Predicting aided outcome using aided word recognition scores measured with linear amplification at above-conversational levels

Fereczkowski, Michal; Neher, Tobias

Published in:
Ear and Hearing

Publication date:
2022

Document version:
Accepted manuscript

Document license:
Unspecified

Citation for published version (APA):

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 29. Aug. 2022
Predicting aided outcome with aided word recognition scores measured with linear amplification at above-conversational levels

Michal Fereczkowski¹,² & Tobias Neher¹,²

¹ Institute of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark
² Research Unit for ORL – Head & Neck Surgery and Audiology, Odense University Hospital & University of Southern Denmark, Odense, Denmark

Financial disclosures/conflicts of interest: This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (‘Better hEAring Rehabilitation’ project) as well as GN Hearing, Oticon, and WS Audiology. There are no conflicts of interest, financial, or otherwise.

All correspondence should be addressed to:
Michal Fereczkowski, Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark, email: mfer@health.sdu.dk
ABSTRACT

Objectives:

Many hearing aid (HA) users receive limited benefit from amplification, especially when trying to understand speech in noise, and they often report hearing-related residual activity limitations. Current HA fitting strategies are typically based on pure-tone hearing thresholds only, even though suprathreshold factors have been linked to aided outcome. Furthermore, clinical measures of speech perception such as word recognition scores (WRSs) are performed without frequency-specific amplification, likely resulting in suboptimal speech audibility and thus inaccurate estimates of suprathreshold hearing abilities. Corresponding measures with frequency-specific amplification (‘aided’) would likely improve such estimates and enable more accurate aided outcome prediction. Here, we investigated potential links between either unaided WRSs or aided WRSs measured at several above-conversational levels and two established HA outcome measures: The Hearing-In-Noise Test (HINT) and the International Outcome Inventory for Hearing Aids (IOI-HA).

Design:

Thirty-seven older individuals with bilateral hearing impairments participated. Two conditions were tested: unaided and aided, with all stimuli presented over headphones. In the unaided condition, the most comfortable level (MCL) for the presented speech stimuli, WRS at MCL+10 dB as well as uncomfortable levels (UCLs) for narrowband noise stimuli were measured. In the aided condition, all stimuli were individually amplified according to the ‘National Acoustic Laboratories – Revised, Profound’ fitting rule. Aided WRSs were then measured using an Interacoustics Affinity system at three above-conversational levels, allowing for the maximum
aided WRS as well as the presence of ‘rollover’ in the performance-intensity function to be estimated. Multivariate data analyses were performed to examine the relations between the HINT (measured using a simulated HA with the NAL-RP amplification) or IOI-HA scores (for the participants’ own HAs) and various potential predictors (age, pure-tone average hearing loss, unaided WRS, aided WRS, rollover presence, and UCL).

Results:

Aided WRSs predicted the HINT scores better than any other predictor and were also the only significant predictor of the IOI-HA scores. Additionally, UCL and rollover presence in the aided WRSs were significant predictors of the HINT scores and competed for variance in the statistical models. Neither age nor pure-tone average hearing loss could predict the two aided outcomes.

Conclusions:

Aided WRSs can predict HA outcome more effectively than unaided WRSs, age or pure-tone audiometry and could be relatively easily implemented in clinical settings. More research is necessary to better understand the relations between rollover presence, UCL and speech recognition at above-conversational levels.
INTRODUCTION

According to the World Health Organisation (Chadha et al., 2021), 5% of the population worldwide suffer from a disabling hearing impairment, with this number being projected to increase to 10% by 2050. Consequently, there is a need for effective and efficient hearing loss treatment. Currently, hearing aids (HAs) are the main type of hearing rehabilitation. In the clinic, HAs are typically fitted based on the pure-tone audiogram, which is a measure of the audibility of sounds. HAs provide frequency-specific amplification to improve the audibility of speech and other signals, and hence they generally improve speech perception in quiet. For example, Duquesnoy and Plomp (1983) showed that aided speech-recognition thresholds (SRTs) in quiet are generally better than corresponding unaided SRTs.

Even though SRTs measured in quiet can be of some clinical value, daily-life human communication predominantly occurs at supra-threshold levels and often in the presence of background noise. In such situations, the speech level can be assumed to be relatively high. In noisy conditions, HA outcome is known to vary widely across hearing-impaired (HI) individuals, with this variability being difficult to predict from the audiogram (Larson et al., 2000; Shanks et al., 2002). Whereas for some individuals HA amplification will be sufficient to restore near-normal speech intelligibility, for many this will not be the case— even under favorable listening conditions. Such listeners will therefore exhibit suprathreshold hearing deficits that are not addressed by current audiogram-based HA fitting strategies. For the development of more effective compensation strategies, a better understanding of the underlying suprathreshold deficits in individual listeners and thus the inter-individual variability is necessary.

In the clinics, suprathreshold speech perception is commonly assessed using monosyllabic words presented in quiet at a comfortable listening level, with the corresponding word recognition
scores (WRSs) being used to quantify the performance. WRSs can vary between 0 and 100% (best possible performance) and are assumed to indicate the maximum possible performance under optimal listening conditions and are used for counselling purposes, despite mixed evidence for its predictive value for aided speech recognition (McRackan et al., 2016). For instance, Dornhofer et al. (2020) reported correlations between WRS, pure-tone average hearing loss (PTA) and SRT in quiet, and patient-reported measures of HA benefit that were low-to-weak at best. In contrast, Brännström et al. (2014) reported that for experienced HA users WRSs – unlike PTA and SRTs in quiet – were a significant predictor of the participants’ responses to items 3, 5 and 6 of the International Outcome Inventory for Hearing Aids (IOI-HA) questionnaire (Cox & Alexander, 2002). These three items quantify a HA user’s perception of residual activity limitations, residual participation restriction and the impact of the hearing loss on other persons, with the sum of the responses to these items being termed the IOI-HA interaction score. Brännström et al. (2014) found that a linear model of the interaction score with WRSs as the only predictor could explain 16% of the variance (adjusted $R^2$).

If WRS indicates the maximum performance for speech presented at an optimal, suprathreshold level, then the difference between 100% and the measured WRS can be attributed to suprathreshold deficits. However, for the effective diagnosis of speech perception-related suprathreshold deficits, good speech audibility is a prerequisite (Humes, 2007). In individual listeners, good audibility can be achieved by presenting speech at relatively high (above-conversational) levels and/or by means of frequency-specific amplification. A limitation of clinical WRSs is that frequency-specific amplification is not provided and that insufficient speech audibility will be the result, for example at high frequencies in listeners with steeply sloping audiograms (Dorfler et al., 2020). Indeed, Brännström et al. (2014) suggested that it may be more
appropriate to compare IOI-HA scores to WRSs obtained with HAs. This is also in line with work by McRackan et al. (2016) who concluded that aided WRSs can more accurately predict HA benefit and can better guide counseling than their unaided counterparts. Consequently, aided WRSs could constitute a clinically feasible measure of suprathreshold hearing abilities that could predict HA outcome more effectively than the WRSs currently used in hearing clinics.

Another potential limitation of current WRS measurements is that they are often performed at a single presentation level and that methods for selecting that level are not well established (Guthrie & Mackersie, 2009). Research has shown that presentation levels above approximately 80 dB SPL lead to poorer speech intelligibility in normal-hearing (NH) and unaided HI listeners (e.g., Studebaker et al., 1999; Dubno et al., 2005a, 2005b; Studebaker et al., 1999) as well as in aided HI listeners (Shanks et al., 2002). Thus, the choice of a single presentation level that will lead to maximal speech intelligibility is a non-trivial task, particularly in listeners with more severe hearing impairments. Measuring WRSs at multiple presentation levels could lead to improved estimates of maximal speech recognition in individual listeners.

In the audiological literature, the reduction in speech recognition with increasing level is termed *rollover*. Clinical studies have traditionally interpreted the presence of rollover in the performance-intensity function as an indicator of retro-cochlear hearing loss, for example due to acoustic neuroma (Jerger & Jerger, 1974). More recently, Shehorn et al. (2020) measured speech recognition performance at three above-conversational levels (74, 89 and 104 dBA SPL) in listeners with normal audiograms. Performance was poorest at the highest level (i.e., the performance-intensity functions exhibited rollover) and related to acoustic-reflex amplitudes and, indirectly, noise exposure history. This suggested that noise-induced synaptopathy in the human auditory nerve may underlie functional speech recognition deficits at higher presentation levels.
(Shehorn et al., 2020), and lead to observable rollover in the performance-intensity function.

Synaptopathy can be defined as a loss of connections between inner hair cells and auditory nerve fibers that does not lead to permanent threshold shifts (Shehorn et al., 2020). It is typically considered selective for fibers with low spontaneous rates (Furman et al., 2013) and can therefore be considered a supra-threshold deficit. Thus, while rollover is common in normal-hearing listeners and thus cannot be considered a dysfunction per se, supra-threshold deficits (e.g., mild in the listeners from Shehorn et al. (2020) or severe in the listeners from Jerger and Jerger (1974)) may lead to additional rollover in the performance-intensity functions of individual listeners.

In view of the above, we hypothesized that WRSs measured with frequency-specific amplification (i.e., aided WRSs) at multiple above-conversational presentation levels would provide more accurate estimates of individual speech recognition abilities than WRSs measured at a single level without frequency-specific amplification. Additionally, rollover presence in the performance-intensity function of individual listeners could be estimated from the aided WRS data and may provide additional diagnostic value. In the current study, we therefore performed such measurements with a group of older experienced HA users. We then investigated whether aided WRSs and rollover presence are better predictors of two established outcome measures – the Hearing-In-Noise Test (HINT; Nielsen & Dau, 2011) and the IOI-HA – than WRSs measured at a single level without frequency-specific amplification. While IOI-HA scores were obtained for listener’s own HAs, the HINT scores were measured with the same frequency-specific amplification as the aided WRSs, to ensure good audibility. In all tested models, we included age and PTA as covariates. Additionally, since the measurements involved speech presented at above-conversational levels, we included a measure of loudness discomfort as another covariate. To
achieve good control, we amplified the stimuli with individually prescribed frequency-specific linear gains and presented all signals via headphones.

METHODS

Ethical approval for the current study was obtained from the Regional Committee on Health Research Ethics for Southern Denmark (S-20162000-64). All participants signed an informed consent form and received a monetary reimbursement for their time (corresponding to 120 Danish crowns/hour).

Participants

Thirty-seven experienced HA users with a mean age of 74.8 years (standard deviation, SD = 6.4 years) participated. They were recruited using a large clinical database set up as part of the Danish ‘Better Hearing Rehabilitation’ (BEAR) project. All participants were bilateral HA users with at least two years of experience. Their own HAs were non-linear, multi-channel devices dispensed in 2017 as part of the BEAR project and fitted according to standard clinical procedures. The participants were chosen to exhibit a rather wide range of clinical WRSs (60-100%). The mean PTA averaged across 500, 1000, 2000 and 4000 Hz and left and right ears was 47.8 dB HL (SD = 9.7 dB HL). The 5th and 95th percentiles of the PTA distribution were 33 and 62 dB HL, respectively. Figure 1 shows the audiometric and age data for all participants.

To exclude participants with clearly asymmetrical hearing losses, an additional inclusion criterion was that the interaural asymmetry measured in terms of the absolute PTA difference across the left and right ears should not exceed 15 dB. The mean absolute PTA difference across left and right ears was 4.5 dB (SD = 3.3 dB). Only in two cases did this difference exceed 10 dB (max. difference = 11.3 dB).
Figure 1: Panel A shows the distribution of audiograms of the 37 participants. The thick lines show means for the seven test frequencies and the error bars show ±1 SD. Left-ear values are shown in blue and right-ear values in red. Panel B shows boxplots of the PTA data for the left (L) and right (R) ears as well as the individual data (values from the two ears of individual listeners are connected with dotted lines). Panel C shows the age distribution of the listeners. In all boxplots, the box edges indicate the 25th and 75th percentiles and the whiskers indicate the range of the data. Datapoints situated further than 1.5 interquartile ranges away from the box edges are shown by the ‘+’ symbol.

**Predictor Measurements**

All predictor measurements were performed monaurally in both ears under unaided and aided conditions.
Unaided Predictors

The unaided measurements included pure-tone audiometry, most comfortable level (‘MCL’) measurements, unaided WRSs and categorical loudness scaling. The MCL measurements were performed in quiet with running speech from the Dantale-I material (Elberling et al., 1989) using the Acceptable Noise Level (Nabelek et al., 2006) procedure implemented in the Interacoustics (Middelfart, Denmark) Affinity 2.0 system. For reasons of accuracy, the MCL was estimated three times per participant. The median of the three values was taken as the unaided MCL estimate (‘uMCL’). The average uMCL was 83.1 dB SPL across right ears and 83.4 dB SPL across left ears. The within-listener, across-ear differences in the uMCL estimates were small (mean: 0.4 dB; SD: 5.0 dB).

The unaided WRS measurements were performed using the monosyllabic word lists from the Dantale-I material. Each of these lists consists of 25 words. Before the measurements, the participants received verbal instructions to make sure they could perform the task. One randomly chosen test list was presented per ear at 10 dB above the uMCL (i.e., at the ‘uMCL+10 dB’ level). This approach was chosen based on the work of Guthrie and Mackersie (2009) who showed that the presentation level that maximizes phoneme recognition in listeners with moderate and moderately-severe hearing losses is, on average, 10 dB above the individual speech MCL. In the current study, the WRS obtained at uMCL+10 dB were taken as an estimate of the maximally achievable WRS under unaided conditions. In the following, they will be referred to as ‘uWRS’.

The adaptive categorical loudness scaling (ACALOS) procedure (Brand & Hohmann, 2002) with narrowband noise stimuli centered at 1 and 4 kHz was used to estimate one loudness function per test frequency. In this test, listeners rate the loudness of the presented stimuli using a scale with 11 categories ranging from “inaudible” to “too loud”. These categories are then
transformed into categorical units (CUs) ranging from 0 to 50. Loudness functions were fitted to
the collected data with the ‘BTUX’ procedure of Oetting et al. (2014), where two straight-line
segments, fitted at low and high levels, are connected by a smooth transition region between 15
and 35 CUs. The resultant loudness functions are defined by three parameters: a low-level slope,
a high-level slope (i.e., the slopes of the two straight-line segments) and the level at which the low-
level and high-level segments of the function meet. Since loudness discomfort may be an important
factor in the interpretation of WRS data obtained at above-conversational levels, two UCL
variables (‘UCL1k’ and ‘UCL4k’) were estimated by finding the level for which a given loudness
function reached 50 CUs. The maximum possible estimate was 140 dB SPL. Since the number of
participants and thus the sample size was limited, a decision was made to restrict the number of
predictors considered in the linear models (see the Data analysis section), and thus no other
ACALOS-based predictors were used. Additionally, the ACALOS curve-fitting procedure set 38
out of 74 high-level slope estimates at 4 kHz to a default value, which is automatically chosen
when the responses to high-level stimuli span less than four CUs (e.g., due to limited maximum
presentation levels). This severely limited the informative value of the UCL4k variable. Therefore,
a decision was made to only use UCL1k as a covariate in the analyses reported below.

227 Aided Predictors

Regarding the aided condition, the MCL and WRS measurements described above were
also performed with the stimuli pre-processed using frequency-specific linear amplification. The
gains were set individually for each ear according to the ‘National Acoustic Laboratories –
Revised, Profound’ fitting rule (NAL-RP; Dillon (2012). Figure 2 shows boxplots summarizing
the prescribed NAL-RP gains for frequencies ranging from 250 to 8000 Hz.
Figure 2: Boxplots summarizing the prescribed NAL-RP gains for frequencies ranging from 250 to 8000 Hz, for left (left panel) and right (right panel) ears.

Following the aided MCL measurements (‘aMCL’), aided WRSs were measured at three levels. By default, these levels were chosen to be aMCL+10 dB, aMCL+20 dB, and aMCL+30 dB, tested in ascending order. One randomly selected Dantale-I test list was used for each level. Before the actual measurements, an already used test list was employed to check whether the next presentation level would be still acceptable for the participants. If this was not the case, the test levels were adjusted such that three levels not lower than aMCL, and with any two levels differing by at least 5 dB, were tested per ear. In this manner, three aided WRSs were obtained per ear, with the highest score being taken as the aided WRS (‘aWRS’). Furthermore, rollover presence was estimated for each ear by calculating the difference between the aWRS and the WRS obtained for
the highest presentation level. Based on the 95% confidence intervals (CIs) of the binomial distribution, it was then checked whether rollover was statistically significant, that is, whether the WRS obtained at the highest presentation level fell outside the 95% CI of the aWRS. If this was the case and the reduction in WRS exceeded 8% (corresponding to at least three incorrect words), rollover presence (‘ROp’) was assumed. The additional >8% criterion was introduced to allow for attentional lapses that could lead to single, incorrect responses. For the collected data, no cases of 4% reductions in WRSs were observed. Thus, the 8% criterion used here was the lowest threshold value having an effect on the estimates of rollover presence. For the subsequent data analyses, the binary ROp variable was used as a predictor. This was because the amount of rollover (in percent) was considered unreliable due to it being the difference between two relatively noisy WRS measurements.

Outcome Measures

**HINT**

The HINT is a widely used test for assessing speech-in-noise performance. In the current study, it was performed with the speech material presented at aMCL+10 dB to facilitate comparability with the aWRS data. While better comparability could have been obtained if the HINT presentation level had been equal to the level at which the aWRSs were measured, this would have led to HINT measurements at very high levels in the case of some listeners with monotonically increasing aWRS performance-intensity functions and was thus avoided. The noise level was adjusted in an adaptive manner to determine the signal-to-noise ratio (SNR) corresponding to 50%-correct sentence recognition. Both speech and noise were amplified according with the same NAL-RP gains as used for the aWRS measurements.
IOI-HA

The IOI-HA questionnaire consists of seven relatively simple questions that assess an individual’s experience with their HAs. The questions relate to HA use, HA benefit and HA satisfaction, and the responses are given on a scale ranging from 1 to 5, with 5 denoting best possible outcome. IOI-HA questions are often grouped into two subscales: introspection and interaction (Brännström et al., 2014). The introspection subscale comprises items 1, 2, 4, and 7 and quantifies the listener’s experiences with the HA, such as use time and self-perceived HA benefit. For instance, question 1 is formulated as follows: “Think about how much you used your present hearing aid(s) over the past two weeks. On an average day, how many hours did you use the hearing aid(s)?”. As for the interaction subscale, which was already introduced in the Introduction, question 5 is a good example and is formulated as follows: “Over the past two weeks, with your present hearing aid(s), how much have your hearing difficulties affected the things you can do?”. In the current study, the IOI-HA was administered in pencil-and-paper form, with the questions taken from the revised Danish translation of the original questionnaire (Thunberg Jespersen et al., 2014). The participants were asked to assess their experiences with their own HAs. If in doubt about the procedure or the questions, they were encouraged to ask for further explanations.

Test Setup and Procedure

All measurements were performed in a soundproof listening booth. The audiogram, MCL and WRS measurements were performed using an Interacoustics (Middelfart, Denmark) Affinity
The system that was connected to a personal computer (PC). The stimuli were presented over RadioEar (Middelfart, Denmark) DD45 headphones. These headphones were chosen because of their ability to produce high output levels with low harmonic distortion (see RadioEar DD45, Technical Specification Sheet). In the unaided condition, the standard procedures implemented in the Affinity system were followed. In the aided condition, the stimuli were pre-processed by filtering the Dantale-I audio files with a 50th-order finite impulse response filter. The filter was implemented in Matlab, with its magnitude response following the gains prescribed by the NAL-RP rule for each ear. Prior to the filtering, the waveforms were attenuated by 35 dB (broadband) to avoid digital clipping due to the amplification. The aided stimuli were then created in Matlab by applying NAL-RP gains to the original sound files from the Interacoustics Affinity system, thereby assuring individual frequency-dependent amplification. The final stimuli, created for the two participants with the largest prescribed gains, were then recorded using the entire experimental setup and an artificial ear simulator. Since no clipping was observable in these recordings, the effectiveness of the chosen approach was confirmed. The ACALOS and HINT measurements were performed on the same PC using custom-made Matlab scripts with the stimuli being presented via Sennheiser (Wennebostel, Germany) HDA200 headphones connected to an RME Fireface UC soundcard. The implementations from the BEAR test battery (Sanchez-Lopez et al., 2021) were used for this purpose.

All measurements were performed during a single 2-hour session that was divided into two blocks. In the first block, the unaided measurements were performed (pure-tone audiometry, uMCL, uWRS, and ACALOS). After the first block, the participants were given a 15-min break, during which they filled in the IOI-HA questionnaire. In the second block, the aided measurements were performed (aMCL, aWRS, and HINT).
Data Analysis

All data analyses were performed in Matlab v2018b (MathWorks, Natick, US) and R v4.1.2 (R Development Core Team, 2010). First, the collected data were screened for outliers, that is, datapoints more than three interquartile ranges away from the first and third quartile of the corresponding dataset (Tukey, 1977). Two outliers were found in the HINT data from a single participant (both ears). In that case, the SRTs reached 60 dB SNR, meaning that the task was effectively a speech-in-quiet test. Thus, the data of this participant were excluded from all subsequent analyses. Further, in a single case, UCL1k could not be reliably estimated from the ACALOS data, and, to maximize the comparability of all tested models (see below), the corresponding HINT datapoint was also removed from all subsequent analyses. The remaining variables were tested for normality by means of the Shapiro-Wilk test. Since several of them (e.g., UCL1k and aWRS) were not normally distributed, a decision was made to use non-parametric Wilcoxon rank tests and Spearman’s correlation coefficient for direct comparison of these variables.

Further, multiple linear regression was used for modeling the HINT and IOI-HA outcomes. In regression analyses, the normality assumption concerns the distribution of the model residuals (Field et al., 2012). To test for deviations from normality, the Shapiro-Wilk test was applied to the residuals of all models reported here, showing no significant deviations (all $p > 0.05$).

The HINT modeling was performed based on the 71 datapoints that remained after removal of three datapoints (see above). Since there were two datapoints per participant and variable (i.e., one per ear), a mixed-effects model was used to allow accounting for within-subject effects. Given the available sample size ($N = 71$), maximally five predictors (excluding the fixed and random
intercepts) were included in the tested models, since 10 datapoints per predictor are considered the minimum needed for avoiding overfitting in linear regression analyses (Hair Jr et al., 2010). The main predictors of interest were age, PTA, uWRS, aWRS, and ROp. Furthermore, because aWRS and ROp were obtained at high presentation levels, it was hypothesized that rollover presence might be related to loudness discomfort, thus leading to a correlation between UCL1k and ROp.

A total of three HINT models were tested. For reference purposes, the first model contained four clinically available predictors (age, PTA, uWRS and UCL1k). In the second model, uWRS was replaced with aWRS and ROp. In other words, there were five predictors (age, PTA, aWRS and ROp and UCL1k), allowing for direct comparison of the predictive power of uWRS on the one hand, and aWRS and ROp on the other hand. The third model corresponded to the second model, except that UCL1k was removed to investigate its potential influence on the results.

Given that for the IOI-HA questionnaire there were half as many datapoints (because these data were participant- rather than ear-specific), the modeling required some modifications to the predictors. The obtained data were either summed (ROp) or averaged (UCL1k) across a participant’s two ears, or the better of the two ear-specific values was selected (PTA, uWRS, aWRS). As a result, the following variables were available for the IOI-HA modeling: age, better_PTA, better_uWRS, better_aWRS, sum_ROp, and mean_UCL1k. Given the available sample size (N = 37), maximally four predictors were included in the tested regression models. Since only one datapoint per participant was available, the models did not include random effects.

All p-values reported below remained significant ($p < 0.05$) after correction for the experiment-wise false discovery rate (Benjamini & Hochberg, 1995), except for a single case of the ROp variable in the second HINT model (see Table 1).
RESULTS

**uWRS and aWRS data**

Panel A in Figure 3 shows boxplots of the uWRS together with individual scores. The medians are 80%-correct and 76%-correct for the left-ear (blue) and right-ear (red) data, respectively. The distributions range from 32%-correct (left ear) and 24%-correct (right ear) to 100%-correct, with the left and right ears showing comparable scores overall (Wilcoxon signed-rank test, \( p = .40 \)). For some individual participants, large differences between the ears are apparent. While the median absolute difference between the ears is just 8%, the distribution of these differences ranges from 0% to 72%.

Panel B in Figure 3 shows boxplots of the aWRS together with individual scores. Here, the median value is 92%-correct for both ears. The distributions range from 68%-correct (left ear) and 64%-correct (right ear) to 100%-correct, with the left and right ears showing comparable scores overall (Wilcoxon signed-rank test, \( p = .96 \)). The absolute across-ear differences range from 0% to 20%, with eight participants exceeding a 10% difference. A Spearman correlation analysis revealed that aWRS was moderately related to PTA (\( r_s = -0.48, p < .001 \)).

Panel C in Figure 3 shows a scatterplot of uWRS (horizontal axis) versus aWRS (vertical axis). The dashed line corresponds to the identity line and the solid lines show the 95% CIs derived from the binomial distribution of the number of successes per 25 trials. The aWRS is, in most cases, larger than the corresponding uWRS. Moreover, in many cases the difference is statistically significant, as reflected by datapoints lying above the upper bound of the 95% CI. The difference between the median uWRS and median aWRS is 12% (Wilcoxon signed-rank test, \( p < .001 \)). A Spearman correlation analysis revealed that uWRS and aWRS were moderately related (\( r_s = 0.52, p < .001 \)).
Figure 3: uWRS (panel A) and aWRS (panel B) data. The blue and red symbols with dotted lines indicate individual listeners. Panel C shows a scatterplot of the uWRS and aWRS data.

Rollover in the aided performance-intensity functions

Figure 4 presents individual performance-intensity functions for the aided WRS data. The horizontal axis shows the sound pressure level before NAL-RP amplification (i.e., the level a HA would measure on the input side). The vertical axis represents the percentage of correctly identified monosyllabic words. Each panel corresponds to one participant, with the participant’s number being shown in the lower left-hand corner of each panel. The main observation that can be made is that in many cases maximum aided WRS performance is not achieved at the highest presentation level. For several ears, the performance-intensity function has a hill-top shape (e.g., the left ear of participant 22) or decreases across the level range (e.g., the right ear of participant 16 and the left ear of participant 19). Rollover presence (as defined in the Methods section) is marked by thick
lines and filled symbols. As can be seen, almost half of the tested ears (35 out of the 74) showed significant rollover at high presentation levels. A correlation analysis showed that ROp was related to neither PTA ($r_S = 0.07$, $p = .57$) nor UCL1k ($r_S = 0.14$, $p = .22$).

Figure 4: Aided WRS as a function of presentation level (in dB SPL) before amplification for the 37 participants. The blue and red symbols correspond to left and right ears. In many cases, maximum performance is achieved at one of the lower levels with performance dropping above that level. Thick lines and filled symbols indicate significant rollover. Some listeners were tested over a narrower level range than the default 20 dB of difference between aMCL + 10 and aMCL + 30 dB. Specifically, out of the 37 subjects, 16 were tested with the 20-dB range in both ears and 21 were tested with the 20-dB range in at least one ear. The minimum tested range was 10 dB, and eight subjects were tested with the 10-dB range in at least one ear.
Another noteworthy aspect is the input level at which maximum performance was achieved in ears with and without significant rollover. In ears with significant rollover, this was the level above which rollover occurred. The median values of the two datasets correspond to 81 dB SPL. In terms of spread they are also similar, as indexed by mean absolute deviations from the mean of 7.1 dB SPL (rollover absent) and 5.2 dB SPL (rollover present). According to a Wilcoxon rank-sum test, the two datasets do not differ in this regard ($p = .46$).

**UCL estimates**

Figure 5 presents the UCL1k data. The estimates cover a 64-dB range. The absolute across-ear differences range from 0 dB to 27.3 dB, with 30 participants showing a difference of less than 10 dB and a median difference of 3.8 dB. An inspection of Figure 5 suggests that the UCL estimates were, in most cases, consistent across left and right ears. A Spearman’s correlation analysis showed a moderate correlation between UCL1k and hearing threshold levels at 1 kHz ($r_S = 0.51$, $p = .0001$). However, UCL1k was not correlated with ROp ($p = .19$).

Figure 5: UCL estimates from the ACALOS procedure obtained with narrowband noise centered at 1 kHz.
**HINT measurements**

The left panel of Figure 6 shows boxplots of the HINT measurements together with the individual data for the 72 test ears. The collected SRTs range over 15 dB (i.e., from -0.4 to 14.5 dB SNR). A Wilcoxon signed-rank test did not show a mean difference between left and right ears ($p = .54$). Spearman’s correlation coefficient showed a moderate correlation between left and right ears ($r_S = 0.46$, $p = .005$). Within-subject differences between the tested ears exceeded 1.3 dB in 20 participants and thus fell outside the 95% test-retest CI for this task, as calculated based on a within-subject SD of 0.7 dB SNR estimated using data from Nielsen and Dau (2011) and Wu et al. (2021). Thus, the majority of the participants showed a small but significant difference in HINT performance across ears.

In terms of speech presentation levels, the median was 78 dB SPL (before amplification). Overall, the presentation levels were therefore somewhat higher than in other studies, where the noise level is typically fixed at 65-70 dB SPL and the speech level is varied.
Figure 6: HINT data for the 71 test ears (left panel) and IOI-HA data for the 37 participants (right panel). The presented IOI-HA scores are based on items 1, 2, 4, and 7 (Introspection), 3, 5, and 6, (Interaction), and 1-7 (Global).

**IOI-HA data**

The right panel of Figure 6 shows boxplots of the IOI-HA scores together with the individual data of the 37 participants. The left-most boxplot shows scores for the introspection subscale, the middle boxplot shows scores for the interaction subscale, and the right-most boxplot shows global scores (i.e., scores obtained by averaging the data from all seven IOI-HA items with equal weights). It is noteworthy that, unlike for the other two sets of scores, the interaction scores span the entire range of possible values (1-5). The median values of the three sets of scores were 4.25 scale points (introspection), 3.33 scale points (interaction), and 3.71 scale points (global). A Wilcoxon signed-rank test showed a significant difference between the interaction and introspection scores ($p = .005$). The median difference between these two sets of scores was 0.5 scale points, that is, the participants scored questions related to residual activity limitations lower (worse) than questions related to use time or overall satisfaction.

**Modeling HINT Outcome**

A Spearman correlation analysis revealed that HINT performance was related to both uWRS ($r_S = -.37, p = .001$) and aWRS ($r_S = -.60, p < .0001$). Partial correlation analysis showed that while aWRS was significantly related to HINT performance when uWRS was controlled for ($r_S = -0.52, p < .001$), the opposite was not the case ($r_S = -0.12, p = 0.33$).
To further investigate the predictive power of the different variables, three linear mixed-effect models were created, as explained above (see Data Analysis section). Four predictors were selected for the first model (Model 1): age, PTA, uWRS and UCL1k. The left-most columns of Table 1 provide a model summary. Only uWRS and UCL1k were significant, with the adjusted $R^2$ value being 0.42 (estimated for the fixed effects only, after removal of age and PTA) and without any collinearity issues being observed (variance inflation factor, VIF $\leq 1.07$). Interestingly, UCL1k explained a larger share of the variance than uWRS (measured in terms of sum of squares).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\beta$</th>
<th>SS</th>
<th>t-statistic</th>
<th>p value</th>
<th>$\beta$</th>
<th>SS</th>
<th>t-statistic</th>
<th>p value</th>
<th>$\beta$</th>
<th>SS</th>
<th>t-statistic</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.01</td>
<td>-0.602</td>
<td>0.551</td>
<td>11.14</td>
<td>-</td>
<td>1.89</td>
<td>.064</td>
<td>19.46</td>
<td>-</td>
<td>3.52</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.02</td>
<td>0</td>
<td>-0.344</td>
<td>0.733</td>
<td>0.03</td>
<td>1</td>
<td>0.59</td>
<td>.558</td>
<td>0.05</td>
<td>4</td>
<td>1.07</td>
<td>.293</td>
</tr>
<tr>
<td>PTA4</td>
<td>0.02</td>
<td>1</td>
<td>0.639</td>
<td>0.526</td>
<td>-0.01</td>
<td>0</td>
<td>-0.30</td>
<td>.762</td>
<td>-0.004</td>
<td>0</td>
<td>-0.14</td>
<td>.892</td>
</tr>
<tr>
<td>uWRS</td>
<td>-0.19</td>
<td>89</td>
<td>-5.64</td>
<td>&lt;.0001</td>
<td>0.076</td>
<td>26</td>
<td>3.05</td>
<td>.004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ROp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.01</td>
<td>12</td>
<td>2.09</td>
<td>.041 (.082)</td>
<td>1.37</td>
<td>24</td>
<td>2.78</td>
<td>.007</td>
</tr>
<tr>
<td>UCL1k</td>
<td>0.12</td>
<td>64</td>
<td>4.678</td>
<td>&lt;0.0001</td>
<td>0.076</td>
<td>26</td>
<td>3.05</td>
<td>.004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>uWRS</td>
<td>-0.07</td>
<td>52</td>
<td>-4.193</td>
<td>&lt;0.0001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>$df = 65$</td>
<td>148</td>
<td>$df = 64$</td>
<td>146</td>
<td>$df = 65$</td>
<td>164</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adj. $R^2$ &amp; -0.42</td>
<td>Adj. $R^2$ &amp; -0.50</td>
<td>Adj. $R^2$ &amp; -0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The first four columns (Model 1) show a summary of the reference HINT model with 4 clinically available predictors. The middle four columns show Model 2, which is similar to Model 1, but with uWRS replaced by aWRS and ROp. The last four columns show a summary of the model, which is similar to Model 2, but without the measure of loudness discomfort (UCL1k) as a predictor. Abbreviations: HINT - Hearing-in-noise test; WRS - Word Recognition Score; aWRS
- maximum aided WRS; uWRS - measured unaided WRS; PTA4 - four-frequency pure-tone average; ROp - rollover presence; UCL1k - uncomfortable level measured with narrowband noise centred at 1 kHz. See supplemental file 1 for scatterplots of each of the predictors plotted against the HINT outcome. Supplemental file 2 comprises the table of scatterplots and Spearman’s correlations for the HINT outcome and its predictors. Supplemental file 3 comprises a scatterplot of predictions from the fixed-effect portion of Models 1 and 2, as well as a comparison of the absolute value of the residuals from the two models, to facilitate comparisons.

The middle columns of Table 1 provide a summary of the second model (Model 2), which included five predictors: age, PTA, aWRS, ROp and UCL1k. As can be seen, only aWRS, ROp, and UCL1k were significant. Once again, no issues with collinearity were observed (VIF ≤ 1.39). The adjusted $R^2$ of the reported model (estimated for the fixed effects only, after removal of age and PTA) was 0.50.

The right-most columns of Table 1 provide a summary of the third model (Model 3), which included four predictors: age, PTA, aWRS and ROp. Only aWRS and ROp were significant, with aWRS accounting for five times more of the variance than ROp (measured in terms of sum of squares). Since the maximum VIF was found to be 1.29, collinearity was not an issue for this model either. After removal of the two insignificant predictors (age and PTA) from the third model, the adjusted $R^2$ was 0.42.

The coefficients for aWRS and ROp show a good correspondence between Model 2 and Model 3. In both cases, HINT performance improves with higher aWRS and deteriorates when rollover is present. Further, the $p$-value for ROp decreased from 0.04 in Model 2 to 0.007 in Model 3, suggesting that ROp and UCL1k both can predict HINT outcome but that they compete for the
same variance. In fact, after correction for multiple testing, the $p$-value for ROp in Model 2 was 0.082 and thus exceeded the 5% significance level. Moreover, Model 1 and Model 2 show that HINT performance decreases with higher UCL1k. Further, the adjusted $R^2$ of Model 2 was 0.50 and thus higher than that of Model 1 and Model 3 (0.42), suggesting that the removal of UCL1k from Model 2 led to a reduction in predictive power (Model 3).

To further compare the predictive power of the tested models, the Akaike Information Criterion (AIC; Sakamoto et al., 1986) was used. The analysis showed that the AIC of Model 2 ($\text{AIC}_M2 = 302.1$) was lower than that of Model 1 ($\text{AIC}_M1 = 311.6$) and the difference ($\Delta \text{AIC}_{M1-M2} = 9.5$) suggests, according to Anderson and Burnham (2004), that Model 2 outperformed Model 1 in terms of predictive power. Further, the AIC for Model 3 ($\text{AIC}_M3$) was 309.6. Model 1 can be considered as having similar predictive power as Model 3 ($\Delta \text{AIC}_{M1-M3} = 2$), while Model 2 was considerably better than Model 3 ($\Delta \text{AIC}_{M2-M3} = 7.5$). Overall, these findings agree well with what the comparison of the adjusted $R^2$ values for the three models already suggested. Finally, the supplementary material provides a figure comparing the HINT-score predictions from Models 1 and 2, as well as the measured score, to help the reader better assess the practical differences between the predictions from the two models.

### Modeling IOI-HA Outcome

Table 2 provides a summary of the two regression models used to predict IOI-HA outcome. In the first IOI-HA model, four predictors were used: age, better_PTA, better_uWRS, and mean_UCL1k. The last predictor was added, based on the results of the HINT modeling. The left-most columns of Table 2 summarize this model (Model 4). None of the tested predictors was significant, and the adjusted $R^2$ value for this model was very close to 0.
Since the the adjusted $R^2$ was very low, and a maximum four predictors per model were allowed in the IOI-HA analyses, UCL1k was not included in Model 5. As an effect, Model 5 predictors were: age, better_PTA, better_aWRS and sum_ROp, which allowed for direct testing of the main research question. The right-most columns of Table 2 summarize this model. The only significant predictor was better_aWRS. The corresponding adjusted $R^2$ value of the pruned model (including only better_aWRS) was .30. Overall, better_aWRS was the only significant predictor of the IOI-HA interaction score.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>β</th>
<th>SS</th>
<th>t-statistic</th>
<th>p value</th>
<th>B</th>
<th>SS</th>
<th>t-statistic</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.27</td>
<td>-</td>
<td>1.33</td>
<td>.192</td>
<td>-6.19</td>
<td>-</td>
<td>-1.72</td>
<td>.0942</td>
</tr>
<tr>
<td>Age</td>
<td>-0.009</td>
<td>0</td>
<td>-0.31</td>
<td>.760</td>
<td>-0.01</td>
<td>0</td>
<td>-0.43</td>
<td>.67</td>
</tr>
<tr>
<td>better_PTA4</td>
<td>-0.024</td>
<td>2</td>
<td>-1.23</td>
<td>.229</td>
<td>0.02</td>
<td>1</td>
<td>0.97</td>
<td>.34</td>
</tr>
<tr>
<td>better_uWRS</td>
<td>0.013</td>
<td>1</td>
<td>1.06</td>
<td>0.299</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>better_aWRS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.10</td>
<td>12</td>
<td>4.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>sum_ROp</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.42</td>
<td>3</td>
<td>-1.98</td>
<td>.057</td>
</tr>
<tr>
<td>mean_UCL1k</td>
<td>-0.003</td>
<td>0</td>
<td>-0.18</td>
<td>0.86</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: The first four columns (model 4) show a summary of the IOI-HA factor 2 model with four clinically-available predictors. The last four columns show a summary of Model 5, which is similar to Model 4, but with better_uWRS and mean_UCL1k variable replaced with better_aWRS and sum_ROp. Abbreviations: IOI-HA – International Outcome Inventory for Hearing Aids; WRS - Word Recognition Score; better_aWRS - maximum aided WRS across both ears; uWRS – maximum measured unaided WRS across both ears; better_PTA4 – lowest four-frequency pure-
DISCUSSION

The main purpose of the current study was to investigate relations between unaided WRS and aided WRSs collected at above-conversational levels on the one hand, and HINT and IOI-HA outcome on the other hand. The results showed that aWRS (maximum aided WRS) was the most important predictor of both HINT and IOI-HA outcome and the only one capable of predicting both outcome measures. The HINT outcome was more closely related to aWRS, than to uWRS, as indicated by the significant partial correlation between aWRS and HINT when uWRS was controlled for ($r_S = -0.52$, $p < .001$). Furthermore, the HINT analyses indicated that rollover presence in the aided performance-intensity function may also influence HA outcome. However, the contribution of ROp to Model 3 was considerably higher than to Model 2, which included UCL1k as a predictor. Further, UCL1k significantly contributed to Model 1, which only comprised predictors currently available in clinical practice. In the following sections, these findings will be discussed.

Predictive Power of Aided WRS

The observed predictive power of aWRS is broadly consistent with the findings of Lopez-Poveda et al. (2017) and Brännström et al. (2014). In the study of Lopez-Poveda et al. (2017),
aided HINT scores and IOI-HA scores could be partly explained based on the speech intelligibility index estimated for aided speech in quiet (SIIq). Although both SIIq and aWRS are measures of aided speech perception in quiet, there are some important differences between them. While Lopez-Poveda et al. (2017) reported that SIIq could be replaced as HINT predictor by the PTA (calculated across 0.5, 1, 2, 4 and 6 kHz), the current study did not find PTA (calculated across 0.5, 1, 2, and 4kHz) to be an effective predictor of aided outcome. This discrepancy could be due to differences in speech presentation levels. In the Lopez-Poveda et al. study, the speech level was fixed at 65 dB SPL (before amplification) while in the current study it was individually adjusted (median value: 78 dB SPL). This means that some speech components (e.g., those located at higher frequencies) could have been less audible in the study by Lopez-Poveda et al. compared to the current study. This possible explanation is supported by the fact that the PTA in the Lopez-Poveda et al. study ranged from 33 to 63 dB HL (5th and 95th percentiles), which corresponds well to the PTA values of the participants tested here (Figure 1).

In the study by Brännström et al. (2014), an adjusted $R^2$ of 16% was reported for their model of IOI-HA interaction scores with maximum unaided WRS measured for the ear with the better PTA as the only significant predictor. In the current study, the second IOI-HA model revealed aWRS as being the only significant predictor with an adjusted $R^2$ of 30% (Table 2). Compared to Brännström et al.’s IOI-HA model, the improved model performance observed here is likely due to the use of individual frequency-specific amplification in the WRS measurements.

Effects of Rollover and Uncomfortable Level on HINT performance

Another predictor that was derived from the aided WRS measurements was rollover presence (ROp). In Model 3, ROp was highly significant (Table 1). As pointed out above, the
median speech presentation level in the HINT measurements was 78 dB SPL, which was very close to the median presentation level for the aWRS data (81 dB SPL). Taken together, this suggests that rollover presence at above-conversational presentation levels may be indirectly related to speech perception at lower presentation levels that are optimal for monosyllabic word recognition.

However, as mentioned in the Introduction, even listeners with normal audiograms can exhibit rollover at high presentation levels. Given that the speech intelligibility index (SII, ANSI, 1997) includes a desensitization factor that reflects this phenomenon, a follow-up analysis was carried out which suggested that the SII decreased at the highest presentation level used in the aided WRS measurements for most of the tested ears (results not shown). However, the SII could not predict ROp on an individual basis, and the calculated SII values could not predict HINT outcome.

On the other hand, Models 1 and 2 showed that UCL1k was significantly related to HINT outcome and that the influence of ROp on HINT was lower in Model 2 (including UCL1k as a predictor) than in Model 3 (not including UCL1k), which suggests that ROp and UCL1k competed for variance in Model 2. This was not unexpected, since UCL1k was introduced to account for the fact that the aWRS data were obtained at above-conversational levels, as it was assumed that rollover presence might be related to low UCL1k. However, contrary to the expectations, no correlation was found between ROp and UCL1k.

Model 1, which only included clinically available predictors, showed similar predictive power to Model 3 (which included aWRS and ROp), and the sums of squares values for Model 1 showed that UCL1k could explain more variance in HINT outcome than uWRS in the same model.
The high significance of UCL1k in Models 1 and 2 may suggest that listening discomfort may have affected the HINT measurements. If this was the case, one could expect that higher UCL1k would lead to less discomfort, and thus better HINT performance. However, in the tested models, HINT performance was decreasing as UCL1k was increasing. The correlation could be due to confounding variables related to hearing thresholds. For instance, the hearing threshold for 1-kHz pure-tones can be positively correlated to both UCL1k and the HINT outcome. In the current data, there was a significant positive correlation between the hearing threshold at 1 kHz (HTL1k) and UCL1k ($r_S = 0.5$, $p < 0.0001$). However, a partial correlation between HINT and UCL1k was significant when controlling for HTL1k. Further, additional linear models (not shown), similar to Models 2 and 3 and including HTL1k as a covariate, did not show HTL1k as a significant factor. Another potential source of elevated UCLs may be conductive hearing loss (Tyler et al., 2014). Since bone conduction data was not collected in these experiments and it is not possible to investigate this potential relation.

An additional, somewhat speculative, explanation for the significance of UCL1k as a predictor and competition for variance with ROp in the HINT models is related to recent findings of Shehorn et al. (2020). In that study, ipsilateral acoustic reflex amplitudes were found to be correlated with speech-recognition performance at high presentation levels in listeners with normal audiograms, leading the authors to suggest a potential connection between synaptopathy and reflex amplitudes. In Olsen (1999), UCLs and acoustic reflex thresholds (ART) were highly correlated, with the observed correlations (Pearson’s $r = 0.88$, $p < 0.001$) being stronger than correlations between UCLs and hearing thresholds or acoustic reflex thresholds and hearing thresholds. In the current study, higher UCL1k was found to be related to poorer HINT score. In light of the positive correlation of ART and UCL, this is a counterintuitive finding, as one could expect that impaired
acoustic reflex will lead to decreased UCL and thus HINT performance at above-conversational levels. However, in the reported analyses UCL1k competed for variance with rollover presence, and thus one could speculate about a relation between UCL1k, rollover presence, acoustic reflex characteristics and synaptopathy. For instance, as suggested in Shehorn et al. (2020), loss of connections between inner hair cells and low-spontaneous-rate fibers could lead to decreased acoustic reflex amplitudes and increased ARTs, which are correlated with UCLs. Further, impaired functionality of the low-spontaneous-rate fibers may lead to degraded speech perception at moderate and high presentation levels, even if the exact nature of the dependence is still under debate (Carney, 2018). Overall, there may exist a common factor underlying speech-perception deficits and affecting measures such as ART and UCL. This could explain the somewhat counterintuitive relation. However, while the correlation reported in Olsen (1999) was high, the methods used to estimate ARTs and UCLs in the different studies were not identical and therefore caution is necessary when drawing conclusions. Further research is necessary to investigate these potential relations more fully.

Quantifying Suprathreshold Distortions with Aided WRS

Due to the use of frequency-specific amplification and relatively high presentation levels, the aided WRS measurements performed here can be assumed to be largely unaffected by audibility issues and thus reflective of suprathreshold hearing abilities. Two important aspects of aided WRS measurements are that they are conducted with low-context speech materials and that the test conditions are, in principle, optimal for speech understanding (due to the lack of noise and the linear stimulus processing). Therefore, these measures can be interpreted as estimates of residual deficits in terms of speech perception after audibility issues have been accounted for.
The fact that the aWRS data were predictive of both HINT and IOI-HA outcome is a noteworthy finding in light of other research showing that effective prediction of objective and subjective aided outcomes typically requires different types of predictors (Lopez-Poveda et al., 2017). Since the aWRS data were measured under optimal acoustic conditions and represent maximal performance on this task, they can be considered as approximating ceiling performance when contextual information is absent in the speech signal. In this case, the observable performance limitations can be considered as originating from suprathreshold auditory distortions (Plomp, 1978; Sanchez-Lopez et al., 2020). Overall, aWRS can therefore be considered a measure of suprathreshold distortions that appears to have some relevance for a HA user’s daily-life listening experiences.

Finally, rollover presence and UCL1k were both related to HINT outcome, independently of the relation with aWRS. Given the potential link between rollover presence, UCL1k and synaptopathy, this finding suggests the existence of other suprathreshold distortions that are independent of those captured by aWRS. These putative distortions at high presentation levels may range from mild, as in Shehorn et al. (2020), to severe, as in Jerger and Jerger (1974). Further research will be required to investigate these ideas.

Clinical Feasibility and Relevance

Considering clinical practice, the method and equipment used here for collecting aided WRS data is very similar to current clinical WRS procedures. Therefore, aWRS and rollover presence could be relatively easily implemented in clinical settings. Further, while the UCL1k data were obtained using the ACALOS procedure, clinical procedures for measuring uncomfortable levels are readily available.
The Acceptable Noise Level test (Nabelek et al., 2006) is another established procedure, and the MCL data collected here were obtained with the help of an implementation available in the Interacoustics Affinity system. Nabelek et al. (2006) reported that while ANL thresholds were not correlated with speech-recognition performance measures, they could be used to distinguish between two groups: the full-time HA users and the remaining users (part-time and nonusers), hence it could be related to IOI-HA introspection subscale. Given that the aided-WRSs quantify speech-recognition performance and are correlated with the IOI-HA interaction subscale, ANL can be considered a complementary measure to the aided-WRS based ones.

Apart from the possible diagnostic value of the presented results, their potential consequences for HA fitting are also of interest. Given the clear relation between the two HA outcome measures and the aWRS data, it seems likely that participants with low aWRSs would be candidates for additional SNR improvement when compared to other participants with higher aWRSs. Even though ROp and UCL1k also showed some relations to HINT outcome, the nature of these relations will have to be understood better before clinical implications can be discussed.

**Study Limitations**

In the current study, all stimuli were presented over headphones (to achieve good control). This type of stimulus presentation limits the degree of realism and therefore also the generalizability of the results, particularly in the case of the aided HINT measurements. Further, the stimuli were amplified with linear frequency-specific gains as calculated using the NAL-RP prescription rule. The use of linear amplification could also be a reason why UCL1k emerged as a predictor in the HINT models, especially that for the HA-input level equal to aMCL + 10 dB, the NAL-RP gains above 500 Hz were typically 5-20 dB higher than the corresponding real-ear
insertion gains (REIG) for the participants’ own hearing aids within two months after first fit. (data not shown; REIGs were measured at 55, 65 and 80 dB SPL input levels and interpolated for this comparison). Thus, it would be relevant to relate aWRS, ROp, and UCLs to HINT scores measured with a more realistic setup, for example with ear-level HAs and free-field sound presentation.

Concerning the IOI-HA, this questionnaire quantifies a user’s experiences with their own HAs, which may employ rather complex signal processing algorithms (e.g., directionality, noise reduction and non-linear amplification). No advanced HA features were included in the aided WRS measurements reported here. It is possible that if a more complex and realistic HA simulator was used instead, the aWRS data could better predict IOI-HA outcome.

Another limitation concerns the unaided WRS measurements, which were performed at 10 dB above the individual speech MCL only. Measurements at several levels would have made it possible to estimate the unaided performance-intensity function, too. However, this was not possible due to the limited number of Dantale-I test lists. In any case, testing at one level only is representative of current clinical practice and therefore a realistic scenario.

Finally, it should be noted that the reported results were obtained with older experienced hearing aid users with symmetrical, mild-to-severe hearing losses. To determine their generalizability, follow-up work with other clinical subpopulations will be needed.

REFERENCES


File 1: Scatterplots of each of the predictors plotted against the HINT outcome.
File 2: Scatterplots and Spearman’s correlations for the HINT outcome and its predictors.
File 3: Scatterplot of predictions from the fixed-effect portion of Models 1 and 2, as well as a comparison of the absolute value of the residuals from the two models.
File 4: Scatterplots of each of the predictors plotted against the IOI-HA interaction score.
File 5: Spearman’s correlations for the IOI-HA outcome and its predictors.