
STANDARDISED HEARING LOSS RISK PROFILES WITH STATE-SPACE MODELS

Marta Campi

Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
mcampi@pasteur.fr

Gareth W. Peters

Department of Statistics and Applied Probability
University of California Santa Barbara
Santa Barbara, CA 93106, United States
garethpeters@ucsb.edu

Perrine Morvan

Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
pmorvan@pasteur.fr

Mareike Buhl

Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
mbuhl@pasteur.fr

Hung Thai-Van

Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
htaivan@pasteur.fr

January 7, 2025

ABSTRACT

Hearing loss is a growing public health concern, affecting millions worldwide and contributing to impaired communication, social isolation, and reduced quality of life. As a hidden condition, hearing loss is typically diagnosed through behavioural tests like pure-tone audiometry, which often fail to capture the full extent of auditory deficits. Additional tests, such as speech-in-quiet and speech-in-noise, provide a more detailed understanding, yet they are underutilised due to limitations in equipment availability and time constraints in clinical settings. To address these diagnostic limitations, we propose an advanced method for profiling hearing loss dynamics by integrating audiogram and speech test data for risk assessment and prevalence estimation. Our approach utilises state-space models (SSMs), a mathematical framework that models hidden variables, accounting for the unobservable aspects of hearing loss and inferring a common trend across population segments. We develop a baseline state-space model relying solely on audiograms and an extended version that incorporates speech tests, allowing us to capture multiple aspects of auditory function. A rigorous inference procedure, using the estimated likelihoods of these models, is employed to test for statistical differences in auditory profiles across population segments, including by degree of hearing loss, age and sex. This procedure thoroughly evaluates how specific population characteristics affect auditory performance. Our ultimate goal is to establish a national benchmark for clinical practice that supports personalised patient care, improving diagnosis, treatment planning, and monitoring of hearing impairment. By integrating audiometric and speech data, we aim to create standardised risk profiles that inform public health policy, guide targeted screenings, and facilitate timely interventions for individuals at risk of progressive hearing loss.

Keywords State-Space Model, Kalman Filter, Likelihood Ratio Test, Hearing Loss

1 Introduction

The World Health Organization’s (WHO) ‘World Report on Hearing’ [1] underscores the substantial global burden of hearing loss, affecting over 430 million individuals, including 34 million children, who require rehabilitation. Projections indicate that by 2050, more than 700 million people—1 in 10—will be affected. This issue not only hampers communication but also poses significant societal and economic challenges, including increased healthcare costs, reduced productivity, and lower educational outcomes. Hearing loss is linked to various adverse effects, such as delayed language development in children, social isolation, depression, cognitive decline, and dementia [2].

Despite not being life-threatening, hearing loss increases the risk of serious health complications and severely affects quality of life. The limited prevalence estimates often rely on small, non-representative studies focusing on specific age groups and self-reports instead of objective audiometric testing [3, 4, 5, 6, 7, 8, 9, 10]. Many data sources date back to the 1990s [11], yet the prevalence of hearing loss has likely increased with the aging population [12, 13, 14]. The need for more representative studies, especially in Europe, remains critical, as shown by a recent French study estimating hearing loss prevalence at 8.5% to 16.1%, based on self-reports [15].

Hearing disorders arise from genetic, environmental, and age-related factors and have been investigated through various methodologies [16, 17, 18, 19, 20]. Sex differences in auditory processing have been identified, indicating that males and females may experience hearing loss differently due to factors like noise exposure and physiological variations [21, 22, 23, 24].

Research in this field focuses on how cochlear damage and central auditory issues affect hearing sensitivity (attenuation) and clarity (distortion), assessed through audiological tests. Plomp (1978) demonstrated that hearing loss impacts speech perception in quiet and noisy environments [25, 26, 27]. Attenuation relates to the inaudibility of speech signals due to pure-tone hearing loss, while distortion concerns clarity in auditory processing [28].

Clinical practice typically employs pure-tone audiometry and speech audiometry to evaluate these components. Pure-tone audiometry measures the lowest detectable sound pressure levels for various frequencies [29]. In contrast, speech audiometry evaluates speech recognition thresholds in quiet and noisy environments [30, 31, 32]. However, clinical focus often prioritizes quiet assessments, although speech recognition in noise is more relevant for daily life communication.

Suprathreshold speech perception tests offer valuable insights into auditory deficits that standard audiometric testing may overlook [33, 34]. Incorporating these assessments into clinical practice can improve diagnostic accuracy and support personalized interventions for individuals with hearing loss [35].

In summary, accurate hearing assessment requires both pure-tone and speech tests to create a nuanced auditory profile [36]. Advanced statistical and machine learning methods can integrate diverse data, including audiometric assessments and cognitive evaluations, leading to improved diagnostic capabilities. Recent research focuses on classifying individuals based on audiometric patterns and understanding the relationships between these tests and speech recognition abilities, ultimately contributing to a more comprehensive understanding of hearing loss [37, 38, 16, 17].

1.1 Risk Assessment in Hearing Loss: Policy Making and Clinical Practice Development

Efficient risk assessment tools are vital in precision medicine, significantly enhancing health and patient care. In contrast to established frameworks like the Framingham Risk Score in cardiology [39] or oncology [40], the field of hearing health currently lacks robust risk assessment mechanisms. Identifying individuals at higher risk for developing hearing impairment and the progression of existing conditions is crucial for enabling timely interventions.

Policy-making for hearing health is essential due to frequent under-treatment, influenced by factors beyond the high cost of hearing aids [41]. Low adoption rates of hearing aids stem from inadequate testing, poor device fitting, and the inability to detect hidden hearing loss (also known as suprathreshold deficits), which standard audiometry often overlooks. Societal challenges, such as limited accommodations, poor access to care, and a lack of awareness, further exacerbate these issues [41]. This underscores the urgent need for comprehensive policies that make hearing aids more affordable and accessible [42].

Developing standardized benchmark risk profiles for various population segments—including degrees of hearing loss, age groups, and sex—can enhance policy-making. These profiles provide insights into demographic-specific needs and risks, enabling personalized treatment plans and informing public health initiatives. Clinically, standardized risk profiles assist healthcare professionals in setting hearing aid parameters more accurately, improving patient satisfaction, and enabling the sharing of risk profiles while preserving data privacy. Furthermore, such profiles ensure early identification of high-risk individuals, facilitating targeted screenings and timely interventions.

Research on hearing loss prevalence underscores the importance of detailed demographic data, highlighting variations across regions and age groups [43, 44, 45]. While primarily focused on prevalence estimation, these studies stress the significance of demographic factors in developing nuanced risk profiles. For example, [46] discusses the burden of hearing loss in the U.S., advocating for ongoing assessments to inform healthcare strategies, while [47] explores the prevalence among older adults, revealing gaps that risk profiles could help address.

The CONSTANCES cohort study [2] illustrates these concepts, estimating that 6 million people—approximately 9% of France’s population—are affected by hearing loss [48]. The study revealed a low rate of hearing aid use among older individuals, emphasizing the need for targeted public health strategies and improved awareness and affordability of hearing health services.

Building on [2], our study utilizes a dataset from Amplifon France, which includes 48,144 individuals aged 40 to 90 with symmetric hearing loss (defined as less than 15 dB difference in pure-tone average (PTA, detailed later) between ears) [49]. For all individuals, data from pure-tone audiogram, a speech test in quiet, a speech test in noise, as well as age and sex were available. Although this sample represents only 0.8% of the estimated hearing loss population in France, it offers valuable insights into the behavior and characteristics of individuals with hearing loss.

1.2 State-Space Models for Standardised Risk Profiling in Hearing Loss

Building on existing literature concerning hearing loss, audiological diagnostic tests, clinical practices, risk assessment, and prevalence estimation methods, our objective is to develop a method for profiling population dynamics related to hearing loss. Additionally, we aim to establish a rigorous inference procedure to test differences between population dynamics characterised solely by audiogram data and those that integrate speech tests. This approach will help assess the added value of speech tests in hearing loss risk assessment. Ultimately, the proposed model seeks to create a national benchmark for clinical practice, guiding policy-making and improving patient care through more accurate and personalised interventions.

It is essential to recognise that hearing loss is an unobserved phenomenon revealed only through various measurements that describe its progression and behaviour. Furthermore, hearing loss is not static; it evolves over time due to factors such as ageing, noise exposure, environmental conditions, and other health issues. The relationships between different frequencies in the pure-tone audiogram further complicate this evolution, introducing additional variability into the process.

Recent advances in statistical methodologies for hearing loss studies offer valuable insights. For instance, [50] proposed a multivariate mixed-effects model with order-restricted inference to analyse the natural progression of hearing loss across different frequencies and age groups. This model acknowledges the monotonic relationship between age and frequency-related hearing threshold elevation, emphasizing the importance of incorporating parameter constraints to enhance the accuracy of hearing loss estimations and reduce misleading inferences. Similarly, [51] developed a Bayesian log-normal distribution model to address complexities in analysing hearing threshold data, such as heavy-tailed distributions, missing data, and censoring mechanisms, which are relevant in contexts like bacterial infections and surgical interventions. Their approach underscores the need for advanced probabilistic methods to accurately quantify risks associated with hearing loss progression.

In multicenter studies, complex correlation structures pose challenges. [52] developed an analytical framework to properly account for multilevel correlations in clustered data—such as those between ears and testing sites—using a linear mixed-effects model. Finally, [53] introduced a regression-based methodology to investigate how treatments influence disease progression over time, providing a means to assess intervention efficacy through an interaction term that measures changes in outcomes relative to historical progression. Collectively, these studies establish a foundation for developing more sophisticated, constraint-aware statistical models that enhance understanding of hearing loss dynamics, quantify associated risks, and guide improved public health strategies and personalized clinical interventions.

Given this complexity, modelling hearing loss necessitates a thoughtful approach that addresses several key challenges. First, the frequency-interactive nature of hearing must be considered, acknowledging the nonlinear interactions between frequencies. Second, the model must manage two primary error sources: observational error, arising from human, environmental, or instrumental factors, and systematic error, encompassing consistent biases such as calibration issues or testing procedure biases. Lastly, age-related hearing loss should be a central component of the model, as ageing significantly impacts hearing deterioration.

To effectively capture the dynamic and stochastic nature of hearing loss, we propose using state-space models (SSMs), a sophisticated modeling approach well-suited for this purpose. SSMs are probabilistic graphical models that describe the relationships between latent state variables and observed measurements [54]. They are particularly effective in situations where standard regression methods may face challenges due to confounding variables. SSM frameworks

enable us to incorporate a range of potential explanatory factors that may not be directly observable or readily available, embedding them within a stochastic latent state evolution that models an interpretable structural component of the observation process. These models are particularly adept at unravelling intricate temporal dependencies and nonlinear behaviours, making them widely used in economics, biology, and epidemiology [55, 56, 57, 58, 59, 60].

SSMs excel in time series analysis by modelling temporal autocorrelation, which helps distinguish between process variation (systematic error) and observational error [61]. In this hierarchical framework, two time series are involved: the unobserved state, reflecting the true hidden state of the latent stochastic trend, and the observation series, representing measurements related to the state series. Systematic error captures variation in the underlying process over time, while observational error accounts for discrepancies due to randomness or measurement imprecision. Fitting an SSM allows us to obtain both observational parameters and hidden state estimates, providing a more accurate reflection of the true state of nature.

These hidden states are crucial for risk profiling, enabling the identification of trends that indicate a higher risk of hearing loss and informing targeted interventions. Moreover, SSMs facilitate the abstraction of numerous difficult-to-observe covariate factors into a stochastic trend factor, capturing the long-term dynamics of hearing loss. This ability to integrate the latent, evolving characteristics of hearing loss—along with frequency interactions, error sources, and age-related changes—makes SSMs an ideal choice for our modelling hearing loss profiles.

Our work develops a class of SSMs that operates across the frequency domain of hearing rather than over time. A visual characterisation of our framework is provided in Figure 1. This model accommodates both observational and systematic errors, with the latent process representing underlying hearing sensitivity. The observed frequency series corresponds to the audiogram, ordered from lowest to highest frequencies, forming a structured sequence of measurements. Central to our approach is the hypothesis of an age based hearing performance age term structure, influenced at a population level by a frequency specific stochastic period effect, denoted as κ_f , capturing a systematic trend in hearing sensitivity across different frequencies and age groups. A characteristic system diagram explaining the structure of the SSM we propose is presented in Figure 2. The proposed SSM framework will be demonstrated to be particularly effective at modelling age-related hearing loss, where the progressive decline and complex interplay between neighbouring frequencies align naturally with the SSM framework. High frequencies, which tend to exhibit earlier and more pronounced declines, further highlight the importance of a model that captures these nuanced changes.

Additionally, the SSM framework excels at capturing the dependencies and nonlinearities inherent in hearing data. We propose starting with the audiogram as the response variable to infer a baseline hearing trend reflecting hearing loss across different frequencies for the overall population. This inferred trend is especially valuable as it represents a latent, unobserved stochastic factor, offering insights into underlying patterns of hearing loss that standard audiometric tests cannot directly measure. The SSM also accounts for autocorrelation in audiogram data, which arises from the repetitive presentation of frequencies at various intensity levels during testing. By leveraging this autocorrelation, our model provides a robust framework to understand age-related progression and frequency-specific variations in hearing sensitivity.

The formulation of this SSM requires transforming raw audiogram data into proportions, where each proportion is determined by setting an empirical quantile derived from the entire sample for each frequency. These quantiles act as cut-off thresholds, and we compute the proportion of individuals exceeding these thresholds, identifying those with poorer hearing performance. By applying several empirical, data-driven quantiles, we can observe how these proportions vary across frequencies and age groups. The proportions serve as the response variables of the SSM model, whose derived latent factors reflect how hearing loss progresses. We refer to this as the *baseline model*, which can then be applied to analyze different population segments (i.e. degree of hearing loss, sex, etc.).

Figure 1 illustrates the concept behind the baseline model. Audiogram measurements from the sample population are converted into proportions for each considered age group. We denote these proportions by π_f , where f indicates the frequency range, with values ranging from 0 to 1. Consequently, the trend κ_f is defined as the underlying progression of hearing loss risk, which we assume to be ordered by frequencies, forming a latent vector in the model’s structure.

Following the model definition, we employ an autoregressive process of order 1. This structure dictates that only the preceding frequency influences the subsequent one, facilitating a natural ordering in the frequency response. The autoregressive assumption establishes that high-frequency responses are influenced by their lower-frequency counterparts, effectively capturing the progressive nature of hearing loss across the frequency spectrum. The resulting estimated parameters will represent the baseline hearing status common across all frequencies, i.e. α , while β will capture sensitivity differences between frequencies. Both α and β can be derived from the latent trend κ_f , which encapsulates systematic changes across frequencies over age.

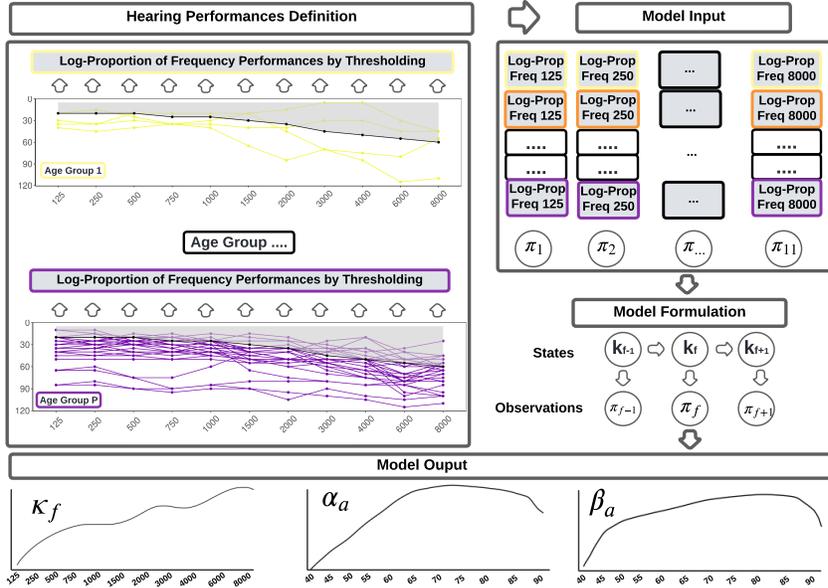


Figure 1: Baseline Model Presentation: We define hearing performance using audiograms categorized by age groups and empirical quantiles for each frequency. We calculate the log-proportion of individuals exceeding these quantiles, which serve as cut-off points for assessing hearing loss severity. The model incorporates a hidden hearing loss state with an autoregressive structure of order 1, where each frequency influences the next, observed through the audiogram. The output includes risk profiles that quantify the prevalence of hearing loss trends across frequencies and age groups (κ_f), a baseline reference for hearing loss by age (α), and the sensitivity of frequencies to the hidden hearing loss state trend (β). Additionally, we analyze the model by degree of hearing loss and sex to reveal differences in risk profiles and patterns.

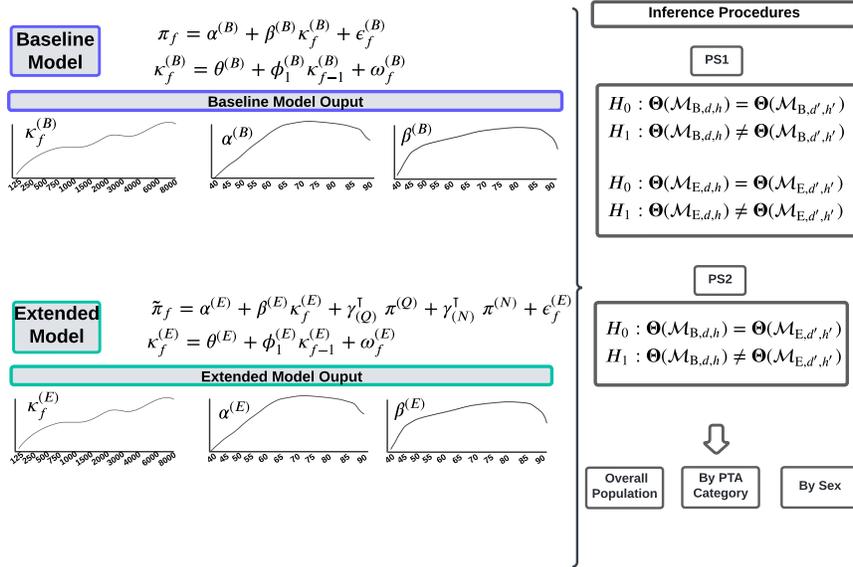


Figure 2: Baseline versus Extended Model: We evaluate two models, i.e., one based solely on audiograms and the other integrating both audiograms and speech tests. The objective is to analyze how risk profiles differ across population segments, including overall, by degree of hearing loss, and by sex. Inference procedures focus on hearing loss severity indices h and h' , with similar considerations for sex. Detailed model specifications are outlined in Subsection 2.3, while problem statements comparing different parameter spaces are introduced in Subsection 2.4.

1.3 Contributions

We formulate the Campi-Peters-Morvan-Buhl-Thai-Van (CPMBT) model, a specific type of state-space structured model. Similar to the widely used Lee-Carter model [62] in demography for mortality rate forecasting and the Nelson-Siegel model in interest rate finance [63], the CPMBT model leverages established methodologies to analyze hearing loss patterns and provides a population-level standardized reference benchmark of age-sex-frequency and age-hearing-loss-degree-frequency specific risk profiles. The formulation of the model involves constructing hearing loss proportions as well as speech-in-noise and speech-in-quiet proportions. Two models are considered: a baseline model based solely on the audiogram and an extended model incorporating the speech tests in addition. Detailed inference procedures are discussed. We rely on a dataset comprising 48,144 individuals (representing approximately 0.8% of the French population affected by hearing loss) with varying degrees of hearing loss, for whom age, sex, audiogram, and two speech tests (in quiet and in noise) were collected. Our contributions are methodological and conceptual at different levels of modelling of state-space models as well as their application in hearing loss. These are multiple and given as follows:

- **Model Formulation.** We formulate the CPMBT model as an innovative state-space structured approach designed to deliver national benchmark references for various risk profiles, including age-frequency, age-hearing-loss-degree-frequency, and age-sex-frequency categories. This model extends traditional state-space frameworks from their conventional time-domain applications to the frequency domain, providing a fresh perspective on hearing loss dynamics across different frequencies. By integrating both the relationship across threshold tests, such as pure-tone audiograms and speech tests, and the risk profiling task, the CPMBT model offers a comprehensive solution. It simultaneously addresses the interplay between different assessment methods and the construction of detailed risk profiles, enhancing both the accuracy of hearing loss assessments and the understanding of underlying auditory mechanisms.
- **Hearing Loss Risk Profiling Definition.** Our approach estimates hearing loss incidence proportions across various age groups, sexes, and degrees of hearing loss severity using empirical quantiles derived from the entire population. By calculating the proportion of individuals within each age cohort who exceed data-driven dB empirical quantiles, we identify those with respective poorer hearing performance. We provide a methodological solution for the derivation of demographic curves descriptive of hearing loss dynamics across age and frequency domains. Our model is designed to be extendable. By incorporating additional data sources, such as survey responses or other audiological tests, we can refine these profiles further. This comprehensive approach enhances our understanding of hearing loss and improves the precision of targeted interventions.
- **Relationships between Pure-Tone Audiogram and Speech Tests.** The CPMBT model incorporates speech tests into the analysis, providing a comprehensive inference framework that accounts for both sensitivity to sound and clarity of speech perception. By exploring the relationship between speech-in-quiet and speech-in-noise scores and pure-tone thresholds, the model identifies the explanatory power of the speech tests, thereby enhancing the precision of tailored interventions and informing public health strategies.

The rest of the paper is organized as follows: Firstly, we present the model and the problem statements. Secondly, we present the partial linear regression framework adapted to the introduced state space model. Following this, a data description section is presented, followed by the results. Lastly, we provide discussion and conclusions.

2 Model Presentation & Problem Statements

This section outlines the baseline and extended state-space models, which are applied to the study population and subgroups based on hearing loss degree and sex. After introducing the necessary notation, the process of constructing hearing loss proportions and deriving empirical quantiles is explained. The models and their formulations are then presented, followed by the problem statements. Finally, the marginal likelihood derivations needed for statistical testing are provided.

2.1 Notation

Our approach frames an audiogram as a time series over the frequency domain, where patients' responses to pure harmonics (or tones) at sequentially ordered frequencies are recorded. We aim to characterise these responses across age, frequency, and intensity by developing a model capturing relationships along these three axes to quantify hearing loss incidence. In this subsection, we introduce the formal definitions required to develop our model framework and the model itself.

Consider a dataset where we group patients into 5-year age intervals defined as $[40-45], [45-50], [50-55], \dots, [85-90]$. Let the index for these age groups be denoted by p , where $p \in \{1, 2, \dots, P\}$. Now, define the random variable $X_{n,p,f}$ as representing the pure tone stimulus response in decibels Hearing Level (dB HL, the standard audiometric measurement scale) for the n -th patient who belongs to age group p at frequency f , where $f \in \{f_1, f_2, \dots, f_F\}$ and $f_1 < f_2 < \dots < f_F$. Here, F corresponds to the total number of frequencies measured, which in this context is $F = 11$, with the specific frequencies typically being the standard audiogram measurements frequencies that exist in a bandwidth defined between 125 Hz through to 8,000 Hz. The total number of people is denoted by N , while the number of people in each age group p is denoted by N_p , with $N = \sum_{p=1}^P N_p$.

For each subject, we consider audiogram measurements as a time series trajectory ordered from lowest to highest frequency (125 Hz to 8,000 Hz). These measurements can be considered as realisations of a stochastic process of the audiogram performance for a given subject of a given age, sex and hearing loss severity diagnosis, denoted by random vector, denoted by $\mathbf{X}_{n,p}(\mathbf{f}) := \{X_{n,p,1}, X_{n,p,2}, \dots, X_{n,p,F}\}$. Define $\mathcal{X} = \{x_1, x_2, \dots, x_I\}$ the set of all possible hearing thresholds (in dB HL), for which random variable $X_{n,p,f}$ takes its values for each frequency f , such that $X_{n,p,f} \in \mathcal{X}$. Then an observed audiogram measurement sequence can be written as $\mathbf{x}_{n,p} \in \mathcal{X}^F$, corresponding to a row vector of dimension F for age group p and patient n . By considering N_p the number of participants in age group p , we obtain the random matrix $X_{\cdot,p,\cdot} \in \mathcal{X}^{N_p \times F}$ whose observed realisation is $\mathbf{x}_{N_p \times F}$, which captures the responses of N_p patients in the age group a_p . All subjects' samples collected in this work are affected by hearing loss and are further categorised according to their pure tone average (also known as PTA). More details on this aspect are given below.

Consider now the two random variables $S_{n,p}^{(Q)}$ and $S_{n,p}^{(N)}$ reflecting the speech reception threshold, SRT_Q , and signal-to-noise-ratio, SRT_N , obtained from the speech-in-quiet and speech-in-noise tests of patient n belonging to age group p , respectively. Define $\mathcal{S}^{(Q)} = \{s_1^{(Q)}, s_2^{(Q)}, \dots, S^{(Q)}\}$ and $\mathcal{S}^{(N)} = \{s_1^{(N)}, s_2^{(N)}, \dots, S^{(N)}\}$ the sets of all possible SRT_Q and SRT_N values, respectively, then $S_{n,p}^{(Q)} \in \mathcal{S}^{(Q)}$ and $S_{n,p}^{(N)} \in \mathcal{S}^{(N)}$. An observed realisation of $S_{n,p}^{(Q)}$ is denoted as $s_{n,p}^{(Q)}$ and a realisation of $S_{n,p}^{(N)}$ is denoted as $s_{n,p}^{(N)}$ and corresponds to the two values of SRT_Q and SRT_N for patient n in the age group p , when each patient's SRT_Q and SRT_N are stacked into matrices one may define the random matrices $\{\mathbf{S}_{N_p \times 1}^{(Q)}\}_{p=1}^P$ and $\{\mathbf{S}_{N_p \times 1}^{(N)}\}_{p=1}^P$.

Through a regression state space model, which accommodates age term structure hearing performance, we aim to analyse the dynamics of the time series $\mathbf{X}_{n,p}(\mathbf{f})$ across age, frequency, and intensity dimensions. The model operates on hearing loss log proportion, defined for each audiological test using quantiles computed over the entire population (introduced in the following subsection). Our research presents a novel approach by defining a baseline model constructed with the audiogram data and an extended model incorporating SRT_Q and SRT_N to observe their statistical significance with respect to age, frequency, and intensity of the audiogram.

2.2 Hearing Loss Proportion

When developing the SSM formulation, the data from the audiogram, SRT_Q and SRT_N measurements is transformed to proportions of exceedence for a given hearing severity level based on total population empirical quantile levels, per age group and per frequency. This allows for dimension reduction in the data, reducing a large sample of patients measurements per frequency to a single scalar proportion measure, treated as a time series over the frequency domain, that characterises the degree of hearing performance for a given category of hearing loss severity by age group and sex. A summary of the procedure is provided in Figure 3. The process creates a standardized measure of performance for each hearing loss severity category, relative to the entire population, that is beneficial when developing a set of standardised hearing loss performance measures. This method further helps address heterogeneity across the frequency domain and age groups by standardising the data into comparable proportions. This reduces the impact of outliers and noise, leading to a more robust and reliable analysis that accounts for variability within the population.

2.2.1 Population Relative Performance Thresholds via Empirical Quantiles

The data set of audiogram performance measurements for a set of N total patients is transformed into hearing loss severity category specific relative proportions by age and frequency, based on a set of D relative performance thresholds. These relative performance thresholds at each frequency are obtained based on the following empirical quantile levels of the total population's performance. Computing the proportion of individuals exceeding these thresholds within each subgroup provides a measure of relative performance. The threshold levels are based on increasing probability levels $d \in \{0.2, 0.4, 0.6, 0.8, 0.9\}$ (here $D = 5$), that in turn will produce a set of performance thresholds based on the empirical quantiles associated with these probability levels.

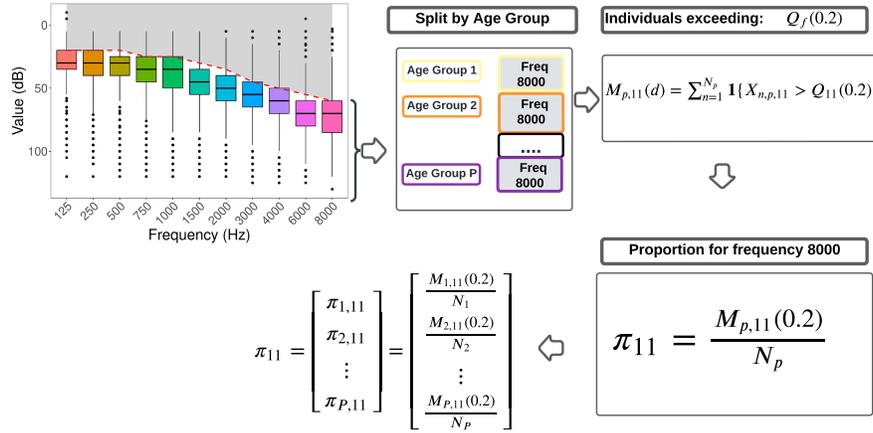


Figure 3: Construction of hearing threshold proportions at frequency $f_{11} = 8,000$ Hz (using same representation as in Figure 4). The proportion construction focuses on individuals whose thresholds exceed the empirical quantile $Q_{11}(0.2) = 60$ dB HL. These individuals are grouped by age, and $M_{p,11}(0.2)$ represents the count of people with thresholds above 60 dB. Dividing this by the total number of individuals in each age group (N_p) gives the proportion π_{11} for each age group.

The empirical quantile $Q_f(d)$ for frequency f of the audiogram data samples is defined using the empirical cumulative distribution function (ECDF) of the data comprising N total number of patients as

$$Q_f(d) = \hat{F}_{X_{\cdot,\cdot},f}^{-1}(d) \quad \text{with} \quad \hat{F}_{X_{\cdot,\cdot},f}(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(X_{\cdot,\cdot},f \leq x)$$

where $\hat{F}_{X_{\cdot,\cdot},f}$ is the empirical cumulative distribution function (ECDF) for the data related to frequency f , and $\mathbf{1}(X_{\cdot,\cdot},f \leq x)$ is an indicator function that equals 1 if $X_{\cdot,\cdot},f \leq x$ and 0 otherwise, and N is the total number of observations. Empirical quantile based thresholds are also selected in an analogous manner for the SRT_Q and SRT_N observed sample data across the total N patients and denoted by $Q^{(Q)}(d) = \hat{F}_{S^{(Q)}}^{-1}(d)$ and $Q^{(N)}(d) = \hat{F}_{S^{(N)}}^{-1}(d)$. The resulting empirical quantiles are summarised in Supplementary Appendix Table 1.

Empirical Quantiles													
Probabilistic Level	Frequencies											SRT_Q	SRT_N
d	Q_{125}	Q_{250}	Q_{500}	Q_{750}	Q_{1000}	Q_{1500}	Q_{2000}	Q_{3000}	Q_{4000}	Q_{6000}	Q_{8000}	$Q^{(Q)}$	$Q^{(N)}$
20%	20	20	20	25	25	30	35	45	50	55	60	36	1
40%	25	25	30	30	35	40	45	55	60	65	70	42	3
60%	30	30	35	40	40	50	55	60	65	75	75	48	4
80%	40	40	45	50	50	60	60	70	75	85	85	56	7
90%	50	50	55	55	60	65	70	75	80	95	95	62	10

Table 1: Empirical quantiles for each individual frequency of the audiogram, the SRT_Q and SRT_N .

Figure 4 presents five panels, each showing boxplots of audiogram frequencies for the entire population. The red lines in each plot correspond to the empirical quantiles, one for each of the five probabilistic levels d , as given in Table 1. The gray areas indicate the portion of individuals who did not exceed the empirical quantiles and are therefore excluded from the computation of the proportion. As we move across the five panels, it becomes evident that the empirical quantiles capture different segments of the underlying population. An equivalent picture considering the hearing loss degree is given in the Supplementary Information, Section 8.

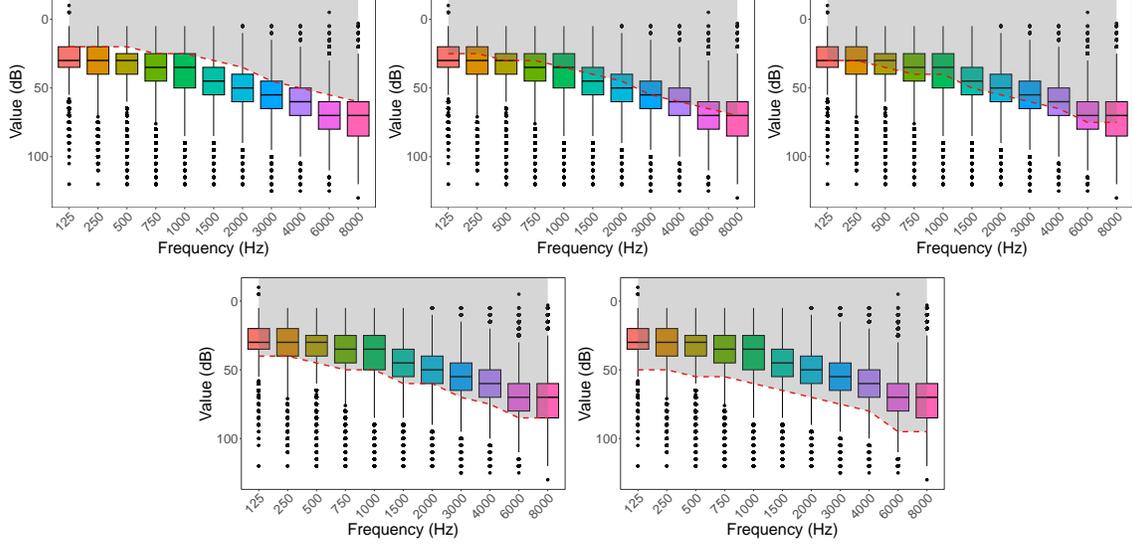


Figure 4: The figure displays boxplots of audiogram thresholds across frequencies, with the x-axis ordered by frequency and the y-axis inverted to show higher dB levels at the top. Each plot corresponds to a different quantile level $Q_f(d)$, marked by red dashed lines, ranging from the 20th percentile in the top left to the 90th percentile in the bottom right, as outlined in Table 1.

2.2.2 Transforming Sample Audiogram Data to Hearing Severity Relative Proportion Time Series

For each of the empirical quantiles $Q_f(d)$ and the audiogram data matrices for each age group $\{\mathbf{X}_{N_p \times F}\}_{p=1}^P$, one counts how many patients within each frequency column exceed $Q_f(d)$, for each of the P matrices, giving relative proportions

$$\pi_{p,f} = \frac{1}{N_p} M_{p,f}(d) = \frac{1}{N_p} \sum_{n=1}^{N_p} \mathbf{1}\{X_{n,p,f} > Q_f(d)\} \quad (1)$$

Thus, $M_{p,f}(d)$ represents the number of participants in age group p for whom the hearing threshold at frequency f exceeds the empirical quantile $Q_f(d)$. Equivalently, for the SRT_Q and the SRT_N one obtains relative proportions using alternatively the counts $M_p^{(Q)}(d) = \sum_{n=1}^{N_p} \mathbf{1}\{S_{n,p}^{(Q)} > Q^{(Q)}(d)\}$ and $M_p^{(N)}(d) = \sum_{n=1}^{N_p} \mathbf{1}\{S_{n,p}^{(N)} > Q^{(N)}(d)\}$. This procedure produces time series vectors over frequencies $\{\pi_f(d)\}_{f=1:F} = \{[\pi_{p,1}(d), \dots, \pi_{p,F}(d)]\}$, and analogous quantities $\pi^{(Q)}(d)$ and $\pi^{(N)}(d)$ that are inputs to the state space modelling.

Note that we apply a log transformation and centering to π_f , resulting in a logarithmic central hearing loss proportion denoted by $\tilde{\pi}_f$. Specifically, this transformation involves taking the natural logarithm of π_f , which helps stabilizing the variance by reducing the effect of outliers and compressing the range of the proportions, thereby mitigating the impact of extreme values. The centered log values then allow for a more symmetric distribution around zero. This transformation ensures homoskedasticity, meaning that the variance of the residuals remains constant across different frequencies, which is crucial for the assumptions underlying many statistical models, including state space models. By transforming the proportions to the real line, we can treat them as continuous variables without the constraint of being bounded between 0 and 1, allowing the model to effectively accommodate differences across frequencies while maintaining consistent statistical properties and improving interpretability. More information is provided in the Supplementary Information.

Regarding the interpretation of these threshold exceedence proportions we note that the successful detection of sounds at higher thresholds $Q_f(d)$ for a given frequency, or at higher thresholds in SRT_Q and SRT_N conditions, indicates that these individuals have more severe hearing loss. Higher detection thresholds correspond to poorer hearing sensitivity. Thus, when computing proportions for these thresholds, we are capturing the ability of individuals with more pronounced hearing loss to detect sounds, relative to the empirical quantiles of the hearing thresholds. Throughout the rest of the paper, we will refer to pure-tone proportion, SRT_Q and SRT_N proportion, respectively.

2.3 The Campi-Peters-Morvan-Buhl-Thai-Van (CPMBT) State-Space Model

We now introduce the proposed SSM formulation, presented generically with the understanding that the model fitting may be performed both for the entire population and for subgroups defined by different degrees of hearing loss and sex. We note that the index for the threshold level d is dropped from notation for the general representation below, however it is understood that each of the Baseline and Extended models will be fit to provide a standardised population profile for each of the $d \in \{1, \dots, D\}$ levels of performance. We seek a dynamic structural explanation of the age term structure proportions, $\tilde{\pi}_f$ of relative hearing loss performance per frequency under the SSM dynamics given by

$$\begin{aligned} \mathcal{M}_{B,d} : \text{Baseline Model.} \quad & \tilde{\pi}_f = \alpha^{(B)} + \beta^{(B)} \kappa_f^{(B)} + \epsilon_f^{(B)}, \\ & \kappa_f^{(B)} = \theta^{(B)} + \phi_1^{(B)} \kappa_{f-1}^{(B)} + \omega_f^{(B)} \end{aligned} \quad (2)$$

$$\begin{aligned} \mathcal{M}_{E,d} : \text{Extended Model.} \quad & \tilde{\pi}_f = \alpha^{(E)} + \beta^{(E)} \kappa_f^{(E)} + \gamma_{(Q)}^\top \pi^{(Q)} + \gamma_{(N)}^\top \pi^{(N)} + \epsilon_f^{(E)}, \\ & \kappa_f^{(E)} = \theta^{(E)} + \phi_1^{(E)} \kappa_{f-1}^{(E)} + \omega_f^{(E)} \end{aligned} \quad (3)$$

The model parameters are each interpreted as specified in the following Table 2. Each model parameter is indexed by B if it belongs to the baseline model or E for the extended one.

Models Interpretation & Hearing Loss Risk Parameter Profile	
Parameter	Description
α	Age-specific ($\alpha = (\alpha_1, \dots, \alpha_P)^\top$) intercepts capturing the baseline hearing loss level for each age group.
κ_f	Common baseline trend in hearing loss proportions across all age groups for frequency f , i.e. a common effect applicable to all ages.
β	Age-specific ($\beta = (\beta_1, \dots, \beta_P)^\top$) sensitivity of hearing loss proportions $\tilde{\pi}_f$ to a change in the general trend of hearing loss κ_f . As such it measures how the pure-tone proportions changes with age across frequencies relative to the common frequency effect.
ϵ_f	Age-specific ($\epsilon_f = (\epsilon_{1f}, \dots, \epsilon_{Pf})^\top$) noise term for observation error assumed distributed as $\epsilon_f \sim \mathcal{N}(0, \text{diag}(\sigma_{\epsilon_1}^2, \dots, \sigma_{\epsilon_P}^2))$
ω_f	Noise term for variation in the latent state κ_f that is an inherent process stochasticity, assumed distributed as $\omega_f \sim \mathcal{N}(0, \sigma_\omega^2)$.

Table 2: CPMBT Model Parameter Interpretations.

The frequency specific state variable capturing the common trend of the population, denoted κ_f is a latent stochastic factor capturing the baseline hearing loss trend across all frequencies and applicable to all ages. For a given age group p , if the corresponding loading β_p is in absolute value large, it implies the pure-tone proportion for that age group p is more responsive to changes in the age term structure trend changes relative to age groups with smaller absolute values of this loading. Furthermore, in the extended model, $\gamma_{(Q)}$ and $\gamma_{(N)}$ are the parameters of the two column vectors $\pi^{(Q)}$ and $\pi^{(N)}$ each of dimension P , containing the proportions for SRT_Q and SRT_N , respectively.

If the above SSMs are instead fit to a sub-population considering hearing loss degree, then we add another index $h \in \{h_1, \dots, h_H\}$ where h_1 represents the index of the slight hearing loss individuals, h_2 the index for mild hearing loss individuals, etc. The models then are denoted as $\mathcal{M}_{B,d,h}$ and $\mathcal{M}_{E,d,h}$. Note that we present our methodology with respect to hearing loss subgroups, but an equivalent reasoning applies to sex, i.e. we add another index $s \in \{s_1, s_2\}$ where s_1 represents female and s_2 male, respectively.

2.3.1 Model Interpretation of Hearing Performances Definition

A relevant point is to understand how to interpret the introduced models and the information we aim to capture with them. Figure 5 presents the idea behind our conceptualization and further supports the need for a state-space model (SSM) in this scenario.

The audiogram thresholds represent the lowest level of sound (measured in decibels, or dB) that can be heard at a given frequency f 50% of the time. Our objective is to quantify and describe hearing performance by characterizing these

thresholds across the age and frequency domains. To do this, we defined proportions by ordering the study participants from the lowest to the highest hearing thresholds and then partitioned the population using empirical quantiles derived at several probabilistic levels. This approach allowed us to isolate the number of individuals who would fail to detect sounds above a given empirical quantile and to observe how these quantities change across the population, by severity of hearing loss and sex.

Consider the diagrams provided in Figure 5, referring to the overall case hence by considering the whole studied sample without any demographic segmentation. For a given age group p and a probabilistic level of 20%, the empirical quantile represents a specific hearing threshold in dB HL. We then consider the number of individuals whose hearing thresholds exceed this empirical quantile—i.e., individuals who would fail to detect sounds at lower levels than this threshold at a given frequency. This “failure” to detect such a sound is incorporated by the indicator function utilized in Eq. (1) and given as

$$M_{p,f}(d) = \sum_{n=1}^{N_p} \mathbf{1}\{X_{n,p,f} > Q_f(d)\},$$

which specifies the number of patients in age group p , $M_{p,f}(d)$, who do not detect sounds at levels lower than the empirical quantile $Q_f(d)$ at frequency f . Here, the indicator function $\mathbf{1}\{X_{n,p,f} > Q_f(d)\}$ takes a value of 1 when a patient’s threshold $X_{n,p,f}$ exceeds the empirical quantile $Q_f(d)$, marking a failure to hear sounds above this decibel level. This indicator function is formally defined as:

$$\mathbf{1}\{X_{n,p,f} > Q_f(d)\} = \begin{cases} 1 & \text{if } X_{n,p,f} > Q_f(d), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $M_{p,f}(d)$ denotes the total count of such failures in the age group. This formulation allows us to construct a measure of hearing performance that directly reflects the proportion of individuals who cannot detect sounds at progressively louder thresholds, with higher proportions indicating poorer hearing ability within that frequency and age group.

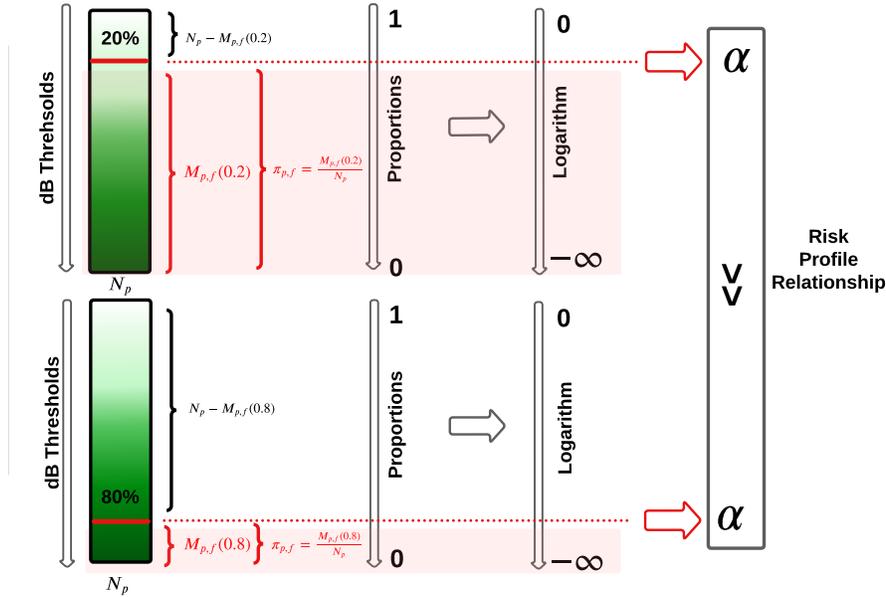


Figure 5: Diagram illustrating the concept of “failure” in detecting sounds, used to define hearing performance. The indicator function counts the number of individuals exceeding a given empirical quantile, indicating a failure to detect sounds below that level. The plot demonstrates how the empirical quantiles increase with increasing dB thresholds, representing worsening hearing loss. Additionally, the diagram shows the resulting proportions and their logarithmic transformations, highlighting the relationship with the risk profile parameters (α) in terms of these empirical quantiles.

As we move to a higher probabilistic level (e.g., from 20% to 80%), the empirical quantile corresponds to a higher sound level (in dB HL), which results in fewer individuals who are unable to detect sounds at this level. Consequently, the proportion of individuals failing to detect sounds at these levels decreases. This dynamic behaviour across

probabilistic levels informs the model’s ability to capture the relative degree of hearing impairment across both age and frequency dimensions, as shown by increasingly negative log-transformed values of the proportions when fewer individuals fail to detect louder sounds.

It is important to note that hearing performance could be defined differently, such as by considering alternative thresholds, cumulative scores, or incorporating multiple measures simultaneously. This flexibility is crucial as it connects the concept of risk profiles underlying different behaviours to the proportions, meaning that the interpretation of risk will change based on how these proportions are constructed. For example, one could define performance based on the combined response at multiple frequencies or in response to different types of stimuli. This nuanced understanding is essential when linking behavioural test outcomes to the constructed proportions and risk profiles.

In our case, the intention is to explain our proportion, which in fact decreases as the empirical quantile increases. For instance, with respect to the α values, this implies that more negative α values will be associated with higher hearing thresholds, thereby indicating poorer hearing performance. As the empirical quantiles increase and represent higher sound levels, the corresponding α values reflect the increasing difficulty individuals face in detecting sounds. Further details on the interpretation of the risk profiles and their behaviour with degree of hearing loss and sex will be given in the Results section.

The relationship between increasing hearing loss degree (i.e., increasing dB thresholds) and increasing empirical quantiles is not simply an inverse linear relationship. If this relationship were strictly inverse and linear, there would be no need to construct our SSM. Instead, we use the SSM because this relationship, which studies (and links with the partial regression framework later introduced) behavioural performance tests, can be influenced by various factors, such as severity of hearing loss and sex, which are included in our analysis, and other demographic factors which are currently not included. This complexity sets the foundation for investigating the proportions of audiological tests responding to more complex stimuli, such as the speech-in-quiet and speech-in-noise tests. For these tests, it is clear that the relationship cannot be just inversely proportional and linear, given the statistical questions we are asking regarding whether the intensity of the sound is the only contributing factor or if there is added information provided by the complexity of the signal in speech tests.

Furthermore, a third important point is that these proportions are highly influenced by the sample size of each segment of the population. Given that our study is observational, we have variability in sample sizes across different age groups and degrees of hearing loss—although sex is less affected. This variability means that risk profile estimates may have increased variance and hence greater uncertainty for segments with smaller sample sizes, though no systematic bias is introduced. Additional details on sample sizes are provided in the Supplementary Information, and further explanation on how this influences the risk profile estimates is given in the Results section.

2.3.2 Model Identification

The baseline and extended models as they are specified will suffer from likelihood based non-identification, since

$$\tilde{\pi}_f = \alpha^{(B)} + \beta^{(B)}\kappa_f^{(B)} + \epsilon_f^{(B)} = \alpha^{(B)} + \beta^{(B)}c + \frac{\beta^{(B)}}{d} \left((\kappa_f^{(B)} - c) d \right) + \epsilon_f^{(B)} = \tilde{\alpha}^{(B)} + \tilde{\beta}^{(B)}\tilde{\kappa}_f^{(B)} + \epsilon_f^{(B)}$$

where c represents a constant shift and d a scaling factor that illustrate the model’s parameter non-uniqueness.

Therefore, when possible it is wise to include an identification constraint. In this case the non-unique identification constraints can be specified as follows for both the baseline (B) and extended (E) models:

$$\sum_{f=1}^F \kappa_f = 0, \quad \sum_{p=1}^P \beta_p = 1.$$

This will ensure that the likelihood will be identified in subsequent estimation procedures. The details about the estimation of the models are provided in Supplementary Information Section 5 and includes the state space filtering derivations, the marginal likelihood formulation and parameter estimation and the partial regression estimation methods explored.

2.4 Testing the Role of SRT_N and SRT_Q in Standardised Hearing Performance Assessment

Two statistical tests are performed in order to assess the significance of SRT_N and SRT_Q in the standardised measures developed by the age-term structure of the CPMBT model formulation in Section 2.3: the Vuong test for the first set of hypothesis testing problem statements (denoted **PS1**) and the Likelihood Ratio Test (LRT) for the second set of hypothesis testing problems (denoted **PS2**). The Vuong test is used to compare non-nested models to assess statistically significant differences in risk profiles. This includes evaluating differences between baseline models for the overall

population versus different PTA categories (giving the severity of hearing loss), as well as between extended models for the overall population versus PTA categories. Note that equivalent tests are performed for sex groups instead of PTA categories; however, we introduce the tests here only for the latter case. Meanwhile, the LRT evaluates how incorporating speech-in-quiet and speech-in-noise tests into the baseline model creates the extended model, determining the added value of these speech tests in explaining variations in hearing loss risk profiles.

These tests are performed by utilising the marginalised likelihood of the formulated SSMs models to check for deviances between them, allowing us to understand differences both within the same class of models (e.g., baseline versus baseline or extended versus extended) and between different classes of models (e.g., baseline versus extended). Each statistical test is formally derived and introduced in the Supplementary Information. Here, we introduce the problem statements to provide an overview of their purpose.

Consider the vectors of the parameters for the baseline and extended model, respectively, defined as

$$\begin{aligned}\Theta(\mathcal{M}_{B,d,h}) &= [\boldsymbol{\alpha}_h^{(B)}, \boldsymbol{\beta}_h^{(B)}, \boldsymbol{\Sigma}_h^{(B)}, \theta_h^{(B)}, \phi_{1,h}^{(B)}, \sigma_{\omega_f,h}^{(B)}] \\ \Theta(\mathcal{M}_{E,d,h}) &= [\boldsymbol{\alpha}_h^{(E)}, \boldsymbol{\beta}_h^{(E)}, \gamma_{h,(Q)}, \gamma_{h,(N)}, \boldsymbol{\Sigma}_h^{(E)}, \theta_h^{(E)}, \phi_{1,h}^{(E)}, \sigma_{\omega_f,h}^{(E)}]\end{aligned}\quad (4)$$

where $\boldsymbol{\Sigma}_h^{(B)} = \text{diag}(\sigma_{\epsilon_1}^{2(B)}, \dots, \sigma_{\epsilon_P}^{2(B)})$ and $\boldsymbol{\Sigma}_h^{(E)} = \text{diag}(\sigma_{\epsilon_1}^{2(E)}, \dots, \sigma_{\epsilon_P}^{2(E)})$.

PS1 inferences are all performed with the Vuong test that treats hypotheses of the type:

$$\begin{aligned}H_0 &: \Theta(\mathcal{M}_{B/E,d,h}) = \Theta(\mathcal{M}_{B/E,d',h'}) \\ H_1 &: \Theta(\mathcal{M}_{B/E,d,h}) \neq \Theta(\mathcal{M}_{B/E,d',h'})\end{aligned}$$

where $d, d' \in \{0.2, 0.4, 0.6, 0.8, 0.9\}$, $h, h' \in \{h_0, h_1, h_2, \dots, h_H\}$ and $h_0 = 0$ corresponds to the case where the model is fit over the entire population and no hearing loss severity is considered. The setting B is using the baseline model and E the extended models. In practice, these tests perform the following comparisons:

- Comparing risk profiles across the overall population versus subpopulation induced by the PTA categories with the baseline and the extended models separately;
- Comparing risk profiles across all the combinations of the PTA categories with the baseline and the extended models separately;
- Comparing risk profiles across all the combinations of the empirical quantiles with the baseline and the extended models separately.

PS2 inferences are all performed with the Likelihood Ratio Test that treats hypotheses of the type:

$$\begin{aligned}H_0 &: \Theta(\mathcal{M}_{B,d,h}) = \Theta(\mathcal{M}_{E,d,h}) \\ H_1 &: \Theta(\mathcal{M}_{B,d,h}) \neq \Theta(\mathcal{M}_{E,d,h})\end{aligned}$$

where $d, d' \in \{0.2, 0.4, 0.6, 0.8, 0.9\}$, $h, h' \in \{h_0, h_1, h_2, \dots, h_H\}$ and $h_0 = 0$ corresponds to the case where the model is fit over the entire population and no hearing loss severity is considered. In practice, these tests perform the following comparisons:

- Comparing risk profiles evaluating the effect of speech tests on hearing loss risk profiles - overall;
- Comparing risk profiles evaluating the effect of speech tests on hearing loss risk profiles - by PTA category (Likelihood Ratio Test).

The test statistics of these sets of hypothesis tests requires the pointwise evaluation of the marginal likelihood. Detailed specifications of the Vuong test and the LRT test are outlined in the Supplementary Information.

2.5 Marginal Likelihood Estimation Recursion

In this subsection, we introduce the steps we performed to derive the marginal likelihood of the baseline and extended CPMBT models. Since the extended model represents an extension of the baseline one, we provide the marginal likelihood for $\mathcal{M}_{E,d}$ and then explain that the one of $\mathcal{M}_{B,d}$ is equivalent by setting the speech parameters to zero, i.e. is a special case of the marginal likelihood of the extended model. All the proofs and derivations for both models are provided in the Supplementary Information, Section 3.

The marginal likelihood of the extended model denoted $f_{(E)}(\tilde{\boldsymbol{\pi}})$ can be formulated as

$$f_E(\tilde{\boldsymbol{\pi}}) = \int f_E(\tilde{\boldsymbol{\pi}}_{1:F} | k_{1:F}) f(k_{1:F}) dk_{1:F} \quad (5)$$

The total law of probability allows us to derive this in a sequence of conditional quantities of the form:

$$f_E(\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}) = \mathcal{N}(\mu_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}})$$

where

$$\mu_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}} = \boldsymbol{\alpha} + \boldsymbol{\beta} \mu_{\kappa_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}} + \gamma_{\mathbf{v}} \mathbf{v} + \gamma_{\mathbf{w}} \mathbf{w}$$

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}} = \sigma_\epsilon^2 \mathbf{1}_P + \boldsymbol{\beta} \left(\phi_{f-1}^2 P_{f-1} + \sigma_{\omega_f}^2 \right) \boldsymbol{\beta}^\top$$

with the filter distributions sufficient statistics given by recursion:

$$\mu_{\kappa_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_i} = \theta + \phi_1 \mu_{\kappa_{f-1} | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_i}$$

$$\sigma_{\kappa_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_i} = \frac{\sigma_\epsilon^2 \sigma_{\kappa_{f-1}}^2}{\sigma_{\kappa_{f-1}}^2 (\boldsymbol{\beta}^\top \boldsymbol{\beta}) + \sigma_\epsilon^2} + \sigma_{\omega_f}^2.$$

Producing a log-likelihood $l(\boldsymbol{\Theta}(\mathcal{M}_E)) = \ln f_E(\tilde{\boldsymbol{\pi}}) = \sum_{f=1}^F \ln f_E(\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1})$ with each component:

$$\begin{aligned} \log \mathcal{L}(\boldsymbol{\Theta}(\mathcal{M}_E)) = & -\frac{FP}{2} \log(2\pi) - \frac{1}{2} \sum_{f=1}^F \left(\log |\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}}| \right. \\ & \left. + (\tilde{\boldsymbol{\pi}}_f - \mu_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}})^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}}^{-1} (\tilde{\boldsymbol{\pi}}_f - \mu_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}}) \right) \end{aligned}$$

Since the CPMBT model is expressed as a linear SSM, it means that optimal estimation can be performed via the Kalman filter [64]. The filter computes the mean and covariance of the state estimates, which correspond to the parameters $\mu_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}}$ and $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}}$ in the predictive distribution. This is done through the use of two matrices: the *innovation covariance* S_f and the *state covariance* P_f , which are reparametrizations of the terms derived earlier in the predictive distribution. Specifically, S_f is the covariance of the prediction error for the model, while, P_f is the posterior variance of the latent state κ_f accounting for the additional variance introduced by the model:

$$\begin{aligned} S_f &= \sigma_\epsilon^2 \mathbf{I}_P + \boldsymbol{\beta} \left(\phi_1^2 P_{f-1} + \sigma_{\omega_f}^2 \right) \boldsymbol{\beta}^\top \\ P_f &= \frac{\sigma_\epsilon^2 \left(\phi_1^2 P_{f-1} + \sigma_{\omega_{f-1}}^2 \right)}{\phi_1^2 P_{f-1} (\boldsymbol{\beta}^\top \boldsymbol{\beta}) + \sigma_\epsilon^2} + \sigma_{\omega_f}^2 \end{aligned}$$

The log-likelihood of the extended model is then re-expressed as

$$\log \mathcal{L}(\boldsymbol{\Theta}(\mathcal{M}_E)) = \sum_{f=1}^F \log f_E(\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}) = -\frac{FP}{2} \log(2\pi) - \frac{1}{2} \sum_{f=1}^F \left(\log |\mathbf{S}_f| + \mathbf{v}_f^\top \mathbf{S}_f^{-1} \mathbf{v}_f \right)$$

where, $\mathbf{v}_f = \tilde{\boldsymbol{\pi}}_f - \boldsymbol{\alpha} - \boldsymbol{\beta} \hat{\kappa}_{f|f-1} - \gamma_{(Q)}^\top \boldsymbol{\pi}^{(Q)} - \gamma_{(N)}^\top \boldsymbol{\pi}^{(N)}$. Substitute \mathbf{S}_f , \mathbf{v}_f into the log likelihood formula

$$\begin{aligned} \log \mathcal{L}(\boldsymbol{\Theta}(\mathcal{M}_E)) = & -\frac{FP}{2} \log(2\pi) - \frac{1}{2} \sum_{f=1}^F \left(\log \left| \sigma_\epsilon^2 \mathbf{1}_P + \boldsymbol{\beta} \left(\phi_1^2 P_{f-1} + \sigma_{\omega_f}^2 \right) \boldsymbol{\beta}^\top \right| \right. \\ & + \left(\tilde{\boldsymbol{\pi}}_f - \boldsymbol{\alpha} - \boldsymbol{\beta} (\theta + \phi_1 \hat{\kappa}_{f-1|f-1}) - \gamma_{(Q)}^\top \boldsymbol{\pi}^{(Q)} - \gamma_{(N)}^\top \boldsymbol{\pi}^{(N)} \right)^\top \left(\sigma_\epsilon^2 \mathbf{1}_P + \boldsymbol{\beta} \left(\phi_1^2 P_{f-1} + \sigma_{\omega_f}^2 \right) \boldsymbol{\beta}^\top \right)^{-1} \\ & \left. \times \left(\tilde{\boldsymbol{\pi}}_f - \boldsymbol{\alpha} - \boldsymbol{\beta} (\theta + \phi_1 \hat{\kappa}_{f-1|f-1}) - \gamma_{(Q)}^\top \boldsymbol{\pi}^{(Q)} - \gamma_{(N)}^\top \boldsymbol{\pi}^{(N)} \right) \right) \end{aligned}$$

An important point to remark on the above discussion on the Kalman Filter is that the innovation covariance and state covariance are equivalently defined in both extended and baseline models. These matrices have the exact same form for the two models (with the different estimated parameters). This means that, the parameters and the estimates related to the speech tests ($\gamma_{(Q)}^\top \boldsymbol{\pi}^{(Q)}$; $\gamma_{(N)}^\top \boldsymbol{\pi}^{(N)}$) play a crucial role in adjusting the expected mean of the observed data in the extended model and they do not directly affect the computation of the two covariance matrices S_f and P_f . Hence, the log-likelihood of the baseline model can then be expressed as

$$\log \mathcal{L}(\boldsymbol{\Theta}(\mathcal{M}_B)) = \sum_{f=1}^F \log f_B(\tilde{\boldsymbol{\pi}}_f | \tilde{\boldsymbol{\pi}}_1, \dots, \tilde{\boldsymbol{\pi}}_{f-1}) = -\frac{FP}{2} \log(2\pi) - \frac{1}{2} \sum_{f=1}^F \left(\log |\mathbf{S}_f| + \mathbf{v}_f^\top \mathbf{S}_f^{-1} \mathbf{v}_f \right)$$

where, this time, the difference is represented by $\mathbf{v}_f = \tilde{\boldsymbol{\pi}}_f - \boldsymbol{\alpha} - \boldsymbol{\beta} \hat{\kappa}_{f|f-1}$. The derivations are provided in the Supplementary Information, Section 3 and details of the Kalman Filter recursive algorithm implementation are given in the Supplementary Information, Section 4.

3 Partial Regression Under the CPMBT Model

In this section we provide further details about the treatment of partial regression structures in the class of extended CPMBT models. Given, estimates for the parameters of the baseline model $\mathcal{M}_{B,d}$ in Equation (2). In the context of partial regression, one may be interested to characterise how the additional two audiological tests, i.e. speech-in-quiet (SRT_Q) and speech-in-noise (SRT_N) contribute to explaining the relative hearing loss performance. To achieve this, we first reformulate the baseline CPMBT model $\mathcal{M}_{B,d}$ in Equation (2) in matrix form as

$$\tilde{\mathbf{\Pi}}_{P \times F} = \mathbf{A}_{P \times F} + \boldsymbol{\beta}_{P \times 1} \mathbf{K}_{1 \times F} + \mathbf{E}_F_{P \times F} \quad (6)$$

where $\tilde{\mathbf{\Pi}}$ is the log central hearing loss proportion matrix considering all eleven frequencies, $\mathbf{A} = \boldsymbol{\alpha} \otimes \mathbf{1}_F$ with \otimes represents the Kronecker product and $\mathbf{1}_F$ is a row vector of dimension F containing ones and hence this produces the repetition of the vector $\boldsymbol{\alpha}$ F times; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_P)^\top$ and $\mathbf{K} = (\kappa_{f_1}, \dots, \kappa_{f_F})$ and \mathbf{E}_F is the error matrix of dimension $P \times F$. Then one can use this formulation to specify the extended model format.

We now reformulate the extended CPMBT model $\mathcal{M}_{E,d}$ provided in Equation (3) in matrix form as follows

$$\tilde{\mathbf{\Pi}}_{P \times F} = \mathbf{X}_1_{P \times 2} \boldsymbol{\gamma}_1_{2 \times F} + \mathbf{X}_2_{P \times 2} \boldsymbol{\gamma}_2_{2 \times F} + \mathbf{E}_F_{P \times F} \quad (7)$$

where

$$\boldsymbol{\gamma}_1_{2 \times F} = \begin{bmatrix} \mathbf{1}_F \\ \mathbf{K} \end{bmatrix}, \quad \boldsymbol{\gamma}_2_{2 \times F} = \begin{bmatrix} \gamma_{(Q)} \\ \gamma_{(N)} \end{bmatrix}, \quad \mathbf{X}_1_{P \times 2} = [\boldsymbol{\alpha} \quad \boldsymbol{\beta}], \quad \mathbf{X}_2_{P \times 2} = \begin{bmatrix} \boldsymbol{\pi}^{(Q)} & \boldsymbol{\pi}^{(N)} \end{bmatrix}$$

with $\gamma_{(Q)}$, $\gamma_{(N)}$ the parameters for SRT_Q and SRT_N and $\boldsymbol{\pi}^{(Q)}$, $\boldsymbol{\pi}^{(N)}$ the proportions referring to SRT_Q and SRT_N , respectively.

In partial linear regression approaches discussed in [65, 66, 67], one undertakes a two stage estimation procedure. Under such a two stage approach when applied in the extended CPMBT model one first regresses $\tilde{\mathbf{\Pi}}$ on \mathbf{X}_1 and then, in a second step, one regresses \mathbf{X}_2 on the residual of the first step. The advantage of this approach is that one can assume that there exists a population factor κ_f that is associated with hearing loss in different age groups and intensity levels. After having estimated this value, one can then observe the complexity of the relationship of speech-in-quiet and speech-in-noise with the pure-tone frequencies hearing loss by taking into account such a factor.

Alternatively, under the conditions of the Frisch-Waugh-Lovell theorem [68, 69] when applied to the model in Equation (7), this theorem states that the ordinary least squares (OLS) estimates of the parameters associated with the predictors $\boldsymbol{\gamma}_2$ obtained after partialing out the predictors \mathbf{X}_1 are identical to those obtained by directly regressing the residuals of the linear regression on the predictors \mathbf{X}_2 . This theorem simplifies the estimation process and allows us to obtain the effect of $\boldsymbol{\gamma}_2$ on the response variable while accounting for $\boldsymbol{\gamma}_1$ effects. By applying the Frisch-Waugh-Lovell theorem, we obtain estimates of the $\boldsymbol{\gamma}_2$ predictors' effects independently of the $\boldsymbol{\gamma}_1$ predictors, enhancing our ability to interpret and make inferences about the relationships between the predictors and the response variable.

To achieve this one defines the projection matrix $\mathbf{H}_{\mathbf{X}_1}$

$$\mathbf{H}_{\mathbf{X}_1}_{P \times P} = \mathbf{X}_1 (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top$$

and the complementary projection matrix defined as

$$\mathbf{M}_{\mathbf{X}_1}_{P \times P} = \mathbf{I}_P - \mathbf{H}_{\mathbf{X}_1}$$

The Frisch-Waugh-Lovell theorem states that the ordinary least square estimates of $\boldsymbol{\gamma}_2$ and the residuals obtained from Equation (7) are identical to those obtained by running the regression

$$\mathbf{M}_{\mathbf{X}_1} \tilde{\mathbf{\Pi}} = \mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2 \boldsymbol{\gamma}_2 + \mathbf{M}_{\mathbf{X}_1} \mathbf{E}_F \quad (8)$$

where \mathbf{E}_F are the residuals of the regression given in Equation (7). In this way, we obtain the effect of \mathbf{X}_2 on $\tilde{\mathbf{\Pi}}$, in a second step which takes into account the non-linearity associated to the relationship of this predictor and the response. Note that, in the Supplementary Information, Section 7 provides some remarks for the computational aspects of this estimation procedure. Results of the partial regression analysis for the extended CPMBT model are presented in Section 5.

4 Data Description and Properties

This section is dedicated to the case study real data description. We begin by outlining the configuration of the tests administered to the participants, including the audiogram and two speech tests conducted in quiet and noise, respectively. Following this, we provide a detailed dataset description, incorporating descriptive statistics and visual representations. Specifically, we present several plots illustrating the distribution of the critical variables as violin plots that offer a deeper insight into the variability and central tendencies within the data.

4.1 Data Acquisition & Testing Procedures

Our study utilizes a dataset from Amplifon France, which contains routine data from hearing aid fitting practices across multiple Amplifon hearing aid acoustician labs in France. For retrospective data analysis, the dataset was provided in pseudonymized form to Institut Pasteur under the BIG DATA AP project. The data protection authority Commission Nationale de l’Informatique et des Libertés (the National Commission on Informatics and Liberty) authorised the processing of BIG DATA AP study data on April 05, 2024.

The dataset included participants’ age, sex assigned at birth, pure-tone audiograms for both ears, and speech recognition thresholds for speech tests in quiet and noise, respectively. The degree of hearing loss was derived by calculating the pure-tone average (PTA) based on individual hearing thresholds at 0.5, 1, 2, and 4 kHz [70], according to the American Speech-Language-Hearing Association (ASHA) classification, detailed in Table 3.

The study focused on participants aged 40-90 years with symmetric hearing loss, defined by a PTA difference of less than 15 dB between ears [49]. This age range was selected based on data availability and completeness. The final dataset comprised 48,144 participants. Data on race or ethnicity were not collected, adhering to French legal restrictions on personal data collection per the 1978 Law on Information and Freedoms [71]. The following sections describe the audiological tests in detail.

Audiogram. All participants completed a half-octave, air-conduction audiogram with measurements in 5 dB steps, for frequencies from 0.125 to 8 kHz, using TDH-39 headphones. Air conduction testing began on the better ear if specified by the patient, otherwise on the right ear. Testing started with an audible pulsed pure tone (1 kHz/60 dB HL or higher, if necessary). The modified Hughson-Westlake procedure [72] was applied to assess hearing thresholds at increasing frequencies from 1 kHz to 8 kHz, afterwards decreasing frequencies from 1 kHz (repeated) until 0.125 kHz

Speech-in-quiet. Patients performed the Lafon test in quiet [73, 74], which evaluates speech intelligibility using test lists of disyllabic words. To estimate the SRT_Q (speech level at which 50% speech intelligibility is obtained), several test lists were presented at fixed signal-to-noise ratios (SNRs), starting at a high level where 100% speech intelligibility was expected for the respective patient, and subsequently decreasing the level until the obtained result fell below 50%. The test was performed binaurally in a free-field condition, with one loudspeaker positioned in front of the listener. The resulting SRT_Q value represents the difference from the normal-hearing speech reception threshold (SRT), thereby providing an interpretation of speech reception threshold loss

Speech-in-noise. Patients performed the HINT-5 min test [75, 76], an adapted European French version of the Canadian French HINT. Meaningful everyday sentences were employed to evaluate speech intelligibility in background noise. As noise, ICRA-1 stationary noise with a long-term average speech spectrum [77] was used, presented at a fixed level of 60 dB SPL. The SRT_N was estimated by adaptively varying the speech level within a test list of 20 sentences, starting from an SNR of 5 dB SNR. Testing was conducted binaurally in a free-field condition, with one loudspeaker positioned in front of the listener. The resulting SRT_N value represents the absolute SNR at which 50% speech intelligibility is obtained. The corresponding normal-hearing reference SRT is -1.0 dB SNR for the binaural free-field condition.

Pure-tone average (PTA) Categories	
Degree of hearing loss	PTA range (dB HL)
Normal	-10 to 15
Slight	16 to 25
Mild	26 to 40
Moderate	41 to 55
Moderately severe	56 to 70
Severe	71 to 90

Table 3: Pure-tone average (PTA) categories were defined in accordance with the American Speech-Language-Hearing Association (ASHA) classification [70]. Note that we do not have any participant with Normal hearing in our dataset.

4.2 Data Description

This subsection describes the dataset, with Table 4 providing a comprehensive summary of descriptive statistics, including age, audiogram frequencies (125 Hz to 8000 Hz), SRT_N , and SRT_Q . The mean participant age is approximately 73 years, with a standard deviation (SD) of 9.73 years, indicating an elderly population with some variability, ranging from 40 to 90 years.

Mean hearing thresholds increase over frequency, from about 30 dB HL at 125 Hz to about 72 dB HL at 8 kHz, with standard deviations ranging from 13 to 19 dB HL. Median values are slightly below the means, suggesting a slight right skew. The database contains hearing thresholds spanning the entire range of measurable audiometric levels. The mean SRT_N is 4.43 dB SNR with a standard deviation (SD) of 3.96 dB SNR, while the mean SRT_Q is 45.97 dB SPL with an SD of 11.56 dB SPL. The SRT_Q median is close to the mean, indicating a symmetrical distribution, while the SRT_N median is slightly lower than the mean, suggesting a minor positive skew. SRT_N values range from -10 dB SNR to 20 dB SNR, and SRT_Q values range from 5 dB SPL to 80 dB SPL.

This analysis includes the entire sample, covering all degrees of hearing loss, ages 40 to 90, and both sexes. Additional descriptive statistics by age group, degree of hearing loss, and sex are provided in the Supplementary Information, Section 10, to better understand variable distributions across these demographic and audiometric subgroups.

Descriptive Statistics Overall														
Statistics	Age	Frequencies (Hz)											SRT_N	SRT_Q
		125	250	500	750	1000	1500	2000	3000	4000	6000	8000		
Mean	72.98	30.48	31.24	33.83	36.40	38.11	45.08	48.48	55.79	61.74	70.35	71.71	4.43	45.97
Median	74.00	30.00	30.00	30.00	35.00	35.00	45.00	50.00	55.00	60.00	70.00	70.00	4.00	45.00
SD	9.73	13.19	14.50	15.14	15.43	15.68	16.18	16.28	16.50	16.97	18.24	18.75	3.96	11.56
Min	40	-10.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	-5.00	3.00	-10.00	5.00
Max	90	120.00	120.00	120.00	120.00	120.00	120.00	120.00	125.00	125.00	125.00	130.00	20.00	80.00

Table 4: Descriptive statistics over the whole sample population. Variables include age, audiogram frequencies, SRT_N and SRT_Q . Note that the unit of measures are dB HL for the audiogram frequencies, dB SNR for SRT_N and dB SPL for SRT_Q .

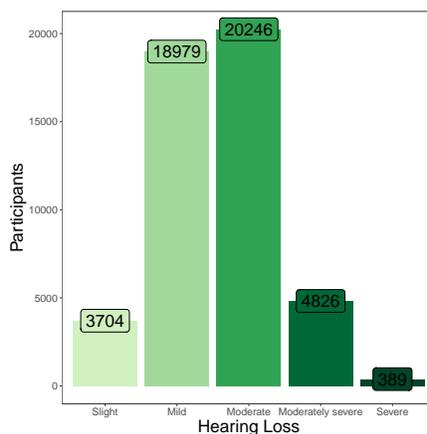


Figure 6: Sample size across Pure Tone Average (PTA) hearing loss categories for the whole database population. Hearing loss degree is classified based on PTA categories.

Figure 6 illustrates the distribution of individuals across different degrees of hearing loss. The categories include slight, mild, moderate, moderately severe, and severe hearing loss. The majority of individuals fall into the Moderate and Mild categories, with nearly equal number of patients (Moderate: 20,246; Mild: 18,979). The moderately severe and slight categories have fewer individuals, with 4,826 and 3,704 people respectively. Finally, the severe category has the smallest number of individuals, with only about 389 people, showing that severe hearing loss is less common within this population. This distribution highlights the varying prevalence of hearing loss severity among the individuals studied. Further details regarding the descriptive statistics split by hearing loss degree are provided in the Supplementary Information, Section 9.

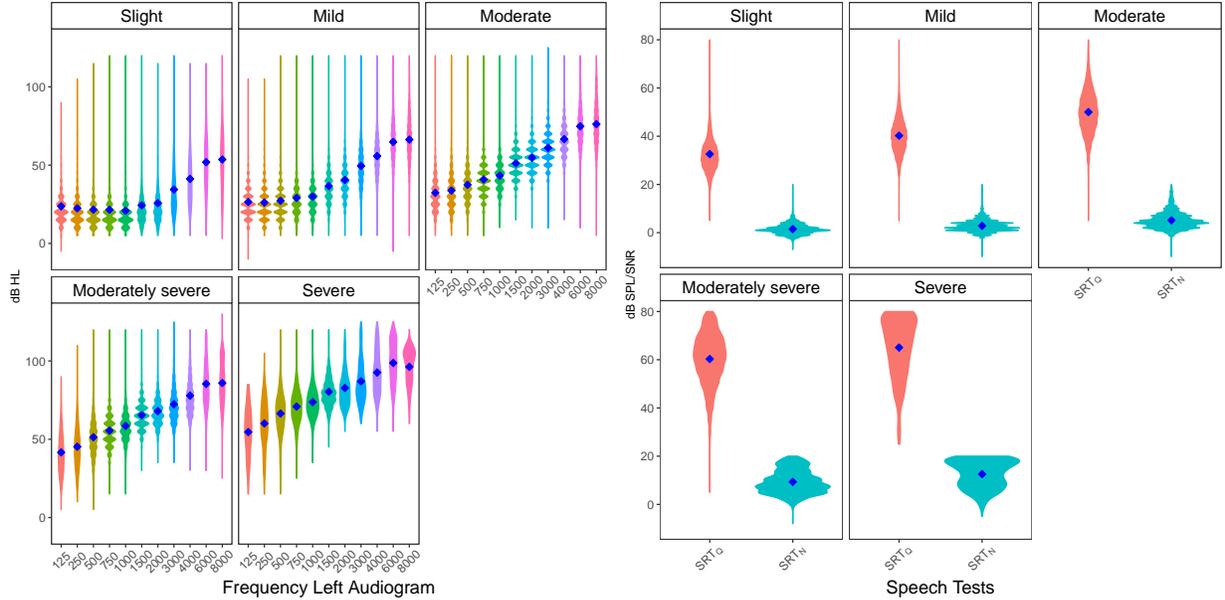


Figure 7: Violin plots of hearing thresholds at different frequencies (left) and SRT_Q , SRT_N (right) by hearing loss degree for the left ear. The x-axis on the left plot represents the measured frequencies from 125 Hz to 8000 Hz, while the x-axis on the right represents the speech tests. The y-axis shows hearing thresholds in dB HL (left) and SRT_Q and SRT_N in dB SPL/SNR (right). Due to the symmetrical nature of hearing loss in the sample, the left ear was selected for this representation. Equivalent results were confirmed when analysing the right ear, ensuring the reliability of the observed patterns. The violin plots of the right ear are in the Supplementary Appendix.

Figure 7 shows data distributions via violin plots, corresponding to descriptive statistics in the Supplementary Information, Section 9. The left panel displays hearing thresholds for the left ear. Although symmetric hearing loss was analyzed, a comparable plot for the right ear is provided in Section 9 for completeness. The right panel illustrates SRT_N and SRT_Q values for the left ear, as these measures were similar for both ears. For brevity, descriptive statistics are reported solely for the left ear.

Sex- and age-specific results are detailed in Section 9, with boxplots and statistics highlighting demographic differences in audiological tests.

In summary, hearing thresholds vary across age groups, with lower thresholds observed from 40-45 to 65-70 years, particularly at lower frequencies (125 Hz to 1000 Hz). However, thresholds increase markedly at higher frequencies with age, rising from 55.44 dB in the 40-45 age group to 81.85 dB in the 85-90 group, indicating age-related hearing loss. SRT_N and SRT_Q values also increase with age, reflecting a decline in speech perception in noisy environments: mean SRT_N rises from 2.79 dB (40-45 years) to 6.94 dB (85-90 years), while SRT_Q increases from 40.70 dB to 53.73 dB.

By sex, females have higher mean thresholds at lower frequencies (125 Hz to 1500 Hz), indicating slightly poorer hearing than males in these ranges. This pattern reverses at higher frequencies (1500 Hz to 8000 Hz), where females generally have lower thresholds, indicating better hearing. Additionally, males show slightly higher mean SRT_N and SRT_Q values than females. Further details are available in the Supplementary Information.

5 Results

This section presents the primary findings of our analysis, beginning with an examination of hearing performances and their interpretation, followed by a description of hearing proportions across the entire sample, categorized by hearing loss degree and sex. The purpose of this analysis is to inform our modelling framework and reveal essential data patterns. The proposed SSM, i.e., the CPMBT model, establishes standardised, data-driven curves summarising data at a population level. This standardisation is achieved by fitting the audiogram responses to a model that transforms them into proportions, reducing inherent variability and heteroskedasticity, enabling a consistent fit. Unlike counts, which are affected by variations in sample size across different population segments, proportions allow for more direct

comparisons, making it possible to observe patterns consistently across subgroups. For consistency, SRT_Q and SRT_N were also transformed to log proportions, as outlined in Subsection 2.3.

A key feature of the model is the definition of standardised curves as risk profiles. The CPMBT model captures relative changes across frequency domains and age groups, with parameters α , β , and κ_f representing these standardised curves. This enables a population-level risk measure by aggregating individual data into a metric generalisable to the population.

In the first part of this section, we explore how the definition of hearing performance influences the derived proportions and the implications of this choice for understanding the relationship with the risk profile. We then explore how the obtained proportions vary across different population segments. This analysis involves setting probabilistic levels to identify empirical quantiles, as explained in Subsection 2.2, which partitions data into groups representing increasing levels of hearing loss. This approach helps us observe whether these segments vary non-linearly and evaluate the extent of this variation. Using proportions instead of counts allows us to mitigate biases introduced by unequal sample sizes, thereby enabling a clearer assessment of how hearing loss patterns emerge across different frequencies, age groups, and severity categories.

Subsequently, we focus on statistical inference results, specifically from two sets of tests addressing our problem statements identified in section 2.4, denoted **PS1** and **PS2**. For **PS1**, the goal was to identify differences within the same model class across population segments. Using the Vuong Test for non-nested models, we assessed model consistency and sensitivity across hearing loss categories and probabilistic levels, helping to identify areas for model refinement. For **PS2**, we compared different model classes, particularly the baseline and extended models, using the Likelihood Ratio Test (LRT), which assesses whether the extended model provides a significant improvement in fit for various population segments.

These statistical tests validate our modelling framework by examining if the extended model more effectively captures data patterns than the baseline model, with consistency across population segments. Remark that further details on these tests are provided in the Supplementary Information.

After identifying significant differences across segments, we evaluate the risk profiles generated by the CPMBT model. These profiles offer both general insights and targeted analyses for specific groups, such as hearing loss levels and sex-based differences. We use partial linear regression to assess SRT_N and SRT_Q 's direct effects on audiogram responses, further informing the factors influencing responses.

A discussion on calibration and model performance is provided in the Supplementary Information, Section 10, to ensure curve accuracy and reliability. We also assess model performance using Mean Square Error with age group intervals of 1-year, 5-year (selected), and 10-year.

The results were fitted using left ear data due to symmetric hearing loss considerations; equivalent right ear results were cross-checked and omitted for space. All research data and analysis scripts are available at the following GitHub repository: <https://github.com/mcampi111/SpiN>.

5.1 Hearing Performances

Figure 8 illustrates how the different $M_{p,f}(d)$, representing the count of individuals who fail to detect sounds above the empirical quantile of interest (i.e., $Q_f(d)$), behave across frequencies f , age groups p , and probabilistic levels d . The x-axis shows the age groups, while the y-axis represents the counts of individuals. The bars are coloured red for those exceeding the empirical quantile and azure for those who do not. Each column of the plot corresponds to a specific audiogram frequency, and each row represents a different probabilistic level d . The last two columns refer to the SRT_Q and SRT_N measures, to which an equivalent approach has been applied—i.e., using the empirical quantiles for these two distributions to count how many individuals exceeded those quantiles.

The results presented in the plot support the interpretation provided in subsection 2.3.1. It is evident that, across all frequencies, increasing the level of the empirical quantile results in fewer individuals failing to detect sounds above that quantile (i.e., fewer individuals exceeding the empirical quantile threshold). Given that the sample sizes remain constant across the empirical quantiles, this means that fewer individuals fail to detect higher sound levels, thus reflecting hearing performance at more severe hearing thresholds—indicating a higher degree of hearing loss.

An equivalent reasoning can be applied to the speech tests variables (SRT_Q and SRT_N). As the empirical quantiles increase, the distribution of speech recognition thresholds shifts, reflecting variations in speech perception performance. One test (SRT_Q) measures speech perception in quiet, while the other (SRT_N) measures it in noise. Unlike pure-tone audiometry, these thresholds represent a continuous measure of speech intelligibility, where higher quantiles indicate progressively more challenging listening conditions with reduced speech recognition performance. At each quantile,

the measure represents the signal level at which an individual can correctly identify approximately 50% of speech signals, either in quiet or noisy environments. This distinction highlights the increased difficulty that individuals may face in challenging acoustic environments, which is captured by the risk profiles at higher quantiles.

Figure 9 illustrates how the different $M_{p,f}(d)$, representing the count of individuals who fail to detect sounds above the empirical quantile of interest (i.e., $Q_f(d)$), behave across frequencies f , hearing loss severity groups, and probabilistic levels d . Compared to Figure 8, this time we added the hearing loss severity category (Slight, Mild, Moderate, Moderately Severe, Severe) arranged in rows alongside the different probabilistic levels d . The columns instead refer to the audiogram frequencies, from the lowest to the highest. The x-axis shows the age groups, while the y-axis represents the counts of individuals. The bars are coloured red for those exceeding the empirical quantile and azure for those who do not. As above, the last two columns refer to the speech tests, i.e. to the SRT_Q and the SRT_N variables.

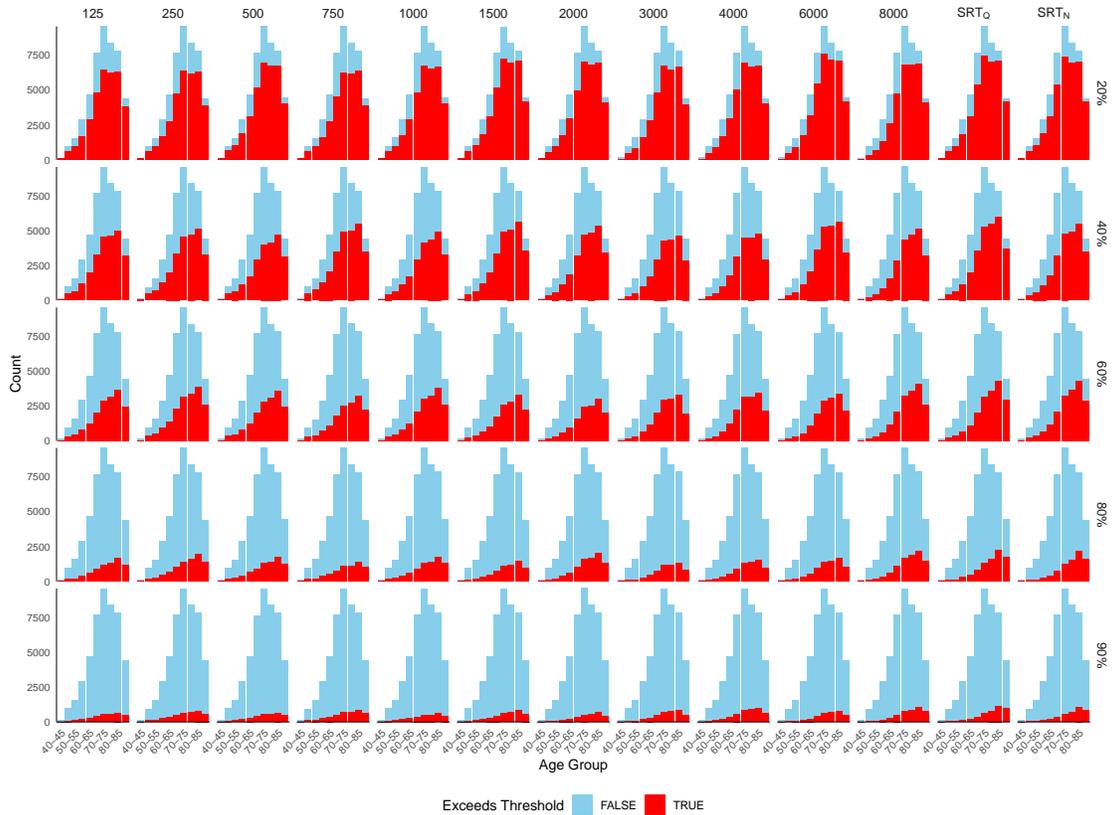


Figure 8: Plot showing the counts of individuals in each age group who either exceed or do not exceed the given empirical quantile thresholds across different frequencies of the pure tone audiograms and the speech tests. The x-axis represents age groups, while the y-axis shows the count of individuals. The coloured segments indicate whether individuals exceed the empirical quantile at each frequency (or speech tests), providing insight into how hearing performance changes with age and increasing empirical quantiles. This visualization helps to interpret the distribution of hearing threshold exceedances and provides a basis for understanding the variability in hearing performance across the population.

An equivalent trend to the one of Figure 8 is identified, i.e., higher counts of individuals failing to detect sounds above the empirical quantiles with increasing quantile levels, is observed in the Slight, Mild, Moderate, and Moderately Severe categories. The only category in which there is almost no one passing the empirical quantiles is the Severe category. This is mainly due to the small sample size of this population segment for which, in fact, it is possible to observe only few individual exceeding the thresholds at the very first probabilistic level.

Although this pattern for the Severe group is different compared to the rest of the categories and Figure 8, it is still rigorous and offers valuable insights into the dynamics of hearing loss across different severity levels. Specifically, the exceedance rate—i.e., the proportion of individuals exceeding the empirical quantile—changes distinctively across the different hearing loss categories as the empirical quantile level increases. This behaviour is influenced by both the

intrinsic definition of these categories, which are defined by ranges of the pure tone average (PTA), and the variations in sample size. This makes the results informative, as it reveals the rate at which exceedance decreases across different hearing loss categories and individual frequencies, apart from the Severe category where the trend remains relatively stable. Notably, a preliminary observation reveals a trend of shifting medians to higher ages from top to bottom, correlating with increasing hearing loss severity categories—a pattern further explored in the risk profile section.

This rate of change in exceedance across hearing loss categories provides important information about hearing loss progression and its impact across different severity levels. It allows us to understand how different categories respond differently to increasing quantile thresholds, giving insights into the dynamics of hearing thresholds and risk profiles for individuals with different levels of hearing loss. This is particularly important for understanding the population's heterogeneity and characterizing hearing loss not just at the level of pure tone average but also at a more granular frequency-specific level. Furthermore, this analysis extends to speech tests that offer additional dimensions of hearing assessment by measuring the ability to understand speech in quiet and noisy environments. The rate of exceedance of these tests follows similar behaviours across the probabilistic levels but not across the hearing loss degree as the linearity provided by the pure tone frequencies is not characterising these tests.

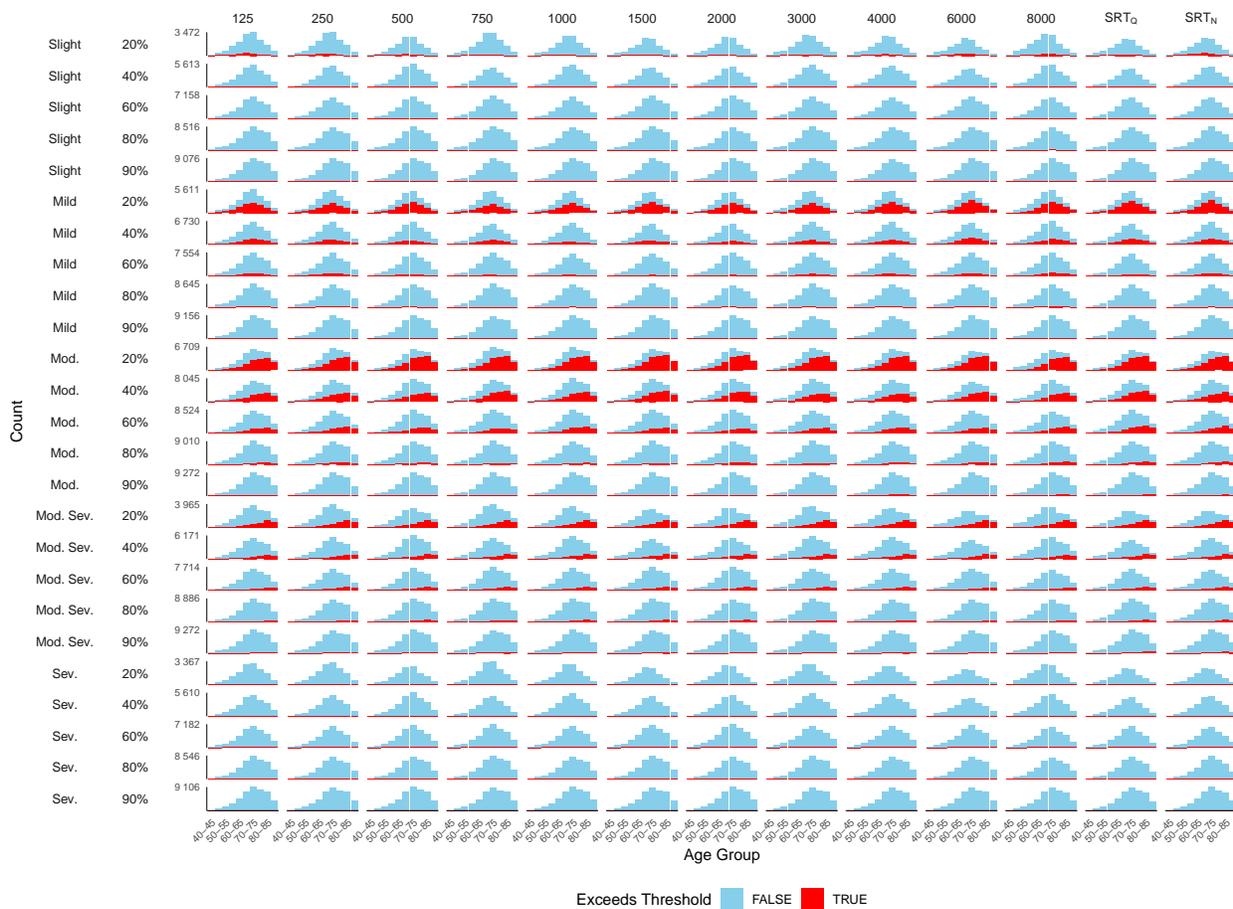


Figure 9: Plot showing the counts of individuals in each hearing loss severity group (Slight, Mild, Moderate, Moderately Severe, Severe) who either exceed or do not exceed the empirical quantile thresholds across different frequencies and speech tests. The x-axis represents age groups, while the y-axis shows the count of individuals, with bars in red for those exceeding the empirical quantile and in azure for those who do not. Each row corresponds to a specific hearing loss severity category combined with a different probabilistic level (20%, 40%, 60%, 80%, 90%). Each column represents one frequency of the audiogram and the last two refers to the speech tests measurements i.e. SRT_Q and SRT_N . This visualization illustrates the relationship between hearing loss severity, frequency, and the counts of individuals exceeding the quantile thresholds, providing insight into how different severities affect hearing performance across the spectrum of empirical quantiles.

At this stage, it is also crucial to note that this is an observational study. We did not attempt to artificially balance the sample sizes across different groups; instead, we aimed to retain all relevant information inherent in the population. The uneven distribution of individuals across severity levels reflects real-world conditions, and thus our analysis deals directly with this variability, rather than trying to mitigate it by excluding data or artificially balancing the sample. This approach ensures that we are capturing the actual dynamics of hearing loss in the population, recognizing that the resulting risk profiles may be dataset-specific. While this means accepting increased variance in the estimates for some groups due to smaller sample sizes, it provides a more direct representation of the population’s heterogeneity. Nevertheless, this choice allows us to avoid introducing bias and provides an authentic representation of the hearing loss landscape. More details on the variability in sample sizes and how this affects the resulting risk profile estimates are provided in the Supplementary Information, and additional discussion in the last section.

5.2 Hearing Proportion Description

Figure 10 visually represents the computed proportions through a series of heatmaps. Each radial segment (or row) corresponds to specific age groups from 40 to 90 years in 5-year increments, while each concentric ring represents distinct audiogram frequencies, SRT_Q , and SRT_N . The figure includes eight main sections of heatmaps: the overall model, sex (females first, then males), and five degrees of hearing loss, ranging from slight to severe. Each section features five heatmaps representing different probabilistic levels (20%, 40%, etc.), corresponding to five empirical quantiles. Higher probabilistic levels indicate elevated quantiles, reflecting increased dB audiogram thresholds and greater hearing loss severity.

These heatmaps show similar patterns to Figures 8 and 9, with the key difference being that they correspond to proportions rather than counts. This allows for direct comparisons across different population segments regardless of sample size. Specifically, each proportion is computed over N_p , the number of people in that age group, enabling a consistent comparison of hearing performance across segments. Furthermore, the heatmaps provide a direct comparison of all population segments considered (including sex) and also incorporate the proportions for SRT_Q and SRT_N , adding an additional dimension to the analysis.

In the overall model, the heatmaps show a trend that is consistent with the quantile estimation method: as the probabilistic level increases (left to right), the proportion of individuals above each quantile systematically decreases—a characteristic inherent to the quantile estimation approach. Notably, for higher frequencies (4000 Hz, 6000 Hz, and 8000 Hz), the proportion of individuals exceeding quantiles increases with age, consistent across all probabilistic levels. In contrast, lower frequencies exhibit more evenly distributed proportions across age groups, suggesting a more pronounced age-hearing loss relationship at higher frequencies. This is visually indicated by lighter colors at the 90% probabilistic level, signaling fewer affected individuals.

For SRT_Q and SRT_N , the heatmaps reveal similar trends. For SRT_Q , proportions exceeding empirical quantiles decrease as probabilistic levels rise. An upward trend with increasing age, particularly at higher probabilistic levels, highlights the impact of aging on speech comprehension. For SRT_N , the proportion of individuals exceeding quantiles also diminishes as probabilistic levels rise, particularly for those over 70, suggesting a comparable age-related decline in speech comprehension in noisy environments.

Examining sex differences, the proportions of women affected by hearing loss gradually increase with frequency, particularly in the lower to mid-range (125 Hz to 1000 Hz), with higher proportions at the first probabilistic level. Females show a consistent upward trend in higher frequencies (3000 Hz to 8000 Hz). In contrast, males start with lower prevalence at lower frequencies but show sharper increases in higher ranges, suggesting greater vulnerability with age.

Overall, both genders show a tendency toward increased hearing loss susceptibility at higher probabilistic levels, especially pronounced in higher frequencies. In SRT_Q assessments, females’ proportions increase with age at higher probabilistic levels, while males exhibit a steeper rise, indicating greater susceptibility to severe speech reception challenges. Similarly, the proportion of women affected by noise-related challenges increases with age, while males show sharper increases at higher levels. This indicates that while females typically show higher proportions at lower probabilistic levels, males exhibit a more significant increase in susceptibility as severity increases.

Analyzing hearing loss proportions across age groups and frequencies reveals consistent trends from slight to severe impairment. Higher frequencies (3000 Hz - 8000 Hz) consistently show greater impairment than lower frequencies (125 Hz - 1000 Hz), highlighting that hearing loss disproportionately affects higher frequency ranges, especially in older individuals. For example, in mild hearing loss cases, the proportion of impairment at higher frequencies can range from about 58% in younger individuals to 97% in older groups, while lower frequencies usually exhibit proportions under 35%. Generally, hearing loss proportions decrease as probabilistic levels rise, suggesting diminished

severity with stricter criteria. Regarding the severe proportions, these are all very close to zero due to the small sample sizes of this population segment.

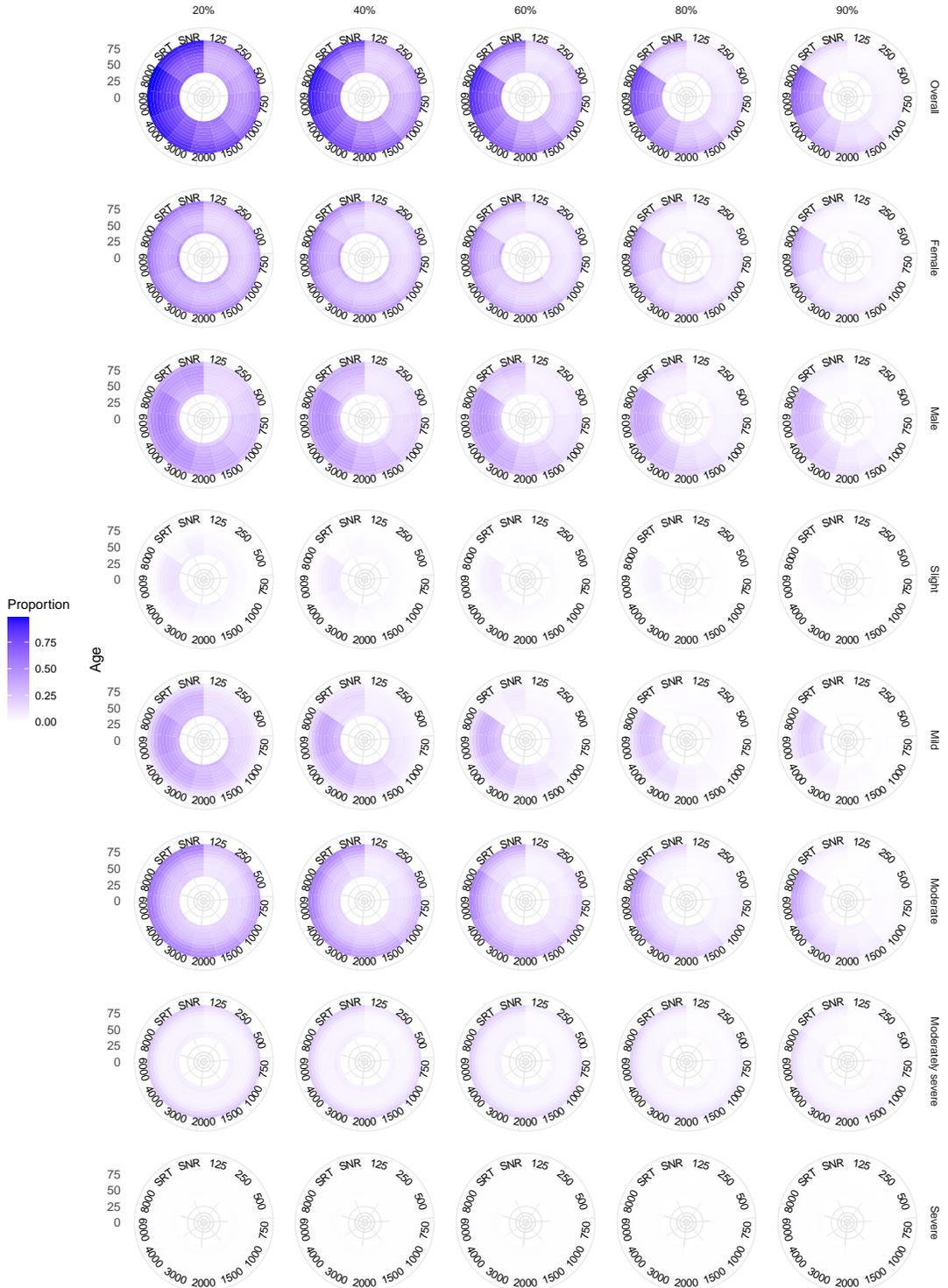


Figure 10: Heatmaps Overall, by sex and by degree of hearing loss.

Examining SRT_Q and SRT_N proportions reveals consistent trends across speech recognition tests. Both SRT_Q and SRT_N proportions generally decrease with hearing loss severity and age, showing a pattern distinct from hearing thresholds. As probabilistic levels rise from 20% to 90%, SRT_Q and SRT_N proportions decrease, indicating stricter classification criteria, especially in slight and mild hearing loss.

Unlike frequency-specific audiogram proportions, SRT_Q and SRT_N reveal distinct patterns of speech recognition across different probabilistic levels. These preliminary observations of speech test proportions suggest varying behaviors that warrant further systematic investigation. While the current visualization provides an initial glimpse, our subsequent analysis will more rigorously examine the nuanced characteristics of speech perception measures.

The heatmaps unveil potential gender-specific variations in hearing loss distribution across frequencies and age groups. These initial visual patterns hint at complex interactions between sex, age, and hearing loss characteristics. The preliminary observations suggest that gender may play a significant role in hearing loss progression, setting the stage for a more detailed exploration in our subsequent analytical approach.

Understanding the proportions of hearing loss is crucial for deciphering population-level auditory health dynamics. These heatmaps provide a foundational visualization of how hearing loss proportions vary across different dimensions—age, frequency, and probabilistic levels. By transforming these initial observations into standardized risk profiles, our proposed CPMBT model aims to move beyond descriptive visualization to a more comprehensive understanding of hearing loss progression.

5.3 PS1 Results with Vuong Test

This section analyzes the results for **PS1** using the Vuong Test, which compares the overall CPMBT model with models segmented by hearing loss categories and sex across various probabilistic levels in both baseline and extended frameworks.

The Vuong Test aims to determine whether segmented population models (by hearing loss degree or sex) fit better than the overall CPMBT model. We performed contrast tests, selecting specific segmented models as references and examining combinations with probabilistic levels and varying degrees of hearing loss or sex. Additionally, we assessed probabilistic levels as references to explore intra-variability, revealing how different thresholds yield varying significance based on the chosen reference model. This highlights the importance of careful reference model selection, as it significantly influences result interpretation and conclusions regarding model fit.

The statistical tests are not symmetric; results from one reference model may differ from those using the switched reference. For instance, comparing slight baseline versus mild baseline is not equivalent to the reverse. We conducted both tests for robustness, with consistent results across both directions, allowing us to report only one case. This consistency suggests approximately symmetric behavior, reinforcing the reliability of our conclusions regarding the fit of segmented population models versus the overall CPMBT model.

Table 5 presents the Vuong test results comparing the overall CPMBT model with segmented models for hearing loss and sex across different probabilistic levels (20%, 40%, 60%, 80%, and 90%).

In the baseline model, significant differences are observed across hearing loss categories. The slight hearing loss category shows significant differences at nearly all probabilistic levels, with p-values as low as 0.000 at 20%, 40%, and 90%. The mild, moderate, and moderately severe categories exhibit significant p-values primarily at 20%, while severe hearing loss demonstrates significance at 20%, 40%, and 90%.

The extended model reveals more nuanced insights. For slight hearing loss, stronger evidence emerges at the 20% and 40% levels. The model consistently shows more significant differences for mild hearing loss across all levels, particularly at 60%. Moderate and moderately severe hearing loss categories also demonstrate robust evidence, especially at lower probabilistic levels. Interestingly, the baseline model shows more significant results for severe hearing loss, consistently outperforming the extended model.

In sex-based comparisons, the baseline model shows limited significance for female models, while male models display sporadic significant results at most levels. The extended model provides broader improvements, with statistically significant differences for all hearing loss degrees at the first three probabilistic levels (20%, 40%, and 60%), and p-values below 0.005.

Detailed results of the Vuong test comparing hearing loss categories and variations within categories are provided in the Supplementary Information, Section 11. Table 5 presents comprehensive comparisons across different hearing loss categories and probabilistic levels. Additionally, Table 6 examines variations within each category by comparing model fit across probabilistic thresholds, illustrating how predictions differ at various levels. These comparisons differ from those in Table 5, which focused on differences between categories (e.g., slight vs. severe). Table 6 highlights how

changes in probabilistic thresholds affect model fit within each category, providing insights into the model’s robustness and consistency across varying degrees of hearing impairment.

PS1 with Vuong Test - Overall CPMBT versus Population Segments CPMBT								
Baseline Model								
Reference	Probabilistic Level	Slight	Mild	Moderate	Contrast		Female	Male
					Mod. severe	Severe		
Overall	20%	0.000	0.010	0.000	0.000	0.003	0.455	0.000
	40%	0.040	0.033	0.029	0.027	0.018	0.074	0.016
	60%	0.902	0.659	0.112	0.136	0.141	0.349	0.000
	80%	0.519	0.837	0.656	0.369	0.128	0.067	0.013
	90%	0.047	0.026	0.003	0.002	0.002	0.576	0.191
Extended Model								
Reference	Probabilistic Level	Slight	Mild	Moderate	Contrast		Female	Male
					Mod. severe	Severe		
Overall	20%	0.014	0.002	0.001	0.000	0.000	0.003	0.000
	40%	0.017	0.004	0.002	0.002	0.002	0.022	0.034
	60%	0.002	0.001	0.001	0.000	0.000	0.006	0.000
	80%	0.081	0.002	0.004	0.003	0.005	0.014	0.006
	90%	0.083	0.023	0.007	0.002	0.002	0.069	0.341

Table 5: Vuong test results. We compare the overall model versus the hearing loss and the sex CPMBT models, for each probabilistic level. The tests are performed for both the baseline and the extended model frameworks.

The results indicate significant differences across most probabilistic levels, especially when comparing the 20% level with higher ones. In the slight hearing loss category, significant differences are found at all levels, with p-values close to 0.000 for both models between the 20% level and the 40%, 60%, and 90% levels. The mild hearing loss category follows a similar pattern, particularly in the extended model, where p-values are consistently low (e.g., 0.000 for the 20% vs. 40% comparison). The extended model shows stronger significance in the moderate hearing loss category, especially at the 40% and 60% levels. The moderately severe category also shows significant differences in the extended model, particularly between the 20% and higher levels. For severe hearing loss, the extended model maintains highly significant differences, with p-values as low as 0.000 for comparisons between the 20% level and the 60%, 80%, and 90% levels. Overall, the extended model provides more consistent significant results, while the baseline model also demonstrates good performance.

PS1 with Vuong Test - Sex				
Reference	Contrast	Prob. Level	P-value Baseline	P-value Extended
Female	Male	20%	0.000	0.000
	Male	40%	0.001	0.000
	Male	60%	0.014	0.008
	Male	80%	0.091	0.011
	Male	90%	0.176	0.178

Table 6: Vuong test results by sex and probabilistic level for sex.

We performed a similar analysis for sex, results are given in Table 6. Significant differences across probabilistic levels were found when comparing males and females. At the 20% level, both models showed highly significant results ($p = 0.000$), indicating strong model performance distinctions. Differences remained significant at the 40% level ($p = 0.000$ for the extended model; $p = 0.001$ for the baseline). At the 60% level, significant differences persisted ($p = 0.014$ for the baseline; $p = 0.008$ for the extended model), although the gap narrowed. The extended model maintained significance at the 80% level ($p = 0.011$), while the baseline model’s significance diminished ($p = 0.091$). By the 90% level, both models lacked significant differences ($p = 0.176$ and $p = 0.178$, respectively), indicating reduced distinguishability.

Table 7 in the Supplementary Information shows results within each sex by probabilistic levels. The extended model exhibits stronger evidence of differences across levels. In females, the baseline model shows significant differences between the 20% level and both 80% ($p = 0.000$) and 90% ($p = 0.021$). The extended model shows significant results for all comparisons involving the 20% level, with additional significant comparisons appearing in both models; however, the extended model provides slightly more consistent differentiation.

A similar pattern is observed for males, where the extended model shows stronger significance for the 20%-40% and 20%-60% comparisons. Other significant comparisons are present in both models, but the extended model yields slightly more significant p-values overall, indicating a more nuanced distinction across probabilistic levels.

Overall, the extended model captures more nuanced differences across probabilistic levels within each hearing loss category, particularly in slight, moderate, moderately severe, and severe categories. This sensitivity extends to sex analysis, with the extended model demonstrating greater statistical significance at various probabilistic thresholds for both females and males. While the baseline model performs well in some instances, the extended model consistently shows more significant results and improved performance, leading to refined interpretations of hearing loss severity within categories and regarding sex-related differences.

5.4 PS2 Results with Log-Likelihood Ratio Test

We now present the results for **PS2**. Tables 7, 8 and 9 summarise the results of the Likelihood Ratio Test (LRT), which compares the baseline and extended models across different empirical quantiles for the overall population, hearing loss subgroups, and sex, respectively. The LRT is used to assess whether the extended model provides a significantly better fit compared to the baseline model.

The results in Table 7 shows that the extended model significantly outperforms the baseline at the 20% and 40% quantiles, with p-values under 0.001, indicating a substantial fit improvement. However, at higher quantiles (60%, 80%, and 90%), p-values exceed 0.05, revealing no significant improvement.

PS2 with LRT - Baseline versus Extended - Overall	
Quantile	p-value
20%	0.000
40%	0.001
60%	0.096
80%	0.089
90%	0.107

Table 7: Likelihood Ratio Test results for the overall population.

In Table 8, the LRT results by hearing loss category indicate that the extended model significantly enhances fit over the baseline at the 20%, 40%, and 60% quantiles for slight, mild, and moderate hearing loss, with p-values below 0.01 and the strongest results at the 60% quantile (all p-values < 0.001). No significant improvement is found for moderately severe and severe categories (p-values of 0.071 and 0.074 at the 60% quantile). At the 80% and 90% quantiles, improvements persist for slight, mild, and moderate hearing loss, but the model is not significant for more severe cases (p-values between 0.079 and 0.087).

Table 9 shows that the extended model significantly improves over the baseline at lower quantiles for both sexes. Specifically, at the 20% and 40% quantiles, p-values are below 0.001, indicating substantial improvement. As quantiles increase, significance diminishes, with no improvements at the 80% quantile. Thus, the extended model is more effective at distinguishing differences in lower levels of hearing loss but less impactful as severity increases.

In summary, the LRT results demonstrate that the extended model offers a significant improvement in fit over the baseline model, particularly at the 20%, 40%, and 60% quantiles across most categories of hearing loss and for both sexes. In the former case, the improvement is most pronounced for slight, mild, and moderate hearing loss categories, while the benefits for moderately severe and severe categories are less consistent. This suggests that the extended model provides better performance and insights, especially for individuals with less severe hearing loss.

PS2 with LRT - Baseline versus Extended - Degree of Hearing Loss		
Quantile	Degree of hearing loss	p-value
20%	Slight	0.009
20%	Mild	0.002
20%	Moderate	0.001
20%	Moderately severe	0.062
20%	Severe	0.063
40%	Slight	0.005
40%	Mild	0.001
40%	Moderate	0.001
40%	Moderately severe	0.068
40%	Severe	0.062
60%	Slight	0.000
60%	Mild	0.000
60%	Moderate	0.000
60%	Moderately severe	0.071
60%	Severe	0.074
80%	Slight	0.000
80%	Mild	0.000
80%	Moderate	0.000
80%	Moderately severe	0.082
80%	Severe	0.079
90%	Slight	0.000
90%	Mild	0.000
90%	Moderate	0.000
90%	Moderately severe	0.083
90%	Severe	0.087

Table 8: Likelihood Ratio Test results by Degree of Hearing Loss.

PS2 with LRT - Baseline versus Extended - Sex		
Quantile	Degree of hearing loss	p-value
20%	Female	0.000
20%	Male	0.000
40%	Female	0.001
40%	Male	0.002
60%	Female	0.005
60%	Male	0.015
80%	Female	0.019
80%	Male	0.026
90%	Female	0.057
90%	Male	0.064

Table 9: Likelihood Ratio Test results by Sex.

5.5 CPMBT Model Risk Profiles

In this section, we discuss the results for the risk profiles of the overall CPMBT and hearing loss models, as well as the overall and sex CPMBT models. All analyses presented herein are based on the baseline model. Figures 11 and 12 provide a summary of the risk profile curves. The rows show α , β , and κ , respectively. Each row presents the profile risk curves across the five probabilistic levels, starting from 20% to 90%. In addition, each panel presents curves overall (in red) and with respect to hearing loss (in green) or sex (in magenta and blue). Note that a higher probabilistic level captures segments of the underlying population with more severe hearing loss. Moreover, as highlighted in Figure 6, the sample size of the hearing loss groups is not uniform and differs greatly between groups.

We start by examining Figure 11 and focusing on the age-specific intercepts α , representing the baseline hearing loss level across age (hence providing a value for each age group). Note that our responses $\tilde{\pi}_f$ are log-proportions, meaning that since the original proportions are between 0 and 1, the log-proportions range from $-\infty$ to 0. This implies that the smaller the proportion, the more negative the log-proportion becomes. As previously discussed, higher empirical quantiles imply lower proportions since fewer individuals fail to detect higher dB levels, which therefore indicates a greater degree of hearing loss for that specific proportion. For the α risk profiles, this means that the more negative the α values, the smaller the proportion, indicating that fewer people in that age group are able to exceed the empirical

quantile. Consequently, as we move across empirical quantiles (from left to right), the α values become increasingly negative, isolating segments with more severe levels of hearing loss.

A second aspect of this analysis is to observe how the α risk profiles behave across age for a fixed empirical quantile. In practice, if the baseline average α decreases, the average proportion also decreases, suggesting that fewer individuals exceed the empirical quantile threshold, which indicates worsening hearing levels for those proportions (and better hearing for the remainder of the population that does not exceed the threshold). With this perspective in mind, we start by looking at the overall baseline level of hearing loss α , shown in red in the first row panels. These age-specific intercepts provide a global measure of baseline hearing loss for each age group and initially show a slight expected decrease, followed by an increase, eventually stabilizing into a relatively flat pattern.

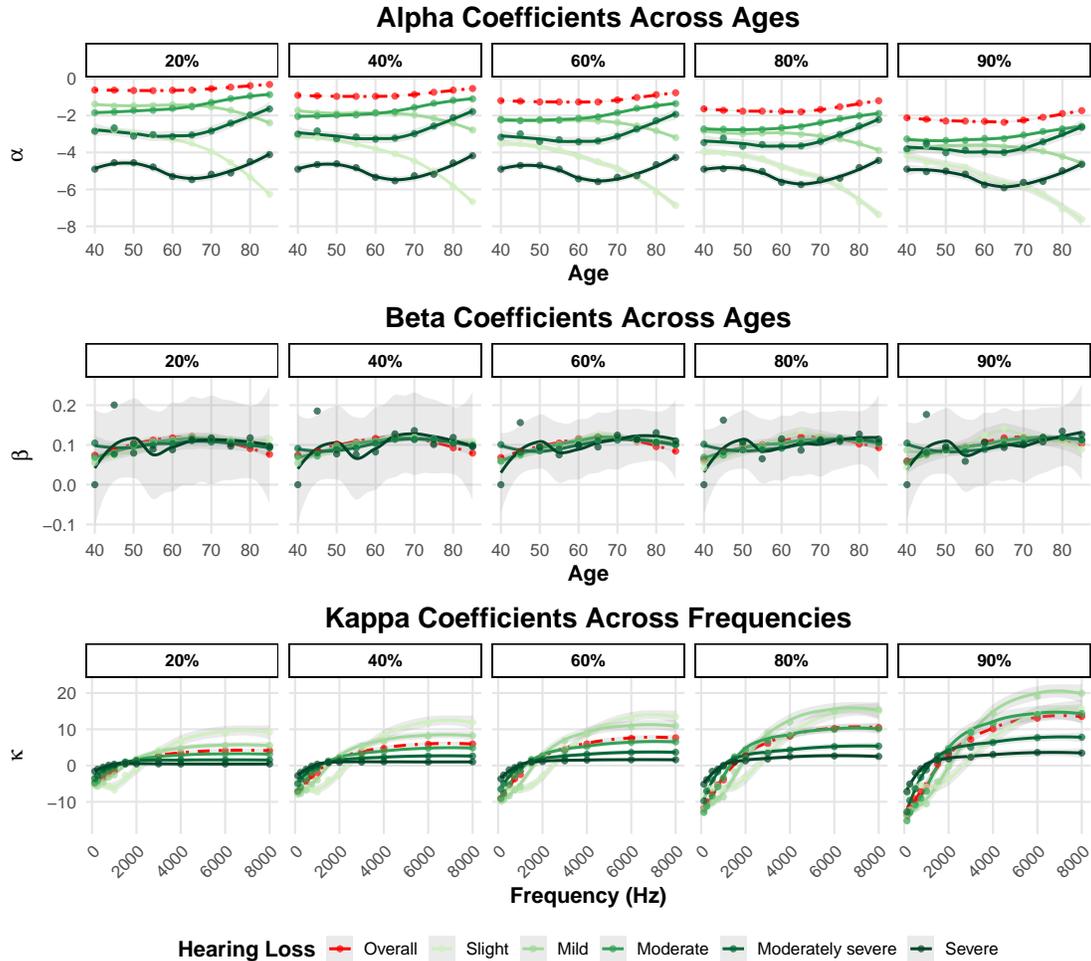


Figure 11: Risk profiles by degree of hearing loss. Each row carries a specific profile, i.e. α , β and κ_f . The columns represent the plots for the different probabilistic levels. While the x-axis of α and β corresponds to age, the panels in the last row for κ_f show frequency on the x-axis. The y-axis corresponds to the range values for each profile. Note that red lines represent the overall population risk profile curves, while green lines represent the profile curves corresponding to degree of hearing loss categories.

One might expect the baseline hearing loss α to show a consistent decrease with age, suggesting higher proportions for younger ages and lower proportions for older ages (possibly highlighting age-related hearing loss dynamics). However, our quantile estimation method, which accounts for sample size variations across age groups, reveals a more nuanced pattern of hearing loss progression. As age increases, the relative proportions do not decrease as expected because the number of individuals exceeding the empirical quantile remains relatively stable relative to the total population segment size. Consequently, the average proportion appears constant, indicating that the distribution of those able to

detect sounds above the empirical quantile does not show a sharp decline. This observed pattern suggests that while hearing loss may progress with age, its representation is complex. The findings are likely influenced by sample size variations, particularly in older age groups, highlighting the critical importance of considering population demographic characteristics in hearing loss analysis.

Next, we examine the α values for each degree of hearing loss. As with the overall α age-specific risk profiles, moving across the probabilistic levels yields lower values due to the decreasing proportions at higher empirical quantiles. Furthermore, the distance between these risk profiles diminishes as we move to higher probabilistic levels because individuals with more severe hearing loss dominate these higher quantiles, leading to convergence in hearing thresholds. This convergence indicates that at higher levels of hearing loss, the differences across age groups become less distinct, reflecting the fact that severe hearing impairment tends to affect all ages similarly once a high threshold of impairment is reached.

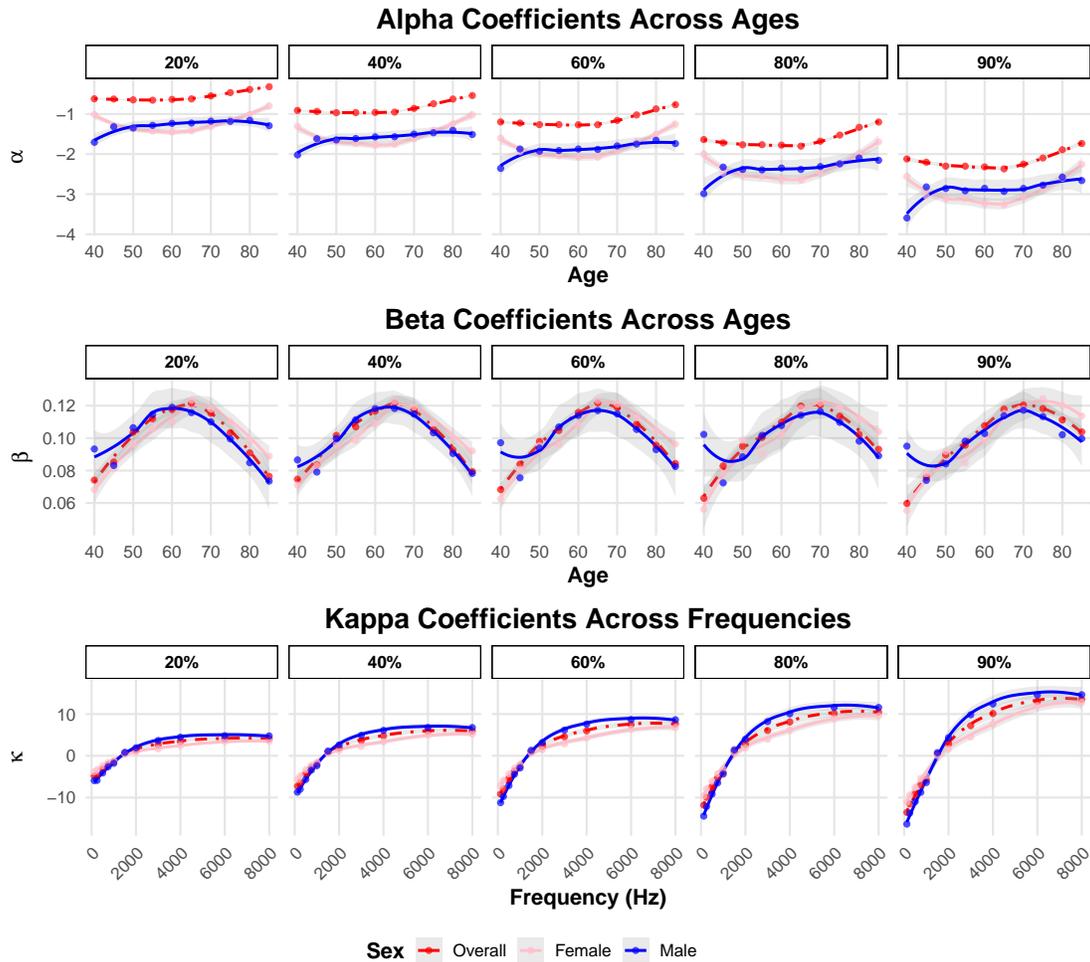


Figure 12: Risk profiles by sex. Each row carries a specific profile, i.e. α , β and κ_f . The columns represent the plots for the different probabilistic levels. While the x-axis of α and β corresponds to age, the panels in the last row for κ_f show frequency on the x-axis. The y-axis corresponds to the range values for each profile. Note that red lines represent the overall population risk profile curves, while coloured lines represent the profile curves corresponding to the sex categories.

If we instead focus on each empirical quantile, it is evident that for higher degrees of hearing loss—specifically moderate, moderately severe, and severe—there is a clear ordering where higher degrees of hearing loss present more negative risk profiles, as expected. This ordering highlights that individuals with less severe hearing loss tend to have better hearing thresholds, even at similar empirical quantile levels, leading to a stratification in the risk profiles. The trend across ages is similar to the overall risk profile curve, where it initially decreases slightly with age before

stabilizing or increasing. This trend is expected to diminish to reflect the gradual decline in hearing performance over time, consistent with the natural progression of age-related hearing loss. As in the overall case, the trend remains almost constant or even increases at higher ages due to the influence of sample size variability, which affects the relative proportions and prevents them from decreasing as expected.

For individuals with slight and mild hearing loss, the α risk profiles exhibit the expected behaviour over age—i.e., a consistent decrease in baseline hearing loss that becomes more pronounced at older ages. In these two hearing loss segments, the sample sizes are considerably more uniform across age groups (see more details in the Supplementary Information), resulting in more consistent relative proportions that diminish across age groups, thereby presenting lower age-intercept risk profiles. This pattern indicates that the relative baseline risk of hearing loss in individuals with slight and mild impairments is more predictable across ages, likely due to the steadier distribution of individuals in these categories compared to the more severe groups, where sample sizes and hearing abilities vary widely.

After examining the α values, we now focus on the second rows of Figure 11, which depict the β parameters. For visualization purposes only, we provided further plots of these profiles in the Supplementary Information. The β parameter is crucial for understanding how sensitive the audiogram proportions $\hat{\pi}_f$ are to changes in the hearing loss trend (κ_f). High β values indicate that the corresponding age group is highly sensitive to changes in the hearing loss trend (κ_f), resulting in more pronounced changes in audiogram proportions. Conversely, smaller β values indicate that the audiogram proportions for those age groups are less affected by changes in the overall hearing loss trend. Thus, the β parameters highlight how different age groups respond to variations in the frequency domain, providing insight into the age-dependent sensitivity to changes in hearing performance.

Next, we examine the β risk profiles for the overall population (shown in red) to understand how sensitivity to changes in the hearing loss trend (κ_f) varies across age groups. Note that we smooth the risk profiles across age and, for the severe category only, certain points were not interpolated. This information is instead captured by the high standard deviation across such a curve provided in the shaded area in the Figure. Overall, β values look fairly constant, with slight increase from ages 40 to around 65-70, where they peak, and then decrease in the 70-85 age range. This slight increase in the pattern suggests that individuals aged 65-70 might be at a critical point in their hearing health, with their audiogram proportions being most sensitive to changes in the overall hearing loss trend (κ_f). For these age groups, changes in κ_f translate to more substantial changes in their audiogram proportions, potentially indicating that they are at an age where hearing health is particularly dynamic and responsive to environmental factors or interventions.

Younger adults (ages 40-50) and the oldest age groups (75-85) exhibit lower β values, particularly at lighter hearing loss levels, indicating less sensitivity in their audiogram proportions to changes in the overall trend. For younger adults, this lower sensitivity suggests greater stability in hearing, making them less affected by overall hearing trends. In contrast, for the oldest age groups, the lower β values may indicate that hearing loss has stabilized to some degree, with further changes in κ_f having a limited impact on their already established hearing status.

Compared to the age-specific intercepts (α), the β parameters demonstrate more consistent values across the probabilistic levels, indicating similar sensitivity to the hearing loss trend across all probabilistic segments. These findings highlight the methodological potential of our risk profile approach in characterizing hearing loss progression. The created risk profiles offer a novel framework for understanding age-related hearing variations, providing a data-driven method to stratify population segments. Such profiles could potentially inform future research directions, such as investigating whether specific age groups or hearing loss profiles might benefit from tailored intervention strategies or hearing aid algorithms. Further investigations would be needed to validate these potential applications, but the current approach demonstrates a promising approach to systematically capturing hearing loss dynamics across different population segments.

We then focus on analyzing β parameters across various degrees of hearing loss and probabilistic levels to understand how sensitivity to the hearing loss trend (κ_f) varies depending on the severity of hearing loss. In general, β values increase from age 40 to a peak around 65-70, then decrease from ages 70 to 85. This trend is most pronounced in the slight and mild hearing loss categories, particularly at higher probabilistic levels (60%, 80%, 90%), likely because these individuals retain more capacity for variation in hearing sensitivity compared to those with more severe loss. The moderate hearing loss category closely follows the overall β trend, showing similar sensitivity patterns.

As hearing loss severity increases, the relationship between age and sensitivity to the trend (κ_f) becomes increasingly variable, particularly in the moderately severe and severe categories. For moderately severe hearing loss, the β risk profile decreases until around age 45, then increases to align more closely with the patterns of other hearing loss categories. This suggests that individuals with moderately severe hearing loss experience a more pronounced sensitivity to the hearing loss trend as they age, especially at higher probabilistic levels.

For the severe category, the β values exhibit a sudden increase around age 45, followed by a drop and stabilization as probabilistic levels increase. This sudden peak might be due to differences in sample sizes, particularly as this category

has the fewest individuals, which could lead to variability in the estimates. Nevertheless, the high sensitivity observed at younger ages and the subsequent stabilization suggest that individuals with severe hearing loss have pronounced changes in sensitivity early on, which then levels off with age. This variability implies that standard age-related patterns may not apply consistently for individuals with severe hearing loss, highlighting the need for individualized assessments to better understand and predict changes in audiogram performance.

Lastly, we focus on the last row of Figure 11, which presents κ_f , the trend in the proportions of hearing loss that is consistent across all age groups for frequency f . κ_f represents a common trend in hearing loss across all age groups, capturing the dynamic changes in hearing sensitivity over different frequencies. A higher level of κ_f corresponds to more severe hearing loss, reflecting a stronger trend in hearing loss. Each risk profile reflects the relative dynamic behaviour of hearing loss within its category, meaning that while slight hearing loss will not exceed the severity of a severe category, it may exhibit faster changes across the frequency domain.

Furthermore, examining higher probabilistic levels shows consistently larger κ_f values, which indicate more severe hearing impairments in these population segments (due to a reduction of their hearing abilities in the proportions considered). This trend is particularly pronounced between 1000 and 3000 Hz, suggesting a critical range for hearing loss. Low frequencies (125-500 Hz) show negative κ_f values, transitioning to positive in the mid-range (1000-2000 Hz), and increasing further at high frequencies (3000-8000 Hz). This pattern underscores the frequency-dependent nature of hearing loss and highlights the importance of comprehensive audiometric testing across a wide frequency range for early detection and intervention.

As the severity of hearing loss decreases, the trend shift within the 1000-3000 Hz range moves towards higher frequencies, particularly around 2000-3000 Hz for milder cases. This frequency range is critical for speech recognition, clarity, and overall communication ability. Hearing loss in this range can significantly affect a person's ability to understand speech, especially in complex listening environments. Therefore, it is not surprising that a shift in the trend pattern occurs precisely within this range. Additionally, in less severe hearing loss cases, higher frequency levels are often more affected, explaining the shift toward higher frequencies. Given the importance of this frequency range for speech recognition and clarity, this emphasizes the need for comprehensive audiometric testing across frequencies to enable early detection and effective intervention, especially in cases of mild to moderate hearing loss.

The slight risk profile for hearing loss shows the most pronounced dynamic change in κ_f among all categories. It starts at the lowest point and reaches the highest point throughout the frequency range, indicating that hearing loss in this category is more pronounced at high frequencies than at low frequencies. As we move across the hearing loss risk profiles, we observe a contrasting dynamic in the trend reference curves before and after the shift within the 1000-3000 Hz range. Although the slight hearing loss profile begins at a lower level compared to the severe category, its steepness and endpoint are considerably higher. This suggests that mild hearing impairments are more affected at high frequencies than at low frequencies. In contrast, more severe hearing loss categories tend to show a more uniform pattern across the entire frequency spectrum. Notably, the overall population κ_f generally falls between the mild and moderate hearing loss profiles, indicating that the overall trend represents an intermediate state between these two categories.

We now perform an equivalent analysis by examining Figure 12. This time, we compare the risk profile curves for the overall population with those of males and females. We begin by focusing on the age-specific α values. As the probabilistic level increases from 20% to 90%, the α values become progressively more negative, indicating worsening baseline hearing loss at more stringent probabilistic levels.

When examining age-related patterns, notable differences emerge between females and males. For females, a decrease in baseline hearing loss is observed from ages 40 to 65, which is consistent with the expected trend—indicating worse hearing with age progression. However, after age 65, there is an increasing trend in α , suggesting that the relative proportions are increasing, leading to a reduced risk in terms of baseline hearing loss for females as age increases. This is largely attributed to sample size variability, which affects the relative proportions and prevents them from decreasing as expected. For males, an opposite trend appears. Between the ages of 40 and 50, there is an increase in α , suggesting an improvement in baseline hearing loss. This is followed by a relatively flat trend until age 70, after which the α values decrease again, indicating a worsening of baseline hearing loss. This fluctuating behaviour becomes more pronounced with higher probabilistic levels, although the overall pattern remains consistent. This trend may suggest that, while baseline hearing loss is generally expected to worsen with age (i.e., more negative α values), the relative proportions for the male group increase initially, then decrease—showing a different trend compared to females. This variability highlights the influence of sample size differences across age groups, which impacts the consistency of hearing loss patterns within the male population.

We then move to the β risk profile, measuring the sensitivity of the audiogram proportions $\tilde{\pi}_f$ to the trend over frequency κ_f . If we focus on the first probabilistic levels, the highest sensitivity for the male occurs at the age of

60, while, for the females slightly later in time, at the age of 65. Moving across the probabilistic levels, these peaks tend to be delayed, i.e. around the ages of 70 and 75, respectively, by the 90% level. Such a sensitivity increases and decreases before and after these peaks, with the exception for the male showing a higher starting point than female which decreases to then increase again, usually between 40 to 50 across higher probabilistic levels in particular. These sex-based differences in β dynamics underscore the importance of considering sex when interpreting and predicting changes in audiogram proportions. They also suggest that interventions and monitoring strategies may need to be tailored not only to the degree of hearing loss but also to the sex of the individual, particularly in the earlier stages of hearing impairment. As for the case considering hearing loss severity, the range of the β values remain fairly constant across the probabilistic levels.

The parameter κ_f , representing the latent baseline hearing loss trend across frequencies, reveals distinct patterns when comparing the overall population to subgroups by sex. For the overall population, κ_f values increase as frequency rises, indicating a more pronounced hearing loss at higher frequencies. At the 20th probabilistic level, κ_f ranges from -0.73 at 125 Hz to 0.61 at 8000 Hz, suggesting a shift towards more severe hearing loss at higher frequencies. This trend is consistent across other percentiles (40%, 60%, 80%, and 90%), with increasingly higher κ_f values corresponding to more severe hearing loss. When comparing between sexes, females tend to have lower κ_f values across most frequencies, particularly at lower thresholds, implying a generally lower baseline hearing loss compared to males. For instance, at the 20th probabilistic level, females have a κ_f of -0.57 at 125 Hz compared to -0.73 for the overall population, while at 8000 Hz, the κ_f for females is 0.52 versus 0.61 for the overall population. These findings suggest that males may experience more severe high-frequency hearing loss compared to females, especially as hearing loss progresses. The increase in κ_f across higher percentiles further confirms that the severity of hearing loss is more substantial in men, particularly at high frequencies, underscoring the importance of considering sex-specific differences when modelling hearing loss dynamics.

5.6 Extended Model with Partial Linear Regression Results

This section presents the results of the extended model, estimated using the partial linear regression method introduced in Subsection 3. After estimating the baseline model, we applied the partial estimation method to calculate the coefficients γ_Q and γ_N , which quantify the relationship between the speech test proportions and each individual audiogram frequency π_f . Subsequently, these coefficients were evaluated using a two-sided t -test to test the null hypothesis that each coefficient is equal to zero. Formally, this consists in

$$H_0 : \gamma_i = 0$$

$$H_1 : \gamma_i \neq 0$$

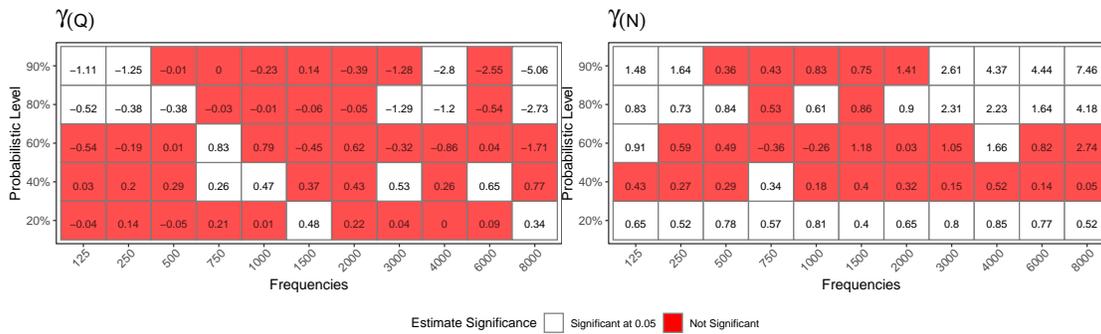


Figure 13: Heatmaps show the extended model speech estimate coefficients γ_Q (left) and γ_N (right) for models fitted on the overall hearing loss population. Each heatmap row represents a model fitted at a specific probabilistic level. For instance, the bottom row of the left heatmap displays partial linear regression coefficients for the speech test in quiet at the 20% probabilistic level, while the corresponding row of the right heatmap shows coefficients for the speech test in noise at the same probabilistic level. The x-axis represents audiogram frequencies. Statistically significant estimates (based on a two-sided t -test) are coloured in white, while non-significant estimates are coloured in red.

The test statistic used is the t -statistic, calculated as

$$t = \frac{\hat{\gamma}_i}{\hat{\sigma}_{\hat{\gamma}_i}}$$

where $\hat{\gamma}_i$ is the estimated value of the coefficient for the i -th speech test, and $\hat{\sigma}_{\hat{\gamma}_i}$ is the standard error of $\hat{\gamma}_i$. The t -statistic follows a t -distribution with $n - k$ degrees of freedom, where n is the sample size and k is the number of predictors (including the intercept) in the model, given as

$$t \sim t(n - k)$$

The significance of each coefficient is assessed using the t -statistic and associated p-value, with coefficients having p-values below 0.05 deemed statistically significant. Results are visualized in two heatmaps: one for speech-in-quiet estimates (γ_Q) and another for speech-in-noise estimates (γ_N). The x-axis represents audiogram frequencies, while the y-axis indicates different probabilistic levels. Each row corresponds to the same probabilistic level, facilitating direct comparison of estimates across tests. Cells display the estimated γ coefficient, with color indicating statistical significance, highlighting variations across frequencies and levels. Findings are first presented for the overall population, followed by breakdowns by degree of hearing loss and sex. Detailed results are in Supplementary Information, Section 13.

Figure 13 shows estimates for the overall population. The left panel presents speech-in-quiet coefficients (γ_Q), while the right panel shows speech-in-noise coefficients (γ_N). As the probabilistic level rises, the model captures more severely impaired populations. Speech-in-quiet coefficients are significant at extreme audiogram frequencies, especially at higher levels, while speech-in-noise coefficients show more overall significance. Notably, all speech-in-noise coefficients are significant at the first level, indicating potential overfitting, as none are significant at the 40% level.

The heatmaps in Figure 14 present results stratified by degree of hearing loss, with speech-in-quiet coefficients on the left ($\gamma_Q^{h_i}$) and speech-in-noise on the right ($\gamma_N^{h_i}$). Each subplot row corresponds to a model for specific hearing loss degrees.

In the slight hearing loss category, significant estimates mainly come from speech-in-noise, with few speech-in-quiet estimates significant, except at the 90% level. Significant speech-in-noise coefficients appear at frequencies like 500, 750, 2000, 3000, 4000, 6000, and 8000 Hz. As the probabilistic level increases, significant estimates decline, suggesting reduced relevance of speech-in-noise.

For the mild hearing loss category, speech-in-quiet shows increased significant estimates compared to the slight category, especially at the 20% level. At 90%, nearly all estimates are significant except at 3000 Hz.

In speech-in-noise for the mild category, significant estimates are concentrated in the higher frequency range (2000 to 8000 Hz), indicating stronger relevance for high-frequency audiogram measures.

In the moderate hearing loss category, speech-in-quiet coefficients show increased significance across most frequencies. However, as severity increases, significant estimates decrease, while speech-in-noise remains significant at higher frequencies (4000 to 8000 Hz) across nearly all levels, with additional significant frequencies at 750, 1000, 1500, and 2000 Hz for the first probabilistic level.

Notably, the 1000 Hz frequency shows statistically significant coefficients across most probabilistic levels, with the exception of the last level. Compared to the mild hearing loss case, lower frequencies tend to show more significance in the moderate category, particularly at 125 Hz. Therefore, in general, for moderate hearing loss, speech-in-quiet estimates are more frequently significant, and a similar trend can be observed for low-frequency estimates in speech-in-noise.

The following set of heatmaps presents the coefficient estimates for speech tests in the moderately severe hearing loss category. Focusing on $\gamma_Q^{h_4}$ (speech-in-quiet estimates), we observe a lower number of statistically significant coefficients, particularly at higher probabilistic levels where most coefficients lack significant p-values. However, as probabilistic levels decrease, this trend reverses, with the majority of coefficients becoming significant. Notably, across almost all probabilistic levels, the 500 Hz audiogram frequency consistently shows significant coefficients.

The heatmap for $\gamma_N^{h_4}$ (speech-in-noise estimates) shows significantly fewer significant coefficients compared to the moderate hearing loss case, indicating reduced statistical significance for both tests, particularly speech-in-noise, in the moderately severe hearing loss group.

For severe hearing loss, almost none of the coefficients are statistically significant across audiogram frequencies. Speech-in-quiet shows fluctuating significance, with nearly none for slight and severe cases and a mixed pattern for moderate cases. In contrast, speech-in-noise coefficients exhibit several significant estimates across slight, mild, and moderate categories, influenced by audiogram frequencies and probabilistic levels. The limited sample size in the severe group may affect these results.



Figure 14: Heatmaps for the extended model speech estimate coefficients $\gamma_Q^{h_i}$ (left) and $\gamma_N^{h_i}$ (right) for the models fitted by degree of hearing loss, with h_i representing the different PTA categories and $i \in \{1, \dots, 5\}$. The first set of heatmaps corresponds to the ones for slight hearing loss, followed by mild, moderate, moderately severe, and severe. Each heatmap row represents a model fitted at a specific probabilistic level. The same logic applies to each row as explained in Figure 13. The x-axis represents the audiogram frequencies. Cells corresponding to statistically significant estimates (based on a two-sided t-test for a coefficient different from zero) are coloured in white, while non-significant estimates are coloured according to the hearing loss colour, i.e. slight, mild, etc.

These findings should be considered in the context of the LRT results. The LRT demonstrated that the extended model significantly improves model fit for slight, mild, and moderate hearing loss categories, particularly at the 20%, 40%, and 60% quantile levels. This aligns with the observed coefficient significances, suggesting that the speech test coefficients for these hearing loss categories are more informative. Conversely, the lack of significant LRT improvements for moderately severe and severe categories corresponds with the limited coefficient significance observed in these groups. The limited sample size in the severe group may further influence these results and should be considered when drawing conclusions.

Next, Figure 15 presents coefficient estimates from extended models stratified by sex, revealing distinct patterns.

In the left heatmaps for speech-in-quiet ($\gamma_Q^{s_1}$ for females and $\gamma_Q^{s_2}$ for males), females show significant estimates mainly at the first probabilistic level (20%), except for 2000 Hz. None are significant at the second level, and fewer remain significant at the third and fourth levels, concentrated at the highest and lowest frequencies. Thus, significance declines for females as the probabilistic level increases.

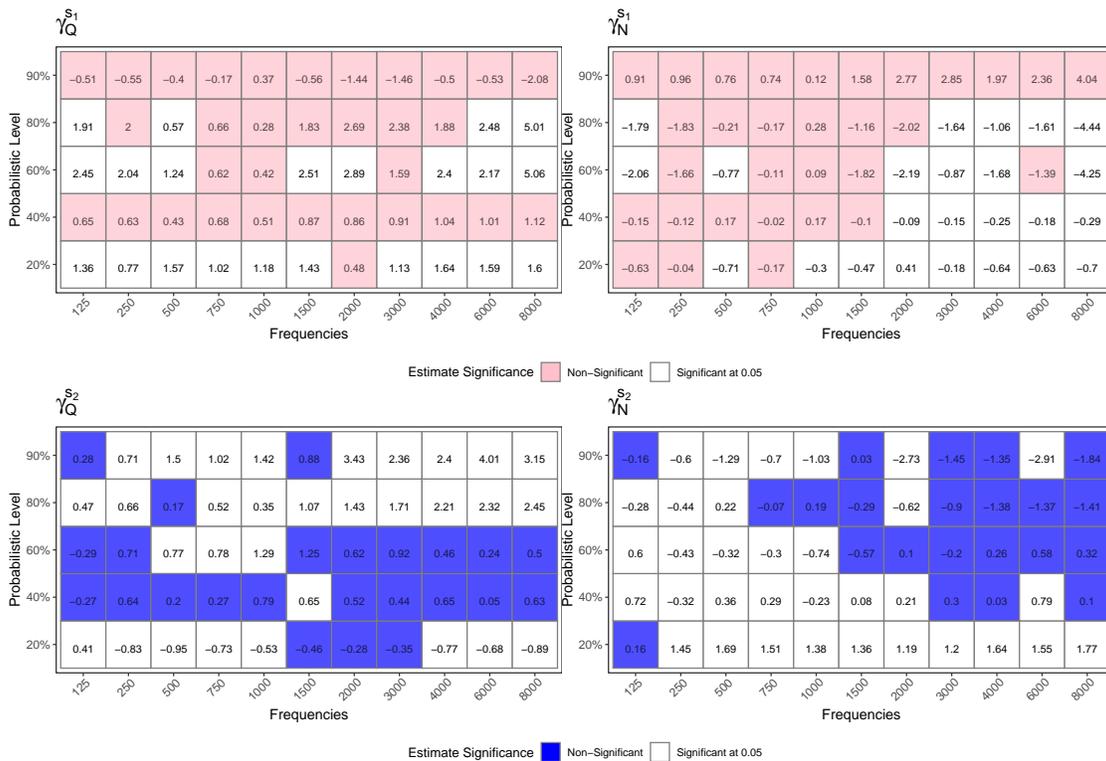


Figure 15: Heatmaps for the extended model speech estimate coefficients $\gamma_Q^{s_i}$ (left) and $\gamma_N^{s_i}$ (right) for the models fitted by sex, with s_i representing females and males, respectively, and $i \in \{1, 2\}$. The top heatmaps refer to the model fitted on the females, while the bottom heatmaps to the ones of the males. Each heatmap row represents a model fitted at a specific probabilistic level. Same logic applies to each row as explained in Figure 13. The x-axis represents audiogram frequencies. Cells corresponding to statistically significant estimates (based on a two-sided t-test for a coefficient different from zero) are coloured in white, while non-significant estimates are coloured according to the selected sex color.

For males, all estimates at the first level are significant except at 1500, 2000, and 3000 Hz. While significance decreases at 40% and 60%, it increases again at higher levels. Therefore, speech-in-quiet becomes more significant for males as hearing loss severity increases, contrary to the trend for females.

In the right heatmaps for speech-in-noise, significant patterns between sexes persist. Females show significant estimates across nearly all probabilistic levels, except the last, with no significance at higher frequencies (2000 to 8000 Hz). Males have nearly all coefficients significant at the first level, except for 125 Hz, but as probabilistic levels rise, significance shifts toward lower frequencies (from 1500 Hz to 125 Hz).

In summary, females and males demonstrate distinct patterns of coefficient significance for speech tests. These differences align with the LRT results, which showed significant model improvements at lower quantile levels. The frequency-specific significance of speech-in-noise coefficients suggests that these tests provide complementary information to audiogram measurements, particularly in characterizing hearing loss variations across different frequency ranges.

6 Discussion & Conclusion

The present study aimed to establish a comprehensive framework for assessing hearing loss through the formulation and application of advanced statistical modelling, specifically the state-space model corresponding to the CPMBT model. This model integrates information from audiograms and speech tests to develop standardized risk profiles and inference procedures designed to characterize differences within population segments. The findings of this study provide significant insights, which we discuss in relation to the two main sets of hypothesis testing challenges that were denoted as problem statements (**PS1** and **PS2**), the statistical inference results, and the outputs of the partial linear regression analyses.

This study developed a framework for assessing hearing loss using a state-space modelling approach, implementing the CPMBT model family, which encompasses both baseline and extended model variants. These models integrate audiogram and speech test data to create standardized risk profiles. The findings are discussed through the lens of two primary problem statements (**PS1** and **PS2**), statistical inference results, and partial linear regression analyses.

We introduced a novel method for defining hearing loss risk profiles based on the regression relationships studied between pure-tone audiometry and speech tests, which allow for personalized assessments and comprehensive population characterization. Our model enables detailed analysis of age and frequency dynamics critical for monitoring hearing loss and quantifies the relationship between speech tests and audiogram frequencies, providing essential insights for hearing loss progression. The methodologies developed here facilitate comparisons across different population segments.

Statistical inference tests, including the Vuong Test and Likelihood Ratio Test (LRT), evaluated model efficacy. **PS1** assesses differences within the same class of models, revealing that both baseline and extended models identify significant differences among overall and subgroup populations. The extended model incorporates speech tests, demonstrating a nuanced understanding of variations, especially at lower quantiles.

PS2 evaluates differences between the baseline and extended models, showing that the extended model significantly outperforms the baseline, particularly for lower quantiles associated with less severe hearing loss. This effectiveness is reflected in both the Likelihood Ratio Test results and the partial regression coefficient significances. The improvements extend to both sexes, with significant results at lower quantiles, although the statistical significance diminishes at higher quantile levels.

The CPMBT model's risk profiles offer insights into hearing loss dynamics across age, frequency, and severity. Parameters α , β , and κ_f reveal that α increases with age, particularly in severe categories, and shows notable gender differences. The parameter β indicates sensitivity to hearing loss trends, showing distinctive variations with a notable peak around ages 65-70, as detailed in the risk profile analysis. The κ_f parameter varies by severity, reflecting greater deterioration at higher severities.

Partial linear regression analysis reveals significant distinctions in frequency-specific coefficient estimates for speech recognition tests (SRT in noise and quiet). These tests provide insights into speech perception across different hearing loss severities, with speech-in-noise tests showing particular significance for slight to moderate hearing loss (lower severity categories), and speech-in-quiet tests becoming more informative for severe hearing loss categories. The analysis highlights the complementary nature of speech tests in characterizing hearing loss, with varying contributions across frequency ranges and hearing loss categories.

Sex-based differences in coefficient significances reveal distinct patterns of speech test estimates across frequencies. While the current analysis suggests differing significance for speech tests between females and males, these observations require further investigation to establish their clinical relevance. The study underscores the potential value of comprehensive, individualized hearing assessments that consider multiple assessment methods.

Importantly, this study is based on an observational design, which introduces variability in sample sizes across different population segments. Such variability has a direct impact on the uncertainty of the estimates obtained, especially where sample sizes are relatively small, leading to increased variance. Despite this, our approach maintains robustness in revealing meaningful trends in hearing performance across the population.

Another key aspect of this study is its methodological nature. The CPMBT models presented here provide a method to derive risk profile curves that can be used to assess hearing loss. To capture evolving population dynamics and reduce estimation uncertainty, these curves should be periodically updated (we recommend constructing these curves annually, semi-annually, or even quarterly to ensure that the derived risk profiles remain relevant and informative) as new data becomes available. This approach allows for more precise tracking of hearing loss trends and refinement of risk profile estimates.

Additionally, while the current study offers an interpretation of general patterns, it is crucial to acknowledge that these patterns are influenced by multiple factors, including age structure, gender distribution, and the severity of hearing loss across the population. The complexity of hearing loss dynamics underscores the necessity of using a state-space model to define risk curves, which provide a structured approach to explaining population segment variations in hearing loss prevalence. The CPMBT model allows us to understand how these dynamics evolve over time, offering a benchmark for risk assessment that could be influenced by various other factors, such as socio-demographic characteristics, comorbidities/other health-related variables or environmental exposures, which are not included in the present work but could be explored in future studies.

In conclusion, this research highlights the need for advanced modelling techniques to enhance hearing loss assessment, advocating for a personalized approach that integrates speech and pure-tone audiogram data. Future research should investigate the interplay between individual factors and hearing outcomes to improve diagnostic accuracy and intervention strategies. Moreover, continuous updates of risk profiles are recommended to ensure their applicability and precision in real-world contexts, as population characteristics evolve over time.

Authors' Contributions

M.C.: Conceptualization, data curation, formal analysis, investigation, methodology, software, original draft preparation, writing—review and editing.

G.W.P.: Conceptualization, formal analysis, investigation, methodology, supervision, original draft preparation, writing—review and editing.

P.M.: resources, data curation, review and editing.

M.B.: formal analysis, investigation, original draft preparation, writing—review and editing.

H.T-V.:supervision, funding, review and editing.

M.C., P.M., M.B., H.T-V. confirm that they had full access to all the data in the study. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by a grant from Fondation Pour l' Audition (FPA) to the Institut de l' Audition (FPA09 to Hung Thai-Van and FPA10 to Paul Avan and the Ceriah), a French government grant managed by the Agence Nationale de la Recherche under the France 2030 program, reference ANR-23-IAHU-0003 and by Amplifon. MB was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project ID 496819293.

Data Statement

The data provider is Amplifon France.

References

- [1] World Health Organization. *World report on hearing*. World Health Organization, 2021.
- [2] Quentin Lisan, Marcel Goldberg, Ghizlene Lahlou, Anna Ozguler, Sylvie Lemonnier, Xavier Jouven, Marie Zins, and Jean-Philippe Empana. Prevalence of hearing loss and hearing aid use among adults in France in the constances study. *JAMA network open*, 5(6):e2217633–e2217633, 2022.
- [3] Lesley M Haile, Kaloyan Kamenov, Paul S Briant, et al. Hearing loss prevalence and years lived with disability, 1990-2019: findings from the global burden of disease study 2019. *The Lancet*, 397(10278):996–1009, 2021.
- [4] David McDaid, A-La Park, and Shelly Chadha. Estimating the global costs of hearing loss. *International Journal of Audiology*, 60(3):162–170, 2021.
- [5] Gill Livingston, Andrew Sommerlad, Vasiliki Orgeta, et al. Dementia prevention, intervention, and care. *The Lancet*, 390(10113):2673–2734, 2017.

- [6] Frank R Lin, Kristine Yaffe, Jing Xia, et al. Hearing loss and cognitive decline in older adults. *JAMA internal medicine*, 173(4):293–299, 2013.
- [7] Quentin Lisan, Thomas T van Sloten, Cedric Lemogne, et al. Association of hearing impairment with incident depressive symptoms: a community-based prospective study. *The American Journal of Medicine*, 132(12):1441–1449.e4, 2019.
- [8] Paul Mick, Ichiro Kawachi, and Frank R Lin. The association between hearing loss and social isolation in older adults. *Otolaryngology–Head and Neck Surgery*, 150(3):378–384, 2014.
- [9] Rui Gong, Xiaoyuan Hu, Chunmei Gong, et al. Hearing loss prevalence and risk factors among older adults in china. *International Journal of Audiology*, 57(5):354–359, 2018.
- [10] Seiji Sugiura, Yasuyuki Uchida, Yukiko Nishita, et al. Prevalence of usage of hearing aids and its association with cognitive impairment in japanese community-dwelling elders with hearing loss. *Auris Nasus Larynx*, 49(1):18–25, 2022.
- [11] Hans Morten Borchgrevink, Kristian Tambs, and Howard J Hoffman. The nord-trøndelag norway audiometric survey 1996-98: unscreened thresholds and prevalence of hearing impairment for adults \geq 20 years. *Noise and Health*, 7(28):1–15, 2005.
- [12] Howard J Hoffman, Robert A Dobie, Karen G Losonczy, et al. Declining prevalence of hearing loss in us adults aged 20 to 69 years. *JAMA Otolaryngology–Head & Neck Surgery*, 143(3):274–285, 2017.
- [13] Wenqing Zhan, Karen J Cruickshanks, Barbara EK Klein, et al. Generational differences in the prevalence of hearing impairment in older adults. *American Journal of Epidemiology*, 171(2):260–266, 2010.
- [14] Blake S Wilson, Debara L Tucci, Michael H Merson, and Gerard M O’Donoghue. Global hearing health care: new findings and perspectives. *The Lancet*, 390(10111):2503–2515, 2017.
- [15] L Haeusler, T de Laval, and C Millot. *Etude quantitative sur le handicap auditif à partir de l’enquête ”Handicap-Santé” : document de travail, série études et recherches*. 2019. Accessed April 11, 2019.
- [16] Judy R Dubno, Mark A Eckert, Fu-Shing Lee, Lois J Matthews, and Richard A Schmiedt. Classifying human audiometric phenotypes of age-related hearing loss from animal models. *Journal of the Association for Research in Otolaryngology*, 14:687–701, 2013.
- [17] Kenneth I Vaden, Lois J Matthews, Mark A Eckert, and Judy R Dubno. Longitudinal changes in audiometric phenotypes of age-related hearing loss. *Journal of the Association for Research in Otolaryngology*, 18:371–385, 2017.
- [18] Harold F Schuknecht and Mark R Gacek. Cochlear pathology in presbycusis. *Annals of Otolaryngology, Rhinology & Laryngology*, 102(1_suppl):1–16, 1993.
- [19] Harold F Schuknecht, Kozo Watanuki, Tadahiko Takahashi, A Aziz Belal Jr, Robert S Kimura, Diane Deleo Jones, and Carol Y Ota. Atrophy of the stria vascularis, a common cause for hearing loss. *The Laryngoscope*, 84(10):1777–1821, 1974.
- [20] Arne W Scholtz, Keren Kammen-Jolly, Edward Felder, Burkhard Hussl, Helge Rask-Andersen, and Anneliese Schrott-Fischer. Selective aspects of human pathology in high-tone hearing loss of the aging inner ear. *Hearing research*, 157(1-2):77–86, 2001.
- [21] Dennis McFadden. Sex differences in the auditory system. *Hearing Research*, 238:34–44, 1998.
- [22] Deborah Kimura. Sex differences in the auditory system: Animal models. *Journal of Comparative Psychology*, 129:189–200, 2015.
- [23] Raymond D. Kent. The biology of speech: Frequency characteristics and perception. *Journal of Speech, Language, and Hearing Research*, 40:1234–1250, 1997.
- [24] Trong T. Le, Laurie A. McCauley, and Oriol Rodrigo. Noise exposure and occupational noise-induced hearing loss in different industries. *American Journal of Industrial Medicine*, 38:80–90, 2000.
- [25] Reinier Plomp. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical society of America*, 63(2):533–549, 1978.
- [26] Raymond Carhart. Basic principles of speech audiometry. *Acta Oto-Laryngologica*, 40(1-2):62–71, 1951.
- [27] Raymond Carhart and Tom W Tillman. Interaction of competing speech signals with hearing losses. *Archives of Otolaryngology*, 91(3):273–279, 1970.
- [28] Mead C Killion and Patricia A Niquette. What can the pure-tone audiogram tell us about a patient’s snr loss?. *The Hearing Journal*, 53(3):46–48, 2000.

- [29] Lidija Ristovska, Zora Jačova, Jasmina Kovačević, Vesna Radovanović, and Husnija Hasanbegović. Correlation between pure tone thresholds and speech thresholds. *Human Research in Rehabilitation*, 11(2):120–125, 2021.
- [30] Kenneth G Shipley and Julie G McAfee. *Assessment in speech-language pathology: A resource manual*. Plural Publishing, 2023.
- [31] Jack Katz, Larry Medwetsky, Robert F Burkard, and Linda J Hood. *Handbook of clinical audiology*. Wolters Kluwer, Lippincott William & Wilkins Philadelphia, 2009.
- [32] Nancy Tye-Murray. *Foundations of aural rehabilitation: Children, adults, and their family members*. Plural Publishing, 2022.
- [33] Van Summers, Matthew J Makashay, Sarah M Theodoroff, and Marjorie R Leek. Suprathreshold auditory processing and speech perception in noise: Hearing-impaired and normal-hearing listeners. *Journal of the American Academy of Audiology*, 24(04):274–292, 2013.
- [34] David R Moore, Mark Edmondson-Jones, Piers Dawes, Heather Fortnum, Abby McCormack, Robert H Pierzycki, and Kevin J Munro. Relation between speech-in-noise threshold, hearing loss and cognition from 40–69 years of age. *PLoS one*, 9(9):e107720, 2014.
- [35] Mareike Buhl. Interpretable clinical decision support system for audiology based on predicted common audiological functional parameters (cafpas). *Diagnostics*, 12(2):463, 2022.
- [36] A Stach Brad. Clinical audiology: An introduction/brad a. stach. *Detroit, Michigan: Delmar, Cengage Learning*, 2010.
- [37] François Charih, Matthew Bromwich, Amy E Mark, Renée Lefrançois, and James R Green. Data-driven audiogram classification for mobile audiometry. *Scientific Reports*, 10(1):3962, 2020.
- [38] Aravindakshan Parthasarathy, Sandra Romero Pinto, Rebecca M Lewis, William Goedicke, and Daniel B Polley. Data-driven segmentation of audiometric phenotypes across a large clinical cohort. *Scientific reports*, 10(1):6704, 2020.
- [39] Peter WF Wilson, Ralph B D’Agostino, Daniel Levy, Albert M Belanger, Harris Silbershatz, and William B Kannel. Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18):1837–1847, 1998.
- [40] Andrew M Wolf, Richard C Wender, Ruth B Etzioni, Ian M Thompson, Anthony V D’Amico, Robert J Volk, Durado D Brooks, Cheryl Dash, Idris Guessous, Kate Andrews, et al. American cancer society guideline for the early detection of prostate cancer: update 2010. *CA: a cancer journal for clinicians*, 60(2):70–98, 2010.
- [41] Dan G Blazer, Sarah Domnitz, Catharyn T Liverman, the Committee on Accessible, Affordable Hearing Health Care for Adults, Engineering the National Academies of Sciences, and Medicine. Hearing health care services: improving access and quality. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*, 2016.
- [42] Adele M Goman and Frank R Lin. Prevalence of hearing loss by severity in the united states. *American journal of public health*, 106(10):1820–1822, 2016.
- [43] David B Rein, Carolina Franco, Nicholas S Reed, Evan R Herring-Nathan, Phoebe A Lamuda, Katelin M Alfaro Hudak, Wen Hu, Alex J Hartzman, Karl R White, and John S Wittenborn. The prevalence of bilateral hearing loss in the united states in 2019: a small area estimation modelling approach for obtaining national, state, and county level estimates by demographic subgroup. *The Lancet Regional Health–Americas*, 30, 2024.
- [44] Karen J Cruickshanks, Ted S Tweed, Terry L Wiley, Barbara EK Klein, Ronald Klein, Rick Chappell, David M Nondahl, and Dayna S Dalton. The 5-year incidence and progression of hearing loss: the epidemiology of hearing loss study. *Archives of Otolaryngology–Head & Neck Surgery*, 129(10):1041–1046, 2003.
- [45] Peishan Li, Kaiyun Pang, Rong Zhang, Lan Zhang, and Hui Xie. Prevalence and risk factors of hearing loss among the middle-aged and older population in china: a systematic review and meta-analysis. *European Archives of Oto-Rhino-Laryngology*, 280(11):4723–4737, 2023.
- [46] Lydia M Haile, Aislyn U Orji, Kelly M Reavis, Paul Svitil Briant, Katia M Lucas, Fares Alahdab, Till Winfried Bärnighausen, Arielle Wilder Bell, Chao Cao, Xiaochen Dai, et al. Hearing loss prevalence, years lived with disability, and hearing aid use in the united states from 1990 to 2019: Findings from the global burden of disease study. *Ear and hearing*, 45(1):257–267, 2024.
- [47] Frank R Lin, Roland Thorpe, Sandra Gordon-Salant, and Luigi Ferrucci. Hearing loss prevalence and risk factors among older adults in the united states. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 66(5):582–590, 2011.
- [48] World Health Organization. Deafness and hearing loss, 2021. Accessed: 2024-08-21.

- [49] Position statement: Red flags-warning of ear disease - american academy of otolaryngology-head and neck surgery (aao-hns). <https://www.entnet.org/resource/position-statement-red-flags-warning-of-ear-disease/>, 2024. Accessed: 2024-08-11.
- [50] Sophia Rosen and Ori Davidov. Order-restricted inference for multivariate longitudinal data with applications to the natural history of hearing loss. *Statistics in Medicine*, 31(16):1761–1773, 2012.
- [51] Byron J Gajewski, Jack D Sedwick, and Patrick J Antonelli. A log-normal distribution model of the effect of bacteria and ear fenestration on hearing loss: a bayesian approach. *Statistics in Medicine*, 23(3):493–508, 2004.
- [52] Yanghui Sheng, Ce Yang, Sharon Curhan, Gary Curhan, and Molin Wang. Analytical methods for correlated data arising from multicenter hearing studies. *Statistics in medicine*, 41(26):5335–5348, 2022.
- [53] Ana M. Ortega-Villa, Martha C. Nason, Michael P. Fay, Sara Alehashemi, Raphaela Goldbach-Mansky, and Dean A. Follmann. Regression approaches to assess effect of treatments that arrest progression of symptoms. *Statistics in Medicine*, n/a(n/a).
- [54] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [55] Masanao Aoki. *State space modeling of time series*. Springer Science & Business Media, 2013.
- [56] KB Newman, ST Buckland, Byron JT Morgan, R King, DL Borchers, Diana J Cole, Panagiotis Besbeas, O Gimenez, and L Thomas. Modelling population dynamics. *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-Space Methods*. Springer New York, New York, USA, pages 169–195, 2014.
- [57] Chang-Jin Kim and Charles R Nelson. *State-space models with REGIME Switching: Classical and gibbs-sampling approaches with applications*. MIT press, 2017.
- [58] Bernard Friedland. *Control system design: an introduction to state-space methods*. Courier Corporation, 2012.
- [59] Vanja Dukic, Hedibert F Lopes, and Nicholas G Polson. Tracking epidemics with google flu trends data and a state-space seir model. *Journal of the American Statistical Association*, 107(500):1410–1426, 2012.
- [60] Matteo Fasiolo, Natalya Pya, and Simon N Wood. A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science*, pages 96–118, 2016.
- [61] Marie Auger-Méthé, Ken Newman, Diana Cole, Fanny Empacher, Rowenna Gryba, Aaron A King, Vianey Leos-Barajas, Joanna Mills Flemming, Anders Nielsen, Giovanni Petris, et al. A guide to state–space modeling of ecological time series. *Ecological Monographs*, 91(4):e01470, 2021.
- [62] Ronald D Lee and Lawrence R Carter. Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671, 1992.
- [63] Charles R Nelson and Andrew F Siegel. Parsimonious modeling of yield curves. *Journal of business*, pages 473–489, 1987.
- [64] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.
- [65] Norman R Draper and Harry Smith. *Applied regression analysis*, volume 326. John Wiley & Sons, 1998.
- [66] David F Hendry and Mary S Morgan. *The foundations of econometric analysis*. Cambridge University Press, 1997.
- [67] David A Belsley, Edwin Kuh, and Roy E Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, 2005.
- [68] Ragnar Frisch and Frederick V Waugh. Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401, 1933.
- [69] Michael C Lovell. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010, 1963.
- [70] John G Clark. Uses and abuses of hearing loss classification. *Asha*, 23(7):493–500, 1981.
- [71] Article 8 - loi n° 78-17 du 6 janvier 1978 relative à l’informatique, aux fichiers et aux libertés - légifrance. https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000037090124/2018-05-25, 2024. Accessed: 2024-08-11.
- [72] Raymond Carhart and James F Jerger. Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders*, 24(4):330–345, 1959.
- [73] Jean-Claude Lafon. Le test phonétique. https://books.google.fr/books/about/Le_test_phon%C3%A9tique.html, 1964.

- [74] Jean-Claude Lafon. Le test phonétique et la mesure de l'audition. https://books.google.fr/books/about/Le_test_phon%C3%A9tique_et_la_mesure_de_l_au.html, 1958.
- [75] Johanna Buisson Savin, Pierre Reynard, Eric Bailly-Masson, Célia Joseph, Charles-Alexandre Joly, Catherine Boiteux, and Hung Thai-Van. Adult normative data for the adaptation of the hearing in noise test in european french (hint-5 min). In *Healthcare*, volume 10, page 1306. MDPI, 2022.
- [76] Véronique Vaillancourt, Chantal Laroche, Chantal Mayer, Cynthia Basque, Madeleine Nali, Alice Eriks-Brophy, Sigfrid D Soli, and Christian Giguère. Adaptation of the hint (hearing in noise test) for adult canadian franco-phone populations: Adaptación del hint (prueba de audición en ruido) para poblaciones de adultos canadienses francófonos. *International Journal of Audiology*, 44(6):358–361, 2005.
- [77] Wouter A Dreschler, Hans Verschuure, Carl Ludvigsen, and Søren Westermann. Icras noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment: Ruidos icra: Señales de ruido artificial con espectro similar al habla y propiedades temporales para pruebas de instrumentos auditivos. *Audiology*, 40(3):148–157, 2001.