Hearing loss prevalence and risk factors among older adults in the United States. *J. Gerontol. A, Biol. Sci. Med. Sci.*, 66(5):582–590.

Litovsky, R., Colburn, H., Yost, W., & Guzman, S. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, 106(4):1633–1654.

Llorach, G., Grimm, G., Hendrikse, M., & Hohmann, V. (2018). Towards realistic immersive audiovisual simulations for hearing research: Capture, virtual scenes and reproduction. . *Pro- ceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia*, 33–40.

Llorach, G., Hendrikse, M., Grimm, G., & Hohmann, V. (2020). Comparison of a head-mounted display and a curved screen in a multi-talker audiovisual listening task. *arXiv preprint*, 2004.01451.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British journal of audiology*, 21(2):131–141.

McDermott, J., Keebler, M., Micheyl, C., & Oxenham, A. (2010). Musical intervals and relative pitch: frequency resolution, not interval resolution, is special. *J Acoust Soc Am.*, 128(4):1943-51.

Medved, D., Cavalheri, L., Coelho, A., Fernandes, A., Silva, E., & Sampaio, A. (2021). Systematic review of auditory perceptual and acoustic characteristics of the voice of co- chlear implant adult users. *J Voice*, 35(6):934.e7–934.e16.

Mendel, L. (2008). Current considerations in pediatric speech audiometry. *Int. J. Audiol.*, 47:546–553.

Monini, S., Banci, G., Barbara, M., Argiro, M., & Filipo, R. (1997). Clarion cochlear implant: short-term effects on voice parameters. *Am J Otol.* , 18(6):719–25.

Moore, A., Green, T., Brookes, M., & Naylor, P. (2022). Measuring audio-visual speech intelligibility under dynamic listening conditions using virtual reality. *Audio Engineering Society Con- ference: AES 2022 International Audio for Virtual and Augmented Reality Conference.* Audio Engineering Society.

Moore, B. (2016). A review of the perceptual effects of hearing loss for frequencies above 3 kHz. *Int. J. Audiol.*, 55(12):707–714.

Mora, R., Crippa, B., Cervoni, E., Santomauro, V., & Guastini, L. (2012). Acoustic features of voice in patients with severe hearing loss. *J Otolaryngol Head Neck Surg.*, 41(1):8–13.

Most, T., Adi-Bensaid, L., Shpak, T., Sharkiya, S., & Luntz, M. (2012). Everyday hearing functioning in unilateral versus bilateral hearing aid users. *Am J Otolaryngol.*, 33(2): 205–11.

Mozzanica, F., Schindler, A., Iacona, E., & Ottaviani, F. (2019). Application of Ambulatory Phona- tion Monitoring (APM) in the measurement of daily speaking-time and voice intensity before and after cochlear implant in deaf adult patients. *Auris Nasus Larynx*, 46(6):844–52.

Neidhardt, A., Schneiderwind, C., & Klein, F. (2022). Perceptual matching of room acoustics for auditory augmented reality in small rooms-literature review and theoretical framework. *Trends in Hearing*, 26:23312165221092919.

Neumann, K., Baumeister, N., Baumann, U., Sick, U., Euler, H., & Weißgerber, T. (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. *Int. J. Audiol.*, 51, 157–163.

Nishio, M., & Niimi, S. (2008). Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatr Logop.*, 60(3):120–7.

Noble, W., & Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abil- ities measured by the Speech, Spatial, and Qualities of Hearing scale (SSQ). *Int J Audiol.*, 45(3):172–81.

Olsen, W., & Carhart, R. (1975). Head diffraction effects on ear-level hearing aids. *Audiology*, 14(3):244-58.

Peelle, J. (2017). Listening Effort: How the cognitive consequences of acoustic challenge are reflected in brain and behaviour. *Ear & Hearing*, 39: 204 – 214.

Peelle, J. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear Hear.*, 39(2):204-214.

Perkell, J., Lane, H., Denny, M., Matthies, M., Tiede, M., & Zandipour, M. (2007). Time course of speech changes in response to unanticipated short- term changes in hearing state. *J Acoust Soc Am.*, 121(4):2296–311.

Peterson, A., & Heil, P. (2020). Phase Locking of Auditory Nerve Fibers: The Role of Lowpass Filtering by Hair Cells. *J Neurosci*, 40(24):4700-4714.

Pietrantoni, L., Bocca, E., & Agazzi, C. (1956). Diagnosi delle sordità centrali. *Relazione XLIV Congresso Società Italiana Laringol. Otol. Rinol.*

Plomp, R. (1976). Binaural and monaural speech intelligibility of con- nected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acta Acustica united with Acustica*, 34(4):200–211.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Speech Hear Res.*, 29(2):146-54.

Preminger, J., & Van Tasell, D. (1995). Measurement of speech quality as a tool to optimize the fitting of a hearing aid. *J. Speech Hearing Res*, 726–736.

Preminger, J., & Van Tasell, D. (1995). Quantifying the relation between speech quality and speech intelligibility. *J. Speech Lang. Hearing Res*, 714–725.

Puglisi, G., Warzybok, A., Astolfi, A., & Kollmeier, B. (2021). Effect of reverberation and noise type on speech intelligibility in real complex acoustic scenarios. *Building and Environment*, 204:108137.

Puglisi, G., Warzybok, A., Hochmuth, S., Visentin, C., Astolfi, A., Prodi, N., & Kollmeier, B. (2015). An Italian matrix sentence test for the evaluation of speech intelligibility in noise. *Int. J. Audiol*, 54:44–50.

Ricketts, T., Dittberner, A., & Johnson, E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *J. Speech Lang. Hearing Res*, 51(1):160–172.

Rodríguez-Valiente, A., Álvarez-Montero, O., Górriz-Gil, C., & García-Berrocal, J. (2020). Prevalence of presbycusis in an otologically normal population. *Acta Otorrinolaringologica Espanola,*, 175–180,.

Rosbach, J. (2017). Entwicklung und Evaluation des “Oldenburger Kinder-Satztests” mit Weiblicher Stimme (Development and Evaluation of “Oldenburg Children Test". In *Jade Hochschule Wilhelmshaven Oldenburg Elsfleth.* Wilhelmshaven, Germany.

Ruff, S., Bocklet, T., Nöth, E., Müller, J., Hoster, E., & Schuster, M. (2017). Speech production quality of co- chlear implant users with respect to duration and onset of hearing loss. *ORL J Othorinolar- yngol Relat Spec.*, 79(5):282–94.

Sawyer, C., Armitage, C., Munro, K., Singh, G., & Dawes, P. (2019). Correlates of hearing aid use in UK adults: self-reported hearing difficulties, social participation, living situation, health, and demographics. *Ear Hear.*, 40(5): 1061–8.

Schenk, B., Baumgartner, W., & Hamzavi, J. (2003). Changes in vowel quality after cochlear implantation. *ORL J Othorinolaringol Relat Spec*, 65(3):184–8.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., . . . Pumford, J. (2005). The Desired Sensation Level multistage input/output algorithm. *Trends Amplif.*, 9(4):159-97.

Selleck, M., & Sataloff, R. (2014). The impact of the auditory system on phonation: a review. *J. Voice*, 28(6):688–93.

Seol, H., Kang, S., Lim, J., Hong, S., & Moon, J. (2021). Feasibility of virtual reality audiological testing: Prospective study. *JMIR Serious Games*, 9(3):e26976.

Shamma, S. (1985). Speech processing in the auditory system. I:The rapresentation of speech sounds in the response of the auditory nerve. *J Acoust Soc Am*, 78:1612 – 21.

Shirley, B., & Oldfield, R. (2015). Clean audio for TV broadcast: An object- based approach for hearing-impaired viewers. *J. Audio Eng. Soc*, 63:245–256.

Silva, A., Blasca, W., Lauris, J., & Oliveira, J. (2014). Correlation between the characteristics of resonance and aging of the external ear. *CoDAS*, 6(2):112–116.

Sjolander, M., Bergmann, M., & Hansen, L. (2009). Improving TV listening for hearing aid users.

Smoorenburg, G., Huiskamp, T., Langereis, M., & Bosman, A. (1994). Effects of cochlear implantation on voice quality and speech production. In L. Hochmair-

Desoyer, & E. Hochmair, *Advances in Cochlear Implantation* (p. p. 374–9.). Wien: Manz.

Standaert, S., Rakita, L., & Strelcyk, O. (2017). Improving television listening for hearing aid users. *AudiologyOnline*, 20728.

Steeneken, H., & Houtgast, T. (1999). Mutual dependence of the octave- band weights in predicting speech intelligibility. *Speech Commun.*, 28(2):109–123.
Steeneken, H., & Houtgast, T. (1999). Mutual dependence of the octave- band weights in predicting speech intelligibility. *Speech Commun.*, 28(2):109–123.
Strelcyk, O., & Singh, G. (2018). TV listening and hearing aids. *PLoS One*, 13(6):1–21.

Studebaker, G. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research*, 28(3):455–462.

Svec, J., Titze, I., & Popolo, P. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *J Acoust Soc Am.*, 117(3 Pt 1):1386–94.

Thoidis, I., Vrysis, L., Markou, D., & Papanikolaou, G. (2020). Temporal Auditory Coding Features for Causal Speech Enhancement. *Electronics*, 9:1698.

Tonelli, L., Panzeri, M., & Fabbro, F. (1998). Un'analisi statistica della lingua italiana parlata. *Studi Ital. Linguist. Teor. Appl.*, 3:501–514.

Tyler, R., Wood, E., & Fernandes, M. (1982). Frequency resolution and hearing loss. *Br J Audiol.*, 16(1):45-63.

Ubrig, M., Tsuji, R., Weber, R., Menezes, M., Barrichelo, V., & da Cunha, M. (2019). The influence of auditory feedback and vocal rehabilitation on prelingual hearing-impaired individuals post cochlear implant. *J Voice*, 33(6):947.e1–9.

Upadhyay, M., Datta, R., Nilakantan, A., Goyal, S., Gupta, A., & Gupta, S. (2019). Voice quality in co- chlear implant recipients: an observational cross-sectional study. *Indian J Otolaryngol Head Neck Surg*, 71(Suppl 2):1626–32.

Van De Par, S., Ewert, S., Hladek, L., Kirsch, C., Schütze, J., Llorca-Bofí, J., . . . Seeber, B. (2022). Auditory-visual scenes for hearing research. *Acta Acustica*, 6:55.

Vernero, I., Gambino, M., & Schindler, O. (1998). *Cartella logopedica. Età Evolutiva.* Torino: Omega.

Vernero, I., Stefanin, R., Gambino, M., & Schindler, O. (1998). *Cartella Logopedica—Età Evolutiva (Italian Phonemic Examination Protocol: Logopedic Folder—Developmental Age).* Torino, Italy: Omega.

Wagener, K., & Kollmeier, B. (2005). Evaluation des Oldenburger Satztests mit Kindern und Oldenburger Kinder-Satztest. *Z. Audiol.*, 44:134–143.

Wagener, K., Kühnel, V., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. 1. Design des Oldenburger Satztests. *Z Audiol/Audiol Acoust*, 38:4-15.

Westhausen, N., Huber, R., H, B., Sinha, R., Rennies, J., & Meyer, B. (2021). Reduction of subjective listening effort for TV broadcast signals with recurrent neural networks. *IEEE/ACM Trans. Audio, Speech, Language Process*, 29:3541–3550.

WHO, W. H. (2021). *World report on hearing.* Geneva: World Health Organization.

Wilson BS, F. C., Wolford, R., Eddington, D., & Rabinowitz, W. (1991). Better speech recognition with cochlear implants. *Nature*, 352(6332):236–8.

Withers, S., Gibson, W., Greenberg, S., & Bray, M. (2011). Comparison of outcomes in a case of bilateral cochlear implantation using devices manufactured by two different implant companies (Cochlear Corporation and Med-El). *Cochlear Implants Int.* , 2(2):124– 6.

Yamasoba, T., Lin, F., Someya, S., Kashio, A., Sakamoto, T., & Kondo, K. (2013). Current concepts in age-related hearing loss: Epidemiology and mechanistic pathways,. *Hearing Res.*, 303:30–38.

Zucca, M., Albera, A., Albera, R., Montuschi, C., Della Gatta, B., & Canale, A. (2022). Cochlear implant results in older adults with post-lingual deafness: the role of "Top-Down" neurocognitive mechanisms. *Int J Environ Res Public Health.*, 19(3):1343.

Zwicker, E., & Schorn, K. (1982). Temporal resolution in hard-of-hearing patients. *Audiology.*, 21(6):474-92.