Trends in Hearing
2016, Vol. 20: 1–17
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516655795
tia.sagepub.com
SAGE

Sentence Recognition Prediction for Hearing-impaired Listeners in Stationary and Fluctuation Noise With FADE: Empowering the Attenuation and Distortion Concept by Plomp With a Quantitative Processing Model

Birger Kollmeier[1], Marc René Schädler[1], Anna Warzybok[1], Bernd T. Meyer[1], and Thomas Brand[1]

Abstract

To characterize the individual patient's hearing impairment as obtained with the matrix sentence recognition test, a simulation Framework for Auditory Discrimination Experiments (FADE) is extended here using the Attenuation and Distortion (A+D) approach by Plomp as a blueprint for setting the individual processing parameters. FADE has been shown to predict the outcome of both speech recognition tests and psychoacoustic experiments based on simulations using an automatic speech recognition system requiring only few assumptions. It builds on the closed-set matrix sentence recognition test which is advantageous for testing individual speech recognition in a way comparable across languages. Individual predictions of speech recognition thresholds in stationary and in fluctuating noise were derived using the audiogram and an estimate of the internal level uncertainty for modeling the individual Plomp curves fitted to the data with the Attenuation (A-) and Distortion (D-) parameters of the Plomp approach. The "typical" audiogram shapes from Bisgaard et al with or without a "typical" level uncertainty and the individual data were used for individual predictions. As a result, the individualization of the level uncertainty was found to be more important than the exact shape of the individual audiogram to accurately model the outcome of the German Matrix test in stationary or fluctuating noise for listeners with hearing impairment. The prediction accuracy of the individualized approach also outperforms the (modified) Speech Intelligibility Index approach which is based on the individual threshold data only.

Keywords

models of hearing, speech perception, hearing impairment

Date received: 15 December 2015; revised: 9 May 2016; accepted: 9 May 2016

Introduction

The adequate modeling of the speech recognition in background noise observed in listeners with impaired hearing and the identification of the most relevant sensory and cognitive factors to be incorporated in correctly predicting the individual speech recognition thresholds (SRT) in quiet and in noise has been a topic of interest for several decades already (e.g., Kollmeier, 1990; Pavlovic, Studebaker, & Sherbecoe, 1986; Plomp, 1978; Rhebergen, Lyzenga, Dreschler, & Festen, 2010). One of the most influential and practically applicable ways of performing a systematic, but comparatively simple model-driven classification of the SRT dependence on background noise level was proposed by Plomp (1978) which has been used by several authors since then to characterize their empirical data (Duquesnoy, 1983;

[1] Medizinische Physik and Cluster of Excellence Hearing4all, Universität Oldenburg, Germany

Corresponding author:
Birger Kollmeier, Cluster of Excellence Hearing4All, Universität Oldenburg, Oldenburg D-26111, Germany.
Email: birger.kollmeier@uni-oldenburg.de

Plomp, 1986; Smoorenburg, 1992; Wagener, 2004). For each type of background noise, the Plomp approach estimates an individual A- (Attenuation) parameter characterizing the average loss in sensitivity due to hearing impairment (largely controlled by the audiogram) separately from a D- (Distortion) parameter characterizing the suprathreshold distortion component across frequencies (see later). However, the relation between these individually fitted parameters characterizing the SRT data for a specific speech material and noise and the characterization of hearing impairment (e.g., the audiogram and measures of individual suprathreshold processing deficits) is still unclear. This lack of a theoretical concept underpinning the A- and D- component with independent evidence from other data is one of the major drawbacks of the Plomp approach. Such a theoretical concept should be able to explain the empirically fitted A- and D-component on the basis of a functional model of the effective signal processing in the auditory system and the factors underlying the individual hearing impairment. The current contribution therefore attempts to provide a computational model that should help to bridge this gap between psychoacoustics, speech recognition in noise, and individualized assessment of the effect of hearing impairment. It is based on the Framework for Auditory Discrimination Experiments (FADE) proposed by Schädler, Warzybok, Hochmuth, and Kollmeier (2015) and is extended here toward predicting SRTs for hearing-impaired listeners incorporating different degrees of individualization.

Traditional modeling approaches for speech recognition have primarily used the individual audiogram as an input parameter without explicitly taking suprathreshold distortions into account. These approaches are either based on predefined features (like an energy increase in a certain auditory band) or on instrumental measures that are calibrated using a set of reference thresholds (like the Articulation Index, French and Steinberg, 1947, or Speech Intelligibility Index [SII]-based methods, see ANSI, 1997; Meyer & Brand, 2013; Rhebergen et al., 2010). Using different extensions of the SII for predicting SRT in stationary and fluctuating noise, Meyer and Brand could demonstrate an increase in prediction accuracy over the original SII if either frequency-independent or frequency-dependent level fluctuations of the noise are considered or if frequency-dependent fluctuations of both speech and noise are taken into account. However, irrespective of the variation of the SII employed, an individual audiogram-based prediction of speech recognition in fluctuating noise can only be obtained with a comparatively low prediction quality (Festen & Plomp, 1990; George, Festen, & Houtgast, 2006; Meyer & Brand, 2013; Rhebergen & Versfeld, 2005).

More elaborated approaches are based on psychoacoustical processing models, such as those used in Holube and Kollmeier (1996), Dau, Kollmeier, and Kohlrausch (1997), and Jürgens and Brand (2009), but require an optimal detector that possesses perfect prior knowledge about the to-be-recognized signals. This strong assumption of a detector with perfect a priori knowledge of the speech signal to be detected provides the model with a strong advantage over human listeners performing the same task. This makes it questionable if the requirements for the auditory-system-inspired processing front end to achieve human-like performance are realistic. This, in turn, could be crucial to accurately model human sound perception.

An alternative way of predicting both sentence recognition thresholds and psychoacoustic performance using automatic speech recognition (ASR) without requiring a predefined task-dependent reference criterion or using an "optimum detector" was recently proposed by Schädler et al. (2015) and Schädler, Warzybok, Ewert, and Kollmeier (2016). In a first study, they predicted the outcome of the German Matrix sentence recognition test (Kollmeier et al., 2015) for normal-hearing listeners in different types of stationary background noise. The ASR-based model uses Mel-frequency cepstral coefficients (MFCCs) as a front end and whole-word Gaussian Mixture or Hidden Markov Models (HMMs) as a back end. By training and testing the ASR system with noisy matrix sentences on a broad range of signal-to-noise ratios (SNRs), they were able to predict SRTs for normal-hearing listeners with a remarkably high precision, outperforming SII-based predictions. In a second study, they extended FADE to successfully simulate basic psychoacoustical experiments as well as more complex matrix sentence recognition tasks with a single set of parameters even when employing a range of feature sets (front ends). Even though the proposed FADE framework uses some a priori knowledge (e.g., training of the HMM for each target item for a range of SNR values, availability of a performance measure across all trained versions to determine the training SNR yielding the best performance), Schädler et al. (2015, 2016) concluded that the proposed FADE framework is able to predict empirical data from the literature with fewer assumptions and less restrictions than the optimal detector approach. Moreover, in comparison to the SII or other traditional modeling approaches, FADE does not rely on an empirical reference condition. This article therefore applies this framework to the individual prediction of sentence recognition in hearing-impaired listeners.

Plomp Curves

A desirable property of any valid sentence recognition model is its compatibility with the A + D approach introduced by Plomp (1978, 1986). This approach assumes that the SRT is available for several levels of the background noise *NL* (including the SRT in quiet, i.e., a noise level below the absolute detection threshold for noise). It then fits a function to the data (denoted as “Plomp curve” in the following) with the parameters A and D as follows:

$$SRT_{Plomp,orig} = 10log_{10}\left(10^{\frac{(L_0+A+D)}{10}} + 10^{\frac{(NL+SRT_N+D)}{10}}\right) \quad (1)$$

where $SRT_{Plomp,orig}$ denotes the speech level at threshold as originally proposed by Plomp, $L_0 = 19.9$ dB is the SRT in quiet and $SRT_N = -7.1$ dB (expressed as SNR) denotes the SRT in suprathreshold noise for the average normal-hearing listener for the German Matrix sentence recognition test employed here. Note that the speech-material-dependent values of L_0 and SRT_N employed here are taken from Brand and Kollmeier (2002) and are consistent with a number of other studies employing the same test in normal-hearing listeners (e.g., Wagener, 2004; Wagener, Brand, & Kollmeier, 2006). The task-dependent parameters A and D are fitted to the individual data from each patient.

Equation (1) was employed here in a slightly modified form: A power-law additivity parameter $0 < P \leqslant 100$ is introduced to better reflect the fluctuating noise case:

$$SRT_{Plomp} = 10\log_{10}\left(10^{\frac{(L_0+A+D)*P}{10}} + 10^{\frac{(NL+SRT_N+D)*P}{10}}\right)\Big/P \quad (2)$$

Note that for $P = 1$, Equation (1) is identical to Equation (2). The general shape of the Plomp curves is depicted in Figure 1: An increase in parameter A (“Attenuation component”) produces an increased SRT in quiet hence characterizing the loss in audibility caused by a hearing impairment without assuming a specific frequency dependence, while the SRT at higher noise levels basically remains unchanged. An increase in parameter D (“Distortion component”), on the other hand, corresponds to an upward shift toward higher SRT values of the whole SRT curve as a function of noise level.

For a given hearing loss, the SRT in quiet is dominated by the sum of A + D (horizontal part of the Plomp curves depicted in Figure 1, A). With increasing noise level NL, a transition region (controlled by *P*) occurs until a constant SNR at SRT is achieved across a wide range of noise levels which reflects the D-value. Note that the value of *P* is critically dependent on very few data points in the vicinity of the transition region.

While in theory only two SRT data points at two different noise levels are sufficient to estimate the fitting parameters A and D, a somewhat higher precision of the estimate is achieved if more data points are available and if both very low and sufficiently high noise levels are covered by the data. Wagener (2004) demonstrated that the Plomp parametrization fits very well the individual data for the Matrix sentence recognition test in German for a number of hearing-impaired listeners. Hence, the Plomp approach appears to provide a valid, but task-dependent description of SRT data from hearing-impaired listeners when employing the German Matrix sentence recognition test. Since an increasing number of highly comparable Matrix sentence recognition tests exists for different languages (Kollmeier et al., 2015), it can be assumed that this Plomp curve representation of individual SRT data might as well be suitable for a number of languages.

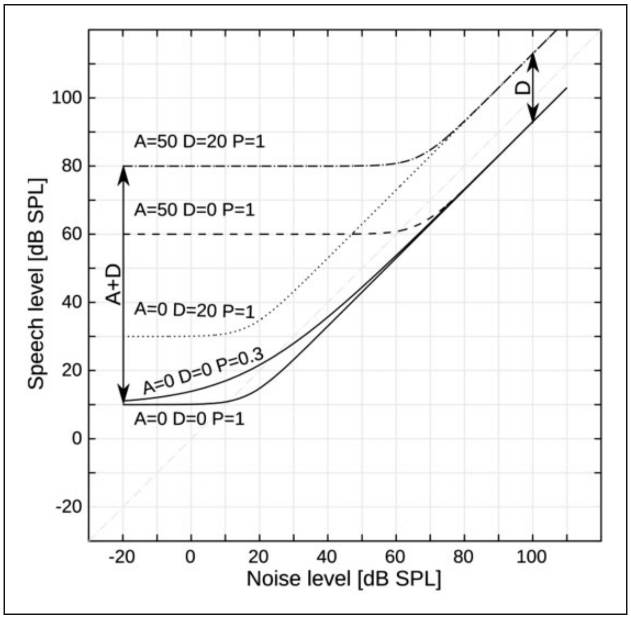


Figure 1. Cartoon of the Plomp curves according to Equation (2): The speech recognition threshold (SRT) in noise is given as speech level on the y-axis, the noise level is plotted on the x-axis. The effect of altering the parameters A, D, and P, respectively, is demonstrated by the five different Plomp curves generated by the respective parameter values listed above the respective curve: The sum A + D produces a shift of the SRT in quiet, D produces an upward shift of the whole Plomp curve, and lowering P below 1 produces a less abrupt transition from the horizontal part of the Plomp curve (i.e., SRT in quiet) to the linear increasing part characterizing the increase of the speech level at SRT with increasing noise level.

Principles of the FADE Approach for Simulating the Effect of Hearing Impairment

The FADE approach outlined earlier has been extended to model the effect of hearing impairment on SRTs as described by the empirically fitted Plomp curves for data obtained with the German Matrix test in stationary and fluctuating noise. The automatic speech recognizer operates on the log-scaled Mel-spectrogram (logMS) of the signal and the noise (see Figure 2, upper two panels for

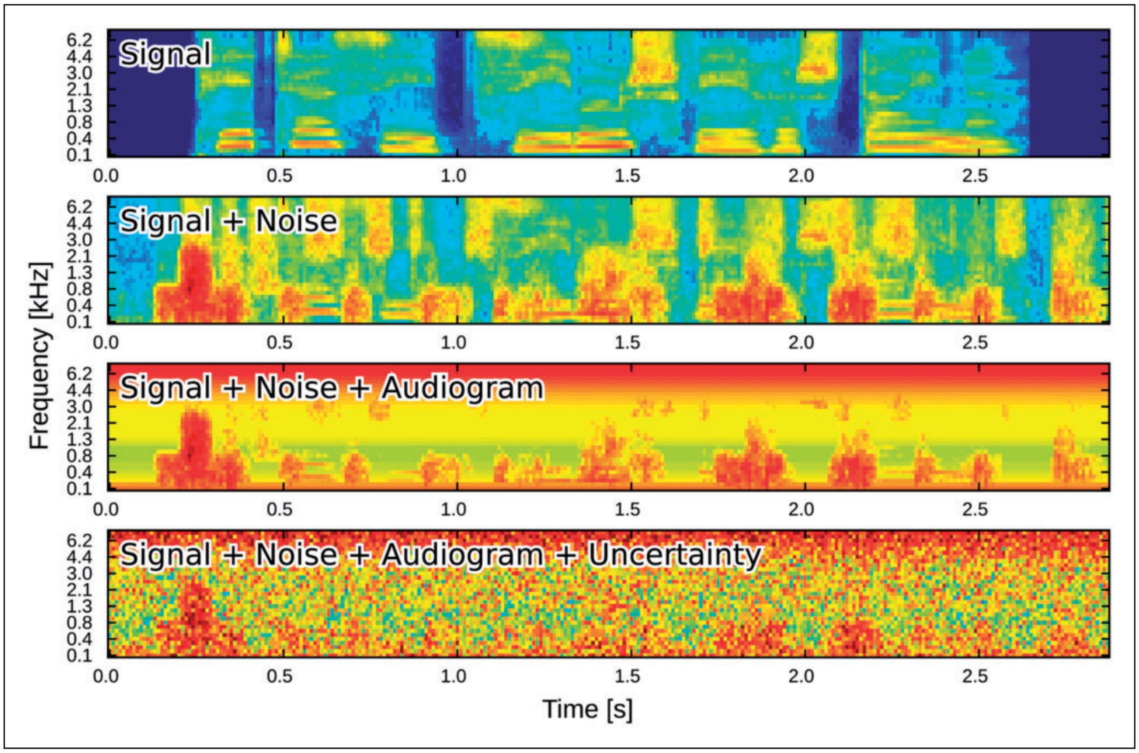


Figure 2. Example for the speech and noise (upper two panels) represented as log-scaled Mel-spectrogram (LogMS) with the modifications introduced by the thresholding procedure (third panel) to represent the individual audiogram and the level uncertainty (fourth panel) introduced to represent individual suprathreshold processing deficits. From these patterns, the Mel-frequency cepstral coefficients (MFCCs) were derived as the input to a Hidden Markov Model-based speech recognizer. In this example, the German matrix test speech sample "Peter sieht sieben schwere Steine" is presented at 75 dB SPL, and the fluctuating noise (ICRA5-250) is presented at 85 dB SPL. The simulated "typical" audiogram N4 was taken from Bisgaard et al. (2010). The value u_L has been set to 10 dB in this example.

an example), from which the MFCCs were derived as the input to a HMM-based speech recognizer. To account for the loss of sensitivity in hearing impairment usually assessed by the audiogram, the individual hearing threshold was applied to the spectro-temporal representation by setting any input spectrogram level to the audiogram level if it is below this value (see, for e.g., Figure 2, third panel). Moreover, to account for any suprathreshold distortion, an increase in the internal noise was considered (see Figure 2, last panel). It was implemented by adding an uncertainty to the representation of levels in the logMS, in the form of an additive Gaussian white noise with a standard deviation of u_L. The effect of an additive noise following a logarithmic compression would be roughly equivalent to a multiplicative noise applied to the input signals. It can be interpreted as central detector noise or individual internal noise due to, for example, insufficiency of the central, internal neural representation or lack of attention or cognitive processing abilities. By setting the hearing threshold and the level uncertainty parameter u_L appropriately, an individual Plomp curve can be generated for each individual hearing-impaired subject which fits the empirical data best.

The following research questions are addressed in this article:

1. Are the model predictions of the FADE approach for hearing-impaired listeners in stationary and fluctuating noise compatible with the A+D approach by Plomp (1986)? How does this compare to the traditional SRT prediction methods like the (extended) SII?
2. How should the individual parameters of the FADE model (i.e., the hearing threshold and the level uncertainty parameter u_L) be set in order to best predict the data (i.e., without using the data to be predicted as a priori knowledge and with a reasonable computational effort)? Is it sufficient to use "typical" parameter values interpolated across the 10 typical audiogram configurations from Bisgaard, Vlaming, and Dahlquist (2010) or is it necessary to set the parameters individually (preferably without computing an individual SRT for each individual audiogram and each value u_L for each background noise type and level)?
3. How does this approach of setting the hearing threshold and the level uncertainty parameter u_L using

FADE compare to traditional SRT prediction approaches where only the individual hearing threshold is accounted for?

To answer these questions, first general predictions were derived both for FADE and the (extended) SII based on typical audiogram shapes from Bisgaard et al. (2010). In a second step, individual predictions were derived and compared with SRT data using the German Matrix test in stationary and fluctuating noise obtained from Brand and Kollmeier (2002). This allows for a comparison across different methods to set the FADE model parameters individually (using a limited computational effort) and across different versions of the SII approach.

Methods

FADE Approach

The simulation FADE from Schädler et al. (2015, 2016) was used to simulate the outcome of the German Matrix test in a stationary and a fluctuating noise condition (cf. Schädler et al., 2015, for details). The speech material of the German Matrix test (see review by Kollmeier et al., 2015) consists of 120 semantically unpredictable sentences with a fixed syntax (name-verb-number-adjective-object, like “Peter sees eight wet chairs”). For each word class, 10 alternative words exist. For each of these words between 8 and 10 alternative recordings exist from which the sentences are constructed. The SRT denotes the SNR that corresponds to 50%-words-correct performance and is usually adaptively measured for human subjects. To obtain SRTs with FADE, an automatic speech recognizer was trained and tested with noisy sentences on a broad range of SNRs (–24 dB to +6 dB), and the lowest SNR which resulted in 50%-words-correct recognition performance was interpolated and used as the predicted SRT. As the front end for the ASR system, modified MFCCs were used. On the back end side, HMMs were used to model speech with whole-word models based on the “parametrically hearing-impaired” acoustical representation provided by the front end. Hearing impairment was modeled in the front end and was implemented in the logMS, from which the MFCC features were derived. A frequency-dependent attenuation was used to model the effect of the elevated threshold in quiet (i.e., the audiogram) by clipping the amplitude values in each channel to the corresponding (interpolated) threshold from the audiogram. To model a suprathreshold distortion component of hearing impairment, a level uncertainty was implemented in the logMS by adding a Gaussian white noise with a standard deviation of u_L.

To generate general predictions based on typical audiogram shapes, simulations were performed for different noise types and noise levels using the 10 typical audiogram configurations from Bisgaard et al. (2010) depicted in Figure 3 without any level uncertainty (i.e., $u_L = 0$). In addition, a normal-hearing audiogram and a systematic variation of u_L between 0 and 50 dB was employed. Plomp curves were fitted to these simulations to estimate the best-fitting A-, D-, and P-values, for each condition. This “database of Plomp curves” plotted in Figures 6(a), (b), and 7 is used for all further steps described later. Note that a complete FADE simulation with all combinations of possible audiograms and all possible values of u_L would have been desirable but was not performed here due to its excessive computational complexity. Also, no individual, iterative fitting of u_L could be performed. Instead, FADE simulations with the individual audiogram and $u_L = 0$ were performed, and an individual estimate of u_L was derived using the simulation results for the normal-hearing audiogram and different values of u_L (“individual” vs. “typical” distortion correction, see later).

Audiological Data

To derive “typical,” that is, nonindividualized predictions of the Plomp curves (i.e., SRT dependence on noise level for a given hearing loss), the 10 “typical” audiograms defined by Bisgaard et al. (2010)

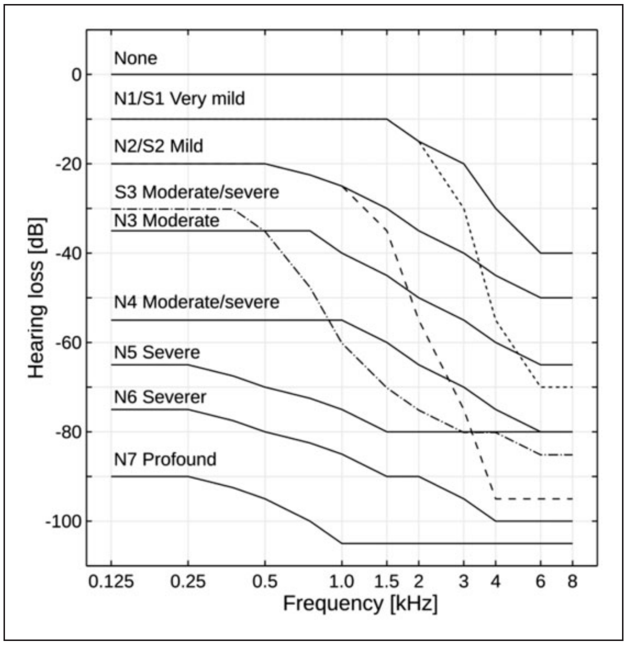


Figure 3. The 10 “typical” audiograms defined by Bisgaard et al. (2010) plotted in an audiogram representation (in dB HL) that are used to predict the Plomp curves in a nonindividualized manner both by the FADE approach and by the SII.

were employed that were originally derived by a vector quantization approach from 28,244 patient cases (cf., Figure 3).

The index parameter Nx/Sx = 1 ... 10 (which is a combined form of N1 ... N7 and S1 ... S3) is used in the following to address these 10 different classes primarily for interpolating the corresponding audiological data and predictions across adjacent audiogram shapes.

In a second step, individual audiometric and speech recognition data of 99 listeners (198 separately measured ears) ranging in age from 23 to 82 years (mean and standard deviation: 61.4 ± 13.2 years) from Brand and Kollmeier (2002) were employed. The patients cover a broad range of hearing loss with the pure tone average varying from 0 to 80 dB HL (mean and standard deviation: 40.5 ± 16.1 dB HL). SRTs were obtained with the German Matrix test in stationary ICRA1 and fluctuating ICRA5-250 noise (Dreschler, Verschuure, Ludvigsen, & Westermann, 2001). The ICRA5-250 noise is a speech-like modulated noise, which simulates the long-term frequency spectrum and modulation properties of a single male speaker with silent intervals limited to 250 ms (Wagener, 2004). An individual noise level was selected between 65 and 85 dB. It was set to the individual medium loudness level $L_{25\mathrm{cu}}$ measured using an adaptive categorical loudness scaling with the ICRA1 noise. The noise level was set to 85 dB if the individual $L_{25\mathrm{cu}}$ (i.e., the level corresponding to 25 categorical loudness units) was larger than 85 dB.

Individual Parameter Selection for the FADE Approach

Figure 4 sketches the *simulation* approach where an individual FADE simulation is performed using the individual audiogram for each subject and $u_L = 0$ to derive an SRT value for each noise type, and those noise levels which were also employed for the audiological data. Subsequently, either a "typical" or an "individual" distortion correction is performed by estimating the appropriate individual D-parameter of the Plomp curve from additional data.

- An "individual" estimation of the value of u_L can be performed by first computing the Plomp curve assuming no level uncertainty ($u_L = 0$) and then using the mismatch between this curve and the actual data for one given background noise condition to derive an individual u_L-estimate. This value can then be used as an input for the respective other noise condition to compute the Plomp curve and to derive the appropriate D-value for correcting the simulated Plomp curve for the individual patient. Hence, independent data for the same individual subject are employed to estimate the individual suprathreshold distortion component.
- For a "typical" estimate of D, the u_L and, successively, the noise-type-specific D-value belonging to the average of the most suitable group of audiograms is employed: For each of the 10 classes of audiograms from Bisgaard et al. (2010), first the deviation between

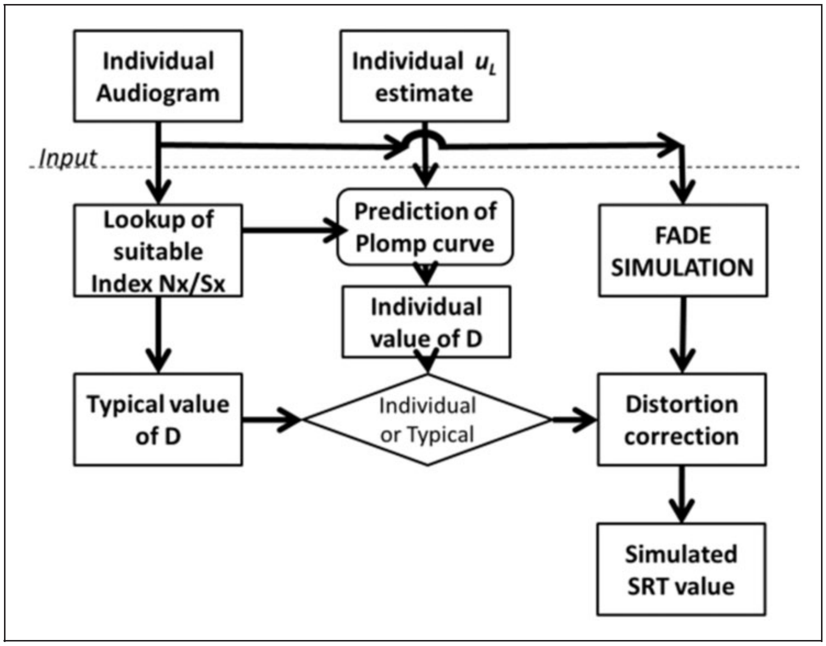


Figure 4. Flow diagram of simulating the individual speech recognition threshold (SRT) from the audiogram with FADE simulation and either an individual or a typical distortion correction (denoted by an a priori selection in the rhomboid box). It is performed by estimating the individual D-value using an interpolation across the index Nx/Sx = 1...10, i.e. the ten typical audiograms by Bisgaard et al. (2010).

prediction and empirical SRT (either for the stationary or for the fluctuating noise) is averaged across all those data sets that belong to the respective class. Second, a value of u_L is determined that provides least deviation between the FADE model output and the empirical averaged data. These estimated typical u_L-values are used to derive and interpolate the appropriate D-values to individually correct the simulated Plomp curve for the individual patient.

Figure 5 sketches the interpolation-based *prediction* approach where the individual audiogram is first approximated by an interpolation across the 10 "typical" audiograms defined by Bisgaard et al. (2010). To do so, the two nearest neighbors of the individual audiogram are looked up, and the value of the index Nx/Sx (ranging from 1 to 10) is interpolated to determine and interpolate across the corresponding Plomp curves from the database. Similar as earlier, the "typical" distortion correction estimate of the value of u_L is obtained and used to perform an individual interpolation to predict the most likely Plomp curve for the individual patient. Alternatively, an "individual" estimation of the value of u_L can be performed by first computing the Plomp curve assuming no level uncertainty ($u_L = 0$) and then using the mismatch between this curve and the actual data for the respective other background noise condition to derive an individual u_L-estimate. This value can then be used as an input to compute the Plomp curve.

Note that the prediction approach depicted in Figure 5 is less computational expensive than the simulation approach depicted in Figure 4, because a complete FADE simulation has only to be performed once for the limited set of Plomp curves in the database (10 audiograms plus several values of u_L), whereas the remainder of the prediction process is straightforward interpolations that do not require much computational effort.

SII Predictions

To compare the nonindividualized FADE simulations based on the typical audiogram shapes with predictions from the standard SII model (ANSI, 1997), an SII prediction of the Plomp curves was also performed using the 10 typical audiograms from Bisgaard et al. (2010). For the stationary noise case, the original SII model was employed, whereas for the fluctuating noise case, the time-dependent eSII-model by Rhebergen and Versfeld (2005) without masking was employed.

For the individualized SII predictions, again the same audiological data were employed as earlier and the three extensions of the SII for predicting SRT in stationary and fluctuating noise methods as described by Meyer and Brand (2013). They used the same noise conditions and a group of 113 listeners (of whom the 99 listeners

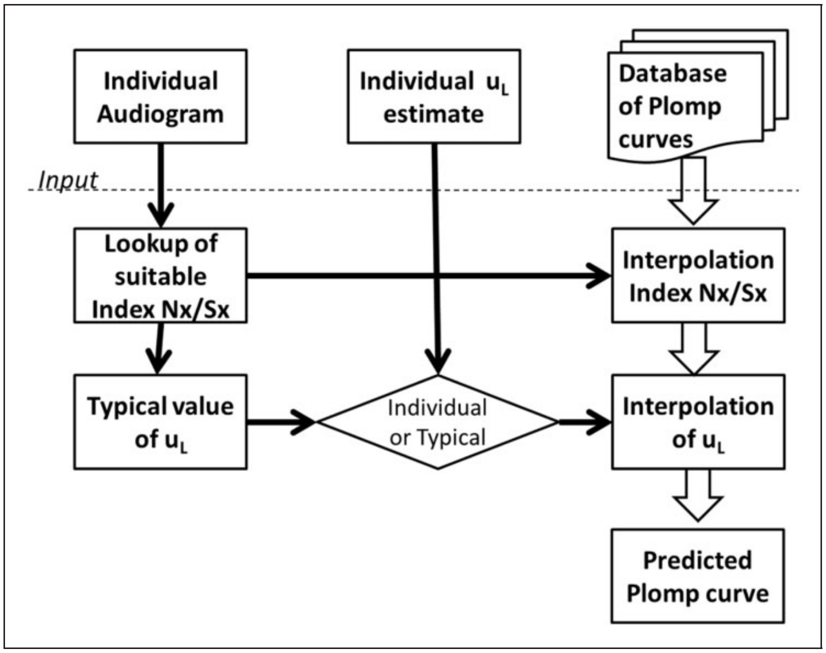


Figure 5. Flow diagram of predicting the individual Plomp curve from the audiogram and either an individual or a typical estimate of the level uncertainty parameter u_L. As a further input, the database of typical Plomp curves is provided generated by the FADE approach using the 10 typical audiograms by Bisgaard et al. (2010, characterized by their index Nx/Sx = 1 ... 10). Note that the computationally expensive individual FADE simulation from Figure 4 is replaced by an interpolation across a database, and that distortion correction is based on the u_L lookup rather than on the fitted D-parameters from the Plomp curve.

considered here are a subgroup) for comparing four SII versions: (a) original SII (ANSI, 1997) which is only considering the long-term frequency spectra of speech and noise, (b) considering frequency-independent level fluctuation of the noise (the SII is calculated using a fixed frequency spectrum of the speech and a fixed shape of the frequency spectrum of the noise, the level fluctuations of the noise are considered using 30 ms time frames and averaging the resulting SII values across frames), (c) considering frequency-dependent level fluctuations of the noise similar to Rhebergen et al. 2010 (like version B, however, the noise level fluctuations are considered independently for each frequency band of the SII), and (d) considering frequency-dependent fluctuations of the speech and the noise (like version C, however, the fluctuations are considered also for the speech). Note that all versions included the same basic calculations as the original SII, that is, the effect of spread of masking on speech intelligibility is accounted for.

Results

General Predictions Based on Typical Audiogram Shapes

(a) FADE simulations without suprathreshold distortion: Figure 6(a) and (b) shows the simulated SRTs for the 10 typical audiograms (N1 ... N7, S1 ... S3) defined by Bisgaard et al. (2010) as a function of level of the stationary noise (solid lines) and the fluctuating ICRA5-250 noise (dashed lines). In general, the curves follow well the general expected shape of the curves according to Plomp (1978) (see earlier).

The A-, D-, and P- values fitted to the simulated curves using the Plomp (1978) formula for the different typical audiograms are given in the insert tables in Figure 6.

Note that most of the variation across the typical audiograms are captured by the variation in the "Attenuation" parameter, whereas only the more severe hearing losses require an additional "Distortion" parameter which also reflects some deviation of the audiogram shape from the standard speech spectrum.

(b) FADE simulations for different levels of level uncertainty: Figure 7 displays the simulated SRT using the FADE approach for a normal-hearing listener with a set of fixed "level uncertainty parameter" u_L-values (between 0 and 50 dB) in order to model an increasing amount of suprathreshold distortions. Note that the curves exhibit a parallel shift to higher SRT values with increasing parameter u_L which is very similar to the effect of the D-parameter of the Plomp model. However, an increase by 10 dB in the level uncertainty parameter u_L does not translate directly into an equally spaced increase of the D-parameter fitted to the curves in Figure 7 (see inlaid table in Figure 7): At low and high u_L-values, the largest resulting difference in D for a 10-dB step in u_L is observed, whereas in the midrange the simulations exhibit a higher robustness against an increase in level uncertainty.

(c) SII predictions: Figure 6(c) and (d) shows the SII-predicted SRTs for stationary noise (solid lines) and the eSII-predicted SRTs for fluctuating noise (dashed lines) for the 10 typical audiograms (N1 ... N7, S1 ... S3) plotted in a similar way as for the FADE approach. Note that curves are similar to the corresponding FADE simulations (Panels (a) and (b)) and follow as well the general expected shape of the curves according to Plomp (1978).

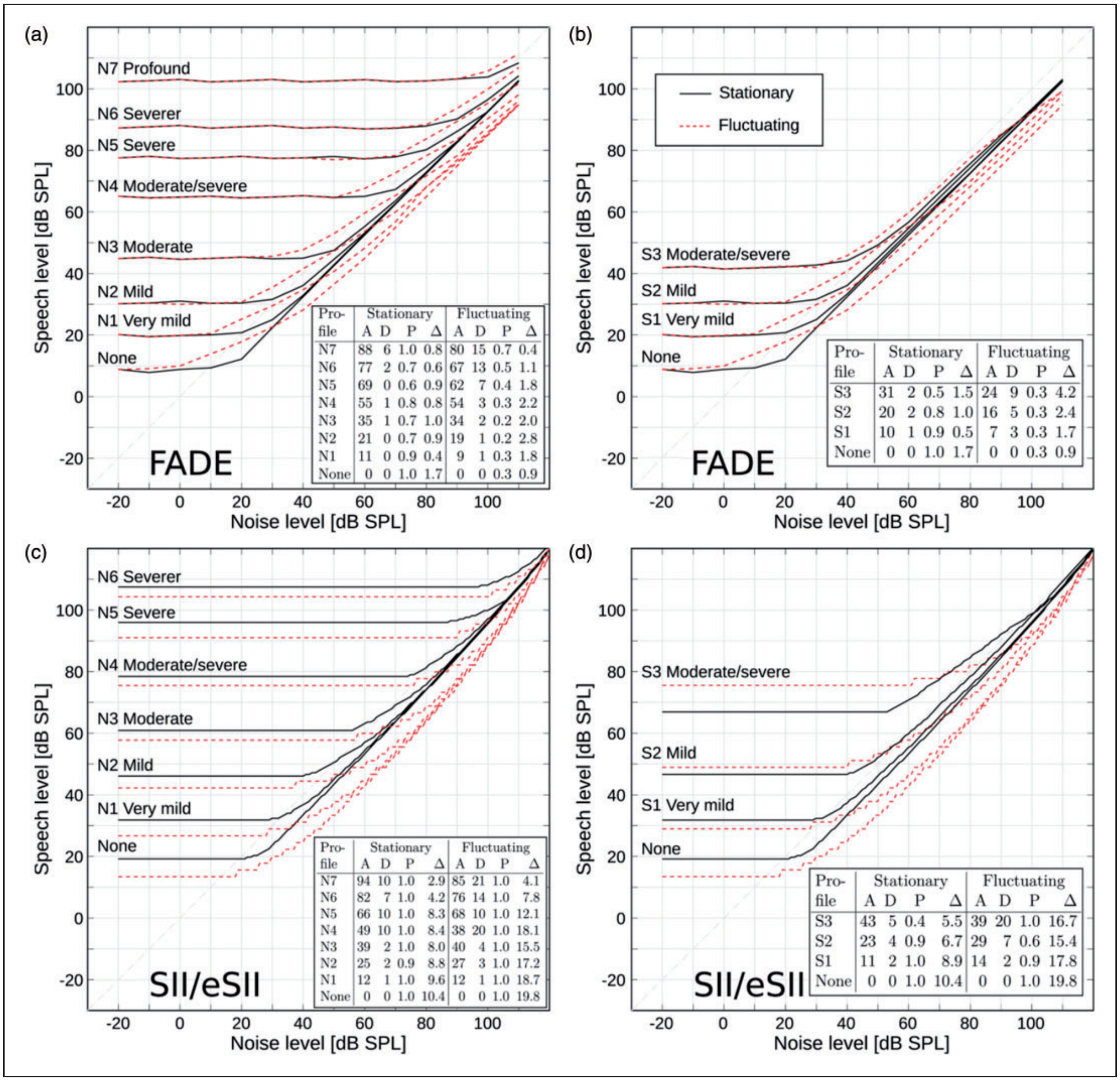
However, the SRT in quiet (reflected by the A-values fitted to the curves) are higher than predicted with the FADE approach and do not coincide between the SII and the eSII approach which is due to different reference data to be used for both approaches while FADE does not require such a reference or calibration curve (see Discussion section).

Moreover, with increasing level of the background noise the SII-curves tend to no longer follow a diagonal line but deviate toward higher SRT values which reflects the "level distortion factor" at higher levels which is included in the SII (ANSI, 1997). This behavior differs from the assumed diagonal line shape of the Plomp curves. As a result, the D-values of the Plomp curves fitted to the SII predictions are increased, and the deviation parameter Δ-indicate a lesser goodness-of-fit in comparison to the FADE approach for the same audiogram class. A similar relative increase in D-values is observed for the Plomp curves fitted to the SII predictions with increasing hearing loss—especially for the three classes of sloping loss in comparison to the nonsloping loss classes with a comparable average loss. This reflects the tendency that an audibility loss at high frequencies leads to a fitted Plomp curve with a high D-factor even though its alteration is more related to a lack in audibility and not necessarily related to a suprathreshold deficit as discussed by Lee and Humes (1993).

However, besides these general factors that produce an increase in the fitted D-value as a fixed consequence of the respective audiogram, the SII model does not offer another independent variable to reflect the suprathreshold distortion component in a way comparable to the parameter u_L from the FADE approach.

Individual Predictions

(a) Individualized FADE predictions and simulations for stationary noise: Figure 8 displays the predictions and simulations for the individual SRT in stationary ICRA1-noise for an increasing degree of individualization. The SRT predictions (black dots) obtained by



Pro- file	Stationary A	 D	 P	 Δ	Fluctuating A	 D	 P	 Δ
N7	88	6	1.0	0.8	80	15	0.7	0.4
N6	77	2	0.7	0.6	67	13	0.5	1.1
N5	69	0	0.6	0.9	62	7	0.4	1.8
N4	55	1	0.8	0.8	54	3	0.3	2.2
N3	35	1	0.7	1.0	34	2	0.2	2.0
N2	21	0	0.7	0.9	19	1	0.2	2.8
N1	11	0	0.9	0.4	9	1	0.3	1.8
None	0	0	1.0	1.7	0	0	0.3	0.9

Pro- file	Stationary A	 D	 P	 Δ	Fluctuating A	 D	 P	 Δ
S3	31	2	0.5	1.5	24	9	0.3	4.2
S2	20	2	0.8	1.0	16	5	0.3	2.4
S1	10	1	0.9	0.5	7	3	0.3	1.7
None	0	0	1.0	1.7	0	0	0.3	0.9

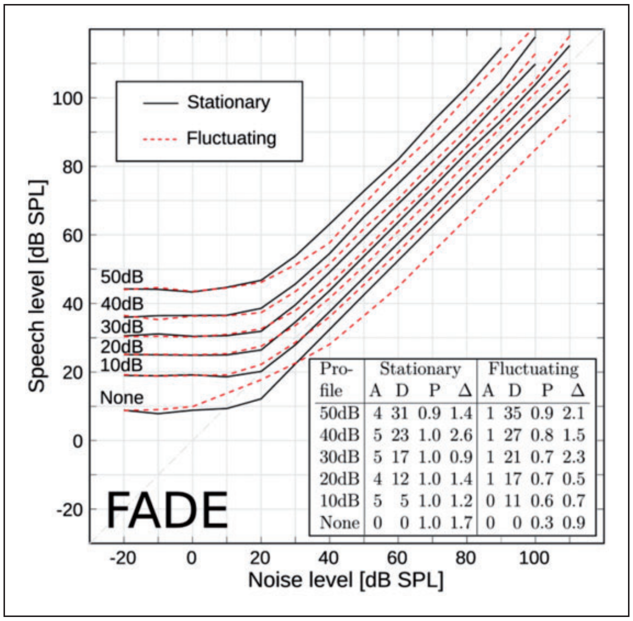
Pro- file	Stationary A	 D	 P	 Δ	Fluctuating A	 D	 P	 Δ
N7	94	10	1.0	2.9	85	21	1.0	4.1
N6	82	7	1.0	4.2	76	14	1.0	7.8
N5	66	10	1.0	8.3	68	10	1.0	12.1
N4	49	10	1.0	8.4	38	20	1.0	18.1
N3	39	2	1.0	8.0	40	4	1.0	15.5
N2	25	2	0.9	8.8	27	3	1.0	17.2
N1	12	1	1.0	9.6	12	1	1.0	18.7
None	0	0	1.0	10.4	0	0	1.0	19.8

Pro- file	Stationary A	 D	 P	 Δ	Fluctuating A	 D	 P	 Δ
S3	43	5	0.4	5.5	39	20	1.0	16.7
S2	23	4	0.9	6.7	29	7	0.6	15.4
S1	11	2	1.0	8.9	14	2	0.9	17.8
None	0	0	1.0	10.4	0	0	1.0	19.8

Figure 6. Speech recognition thresholds (SRT) as a function of the noise level for the German matrix sentence test in the test-specific, stationary noise condition (solid lines) and for the fluctuating ICRA5-250 noise (dashed lines) simulated by the FADE approach (Panels (a) and (b)) and by the (extended) SII model (Panels (c) and (d)). The curves correspond to different grades of hearing impairment based on the 10 standard audiograms (N1 . . . N7 in Panels (a) and (c), and S1 . . . S3 in Panels (b) and (d)) from Bisgaard et al. (2010) displayed in Figure 3. The embedded tables report the best fitting A and D parameters (in dB) and the power coefficient P of the best-fitting Plomp curves as well as the maximum deviation between the fitted curve and the data (denoted as Δ in dB).

interpolating across the 10 prototype audiograms (according to Figure 5) are plotted against the empirical values (given on the x-axis). For comparison, the individualized FADE simulations (according to Figure 4) are given as gray symbols using the individual audiogram. The connection lines between the predicted values (that require only a very small computational load) and the simulated values (that are computationally more expensive) indicate already a high coincidence in SRT prediction between both methods.

Panel (a) denotes the predictions based on the audiogram alone. Note that neither method is able to model the empirical SRT in stationary noise in a satisfactory way since the large spread in the empirical data (ranging from −9 to + 7 dB in SNR) is not reflected in the predictions based on the audiogram alone. The correlation coefficient (Pearson's R^2 with 95% confidence intervals) between data and predictions or simulations is provided in Table 1. A somewhat (but not significantly) higher correlation between data and predictions is achieved for the "typical" distortion correction method employing the distortion correction either from the stationary noise (Figure 8(b)) or the fluctuating noise (Figure 8(c)) than for the predictions based on the audiogram alone. The highest correlation (which significantly exceeds the correlation for all the remaining cases) is achieved for the individual distortion correction (Figure 8(d)) where for each patient the deviation between the prediction and the data for the fluctuating noise is used to correct for the distortions to be expected in the stationary noise case considered here.



Pro- file	Stationary A	 D	 P	 Δ	Fluctuating A	 D	 P	 Δ
50dB	4	31	0.9	1.4	1	35	0.9	2.1
40dB	5	23	1.0	2.6	1	27	0.8	1.5
30dB	5	17	1.0	0.9	1	21	0.7	2.3
20dB	4	12	1.0	1.4	1	17	0.7	0.5
10dB	5	5	1.0	1.2	0	11	0.6	0.7
None	0	0	1.0	1.7	0	0	0.3	0.9

Figure 7. Speech recognition thresholds (SRT) for a normal-hearing listener with different values of level uncertainty u_L in the test-specific, stationary noise condition as a function of the noise level from simulations with FADE. The dashed lines show the same results for the fluctuating ICRA5-250 noise. The embedded table reports the best-fitting parameters for the fitted Plomp curves according to Equation (2).

(b) Individualized FADE predictions and simulations for fluctuating noise: The same representation as in Figure 8 for the stationary noise is presented in Figure 9 for the fluctuating noise. Table 1 shows the correlation coefficients (Pearson's R^2) between modeled SRTs for the fluctuating noise and the empirical data (last column). Again, with increasing degree of individualization, the correlation between data and predictions or simulations is improved without exhibiting significant differences between the cases without distortion correction and with the two typical distortion corrections. However, a significant increase in correlation between data and model predictions for the fluctuating noise case is reached for the individualized distortion correction obtained from the stationary noise data (displayed in Figure 9(d)).

Note that the R^2 scores for the stationary noise case reported here for the four versions of the SII prediction methods from Meyer and Brand (2013) do not deviate significantly from most of the R^2 values reported for the different FADE predictions and simulations with typical distortion corrections. The SII versions A to C provide even a higher prediction accuracy than the FADE without distortion correction. Only the FADE simulations with individual, fluctuating-noise-based distortion correction provide a significantly higher prediction accuracy than all versions of the SII.

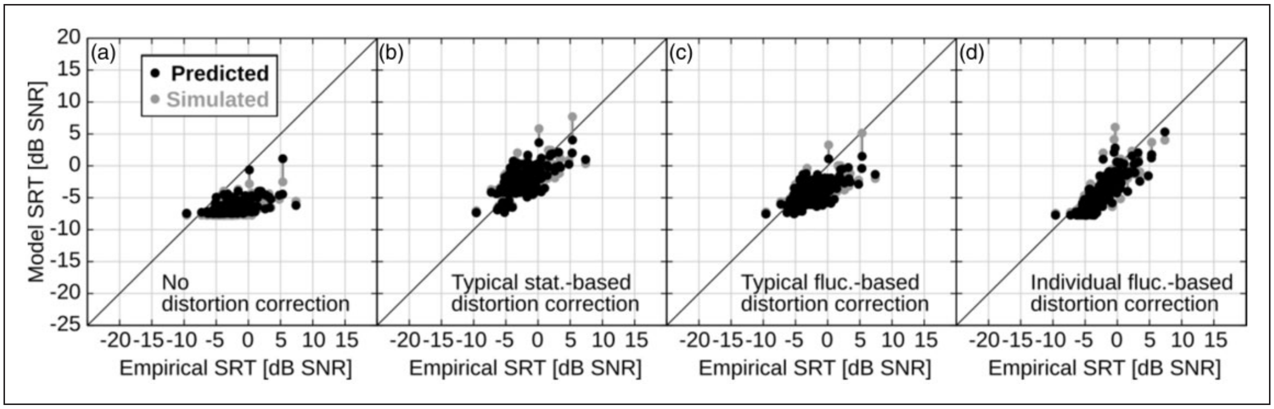


Figure 8. Modeled SRT for 198 ears from 99 subjects plotted against the empirical data (x-axis) for *stationary noise*. The predicted SRTs (according to Figure 5, employing typical audiograms) are given as black dots, the simulated data (according to Figure 4, employing the individual audiogram) are given as grey dots. The simulations are performed for different degree of distortion correction: No distortion correction (Panel (a)), stationary-noise-, typical-audiogram-based distortion correction (Panel (b)), fluctuating-noise-, typical-audiogram-based distortion correction (Panel (c)), and fluctuating noise-based, individual distortion correction (Panel (d)).

Table 1. Statistical Analysis of the Predicted or Simulated SRT.

		Stationary noise				Fluctuating noise			
Model	Distortion correction	R^2	Interval	B	RMS	R^2	Interval	B	RMS
FADE prediction	None	.31	[.21 .42]	–4.1	4.6	.48	[.38 .58]	–4.6	6.1
	Typical stationary based	.44	[.33 .54]	0.0	1.9	.57	[.47 .65]	3.7	5.4
	Typical fluctuation based	.42	[.31 .52]	−1.9	2.7	.56	[.46 .64]	0.0	3.8
	Individual stationary based	–	–	–	–	.78	[.72 .83]	3.4	4.3
	Individual fluctuation based	.63	[.54 .71]	−1.6	2.3	–	–	–	–
FADE simulation	None	.36	[.25 .46]	–4.3	4.7	.57	[.47 .65]	–4.5	5.9
	Typical stationary based	.49	[.38 .58]	–0.1	1.8	.63	[.55 .71]	3.8	5.2
	Typical fluctuation based	.46	[.35 .56]	–2.0	2.7	.63	[.54 .70]	0.1	3.5
	Individual stationary based	–	–	–	–	.83	[.78 .87]	3.8	4.5
	Individual fluctuation based	.70	[.62 .76]	−1.9	2.4	–	–	–	–
SII version A		.55	–	–	–	.24	–	–	–
SII version B		.59	–	–	–	.42	–	–	–
SII version C		.51	–	–	–	.42	–	–	–
SII version D		.35	–	–	–	.52	–	–	–

Note. Pearson's correlation coefficients (R^2) are reported (including their 95% confidence intervals according to Fisher, 1958) along with the RMS prediction error and the bias (B) for predicted or simulated SRTs with different distortion correction methods and SII-based predictions from Meyer and Brand (2013). SRT = speech recognition threshold; RMS = root-mean-square; FADE = Framework for Auditory Discrimination Experiments; SII = Speech Intelligibility Index.

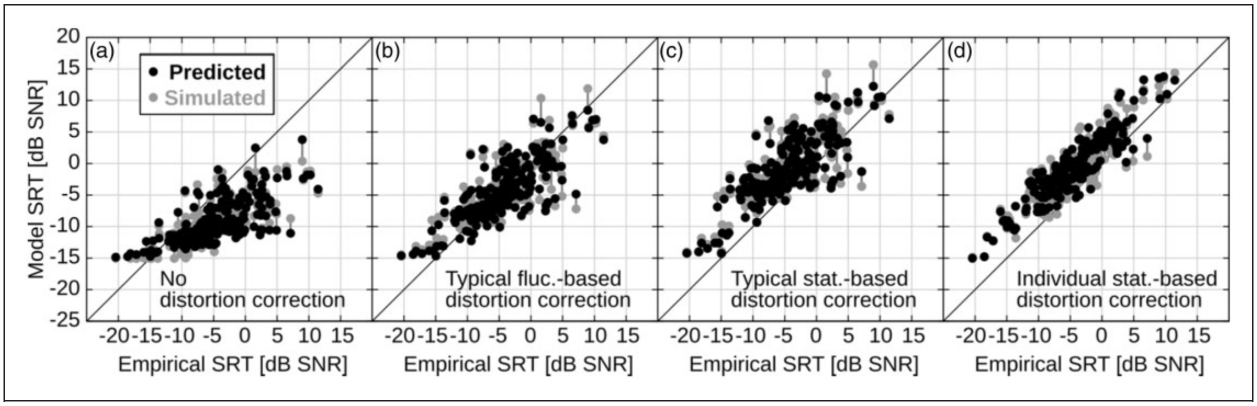


Figure 9. Modeled SRT for 198 ears from 99 subjects plotted against the empirical data (x-axis) similar to Figure 8 but for *fluctuating noise*. SRT = speech recognition thresholds.

For the fluctuating noise case, the SII-based correlations are generally lower or their R^2 deviate not significantly from the FADE predictions without distortion correction. With the exception of SII version D, all FADE predictions and simulations with distortion correction exhibit a significantly higher prediction accuracy. The FADE simulations with all kinds of distortion corrections even surpass SII version D with respect to its R^2 value.

Taken together this indicates that the FADE simulations with individualized distortion corrections provide superior prediction accuracy against all other FADE approaches and all SII versions considered here. In the stationary noise case, FADE with distortion correction and SII perform approximately equally well, whereas in the fluctuating noise case the SII predictions perform generally worse than the FADE approach (with some exceptions).

Discussion

A concept based on the broadly applicable FADE approach has been introduced here for individualized SRT prediction in noise for hearing-impaired listeners. The nonindividualized, typical-audiogram-based version was demonstrated to predict well the curves fitted with the A + D approach from Plomp (1986) using the model parameter u_L to model the apparent suprathreshold processing deficits (which is also reflected by the fitted D-component) independently from the audiogram-data-driven audibility component reflected by the fitted A-component. This positively answers research question 1 from the introduction. In comparison to the SII model and its modifications, the independent control of the model parameter u_L appears to be advantageous.

Second, the FADE model has been evaluated with data from the literature for 198 ears in comparison to the SII and modified SII. The individualization of SRT prediction is based either on simulations using the individual audiogram which requires more computational effort than an interpolation approach utilizing the "typical" audiograms provided by Bisgaard et al. (2010) to create a precomputed database. These two versions do not differ significantly in their prediction accuracy which is also comparable to the accuracy achieved with appropriately modified SII model versions for the stationary noise case. In the fluctuating noise condition, most versions of the FADE approach outperform the modified SII model versions.

Moreover, the possibility of a "typical" or "individual" distortion correction is explored. The correction is based on the assumption of an individually tailored central noise which may be interpreted as level uncertainty or multiplicative, random detector noise. The individual amplitude u_L of this noise can be estimated from the difference between prediction and actual data for typical audiograms or from a different noise condition than the one under test. This answers research question 2 from the introduction by providing different ways to set the individual parameters of the FADE model with a reasonable computational effort and a high prediction accuracy.

Third, using the FADE approach with an individual audiogram and an individual distortion correction outperforms both the SII prediction accuracy and the performance for all other individualization approaches for the FADE model both for the stationary and fluctuating noise case, thus answering research question 3 from the introduction. Obviously, the highest prediction accuracy is achieved if not typical parameter sets, but individualized audiogram and u_L-values are employed.

Advantages of the FADE Approach Presented Here

- It is based on a reference-free simulation principle that primarily relies on basic principles employed in machine learning and ASR. Hence, this approach does not require any normative data or reference speech intelligibility curves (as required, e.g., by the SII or related SNR-derived estimates of speech intelligibility). Moreover, the FADE approach uses only a mixture of signal and noise at several SNRs for training and replaces the optimum detector used in the "effective" psychoacoustic processing models (like PEMO [Dau et al., 1997] and CASP [Jepsen, Ewert, & Dau, 2008]) by a HMM Detector which allows for local stretching and compressing in time. Hence, the FADE approach does not require the exact a priori knowledge of the desired signal and the background noise (as required by the "effective" psychoacoustic models). Instead, it derives an abstraction of this a priori knowledge by training of the HMM and by deciding which training condition provides the lowest threshold estimate. This still assumes a certain amount of a priori knowledge but makes the recognition back end more versatile than a fixed cross-correlation-based optimum detector.
- As already outlined by Schädler et al. (2015), the same basic processing approach can be used to predict both the outcome of speech recognition data and psychoacoustic discrimination experiments using the same processing frontend and back end. Hence, the simultaneous prediction of psychoacoustic data and speech recognition data with hearing-impaired listeners using the same parameter set for individualization may become possible. This yields a high flexibility for inserting the putative causes of a hearing impairment into the prediction process both of speech recognition and psychoacoustic data. Currently, the simplest assumption was employed which utilizes a thresholding procedure to account for the loss in sensitivity and the addition of a level uncertainty at a central stage. Such a level uncertainty on a logarithmically compressed internal representation stage can be interpreted as an "effective multiplicative" noise if projected back on the linear, peripheral signal processing stage. Such a multiplicative noise is a common, most simple "Ansatz" for a central detector noise or individual internal noise due to, for example, insufficiency of the central, internal neural representation or the lack of attention or cognitive processing abilities.
- The approach offers a way to investigate the mechanisms involved in speech recognition in fluctuating noise for normal hearing and hearing-impaired listeners which has been one of the unsolved problems in audiology since many years (Dreschler et al., 2001; Festen & Plomp, 1990; George et al., 2006; Meyer & Brand, 2013; Wagener et al., 2006). While even the most sophisticated version D of the modified SII from Meyer and Brand (2013)—using a time- and frequency-dependent SNR measure to derive the SRT estimate—provides only a moderate prediction accuracy (Pearson's R^2 of .52, see Table 1)—which is comparable to the best performing SII version B for the stationary noise case (R^2 of .59)—it is already outperformed by the FADE approach with a comparable

simple set of assumptions (e.g., FADE simulation with the individual audiogram and typical distortion correction). This provides evidence that the model structure underlying the FADE approach is more appropriate than even the most sophisticated SII version D to correctly model speech recognition in fluctuating noise (see later). A more detailed analysis appears necessary to pinpoint the exact element of the FADE model which provides this advantage.

- The approach to combine the predictions from the FADE model with the parameter fitting of the $A + D$ model by Plomp has the advantage of limiting the computational load for the FADE simulations as not every combination of audiogram and u_L has to be computed. Instead, the Plomp curves are utilized as a patient-specific interpolation method to compute the SRT for each possible noise level characterized by only few fitting parameters (A, D, and P) that are individually determined by only few empirical data points and predicted by only few model simulations. The elegance of this approach is highlighted by the fact that the FADE simulations for a given audiogram produce SRT values that are fit very well with the Plomp curves (as can be seen by the low Δ– values in Figures 6(a), (b), and 7) and that changes in the FADE model parameter u_L does only produce a change in the D-value of the fit without much affecting the A- and P-values—at least for the normal-hearing listener case displayed in Figure 7. The basic assumption for the different individualization methods considered here is that this independence also holds for complete simulations of hearing impairment with FADE. This is supposed to be less the case for the fluctuating noise in comparison to the stationary noise conditions and will have to be checked by more extensive computations in the future.
- Building on this independence assumption, the FADE prediction and individualization concept introduced here yields the advantage that only very little computational costs are required to perform the predictions. It utilizes the "typical," vector-quantized-audiograms provided by Bisgaard et al. (2010) to firstly compute a limited database of only few computational expensive ASR modeling results and then combines them with an interpolation method utilizing the general relations provided by the Plomp curves. This makes the prediction method very easy to use for practical applications. Both the "typical" and the "individual" distortion correction methods also require little computational effort. However, the latter method provides the additional prediction accuracy at the cost of requiring an independent set of measurements for a second noise condition. Since the improvements in prediction accuracy by using the complete, individual-audiogram-based simulation (when employing the same respective distortion correction method) is only very small and statistically not significant, this added effort in individual modeling should only be performed if sufficient computational resources are available.

Comparison of the FADE Approach to SII

The concepts underlying SII and FADE differ considerably: SII considers spectrally (and in its modifications also temporally) resolved SNR estimates that are transformed with a reference curve to average human speech recognition with the respective speech material, thus automatically taking into account many factors involved in speech perception without explicitly modeling them. FADE, on the other hand, explicitly models human speech recognition as a pattern recognition process with a certain internal representation of the speech material obtained from ASR (like the MFCC features used here) or from auditory "effective" signal processing models (Schädler et al., 2016) and a recognizer back end which has evolved from progress in ASR research (like the HMM used here). The training procedure and the test setup employed aims at predicting not the average, but the best possible performance under these circumstances. Given the complexity of the task to be simulated—especially for individual hearing-impaired patients—it is surprising that such a simple approach performs so well and satisfying that it performs about equally with the SII approach for a number of versions provided in Table 1. However, a few differences exist that need attention:

- Predictions at low noise level (thresholds in quiet): The predicted thresholds for a given typical audiogram (Figure 6) are higher for the SII than for FADE and do not coincide for the SII and the eSII predictions (which is due to a different calibration process or different reference curves for the latter). The low thresholds for the FADE—which are about 10 dB lower than the average for listeners with normal hearing, may partially result from the high variability in absolute thresholds even for normal listeners (amounting to at least 10 dB) and the property of FADE to predict the lowest achievable threshold. Moreover, SII tends to overestimate the effect of the hearing loss at high frequencies which leads to an increasingly higher difference to the FADE simulations especially for the classes of sloping hearing loss. This may be due to the approximately uniform weighting across frequency channels in the main speech frequency range by the SII whereas sentence materials—as employed here—tend to be better predicted with higher weighting of lower frequency channels. One advantage of FADE is that it does not require such a frequency weighting but rather learns the salience of different speech cues available above the audiogram-based threshold from the statistical training procedure. This might be a more adequate model of human perception than a predetermined frequency weighting resulting from a compromise between different speech materials as used in the SII.

- Predictions at very high noise levels (above 80 dB SPL): The predictions by SII and eSII do not converge toward a constant SNR (as with FADE and as assumed by the Plomp curves), but rather show a slight increase in SRT with increasing noise level which appears to model human behavior appropriately (level distortion factor due to, e.g., larger spread of masking at high levels). A more realistic, level-dependent front end processing would be required for FADE to be more accurate at these high levels.
- Transition region (40–80 dB SPL): FADE consistently predicts lower thresholds for stationary noise (which is either below or barely above the absolute threshold) than for fluctuating noise which occasionally exceeds the audibility threshold and hence disturbs speech cue detection more than in the stationary case. At higher noise levels, the situation reverses, that is, the SRT for fluctuating noise is well below the corresponding SRT for stationary noise at the same noise level. This expected intersection between the Plomp curves for fluctuating and stationary noise can be observed for the FADE predictions in Figure 6(a) and (b) but not for the SII and eSII in Figure 6(c) and (d), respectively. One reason for this deviation from expectation for the SII is the different SRT predictions in quiet for SII (used for the stationary noises) and eSII according to Rhebergen and Versfeld (2005, used for the fluctuating noise, see earlier). However, even if this effect would be compensated for by an appropriate vertical shift of the eSII-based curves for modulated noise, still no intersection with the curves for stationary noise would result because the knee point of the curves is located at much higher noise levels than for the FADE approach. Apparently the eSII version from Rhebergen and Versfeld (2005) misses the negative effect of the short intervals where the fluctuating noise exceeds the stationary threshold level. This is to a lesser degree the case for the SII version D by Meyer and Brand and not the case for the FADE approach.
- It is unclear why the FADE approach (without the individual distortion correction which is not available for the SII and eSII) works so much better in the case of the fluctuating noise and relatively poorly in the stationary noise case as compared with the best SII version (cf. Table 1). One reason is that the SII was especially constructed and appropriately calibrated to predict the effect of stationary masking noise as correct as possible whereas the FADE approach has a much harder problem to solve right from the start by having to predict the correct range of SRT based on simple processing principles. Hence, it is not surprising that FADE does not outperform the SII for those prediction tasks for which it is best at.
- Even though FADE predictions and simulations with individual distortion correction outperform the respective SII and eSII predictions (cf. Table 1), this is not a fair comparison because extra individual information has been added to the FADE predictions from a separate measurement that is not accessible to the SII. In theory, the SII could be modified as well by adding a simple individual bias (to be estimated from the respective other noise condition). Such a procedure has been suggested by Brand and Kollmeier (2002) and by Rhebergen et al. (2010) in a similar way and is expected to increase the prediction accuracy of the SII based on a pure heuristic approach. More work should be invested to provide the SII and its modifications with some of the properties of the current FADE approach and to perform an unbiased, fair comparison with an equal amount of information being released to all model variations.

Limitations of the Approach Presented Here

Some serious *limitations* have to be kept in mind when applying the method:

- The FADE approach shows a bias which clearly underestimates the individual thresholds for the version without distortion correction (amounting to a bias of −4.1 to −4.6 dB, cf. Table 1) and for the stationary noise using fluctuating noise-based distortion corrections (bias between −1.6 and −2.0 dB). This can partially be attributed to the property of FADE to model the best possible performance rather than the average across human listeners. According to expectations, this bias is removed if a “typical” distortion correction based on the average performance in the same task for the group of listeners within the same audiogram class is performed. Similarly, FADE overpredicts the threshold for the fluctuating noise condition if a distortion correction from the typical or individual performance in stationary noise is performed (bias between 3.4 and 3.8 dB), that is, performance of the individual subjects is better than predicted. This overcorrection of the bias based on the stationary noise data (in combination with an undercorrection of the bias for the stationary noise predictions based on the fluctuating noise data) may be a consequence of the wider range of SRT values across listeners for the fluctuating noise case (approx. 30 dB, see Figure 9) as opposed to the range of SRT values for the stationary noise (approx. 18 dB).

- The model assumptions employed here to model the effect of sensorineural hearing impairment are only very basic: The individual audiogram is simply represented by a frequency-specific thresholding procedure (i.e., setting the output to a constant value if the input does not exceed this predefined threshold). Moreover, the limitations of the central human sound recognition process are simply modeled by a noise representing a certain level uncertainty or central blurring of the internal representation of the input signal. This representation is very simplistic and does not model all aspects of human speech perception and signal processing in the impaired auditory system as known from physiology and psychoacoustics. For example, the loss of compression due to malfunction of the outer hair cells is ignored as well as the loss of temporal fine structure or other signal representation at the brainstem level as produced by a putative loss of inner hair cells. Any deterioration of binaural interaction is also ignored. No differentiation between a central "detector-degradation-simulating" noise as opposed to a more peripherally located noise is performed which could indicate any deterioration in signal processing at the auditory nerve and brainstem level. More physiology-inspired front ends than the MFCC features employed here would be required to adequately represent a more sophisticated modeling of the auditory periphery.
- The listeners considered in the study have a large age range (23–82 years) leading to the assumption that besides the audiogram further suprathreshold processing deficits as well as cognitive factors play a significant role that covary with age as expected from the literature (e.g., Patterson, Nimmo-Smith, Weber, & Milroy, 1982). The limited success of the "typical" distortion correction (where all audiograms for a certain class were considered despite the age range) may be due to this simplification of not considering age and specific cognitive factors. On the other hand, the significant improvement in prediction accuracy by the individual distortion correction may be due to this large variation in individual factors in our patients not covered by the audiogram. It has to be noted though, that a large remaining variance across subjects remains even after individual distortion correction. This hints to the fact that our simple model approach with a central level uncertainty characterized by only one parameter u_L is not sufficient to completely capture the multidimensional effect of ageing and hearing impairment on speech recognition in noise.
- The "individual" distortion correction method proposed here utilizes information from a different speech recognition experiment in noise in order to set the individual distortion correction. Even though this eventually leads to a superior performance of the model in comparison to the SII, this utilization of additional data from a different experiment may be considered as an unfair advantage of the current approach. However, the underlying model assumption is that a single, individual distortion correction is universally applicable to several speech recognition and psychoacoustic experiments with the same subject—an assumption that still has to be tested using data from more experiments.

Potential Applications of the Approach Presented Here

If one accepts these limitations and even more practical limitations of the approach presented here (e.g., computational complexity of the FADE model, sophisticated mixture of a priori computation, interpolation, and comparatively imprecise estimation of threshold and distortion components), it nevertheless may show the following potential applications in the future:

- Even though the ASR method employed here was especially tailored to the German Matrix test OLSA (Wagener, Kühnel, & Kollmeier, 1999), the same general method can be used for all closed-set sentence recognition tests that follow the same format. This closed-set format has the advantage of being able to test a patient in his or her own language by touching the appropriate words on a response device thus eliminating the requirement for the test conductor to understand the language and to assess the correctness of the (verbal) response. Fortunately, 16 of the matrix tests already exist to date in different languages (review by Kollmeier et al., 2015, recommendations for development a test in a given language by Akeroyd et al., 2015) which will make it possible to adapt the prediction method proposed here to an increasing number of languages with comparatively little extra effort. This will help to compare audiological study results across languages by the prediction methods presented here as a kind of objective yardstick. This will enable research to concentrate on those effects that cannot already be predicted from the acoustics of the language-specific test materials.
- Using various possible alternatives for the preprocessing or feature extraction of the FADE model, a straightforward route becomes possible to compare a number of model assumptions about hearing impairment against each other: While for the current computations, a simple MFCC was used, Schädler et al. (2016) already used a more sophisticated,

"effective" auditory model-based preprocessing after Dau et al. (1997) which may also be replaced by more sophisticated psychoacoustical models and appropriate model assumptions about how to change the processing and its parameters as a consequence of sensorineural hearing loss.

- Such a comparative approach of models and the required modifications of processing parameters for predicting most of the observable deficits in hearing-impaired listeners by a minimum set of assumptions and parameters is a very good candidate to find a connection between several audiological outcome measures within the same audiological patient. One might eventually be able to check the consistency across the different performance measures and to determine the minimum number of assumptions and free parameters to be used to completely characterize the individual hearing impairment. This may prove to be a new road toward modeling sensorineural hearing loss with as few parameters and as few prior assumptions as possible.
- Moreover, the fact that the current model does not need the desired speech signals and the background noise as separate signals, but only requires the mixed, complete signal with a set of SNRs for training qualifies this approach for aided performance prediction: Performing the same speech recognition prediction both for the unaided and the aided input signal to be presented to a patient, the effect of a hearing device (or other acoustical assistive listening device) may eventually be assessed in an objective way. This opens a completely new path toward rehabilitative audiology in connection with modern methods of machine learning.

Conclusions

- The ASR-based, broadly applicable FADE approach can predict the empirically found relation between the SRT and noise level as parameterized by Plomp (1978). This quantitatively describes the effect of hearing impairment on SRTs in stationary and fluctuating noise.
- In comparison to using the SII for stationary noise and the eSII (Rhebergen & Versfeld, 2005) for fluctuating noise for the same purpose, a more consistent SRT estimate across both noise types is achieved at least for low and intermediate noise levels.
- Suprathreshold processing deficiencies can be modelled by the level uncertainty parameter u_L which should be individually determined for high prediction accuracy. The highest prediction accuracy (expressed by Pearson's R^2) across all conditions and models is achieved with FADE if an independent data set is used for an individual distortion correction.
- The prediction accuracy achieved with the optimized modifications of the SII (data from Meyer & Brand, 2013) is roughly the same for the stationary noise case and in most cases worse for the fluctuating noise than for the FADE approach if a "typical" distortion correction is employed. This approach utilizes an average across all audiograms belonging to the same class of hearing loss.
- Interpolating from FADE simulations using a "typical" audiogram is not only much less computationally expensive but also not significantly different in prediction accuracy from using the individual audiogram for FADE if the same kind of distortion correction is used.
- Hence, for practical purposes, the typical audiogram interpolation approach with an individual distortion correction (with input from independent data) is recommended which requires only minimal computational effort and yields a higher prediction accuracy than all modifications of the SII employed here—especially for the fluctuating noise case.
- Taken together, the FADE approach is not only more versatile and makes much less assumptions than the SII but also yields a higher prediction accuracy if appropriate independent data for estimating the individual distortion correction is available.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by DFG (SFB/TRR 31 and EXC Hearing4all).

References

Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A.,... Kollmeier, B. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests. *International Journal of Audiology*, *54*(Suppl 2): 17–22.

ANSI S3.5. (1997). *American National Standard Methods for Calculation of the Speech Intelligibility Index*. American National Standards Institute, New York.

Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, *14*(2), 113–120.

Brand, T., & Kollmeier, B. (2002). Vorhersage der Sprachverständlichkeit in Ruhe und im Störgeräusch aufgrund des Reintonaudiogramms (Prediction of speech intelligibility in quiet and noise based on the audiogram).

Proceedings 5. Jahrestagung der Deutschen Gesellschaft für Audiologie, Zürich, 1–3.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow band carrier. *The Journal of the Acoustical Society of America*, *102*, 2892–2905.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artifical noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology*, *40*(3), 148.

Duquesnoy, A. J. (1983). The intelligbility of sentences in quiet and in noise in aged listeners. *The Journal of the Acoustical Society of America*, *74*(4), 1136–1144.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736.

Fisher, R. A. (1958). *Statistical methods for research workers* (13th ed.). London, England: Oliver and Boyd.

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America*, *19*, 90–119.

George, E. L. J., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *120*, 2295–2311.

Holube, I., & Kollmeier, B. (1996). Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. *The Journal of the Acoustical Society of America*, *100*(3), 1703–1716.

Jepsen, M., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, *124*, 422.

Jürgens, T., & Brand, T. (2009). Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model. *The Journal of the Acoustical Society of America*, *126*(5), 2635–2648.

Kollmeier, B. (1990). Meßmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache (*Measurement methods, models and improvement of the intelligibility of speech*), Habilitationsschrift (*Habilitation treatise*). University of Goöttingen, Fachbereich Physik, D-Goöttingen.

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M., Uslar, V. N., Brand, T., . . . Wagener, K. C. (2015). The multilingual matrix test: Principles, applications and comparison across languages - A review. *International Journal of Audiology*, *54*(Suppl 2): 3–16.

Lee, L. W., & Humes, L. E. (1993). Evaluating a speech-reception threshold model for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *93*, 2879–2885.

Meyer, R. M., & Brand, T. (2013). Comparison of Different Short-Term Speech Intelligibility Index procedures in fluctuating noise for listeners with normal and impaired hearing. *Acta Acustica united with Acustica*, *99*(3), 442–446.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788–1803.

Pavlovic, C. V., Studebaker, G. A., & Sherbecoe, R. L. (1986). An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. *The Journal of the Acoustical Society of America*, *80*(1), 50–57.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, *63*(2), 533–549.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech Language and Hearing Research*, *29*(2), 146–154.

Rhebergen, K. S., Lyzenga, J., Dreschler, W. A., & Festen, J. M. (2010). Modeling speech intelligibility in quiet and noise in listeners with normal and impaired hearing. *International Journal of Audiology*, *127*(3), 1570–1583.

Rhebergen, K. S., & Versfeld, N. J. (2005). A speech intelligibility index based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *117*, 2181–2192.

Schädler, M., Warzybok, A., Hochmuth, S., & Kollmeier, B. (2015). Matrix sentence intelligibility prediction using an automatic speech recognition system. *International Journal of Audiology*, *54*(Suppl 2): 100–107.

Schädler, M. R., Warzybok, A., Ewert, S. F., & Kollmeier, B. (2016). A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. *The Journal of the Acoustical Society of America* , *139*, 2708–2722.

Smoorenburg, G. F. (1992). Speech reception *in quiet and in noisy conditions by individuals with noise*-induced hearing loss *in relation to* their tone audiogram. *The Journal of the Acoustical Society of America*, *91*(1), 4221–4237.

Wagener, K. C. (2004). Factors influencing speech intelligibility in noise. *BIS - Verlag Oldenburg*. ISBN 3-8142-0897-8.

Wagener, K. C., Brand, T., & Kollmeier, B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, *45*, 26–33.

Wagener, K. C., Kühnel, V., & Kollmeier, B. (1999). Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. *Zeitschrift für Audiologie*, *38*(1), 4–15.