Trends in Hearing
Volume 21: 1–28
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216517730526
journals.sagepub.com/home/tia
SAGE

Predictors of Hearing-Aid Outcomes

Enrique A. Lopez-Poveda[1,2,3], Peter T. Johannesen[1,2], Patricia Pérez-González[1,2], José L. Blanco[1], Sridhar Kalluri[4], and Brent Edwards[4]

Abstract

Over 360 million people worldwide suffer from disabling hearing loss. Most of them can be treated with hearing aids. Unfortunately, performance with hearing aids and the benefit obtained from using them vary widely across users. Here, we investigate the reasons for such variability. Sixty-eight hearing-aid users or candidates were fitted bilaterally with nonlinear hearing aids using standard procedures. Treatment outcome was assessed by measuring aided speech intelligibility in a time-reversed two-talker background and self-reported improvement in hearing ability. Statistical predictive models of these outcomes were obtained using linear combinations of 19 predictors, including demographic and audiological data, indicators of cochlear mechanical dysfunction and auditory temporal processing skills, hearing-aid settings, working memory capacity, and pretreatment self-perceived hearing ability. Aided intelligibility tended to be better for younger hearing-aid users with good unaided intelligibility in quiet and with good temporal processing abilities. Intelligibility tended to improve by increasing amplification for low-intensity sounds and by using more linear amplification for high-intensity sounds. Self-reported improvement in hearing ability was hard to predict but tended to be smaller for users with better working memory capacity. Indicators of cochlear mechanical dysfunction, alone or in combination with hearing settings, did not affect outcome predictions. The results may be useful for improving hearing aids and setting patients' expectations.

Keywords

hearing impairment, hearing loss, auditory masking, auditory spectral processing, auditory temporal processing, aging, working memory

Date received: 9 December 2016; revised: 1 August 2017; accepted: 3 August 2017

Introduction

Disabling hearing loss affects over 5% of the world's population—360 million people—and one third of people older than 65 years of age. Hearing loss impacts on individuals' ability to communicate with others. This limits their access to services and can cause feelings of loneliness, isolation, and frustration (Palmer, Newsom, & Rooks, 2016). The social and economic impact of hearing loss is large (World Health Organization, 2015). Fortunately, most people with hearing loss may be treated with hearing aids. Unfortunately, hearing performance when aided varies widely across hearing-aid users, and hearing-aid owners report unequal benefits from using their hearing aids. Three percent of owners never use their hearing aids and 25% of owners use their hearing aids less often than weekly. In addition, 9% of owners are dissatisfied with their hearing aids and 10% are neutral about them (Abrams & Kihm, 2015). This raises several questions: (a) Why do some hearing-aid users perform better or benefit from their hearing aids more than others do? (b) Would it be possible to predict these outcomes for a given individual at the time when the hearing aid is prescribed? (c) What are the

---

[1]Instituto de Neurociencias de Castilla y León, University of Salamanca, Spain
[2]Instituto de Investigación Biomédica de Salamanca, University of Salamanca, Spain
[3]Departamento de Cirugía, Facultad de Medicina, University of Salamanca, Spain
[4]Starkey Hearing Research Center, Berkeley, CA, USA

Brent Edwards is presently at the National Acoustics Laboratory, Sydney, Australia.

Corresponding author:
Enrique A. Lopez-Poveda, Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Calle Pintor Fernando Gallego 1, 37007 Salamanca, Spain.
Email: ealopezpoveda@usal.es

Table 1. Hearing-Aid Outcome Measures.

Acronym	Description
SRT_N	*Speech reception threshold in noise* (dB SNR). The signal-to-noise ratio (SNR, in decibels) at which the hearing-aid user understands 50% of the sentences presented in a (time-reversed) two-talker background. It was measured with the user wearing his/her two hearing aids. Lower values indicate better performance.
SSQ_B	*Self-reported improvement in hearing ability obtained from using hearing aids*. It was assessed using the speech, spatial, and qualities questionnaire (Gatehouse & Noble, 2004) slightly modified to assess hearing *improvement* rather than hearing ability (Jensen et al., 2009). Three outcome measures were obtained, one for each of the speech, spatial, and qualities sections of the questionnaire. Larger values indicate greater improvement.
COSI	*Client Oriented Scale of Improvement questionnaire* (Dillon et al., 1997). Aimed at assessing the improvements obtained from hearing-aid treatment tailored to the deficits considered important for and by each individual. Larger values indicate greater improvement.
IOI-HA	*International Outcome Inventory for Hearing Aids*. A minimum core set of questions for assessing hearing-aid user outcomes (Cox et al., 2000, 2002). Larger values indicate greater levels of satisfaction.

Note. For detailed information, see Appendix A.

factors determining aided performance and benefit? The present study addresses these questions.

Speech-in-noise intelligibility is the most sought-after improvement among hearing-aid users (Kochkin, 2002). Indeed, hearing-aid users vary widely in their ability to understand speech in noise (Löhler et al., 2015) and rate "trying to follow a conversation in the presence of noise" as the listening situation with the lowest level of satisfaction (Abrams & Kihm, 2015). Satisfaction may be related to the subjective improvement in hearing ability experienced by the hearing-aid user. For these reasons, in the present study, hearing-aid outcome is assessed by measuring aided speech intelligibility in noise and self-reported improvement in hearing ability as measured with standardized questionnaires (Table 1). We note that the first of these outcome measures is a *state* variable that describes auditory performance when using hearing aids, while the others are *difference* variables intended to capture the subjective benefit obtained from using hearing aids.

Multiple factors might contribute to the wide variability in these hearing-aid outcomes across users. First, outcomes might be better when the hearing aid compensates for the individual's loss of cochlear mechanical amplification. The typical hearing-aid user cannot detect low-intensity sounds and yet experiences high-intensity sounds as loud as normal-hearing listeners do, a phenomenon commonly referred to as *recruitment* (e.g., Marozeau & Florentine, 2007). Loudness recruitment is thought to be due to a loss or dysfunction of cochlear outer hair cells. Outer hair cell dysfunction reduces the cochlear mechanical amplification to soft sounds without altering cochlear mechanical responses to high-intensity sounds (Robles & Ruggero, 2001; Ruggero, Rich, & Recio, 1996), hence the notion that the rapid growth of loudness with increasing sound intensity typically experienced by listeners with hearing loss is due to loss of cochlear amplification (see also Moore & Glasberg, 1997). Not every listener with a hearing loss, however, shows a rapid growth of loudness with increasing sound intensity (Marozeau & Florentine, 2007). Some listeners with hearing loss cannot hear soft sounds but their loudness for sounds at and just above their audiometric thresholds is normal or close to normal (Buus & Florentine, 2001). This phenomenon is possibly due to retro-cochlear dysfunction of an uncertain nature, such as, for example, an inefficient mechano-electrical transduction at the inner hair cell.

A basic function of modern nonlinear hearing aids is to counteract the effect of reduced audibility and loudness recruitment, with the additional possible goal of restoring specific loudness across the audible spectrum of sound (Edwards, 2003). There exist several amplification prescription rules for any given audiometric loss (Byrne, Dillon, Ching, Katsch, & Keidser, 2001; Keidser, Dillon, Carter, & O'Brien, 2012; Keidser, Dillon, Flax, Ching, & Brewer, 2011; Moore, Glasberg, & Stone, 2010; Scollie et al., 2005). All of them apply greater amplification at frequencies where the hearing loss is larger, and greater amplification for low- than for high-intensity sounds, thus compressing a wide range of sound intensities at the hearing aid input into a narrower intensity range at the output. Importantly, the sound intensity at which hearing aid amplification starts decreasing (termed the compression threshold) and the rate of amplification decrease with increasing intensity (the compression ratio) are always determined based on average data (Byrne et al., 2001). That is, all hearing-aid users with identical audiometric losses are prescribed identical amplification-compression schemes in their hearing aids. This average-based approach, however, might be suboptimal for each individual, depending on the extent that his or her hearing loss is due to a loss of cochlear mechanical amplification.

On average, about 60% to 70% of the audiometric loss is due to a loss of cochlear amplification

(Johannesen, Pérez-González, & Lopez-Poveda, 2014; Lopez-Poveda & Johannesen, 2012. After Moore & Glasberg, 2004), and for convenience, we will refer to this contribution also as outer hair cell loss and to the residual audiometric loss as inner hair cell loss (see also Moore, 2014). Individually, however, cochlear amplification loss can account for as little as 30% of the audiometric loss (e.g., Johannesen et al., 2014; Lopez-Poveda & Johannesen, 2012). Therefore, it is conceivable that listeners with greater cochlear amplification loss prefer comparatively more compression in their hearing aids to compensate for the reduction (or lack) of compression in their ears. Conversely, listeners with the same audiometric loss but greater inner hair cell loss would prefer more linear amplification because they have substantial residual cochlear compression in their ears. Significant residual cochlear compression combined with hearing-aid compression might cause excessive distortion and diminish intelligibility (Bode & Kasten, 1971). If this were the case, both aided speech-in-noise intelligibility and self-reported improvement in hearing ability should be correlated with the loss of cochlear mechanical amplification alone or combined with the preferred hearing-aid settings.

Aided speech-in-noise intelligibility might also be determined by aspects unrelated to the individual's loss of cochlear amplification. Speech is a dynamic signal, and much of the information in speech is encoded in the changes of speech energy over time. Indeed, the intelligibility of speech in noise diminishes with manipulations of the temporal information in speech for both normal-hearing (Lopez-Poveda & Barrios, 2013; Lorenzi, Gilbert, Carn, & Moore, 2006; Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007) and hearing-impaired listeners (Lorenzi et al., 2006) and correlates with performance in temporal discrimination tasks (e.g., Füllgrabe, Moore, & Stone, 2015; Johannesen, Pérez-González, Kalluri, Blanco, & Lopez-Poveda, 2016; Strelcyk & Dau, 2009). On the other hand, elderly listeners with close-to-normal audiometric thresholds tolerate lower noise levels than young normal-hearing listeners to achieve identical levels of intelligibility in noise (Johannesen et al., 2016; Peters, Moore, & Baer, 1998) even when they are audiometrically matched to the younger controls (Füllgrabe et al., 2015). Therefore, aided speech-in-noise intelligibility could vary across users depending on their age and auditory temporal processing capacity. These two factors, however, would affect aided and unaided listening equally (unless amplification alters temporal processing capacity). Therefore, they might not affect the self-perceived benefit from using hearing aids.

Of course, many additional factors might affect the hearing-aid outcomes considered here. For example, some aspects of cognition might affect aided speech-in-noise intelligibility (Sommers, 1997; reviewed by Akeroyd, 2008). In particular, aided speech-in-noise intelligibility appears to be worse for listeners with low than with high working memory capacity (Souza, Arehart, Shen, Anderson, & Kates, 2015), at least in the elderly (Füllgrabe & Rosen, 2016b). Another factor is audibility. This might seem paradoxical considering that a main function of a hearing aid is to counteract the effects of reduced audibility. Often, however, hearing-aid users prioritize listening comfort over audibility and request less amplification than would be prescribed based on their audiometric losses (Humes, 2002). Reduced amplification might reduce audibility and hence intelligibility (Woods, Kalluri, Pentony, & Nooraei, 2013).

Here, we attempted to pinpoint some of the factors determining aided intelligibility in a (time-reversed) two-talker background and self-reported benefit from using hearing aids. The main aim was to investigate to what extent these outcomes depend on the hearing aid compensating for the individual degree of cochlear mechanical dysfunction or on aspects such as temporal processing capacity, age, or working memory capacity. A second aim was to develop statistical models that might help audiologists predict outcomes at the time when the hearing aid is prescribed.

Methods

We measured six outcome measures (Table 1) and 28 variables, including demographic data (Table 2), audiological data (Table 3), behavioral and physiological indicators of cochlear mechanical dysfunction (Table 4), indicators of auditory temporal processing abilities (Table 5), hearing-aid settings (Table 6), working memory capacity, aided speech-spectrum audibility in quiet, and pretreatment self-perceived hearing ability (Table 7). Each variable and outcome measure is described in the corresponding table together with a justification for its use in the study. Specific methods can be found in Appendix A.

Of the 28 measured variables, only 19 of them were used as candidate predictors of outcomes in the present study. Among the omitted variables were the *air-bone gap* (Table 3), which was used as a criterion for inclusion in the study, and *uncomfortable loudness levels* (ULL, Table 3), which was used to guide hearing-aid fitting. The statistical procedures required predictors to be scalar. Therefore, categorical variables such as *sex*, *prior hearing-aid use*, *noise exposure*, or *tinnitus sufferer* (Table 2) were measured for completeness but omitted from the analyses. Also omitted was the *dynamic binaural-masking level difference* (BMLD, Table 5) because it could be measured for 54 of the 68 participants only (see Appendix A). Lastly, two of the five

Table 2. Demographic Variables.

Name	Description (units)
Age	*Age* (years). It is a significant factor to hearing-aid outcome (Humes, 2002). In addition, age can affect speech-in-noise intelligibility even in the absence of significant audibility deficits (CHABA, 1988; Kim, Frisina, Mapes, Hickman, & Frisina, 2006; Peters et al., 1998).
Sex	This variable was included for completeness.
Prior hearing-aid use	*Experienced hearing-aid user* (yes, no). Outcomes could be biased depending on whether participants had used hearing aids before entering the study or not. Experienced hearing-aid users may prefer greater amplification than first-time users (e.g., Keidser, O'Brien, Carter, McLelland, & Yeend, 2008), and higher amplification might improve aided speech perception.
Noise exposure	*Self-reported history of noise exposure* (yes, no). In rodents, noise exposure can damage the inner ear in multiple ways, including reductions in the number of inner or outer hair cells, damage to their stereocilia (e.g., Liberman & Dodds, 1984), or reductions in the number of cochlear synapses or auditory nerve fibers (Kujawa & Liberman, 2009), all of which can hinder speech-in-noise intelligibility (reviewed by Lopez-Poveda, 2014).
Tinnitus	*Tinnitus sufferer* (yes, no). Tinnitus sufferers show poorer intelligibility in noise than nonsufferers (Newman, Wharton, Shivapuja, & Jacobson, 1994; Ruy, Ahn, Lim, Joo, & Chung, 2012). On the other hand, hearing-aid users with tinnitus may benefit from the masking that amplification with hearing aids provides, thus increasing hearing-aid benefit (Searchfield, Kaur, & Martin, 2010)

Note. For detailed information, see Appendix A.

Table 3. Audiological Variables.

Acronym or name	Description (units)
PTT	*Pure tone thresholds* (dB HL). Clinically measured audiometric thresholds averaged across test frequencies.
Air-bone gap	*Air-bone conduction gap* (dB). The difference between air- and bone-conduction audiometric thresholds averaged across test frequencies.
ULL	*Uncomfortable loudness levels* (dB HL). The lowest sound levels for monaurally played pure tones of one-second duration that were judged as uncomfortably loud averaged across test frequencies.
$\mathrm{SRT_Q}$	*Unaided speech reception threshold in quiet* (dB HL). The sound pressure level at which the listener was able to correctly reproduce 50% of the disyllabic words presented in quiet unaided. Lower values indicate better performance.

Note. For detailed information, see Appendix A.

Table 4. Variables Indicative of Cochlear Mechanical Dysfunction.

Acronym	Description (units)
$\mathrm{HL_{OHC}}$	*Outer hair cell loss* (dB). The contribution of cochlear mechanical amplification loss to the audiometric hearing loss averaged across test frequencies. Data were taken from Johannesen et al. (2014).
$\mathrm{HL_{IHC}}$	*Inner hair cell loss* (dB). The difference between the total audiometric loss (in dB HL) and $\mathrm{HL_{OHC}}$ averaged across test frequencies. Data were taken from Johannesen et al. (2014).
BMCE	*Residual cochlear compression* (dB/dB). The slope of behaviorally inferred cochlear input/output curves over the range of input levels where compression occurred averaged across test frequencies. Data were taken from Johannesen et al. (2014).
$\mathrm{DPOAE_N}$	The number of primary L_2 levels in the range from 35 to 70 dB SPL (5-dB steps) where distortion-product otoacoustic emissions (DPOAEs) were observed averaged across test frequencies.
$\mathrm{DPOAE_{\mu Pa}}$	*Total DPOAE pressure* (dB SPL). The sum of the DPOAE amplitudes (in μPa) recorded at the eight L_2 levels and converted back into decibels (Reavis et al., 2011). It is an overall measure of the total DPOAE amplitude.

Note. For detailed information, see Appendix A. BMCE = basilar membrane compression exponent.

Table 5. Variables Indicative of Temporal Processing Abilities.

Acronym	Description (units)
BMLD	*Dynamic binaural masking level difference* (dB). A measure of the improvement in the detection threshold of a pure tone signal masked by dynamic noise that occurs in binaural listening when the pure tone has opposite phases at the two ears. It appears to be correlated with self-reported hearing disability (Gatehouse & Akeroyd, 2006). Larger values indicate greater improvements.
FMDT	*Frequency modulation detection threshold* ($\log_{10}$[Hz]). The minimum amount of frequency modulation that can be detected for a 1500 Hz tone carrier. It is a measure of the sensitivity to the temporal fine structure of the stimuli. It is correlated with unaided intelligibility in a two-talker masker condition (Strelcyk & Dau, 2009). Data were taken from Johannesen et al. (2016).
FMRR	*Forward-masking recovery rate* (dB/ms). A measure of the ability to detect a signal preceded by an intense masker sound. It is thought to be related to the ability to perceive weak phonemes preceded by more intense phonemes in running speech (Gregan et al., 2013), and to reflect neural recovery from previous stimulation (Oxenham, 2001). Data were taken from Johannesen et al. (2014).
TI	*Temporal integration* (dB). The improvement in the detection threshold of a sound as the sound duration increases. It is reduced for listeners with hearing loss and the reduction appears unrelated to cochlear mechanical damage (Plack & Skeels, 2007). Steeper-than-normal threshold-duration functions could be indicative of disrupted auditory nerve activity (Zeng et al., 2005) and of primary deafferentation (Marmel et al., 2015), both of which might hinder intelligibility in noise (Lopez-Poveda, 2014).

Note. For detailed information, see Appendix A.

Table 6. Hearing-Aid Related Variables.

Acronym	Description (units)
REIG_{50dB}	*Real-ear insertion gain at 50 dB SPL* (dB). Hearing-aid amplification for an international speech test signal (ISTS, Holube et al., 2010) at 50 dB SPL, as measured with the hearing aid placed in the user's ear.
REIG_{65dB}	*Real-ear insertion gain at 65 dB SPL* (dB). Hearing-aid amplification for an ISTS at 65 dB SPL, as measured with the hearing aid placed in the user's ear.
REIG_{80dB}	*Real-ear insertion gain at 80 dB SPL* (dB). Hearing-aid amplification for an ISTS at 80 dB SPL, as measured with the hearing aid placed in the user's ear.
RECE_{LO}	*Real-ear compression exponent at low intensities* (dB/dB). It was calculated as follows: $1-(\mathrm{REIG}_{50dB}-\mathrm{REIG}_{65dB})/15$. Larger values indicate less compression.
RECE_{HI}	*Real-ear compression exponent at high intensities* (dB/dB). It was calculated as follows: $1-(\mathrm{REIG}_{65dB}-\mathrm{REIG}_{80dB})/15$. Larger values indicate less compression.

Note. For detailed information, see Appendix A.

measured hearing-aid settings (Table 6) were redundant (i.e., they could be inferred from the other three). We chose to omit the real-ear insertion gain (REIG) at 80 dB SPL (REIG_{80dB}) and the real-ear compression exponent at low-to-moderate levels (RECE_{LO}) from the analysis.

Statistical, predictive models of each outcome measure were obtained as follows. First, principal component analysis (PCA) was applied to reduce the 19 predictors into a smaller number of largely independent components. Second, stepwise multiple linear regression (MLR) was applied to express each outcome measure as the sum of the resulting components, each multiplied by a coefficient (linear combination). The model coefficients were optimized for the linear combination of components to explain the largest possible amount of variance in the predicted outcome. Given the abstract nature of some of the components and with the aim of obtaining meaningful predictive models, the stepwise MLR procedure was applied again but using the subset of measured predictors with the larger loadings in each component. We note that this procedure took into account the potential covariance between predictors. The PCA procedures for obtaining the components and the stepwise MLR approach for selecting and prioritizing components or predictors are described in Appendix A.

The PCA-MLR procedures required predictors to be numerical and single valued. Some predictors, such as pure tone thresholds (PTTs), were typically obtained at test frequencies of 0.5, 1, 2, 4, and 6 kHz. Multifrequency valued predictors were combined into a single value by

Table 7. Additional Variables Used in the Present Study.

Acronym	Description (units)
RSpan	*Score in the Reading Span Test* (Daneman & Carpenter, 1980). A measure of working memory capacity. Scores in this test have been linked with hearing-aid outcomes (reviewed by Akeroyd, 2008), and with speech recognition in fluctuating noise with and without hearing aids (Lunner, 2003), at least for older listeners (Füllgrabe & Rosen, 2016b). Larger values indicate better working memory capacity.
SSQ	*Baseline self-reported hearing ability (unaided)* at the time of entering the study and measured using the Speech, Spatial, and Qualities questionnaire (Gatehouse & Noble, 2004). Three independent scores were used as predictors, one for each of the speech, spatial, and qualities sections of the questionnaire. Larger values indicate better hearing abilities.
SII_Q	*Bilateral speech intelligibility index in quiet wearing hearing aids*. The proportion of the amplified speech spectrum above the audiometric threshold of the listener (ANSI, 1997). It was included because parts of the speech spectrum may be inaudible even when wearing the hearing aid, especially if the user requested less gain than recommended by the gain prescription rule. Larger values indicate greater spectrum audibility.

Note. For detailed information, see Appendix A.

weighting the value at each frequency according to the importance of that frequency for speech perception (American National Standards Institute [ANSI], 1997) and summing the weighted values across frequencies. The weights were 0.18, 0.25, 0.28, 0.23, and 0.06 for the test frequencies 0.5, 1, 2, 4, and 6 kHz, respectively (from Tables 3 and 4 of ANSI, 1997).

The study was approved by the Ethics Committee of the University of Salamanca (Salamanca, Spain). The same 68 volunteers (43 men, age range: 25–82 years, mean age = 61 years) with symmetrical sensorineural hearing loss who had participated in earlier related studies (Johannesen et al., 2014, 2016) participated in the present study. They gave their consent to participate prior to their inclusion in the study. In the first phase of the study—prior to hearing-aid fitting—participants carried out multiple clinical and laboratory tests to measure all of the predictors except for the hearing-aid settings and the aided speech intelligibility index in quiet (SII_Q; see Tables 2–7). In the second phase of the study, participants were fitted with two experimental hearing aids using standard clinical procedures (see Appendix A). After a 2-month trial period, during which hearing aids were fine-tuned as per user request (using mainly adjustments of the hearing-aid compressor gain settings), the final hearing-aid settings (Table 6) were recorded, and outcome measures (Table 1) were assessed. Outcomes were assessed for bilateral hearing-aid use. All but two of the predictors, however, were measured for one ear only. In most cases, this was the ear with better (lower) audiometric thresholds in the 2 to 6 kHz frequency range (30 left ears, 38 right ears). The two bilateral predictors were the BMLD (Table 5) and the aided SII_Q (Table 7). Each participant devoted approximately 50 hr of testing. Participants received a hearing aid free of charge in compensation for their services.

Results

Fifteen participants were using hearing aids at the time of entering the study, and 53 had not used hearing aids before. Twenty-five participants reported to have been exposed to high-intensity sounds at some point in their lives, and 26 participants reported to suffer from tinnitus. Most participants had high-frequency hearing losses (Figure 1). On average, they had about 10 dB more hearing loss at low frequencies (0.5 and 1 kHz) than the participants in a similar hearing-aid study (Humes, 2002). Otherwise, participants had hearing losses and audiological profiles typical of hearing-aid users or candidates.

The contributions of cochlear amplification loss and inner hair cell loss to the total audiometric loss (Table 4), the rate of recovery from forward masking, and the absolute detection thresholds used to calculate the temporal integration (TI) data (Table 5) have been reported in Johannesen et al. (2014). Frequency modulation detection thresholds (FMDTs; Table 5) have been reported in Johannesen et al. (2016). All other variables and outcome measures had values that were broadly consistent with those reported in the literature, as described in Appendix A.

Pairwise Correlations Between Outcome Measures and Predictors

Table 8 gives pairwise Pearson correlation coefficients (R) between each outcome measure and each predictor. Statistically significant (two-tailed t test, $N = 68$) values without ($p \leqslant .05$) and with Bonferroni correction[1] for multiple (20) comparisons ($p \leqslant .05/20$) are indicated by asterisks and bold font, respectively. Although the dynamic BMLD was not used in the PCA-MLR analysis as a predictor of outcome (see Methods section), it is

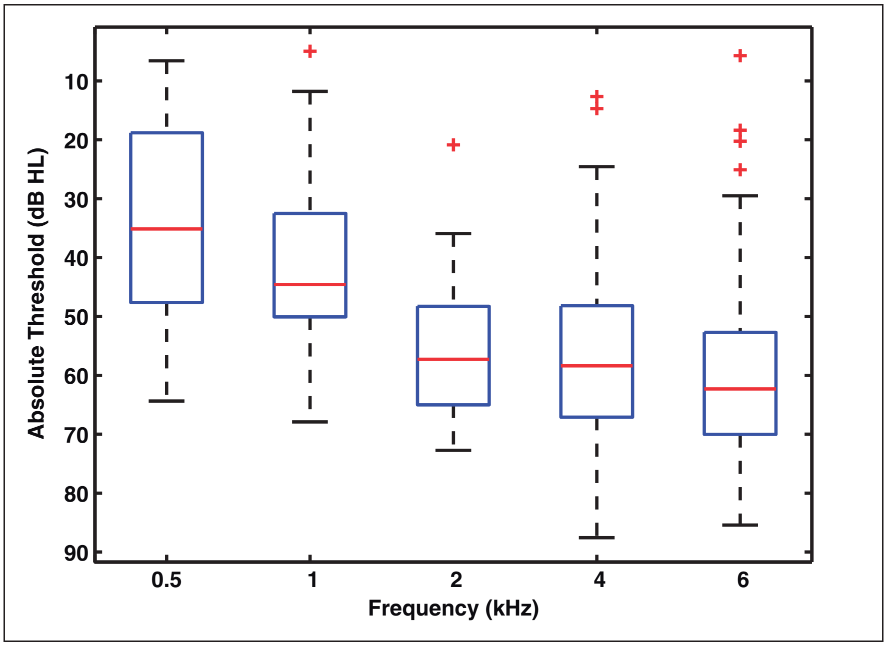

Figure 1. Distribution of hearing losses for each test frequency. Data replotted from Johannesen et al. (2014).

included here for completeness. Overall, the values show that aided intelligibility in a (time-reversed) two-talker background tended to be better (speech reception threshold [SRT_N] was lower) for younger participants (lower age), for participants with smaller hearing losses (smaller PTT), better unaided speech intelligibility in quiet (lower SRT_Q), smaller cochlear amplification losses (smaller HL_{OHC}), smaller inner hair cell losses (smaller HL_{IHC}), and stronger otoacoustic emissions (larger $DPOAE_N$). Intelligibility also tended to be better for participants with better FMDTs (smaller FMDTs), slower rate of recovery from forward masking (smaller forward-masking recovery rate [FMRR]), and for participants who had a greater proportion of audible speech spectrum with their hearing aids in quiet (larger SII_Q).

Figure 2 shows scatter plots of the aided SRT_N against each of the predictors that came up as significant predictors in the MLR models of aided SRT_N (described later). Also shown are linear regression functions (fitted by least squares) and corresponding statistics. The plots suggest a linear relationship between each of the predictors and the aided SRT_N. Audiometric thresholds (PTT) explained around 13% of the SRT_N variance ($R^2 = 0.13$). This value is consistent with that reported by Peters et al. (1998), who found R^2 values in the range 0.11 to 0.25 for fluctuating maskers, although they did not use a (time-reversed) two-talker masker such as the one employed here (see their Table 4).

Self-perceived improvement in hearing ability obtained from using hearing aids (as assessed by each of the speech, spatial, and qualities sections of the SSQ_B questionnaire) was smaller overall (SSQ_B scores were lower) for participants with better working memory (higher RSpan scores). Self-perceived improvement in speech perception tended to be smaller (SSQ_B-speech scores were lower) for participants with faster rates of recovery from forward-masking (larger FMRR values). Self-perceived improvement in spatial hearing and hearing quality was smaller (SSQ_B-spatial and SSQ_B-qualities scores were lower) for participants with greater BMLDs (larger BMLDs). For each of the three aspects measured by the SSQ questionnaires, self-perceived improvement obtained from hearing-aid treatment was positively correlated with the self-perceived hearing ability at the time of entering the study, although the correlation between SSQ-speech and SSQ_B-speech just missed statistical significance.

The scores for the client-oriented scale of improvement (COSI) questionnaire were higher the greater the hearing-aid amplification at low and moderate levels (larger $REIG_{50dB}$ and $REIG_{65dB}$), and the greater the audibility provided by the hearing aids in quiet (larger SII_Q; Table 8). This result indicated that self-perceived hearing improvement, as assessed by the COSI questionnaire, was proportional to hearing-aid amplification and audibility provided by hearing aids. The scores for the international outcome inventory for hearing aids (IOI-HA) were positively correlated with the SII_Q, indicating that hearing-aid user satisfaction, as assessed by the IOI-HA, was positively affected by good audibility.

Table 8. Pairwise Pearson Correlation (*R*) Between Each Predictor and Each Outcome Measure.

Predictor	SRT$_{\text{N}}$	SSQ$_{\text{B}}$-speech	SSQ$_{\text{B}}$-spatial	SSQ$_{\text{B}}$-qualities	COSI	IOI-HA
Age	0.28*	0.07	0.05	0.01	−0.03	0.01
PTT	**0.36***	−0.08	0.01	−0.03	−0.04	−0.11
SRT$_{\text{Q}}$	**0.56***	−0.04	−0.04	−0.08	−0.10	−0.14
HL$_{\text{OHC}}$	0.32*	−0.01	−0.03	−0.00	−0.05	−0.02
HL$_{\text{IHC}}$	0.26*	−0.11	0.03	−0.07	0.01	−0.13
BMCE	0.21	−0.19	−0.16	−0.08	−0.19	−0.14
DPOAE$_{\text{N}}$	−0.25*	0.03	0.00	0.06	0.15	0.13
DPOAE$_{\mu\text{Pa}}$	−0.11	0.04	−0.02	0.06	0.13	0.04
BMLD	−0.19	−0.13	−0.29*	−0.27*	−0.18	−0.10
FMDT	**0.49***	−0.08	0.02	−0.06	−0.15	−0.12
FMRR	0.24*	−0.27*	−0.19	−0.17	−0.22	−0.13
TI	−0.15	−0.08	−0.22	−0.11	−0.04	−0.15
REIG$_{\text{50dB}}$	0.16	0.06	−0.00	0.07	0.23	0.20
REIG$_{\text{65dB}}$	0.15	0.04	−0.04	0.06	0.25*	0.18
RECE$_{\text{HI}}$	−0.23	−0.04	−0.09	−0.10	−0.09	−0.10
RSpan	−0.15	−0.26*	**−0.42***	−0.29*	−0.15	−0.14
SSQ-speech	−0.20	0.21	0.02	0.00	0.17	0.14
SSQ-spatial	0.04	0.27*	**0.37***	0.31*	0.02	0.03
SSQ-qualities	0.03	0.18	0.26*	0.31*	0.08	0.10
SII$_{\text{Q}}$	**−0.40***	0.20	0.04	0.12	0.30*	**0.39***

Note. Asterisks indicate significant correlations (two-tailed *t* test, $p < .05$); bold font indicates significant correlations with Bonferroni correction for multiple comparisons (two-tailed *t* test, $p < .0025 = 0.05/20$). Although BMLD was not used as a predictor in the PCA-MLR analysis, it is shown here for completeness. See Table A1 in Appendix A for *N* values. PTT = pure tone threshold; SRT = speech reception threshold; HL$_{\text{OHC}}$ = outer hair cell loss; HL$_{\text{IHC}}$ = inner hair cell loss; BMCE = basilar membrane compression exponent; DPOAE = distortion product otoacoustic emission; BMLD = binaural-masking level difference; FMDT = frequency-modulation detection threshold; FMRR = forward-masking recovery rate; TI = temporal integration; REIG = real-ear insertion gain; SSQ = speech, spatial, and qualities; SII$_{\text{Q}}$ = speech intelligibility index in quiet; COSI = client-oriented scale of improvement; IOI-HA = international outcome inventory for hearing aids.

Principal Component Analysis

We identified seven principal components (PCs). Table 9 shows the predictors in each component and their loadings.^{2} Predictors with loadings smaller than 0.3 were regarded as negligible and were omitted from further analysis and in Table 9.

Although an interpretation of the identified components is not straightforward, some analysis is useful. The signs of the loadings in Table 9 indicate that PC1 increased with decreasing the proportion of the speech spectrum that is audible when aided and in quiet (SII$_{\text{Q}}$). PC1 also increased with increasing the audiometric thresholds (PTT), the amount of inner hair cell loss (HL$_{\text{IHC}}$), and the unaided SRT in quiet (SRT$_{\text{Q}}$). This suggests that PC1 broadly reflects speech-spectrum audibility deficits: PC1 was larger with greater deficits.

PC2 increased with increasing hearing-aid gain at low and moderate levels (REIG$_{\text{50dB}}$ and REIG$_{\text{65dB}}$) and decreased with increasing hearing-aid compression exponent at moderate-to-high levels (RECE$_{\text{HI}}$; recall that a smaller compression exponent implies greater compression). This relationship is reasonable considering that hearing aids compress more when they provide larger amplification for low-level sounds. Overall, PC2 reflected hearing-aid settings: PC2 was larger with greater amplification and compression.

PC3 increased with increasing scores in the three speech, spatial, and qualities subscales of the pretrial SSQ questionnaire. Higher SSQ-questionnaire scores imply better baseline hearing ability. Thus, overall, PC3 reflected self-reported baseline hearing abilities: PC3 was larger with better abilities. Interestingly, PC3 decreased with increasing working memory capacity (i.e., with increasing RSpan scores). The interpretation for this result is uncertain.

PC4 increased with increasing basilar membrane compression exponent (BMCE; recall that larger BMCE implies less compression) and with increasing the rate of recovery from forward masking (FMRR; larger FMRR implies faster recovery). PC4 also increased with increasing cochlear mechanical gain loss (HL$_{\text{OHC}}$) and with decreasing inner hair cell loss (HL$_{\text{IHC}}$). Increased gain loss combined with decreased inner hair cell loss and increased BMCE is indicative of linearized

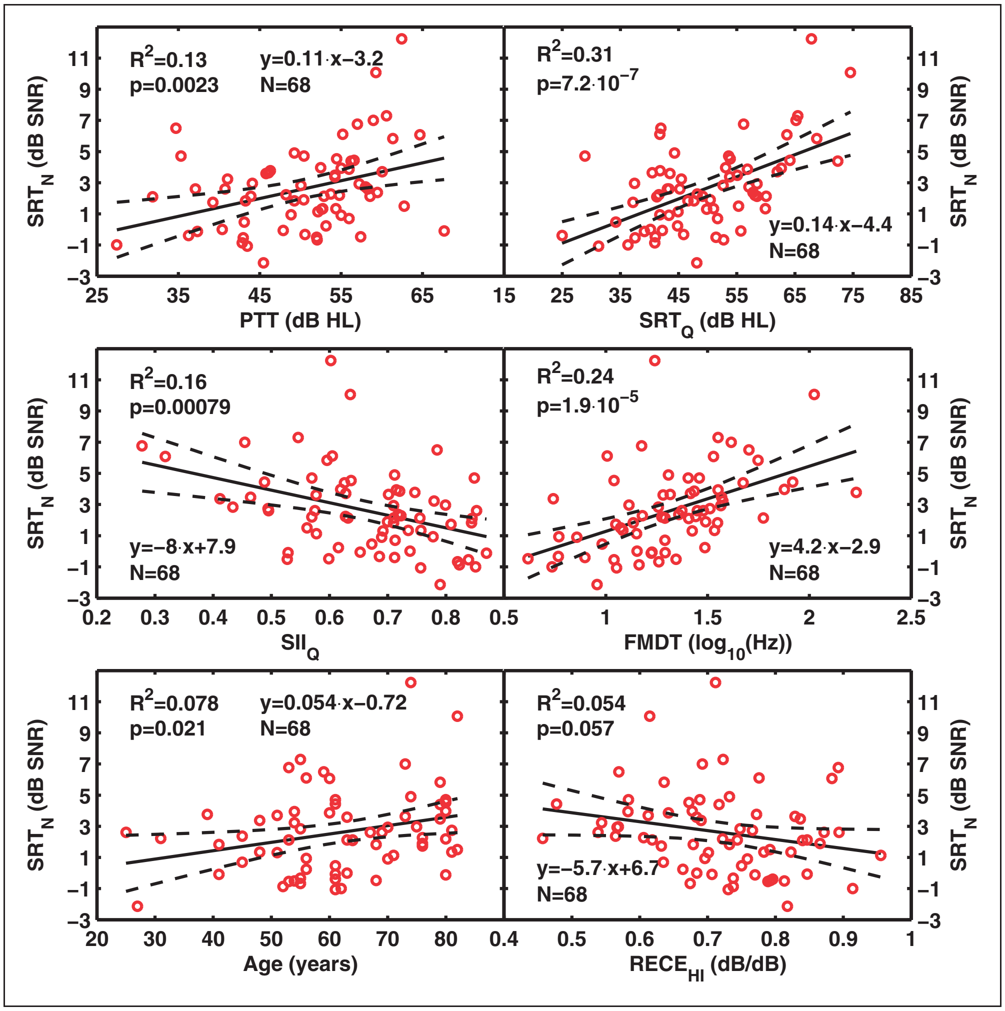

Figure 2. Aided speech-in-noise reception thresholds (SRT_N) against the most significant predictors. Each panel is for a different predictor (PTT, SRT_Q, SII_Q, FMDT, age, and $RECE_{HI}$) as indicated in the abscissa of each panel. Solid lines depict linear regression lines; dashed lines depict the 5 and 95% confidence intervals for the regression line. The inset in each panel gives the proportion of variance of the aided SRT_N (R^2) explained by the different predictors and the probability (p) for the value to occur by chance. Also shown are the regression equation and the number of participants (N).

FMDT = frequency-modulation detection threshold; PTT = pure tone threshold; SRT = speech reception threshold; SII = speech intelligibility index; SNR = signal-to-noise ratio; RECE = real-ear compression exponent.

cochlear input/output curves. Therefore, PC4 possibly reflects cochlear mechanical gain loss: PC4 was larger with greater gain losses.

PC5 increased with increasing age and decreased with increasing working memory capacity (i.e., with increasing RSpan scores) or with increasing self-reported baseline ability in hearing speech (with increasing SSQ-speech scores). This relationship seems reasonable considering that working memory capacity declines with age and so does speech intelligibility even in listeners with normal audiometry. Thus, it seems that PC5 reflected aging effects unrelated to threshold.

PC6 increased with increasing FMDTs and with increasing the unaided SRTs in quiet (SRT_Q). FMDTs index temporal processing capacities, and the fact that SRT_Q was split between PC6 and PC1 (speech-spectrum audibility deficits) suggests that the unaided SRT in quiet (SRT_Q) was concomitantly affected by spectrum audibility deficits and temporal processing capacity, something reasonable. Therefore, PC6 may be interpreted to broadly reflect temporal processing deficits, with larger PC6 values indicating greater deficits. PC6 also increased with increasing hearing-aid compression at moderate-to-high levels ($RECE_{HI}$). The interpretation for this latter result is uncertain.

Lastly, PC7 increased with increasing TI and with increasing the magnitude and extent of $DPOAE_{\mu Pa}$ and $DPOAE_N$, respectively. Conventionally, strong and

Table 9. Principal-Component Factor Loadings for the 19 Predictors.

	PC1 Spectrum audibility deficits	PC2 Hearing-aid settings	PC3 Baseline hearing ability	PC4 Cochlear gain loss	PC5 Aging and cognition	PC6 Temporal processing deficits	PC7 Cochlear compression
Age					0.68		
PTT	0.44						
SRT_Q	0.37					0.34	
HL_{OHC}				0.32			
HL_{IHC}	0.42			−0.32			
BMCE				0.60			
$DPOAE_N$							0.40
$DPOAE_{\mu Pa}$							0.58
FMDT						0.77	
FMRR				0.59			
TI							0.61
$REIG_{50dB}$		0.60					
$REIG_{65dB}$		0.61					
$RECE_{HI}$		−0.30				−0.44	
RSpan			−0.33		−0.53		
SSQ-speech			0.42		−0.43		
SSQ-spatial			0.60				
SSQ-qualities			0.55				
SII_Q	−0.51						

Note. Loadings < 0.3 are omitted. For detailed information, see Appendix A. PTT = pure tone threshold; SRT = speech reception threshold; HL_{OHC} = outer hair cell loss; HL_{IHC} = inner hair cell loss; BMCE = basilar membrane compression exponent; DPOAE = distortion product otoacoustic emission; BMLD = binaural-masking level difference; FMDT = frequency-modulation detection threshold; FMRR = forward-masking recovery rate; TI = temporal integration; REIG = real-ear insertion gain; SSQ = speech, spatial, and qualities; SII_Q = speech intelligibility index in quiet; RECE = real-ear compression exponent; PC = principal component.

extensive DPOAEs are indicative of healthy outer hair cells, an aspect seemingly related to PC4 (cochlear gain loss). Considering, however, that two components cannot reflect the same underlying factor, PC7 must reflect aspects different from cochlear gain loss. The slope of cochlear input/output curves over their compressive region is uncorrelated with cochlear gain loss (Johannesen et al., 2014; Plack, Drga, & Lopez-Poveda, 2004), and DPOAEs appear related to cochlear compression more than to cochlear gain (Shera & Guinan, 1999). Therefore, it is tempting to speculate that PC7 might be related with residual cochlear compression: Larger PC7 values would indicate greater compression. This interpretation, however, appears to clash with BMCE contributing to PC4 rather than PC7.

Predictive Models of Aided Intelligibility in Noise

Stepwise MLR was used to develop a predictive model of aided intelligibility in noise (SRT_N) using the identified components (PC1 to PC7) as predictors. Table 10 gives the resulting model. The model predicted 41% of the SRT_N variance and included (in order of importance) PC6 (temporal processing deficits), PC1 (speech-spectrum audibility deficits), and PC5 (aging and cognition), each of which contributed 28%, 7%, and 6% to the SRT_N variance, respectively. The signs of the coefficients in the model indicate that good aided intelligibility in noise (lower SRT_N) was proportional to (a) good temporal processing (lower PC6), (b) good spectral audibility (lower PC1), and (c) younger age (lower PC5). Strikingly, the model also reveals that the most significant predictor of aided SRT_N was temporal processing deficits (PC1), and that hearing-aid settings (PC2), cochlear mechanical gain loss (PC4), and cochlear compression (PC7) did not contribute significantly to the SRT_N variance.

For practical purposes, it is useful to express the model in Table 10 in terms of the measured predictors rather than the components. To this end, an alternative model was developed using the main predictors in PC6, PC1, and PC5. In developing this model, we omitted predictors with loadings smaller than 0.4 (Table 9). That is, the model was developed using FMDT and $RECE_{HI}$ (for PC6), SII_Q, PTT, and HL_{IHC} (for PC1), and age, RSpan, and SSQ-speech (for PC5).

Table 10. A Predictive Model of Aided Speech Reception Threshold in a (Time-Reversed) Two-Talker Background (SRT_N) Developed Using the Identified Principal Components.

Priority	PCA component	Coefficient	t	p	Accum. R^2
N/A	Intercept	$7.9 \cdot \times 10^{-17}$	$8.2 \cdot \times 10^{-16}$	1.0	N/A
1	PC6 (temporal processing deficits)	0.353	4.4	$4.6 \cdot \times 10^{-5}$	0.28
2	PC1 (spectrum audibility deficits)	0.162	2.9	$5.7 \cdot \times 10^{-3}$	0.35
3	PC5 (aging and cognition)	0.209	2.6	.013	0.41

Note. Columns indicate the component's priority order and name, the regression coefficient, the t value, and corresponding probability for a significant contribution (p), and the accumulated proportion of total variance explained (Accum. R^2), respectively. The priority order was established according to how much the corresponding component contributed to the predicted variance (higher priority was given to larger contributions). The accumulated R^2 is the predicted variance adjusted for the number of variables included in the regression model. PCA = principal component analysis.

Table 11. A Model of SRT_N Equivalent to the Model Shown in Table 10 but Obtained Using Measured Predictors.

Priority	Predictor	Coefficient	t	p	Accum. R^2
N/A	Intercept	5.64	1.8	.080	N/A
1	FMDT	2.97	3.5	.00078	0.23
2	SII_Q	−8.09	−4.1	1.1×10^{-4}	0.34
3	Age	0.050	2.7	.0079	0.38
4	$RECE_{HI}$	−6.51	−2.6	.0123	0.43

Note. The layout is as in Table 10. FMDT = frequency-modulation detection threshold; RECE = real-ear compression exponent; SII_Q = speech intelligibility index in quiet.

Table 12. A Clinical Model of SRT_N.

Priority	Predictor	Coefficient	t	p	Accum. R^2
	Intercept	−1.35	−0.5	.62	–
1	PTT	0.106	3.1	.0026	0.12
2	Age	0.056	2.7	.0094	0.17
3	$RECE_{HI}$	−6.79	−2.5	.0132	0.23

Note. This model was obtained using only the subset of predictors in PC6, PC1, and PC5 with loadings > 0.4 and that may be easily available in a clinical context. The layout is as in Table 10. PTT = pure tone threshold; RECE = real-ear compression exponent.

The resulting model is shown in Table 11. The FMDT was the most significant predictor and explained 23% of the SRT_N variance, followed by the SII_Q, age, and $RECE_{HI}$, which contributed an additional 11%, 4%, and 5% to the predicted variance, respectively. The signs of the coefficients in the model indicate that good aided intelligibility in noise (lower SRT_N) is positively correlated with (a) good frequency modulation detection (lower FMDTs), (b) good aided spectral audibility in quiet (higher SII_Q), (c) younger age, and (d) less hearing-aid compression at moderate-to-high levels (higher $RECE_{HI}$). Again, the model in Table 11 reveals that the most significant predictor of aided intelligibility in a (time-reversed) two-talker masker (SRT_N) was frequency modulation detection, a measure of supra-threshold temporal processing capacity.

Alternative, Clinical Models

One aim of the present study was to help clinicians predict aided speech-in-noise performance for a given individual at the time when the hearing aid is prescribed. The models in Tables 10 and 11 would be hardly useful for this purpose because they involve measuring variables such as FMDT or the aided SII_Q that are not readily available in the clinic. We obtained an alternative predictive model of SRT_N using only the main predictors in PC6, PC1, and PC5 that are or could be reasonably easily available to audiologists: $RECE_{HI}$, PTT, age, RSpan, and SSQ-speech. Table 12 shows the resulting model. The most significant predictor was the mean audiometric threshold (PTT), followed by age, and hearing-aid compression exponent at mid-to-moderate levels ($RECE_{HI}$), each of which contributed 12%, 5%, and 6% to the predicted variance, respectively. In total, the model explained 23% of the SRT_N variance. The signs of the coefficients in the model indicate that good aided intelligibility in noise (lower SRT_N) is related to better thresholds (lower PTT), younger age, and less hearing-aid compression at moderate-to-high levels (larger $RECE_{HI}$).

In developing the clinical model in Table 12, we disregarded the unaided SRT in quiet (SRT_Q) as a predictor because its loadings in PC6 and PC1 were less than the chosen cutoff value of 0.4 (Table 9). SRT_Q, however, was split in two components and its loadings just missed our criterion (0.37 in PC1 and 0.34 in PC6). Also, SRT_Q is routinely available in the clinic. For these reasons, we tried developing an alternative clinical model considering SRT_Q in addition to the other clinical predictors. Interestingly, the resulting model only had SRT_Q as a predictor (coefficient = 0.134, $t = 5.2$, $p = 1.9 \cdot \times 10^{-6}$) and explained 30% of the SRT_N variance. In other words, the unaided intelligibility in quiet (SRT_Q) was

Table 13. Pairwise Pearson Correlations Between Self-Reported Improvement in Hearing Abilities as Assessed by the SSQ_B-, COSI- and IOI-HA-Questionnaire Scores.

	SSQ_B-spatial	SSQ_B-qualities	COSI	IOI-HA
SSQ_B-speech	0.70*	0.67*	0.71*	0.73*
SSQ_B-spatial	–	0.83*	0.37*	0.46*
SSQ_B-qualities	–	–	0.46*	0.51*
COSI	–	–	––	0.75*

Note. Asterisks indicate significant correlations with Bonferroni correction for multiple comparisons ($N = 68$, two-tailed t test, $p < .05/10$). SSQ = speech, spatial, and qualities; COSI = client-oriented scale of improvement; IOI-HA = international outcome inventory for hearing aids.

Table 14. Predictive Models of Self-Reported Improvement in Hearing Abilities as Assessed by the SSQ_B-Questionnaire Scores.

Priority	Predictor	Coefficient	t	p	Accum. R^2
SSQ_B-spatial					
N/a	Intercept	1.18	1.3	.20	N/A
1	RSpan	−0.055	−2.9	.0058	0.16
2	SSQ-Spatial	0.244	2.2	.029	0.21
SSQ_B-qualities					
N/A	Intercept	−1.11	−0.95	.34	N/A
1	SSQ-Qualities	0.397	2.7	.0096	0.08

Note. A model is not shown for SSQ_B-speech scores because SSQ_B-speech scores were not correlated with any of the principal components (see main text for details). The layout is as in Table 10. SSQ = speech, spatial, and qualities.

the best single *clinical* predictor of aided intelligibility in a (time-reversed) two-talker background (SRT_N).

Predictive Models of Subjective Outcome Measures

Scores for the COSI, IOI-HA, and SSQ_B questionnaires were significantly correlated with each other, particularly with the scores for the speech section of the SSQ_B questionnaire (Table 13). For this reason and for conciseness, we only investigated predictive models of SSQ_B scores.

First, we investigated MLR models using the identified components as predictors. Self-reported improvement in hearing speech (SSQ_B-speech scores) was not correlated with any of the components and so it was not possible to obtain a predictive model for SSQ_B-speech scores. The scores for both SSQ_B-spatial and SSQ_B-qualities were correlated with PC3 only (pretrial self-reported hearing ability). PC3 explained 10% ($t = 2.8$, $p = .0072$) and 5% ($t = 2.1$, $p = .045$) of the variance in SSQ_B-spatial and SSQ_B-qualities scores, respectively. The coefficients in the two models were positive (0.225 for SSQ_B-spatial and 0.170 for SSQ_B-qualities), indicating that self-reported improvement in spatial and qualities hearing increased with increasing PC3 (self-reported pretrial hearing ability).

Second, we investigated models of SSQ_B-spatial and SSQ_B-qualities scores using the four measured variables that contributed to PC3, regardless of their loadings (RSpan, SSQ-speech, SSQ-spatial, and SSQ-qualities; Table 9). The obtained models are shown in Table 14. The models predicted 21% and 8% of the SSQ_B-spatial and SSQ_B-qualities scores, respectively. Working memory capacity (RSpan scores) was the most significant predictor of SSQ_B-spatial scores (explaining 16% of its variance) but was not a predictor of SSQ_B-qualities scores. The sign of its coefficient in the model indicates that hearing-aid treatment provided less benefit for spatial hearing to participants with better working memory capacity. The models also suggested a weak trend for self-perceived hearing-aid benefit to be greater for participants with better self-perceived baseline hearing abilities.

Discussion

Models of Aided Intelligibility in Noise

Speech-in-noise intelligibility is the most sought improvement among hearing-aid users (Kochkin, 2002). For this reason, we have used aided intelligibility in a (time-reversed) two-talker background (SRT_N) as a measure of hearing-aid outcome. We have proposed a predictive MLR model for this outcome (Table 10) based on seven PCs inferred from 19 measured predictors (Table 9). In the model, the component interpreted as "temporal processing deficits" (PC6) explained 28% of SRT_N variance, followed by a component interpreted as "speech-spectrum audibility deficits" (PC1) and by a component interpreted as "aging" (PC5), each of which explained 7% and 6% of the SRT_N variance, respectively. The other four components (PC2: hearing-aid settings, PC3: baseline self-reported hearing ability, PC4: cochlear gain loss, and PC7: cochlear compression) did not contribute significantly to the SRT_N variance.

Given the somewhat subjective interpretation of the identified components, we have also proposed two MLR models based on the subset of measured predictors with higher loadings on the components. One model (Table 11) was obtained using all of the predictors in an attempt to pinpoint the main factors affecting SRT_N. An alternative model (Table 12) was aimed as a guide for audiologists in setting patients' expectations about this outcome and was thus obtained using the subset of predictors that would be easily available in the clinic.

The model obtained considering all predictors explained 43% of the SRT_N variance (Table 11).

In this model, sensitivity to frequency modulation (FMDT) came up as the most significant predictor of aided intelligibility in noise: the greater the sensitivity to frequency modulation, the better the aided intelligibility in noise. FMDT alone explained 23% of the SRT_N variance. Assuming that FMDT indexes supra-threshold processing of temporal fine structure (Moore & Sek, 1996) then the present finding suggests that supra-threshold temporal fine structure processing abilities is an important factor for aided speech intelligibility (in a time-reversed two-talker background at least). This finding is consistent with earlier studies (e.g., Hopkins & Moore, 2011; Johannesen et al., 2016; Lorenzi et al., 2006; Pichora-Fuller et al., 2007; Strelcyk & Dau, 2009).

The aided SII_Q was the second most significant predictor and explained 11% of the SRT_N variance (Table 11). SII_Q estimates the proportion of the speech spectrum that is audible in quiet when wearing hearing aids, and the sign of its coefficient in the model indicated that aided intelligibility in noise improved (SRT_N was lower) with increasing spectral audibility in quiet (larger SII_Q). This seems reasonable considering that some participants could have suffered from reduced (spectral) audibility despite their wearing a hearing aid. This might have happened if some participants preferred less amplification than recommended by the hearing-aid gain prescription rule, possibly to improve listening comfort at the expense of reducing intelligibility in noise (Humes, 2002), or if the amount of amplification for low-intensity sounds was insufficient for some hearing-aid users despite our efforts to provide all users with sufficient amplification.

After accounting for temporal processing deficits and spectral audibility deficits, age came up as the third significant predictor of aided intelligibility in noise and explained 4% of the SRT_N variance (Table 11). Elderly listeners tended to perform worse in noise with their hearing aids than younger listeners did. Age can degrade hearing and cognition in multiple ways. We tried to isolate some of those ways by using well-defined predictors representing cochlear mechanical dysfunction, auditory temporal processing, and working memory capacity. The fact that age remained a significant predictor of aided intelligibility once all those predictors were factored out indicates that age represents aspects not accounted for by the other predictors. The aspects in question are uncertain. Perhaps, age represents cognitive decline different from working memory capacity, such as processing speed (Salthouse, 1996) or attention (Craik & Byrd, 1982), both of which might affect performance in demanding speech-in-noise situations (e.g., Cahana-Amitay et al., 2016; Oberfeld & Klöckner-Nowotny, 2016). Alternatively, age might represent temporal processing deficits not captured by the temporal processing predictors employed here.

Lastly, aided intelligibility in noise tended to be better with more linear amplification for high-intensity sounds (i.e., SRT_N was lower—better—with increasing $RECE_{HI}$; Table 11). This may be interpreted to reflect that hearing-aid users preferred more linear amplification at high intensities. The reason is uncertain. Perhaps excessive hearing-aid compression alone or combined with residual cochlear compression at high intensities (65–80 dB SPL) generated excessive distortion that degraded the temporal representation of speech (recall that $RECE_{HI}$ was a contributor to component PC6: temporal processing deficits) and hindered intelligibility (Bode & Kasten, 1971; Boothroyd, Springer, Smith, & Schulman, 1988; Marriage, Moore, Stone, & Baer, 2005).

The clinical model included three predictors: PTT, age, and $RECE_{HI}$ (Table 12). Thus, it was similar to the model obtained with all predictors except that it included the audiometric thresholds (PTT) instead of the aided SII_Q. The PTT and SII_Q, however, are broadly equivalent. Indeed, they contributed to the same factor (PC1, spectral audibility deficits, Table 9) and explained approximately the same amount of SRT_N variance (11%–12%). However, a most important difference between the clinical and the all-predictor models is that the clinical model explained only half the SRT_N variance of the all-predictor model (23% vs. 43%). This is because FMDT was disregarded as a candidate predictor in developing the clinical model (because FMDT is not a clinical measure) and suggests that it would be useful to include some index of temporal processing capacity to help audiologists predict aided performance in noise.

Interestingly, we found an alternative clinical model where the unaided speech intelligibility in quiet (SRT_Q) alone explained 30% of the variance in aided intelligibility in noise (SRT_N). That is, hearing-aid users who required lower intensities to achieve 50% word recognition in quiet without their hearing aids tended to tolerate higher noise levels to achieve 50% sentence recognition with their hearing aids. This result might seem trivial considering that both the predictor (SRT_Q) and outcome (SRT_N) variables are measures of intelligibility. However, SRT_Q is a measure of *unaided intelligibility in quiet* while SRT_N is a measure of *aided intelligibility in a (time-reversed) two-talker background*. Therefore, this finding suggests that the limiting factors for good intelligibility in noise with hearing aids may be related to the limiting factors of intelligibility in quiet without hearing aids. The unaided SRT in quiet (SRT_Q) is routinely measured in audiology clinics. Therefore, this finding demonstrates that the unaided SRT_Q is actually more helpful than the model in Table 12 as a rough guide for audiologists to predict aided performance in noise at the time when the hearing aid is prescribed.

We note that aided intelligibility in a (time-reversed) two-talker background (SRT_N) was not correlated with working memory capacity (Table 8). This was surprising at first given the body of evidence in support for a link between those two variables (reviewed by Akeroyd, 2008). However, all evidence of a link between working memory capacity and speech-in-noise intelligibility comes from studies using older, hearing-impaired listeners. Füllgrabe and Rosen (2016b) failed to find evidence for such a link in young and normal-hearing listeners. In addition, Füllgrabe and Rosen (2016a) demonstrated that age can be a moderating variable of the relationship between working memory capacity and speech-in-noise intelligibility. Given the wide age range of the present sample (25–82 years), and in the light of these more recent studies, the absence of a significant correlation might not be entirely unexpected.

Models of Subjective Outcome Measures

In addition to intelligibility in a (time-reversed) two-talker background, we also assessed the improvement in hearing ability obtained from using hearing aids using three popular questionnaires (SSQ_B, COSI, and IOI-HA). Most of the scores for the three questionnaires were reasonably highly correlated with each other (Table 13). Therefore, for conciseness, we developed predictive models for SSQ_B questionnaire scores only. In general, we found it difficult to predict self-reported hearing improvement based on the present set of PCs or predictors. Notably, we could not find a model to predict improvement in hearing speech (SSQ_B-speech scores). Working memory capacity (as assessed by the reading span test) was found to be a predictor of improvement in spatial hearing (Table 14). The sign of the regression coefficient indicates that good working memory is related with smaller improvement. Perhaps the younger listeners (who probably had larger working memory capacity) might have had milder hearing losses, yielding less room for improvement. We also found pretrial self-reported hearing ability in spatial and qualities hearing to be weak (but significant) predictors of self-reported improvement in spatial and qualities hearing, respectively. This finding indicates that hearing-aid users reporting better baseline (unaided) hearing tended to report comparatively larger (but nevertheless small) benefits from using their hearing aids. The reason is uncertain. Perhaps, the positive correlation is the result of differences in participants' optimism. That is, people who generally view things favorably might be more inclined to rate positively their hearing abilities while unaided and also the degree of improvement that hearing aids offer them. In any case, we note that this result is broadly in line with the finding that intelligibility in noise when aided tended to be better for hearing-aid users with good baseline (unaided) intelligibility in quiet (Table 8).

Perez, McCormack, and Edmonds (2014) concluded that hearing-aid candidates with good sensitivity to temporal fine structure (as measured binaurally) reported larger improvements from using their hearing aids. The present results appear inconsistent with those of Perez et al., as we found no correlation between our (monaural) measure of temporal fine structure sensitivity (FMDT) and SSQ_B scores (Tables 8 and 14). We also found a negative correlation between BMLDs (a binaural measure of temporal fine structure sensitivity) and SSQ_B scores (Table 8). We note, however, that the study of Perez et al. focused on older listeners (51–85 years, mean age = 72.2 years) while our participants were, on average, younger and covered a wider age range (25–82 years, mean = 61 years). Perhaps, the measures of temporal fine structure sensitivity used in the two studies are not equivalent, or the conclusion of Perez et al. holds for older listeners only.

On the Unexplained Variance

Even the best of the present models accounted for only 43% of the variance in aided speech-in-noise intelligibility (Table 11). Also, the best models accounted for only 0%, 21%, and 8% of the variance in self-reported improvement in the speech, spatial, and qualities subscales of the SSQ_B questionnaire, respectively (Table 14). These values seem small even after allowing for the test–retest variability in both predictor and outcome values (as an example, the correlation between test–retest SRT_N estimates was 0.86). Our main aim was not to develop fully predictive models for these outcomes. Nonetheless, these values are admittedly smaller than expected considering the number and diversity of predictors and the methodological care exercised in measuring them. The reason is uncertain, but one limitation of the present study is that we did not measure the initial motivation to wear hearing aids, which is an important factor in self-perceived benefit from using hearing aids (e.g., Brooks, 1989; Lewsen & Cashman, 1997).

Humes (2002) proposed a model that explained a larger proportion (68%) of the variance in aided speech recognition using five components (hearing loss, nonverbal IQ+aging, verbal IQ, DPOAEs, and miscellaneous). Methodological differences possibly explain the difference in predictive power between the model of Humes and the present model. Notably, Humes employed linear amplification and his speech recognition outcome combined measures in quiet and in competition with multitalker babble. Here, we used multiband, nonlinear amplification, and the SRT_N was measured using a more fluctuating competitor (two-talker babble) played in reverse to minimize informational masking (see

Appendix A). This could have made the outcome used by Hume more susceptible to spectral audibility deficits and ours more susceptible to temporal processing deficits. This is supported by the fact that ''hearing loss'' was the most significant predictor of Humes' outcome (accounting for 54% of its variance), while temporal processing deficits was the most significant predictor of our outcome (accounting for 23% of $\mathrm{SRT_N}$ variance).

Of course, speech-in-noise performance with hearing aids may be assessed in many different, and not always equivalent, ways. Hence, the relative importance of the myriad of possible predictors will almost certainly vary depending on the chosen measure. It is possible that the present set of predictors would account for a larger amount of variance in the scores for different aided speech recognition tasks. However, insofar as temporal processing is important for speech-in-noise recognition (e.g., Lopez-Poveda, 2014; Lopez-Poveda & Barrios, 2013; Lorenzi et al., 2006; Pichora-Fuller et al., 2007), hearing-aid outcome studies should include some tests or conditions that are sensitive to supra-threshold frequency modulation detection deficits.

Even larger was the proportion of variance unaccounted for in self-perceived improvement in hearing ability (Table 14). Cox, Alexander, and Gray (2007) showed that the audiogram is a negligible predictor of subjective fitting outcomes and that 20% to 30% of the variance in subjective outcomes can be accounted for by patient variables that can be measured before the fitting, such as reported hearing problems. Among the present predictors (Table 8) were several objective measures of hearing capacity (including audiometric thresholds) and yet the only relevant predictors for this outcome were working memory capacity and self-perceived (unaided) hearing ability at the time of entering the study. Therefore, the present results are broadly in line with the findings of Cox et al. (2007).

On the (Un)Importance of Compensating for Individualized Cochlear Mechanical Dysfunction

The main aim of the present study was to test the hypothesis that the variability in outcomes across hearing-aid users is related to the extent that the hearing aid compensates for the individual degree of cochlear mechanical dysfunction. The results do not support this hypothesis. The PCs interpreted as hearing-aid settings (PC2), cochlear mechanical gain loss (PC4), or residual cochlear compression (PC7) did not contribute significantly to either of the two main hearing-aid outcomes considered here ($\mathrm{SRT_N}$ or $\mathrm{SSQ_B}$-questionnaire scores). This suggests that specific knowledge about the contribution of inner ($\mathrm{HL_{IHC}}$) or outer hair cell ($\mathrm{HL_{OHC}}$) dysfunction to the audiometric loss, or the amount of residual cochlear compression (BMCE) adds little to the information provided by the audiogram with respect to predicting the hearing-aid outcomes considered in the present study. This finding contradicts the opinion put forward elsewhere (e.g., Johannesen & Lopez-Poveda, 2008; Mills, Feeney, & Gates, 2010; Müller & Janssen, 2004) that hearing aids might be better if they were to compensate for the *individual* loss of cochlear nonlinearity.

Instead, the present results demonstrate that self-reported hearing-aid benefit is hardly predictable (by the present set of predictors at least; Table 14) and that auditory temporal processing deficits (PC6), as indexed by FMDT, is the most significant limiting factor for good aided intelligibility in a (time-reversed) two-talker background (Tables 10 and 11). The latter conclusion is broadly in line with that of Johannesen et al. (2016), who reported on the relative importance of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of speech in noise for the same group of hearing-impaired listeners tested here but treated with linear amplification rather than with nonlinear hearing aids. We note, however, that Johannesen et al. (2016) concluded that residual cochlear compression was a significant factor for intelligibility in steady-state noise, even though it was not a predictor of intelligibility in a (time-reversed) two-talker background.

A Final Remark on Possible Effects of Hearing-Aid Compression Speed

Hearing aids can apply dynamic range compression at different speeds, and the speed of compression might affect speech outcomes. Theoretically, slow-acting compression grants a near constant hearing-aid gain-frequency response (over the duration of a few words) and preserves the differences between the short-term spectra in speech, while fast-acting compression can improve speech intelligibility by amplifying very weak speech segments in the temporal dips of the background noise while maintaining louder speech segments at a comfortable loudness (reviewed by Lunner, Rudner, & Rönnberg, 2009). Experimentally, the significance of compression speed for speech intelligibility is still a matter of debate. Some studies have reported no effect of compression speed on speech-in-noise intelligibility (e.g., Moore, Stainsby, Alcántara, & Kühnel, 2004), while others have found fast-acting compression to be superior to slow-acting compression (e.g., van Toor & Verschuure, 2002), and yet others have found slow-acting compression to be superior to fast-acting compression (Hansen, 2002). Other studies have reported large individual differences in the relative benefit of slow- and fast-acting compression for intelligibility in competing speech tasks (Moore, Füllgrabe, & Stone, 2010), or the differential benefit from using fast- versus slow-acting compression to be correlated with cognitive performance (Lunner & Sundewall-Thorén, 2007; Lunner et al., 2009).

The present hearing aids applied slow-acting compression (see Appendix A). Therefore, it is uncertain to what extent the present conclusions would hold for hearing aids with fast-acting compression, or for compression speed optimized individually.

Conclusions

1. Aided intelligibility in a (time-reversed) two-talker background tends to be better for hearing-aid users with good sensitivity to frequency modulation and younger age. Intelligibility tends to improve by increasing amplification for low-intensity sounds and by using more linear amplification for high-intensity sounds.
2. Of these predictors, supra-threshold temporal processing deficits (as indexed by FMDTs) is the most significant limiting factor for good aided intelligibility in a (time-reversed) two-talker background.
3. The unaided SRT in quiet is the single best *clinical* predictor of aided intelligibility in a (time-reversed) two-talker background. This information may be useful for clinicians in setting hearing-aid user expectations.
4. The amount of audiometric loss attributable to loss of cochlear amplification, or the amount of residual cochlear compression, alone or in combination with hearing settings, do not help predict aided speech intelligibility in a (time-reversed) two-talker background.
5. Hearing-aid users that report better baseline (unaided) hearing abilities tend to report greater benefits from using their hearing aids, although this relationship is weak.
6. Hearing-aid users with better working memory capacity tend to report smaller improvement from using hearing aids, although this relationship is weak.

Appendix A: Detailed Methods

Participants, Demographic, and Audiological Predictors

We screened 253 volunteers as candidates to take part in the study. Each of them had a medical interview to provide demographic data (including age and sex), information about their health and drug treatments (current and past), and information about the etiology and time-course of their hearing losses. Volunteers were specifically asked whether they (a) had used hearing aids at any time before entering the study, (b) been regularly exposed to high-intensity sounds, and (c) were suffering from tinnitus. For each of their ears, air- and bone-conduction thresholds as well as uncomfortable loudness levels at frequencies of 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 4.0, and 8 kHz (ANSI, 2004) were measured using a clinical audiometer (Interacoustics AD229e). Tympanometry and acoustic stapedial reflexometry were also applied using a clinical tympanometer (Interacoustics AT235h). We followed standard clinical audiological procedures (British Society of Audiology, 2011).

Only 68 volunteers with symmetric, sensorineural hearing loss requiring hearing-aid treatment were admitted to the study. A hearing loss was regarded as sensorineural when tympanometry was normal and air-bone gaps were $\leqslant 15$ dB at one frequency and $\leqslant 10$ dB at any other frequency. A hearing loss was regarded as symmetrical when the mean air-conduction thresholds at 0.5, 1, and 2 kHz differed by less than 15 dB between the two ears, and the mean difference at 3, 4, and 6 kHz was <30 dB (AAO-HNS, 1993). Participants were regarded as hearing-aid candidates when their hearing loss exceeded 35 dB HL averaged across 0.5, 1, and 2 kHz or exceeded 45 dB HL averaged across 3, 4, and 6 kHz.

The unaided SRT in quiet ($\mathrm{SRT_Q}$) was measured for each ear separately using a standard clinical procedure. Lists of 25 disyllabic Spanish words were presented at several sound intensities (in dB HL), and the percentage of correctly repeated words at each intensity was noted to obtain a performance-intensity function (Cárdenas & Marrero, 1994). At least one word list was presented at an intensity lower than that required for 50% correct identification. The unaided $\mathrm{SRT_Q}$ was estimated as the 50% correct point in a sigmoidal function fitted (by least squares) to the performance-intensity function.

Although audiometric thresholds, air-bone gaps and uncomfortable loudness levels were available at 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 4.0, and 8 kHz, only the values at 0.5, 1, 2, 4, and 6 kHz were used to obtain the corresponding predictors of hearing-aid outcomes in the PCA-MLR analysis.

Hearing Aids

All participants wore receiver-in-canal style Starkey™ hearing aids in both ears throughout the duration of the experiment. The hearing aids were worn with molds that were customized to each participant's ears. These were precommercial release devices at the start of the experiment, but which subsequently became available commercially with little or no modification as Starkey™ X-Series hearing aids. Other than multiband compression, expansion, and output compression limiting and feedback cancellation, all other features in the hearing aid were inactivated for the trial. The time constants for compression were 5 milliseconds for the attack time and 2 seconds for the release time.

Indicators of Cochlear Mechanical Dysfunction

Behaviorally inferred indicators. Cochlear mechanical dysfunction was assessed from cochlear input/output (I/O) curves inferred using a behavioral technique known as the temporal masking curve method (Nelson, Schroder, & Wojtczak, 2001; Lopez-Poveda, Plack, & Meddis, 2003). For each participant, I/O curves were inferred for test frequencies of 0.5, 1, 2, 4, and 6 kHz to assess mechanical dysfunction over a wide range of cochlear regions. The procedure and the resulting I/O curves have been published and analyzed elsewhere and so the reader is referred to the relevant study for detailed information (Johannesen et al., 2014).

From the analysis of the I/O curves, three predictors were obtained:

1. Outer hair cell loss (HL_{OHC}, in dB) defined as the contribution of cochlear mechanical amplification loss to the audiometric loss. It was estimated from comparisons of the compression threshold of each I/O curve with a corresponding reference I/O curve for normal-hearing listeners (see Johannesen et al., 2014; Lopez-Poveda & Johannesen, 2012).
2. Inner hair cell loss (HL_{IHC}, in dB) defined as the difference between the total audiometric loss (in dB HL) and HL_{OHC}.
3. The cochlear compression exponent (BMCE, in dB/dB) defined as the slope of the I/O curve over the range of input levels where compression occurred.

Johannesen et al. (2014) reported that they could not measure I/O curves for participants and test frequencies where the audiometric loss was too high. For the present analysis, we assumed that those cases were indicative of total cochlear amplification loss. Therefore, for those cases, BMCE was set equal to 1 dB/dB, corresponding to a linear I/O curve, and HL_{OHC} was set equal to the maximum possible cochlear gain observed for NH listeners. The latter values were 35.2, 43.5, 42.7, 42.7, and 42.7 dB at 0.5, 1, 2, 4, and 6 kHz, respectively, as reported in Johannesen et al. (2014, p. 11).

For each participant, HL_{OHC}, HL_{IHC} and BMCE were inferred for test frequencies of 0.5, 1, 2, 4, and 6 kHz and across-frequency ANSI-weighted mean values used as predictors in the PCA-MLR analysis.

Physiological Indicators: DPOAEs. Cochlear mechanical dysfunction was also assessed physiologically using DPOAEs. The healthy cochlea is a nonlinear system (Robles & Ruggero, 2001) and generates intermodulation distortion products when stimulated with two simultaneous pure tones (with frequencies f_1 and f_2). These distortion products propagate back to the ear canal, where they can be measured as weak sounds (Kemp, 1978). Here, we measured the ear-canal level of the $2f_1-f_2$ component of the emitted distortion to assess the physiological status of the cochlear mechanical nonlinearity. For convenience, we will refer to this level as the DPOAE.

For each participant, DPOAEs were measured for four f_2 frequencies (0.5, 1, 2, and 4 kHz), and for eight primary tone L_2 levels (from 35 to 70 in 5-dB steps) to assess cochlear dysfunction over a wide range of cochlear regions and stimulus levels. The primary frequency ratio was $f_2/f_1 = 1.2$. Whenever possible, DPOAEs were measured using individually optimal L_1 levels; that is, L_1 levels that produced the largest DPOAE for each L_2 and for each participant. When no optimal L_1 - L_2 relationship was found, we tried to measure DPOAEs using the L_1 - L_2 rule of Neely, Johnson, and Gorga (2005). A DPOAE response was considered present when it exceeded the level of distortion generated by the equipment and the noise floor by at least 6 dB (Johannesen & Lopez-Poveda, 2008). Each DPOAE measurement was attempted three times, and a reading was regarded as valid only when the DPOAE was present at least twice. The mean DPOAE response was noted for further analysis. For $f_2 = 0.5$ kHz, it was additionally required that the standard deviation across measurements was less than 5 dB. For detailed information about equipment and procedures, see Johannesen and Lopez-Poveda (2008).

The potential 32 DPOAE readings (4 frequencies × 8 levels) were collapsed into a single value for use in the PCA-MLR analysis. The collapsing was first done across L_2 levels to obtain a single DPOAE-related metric per test frequency. Next, the resulting frequency-specific values were weighted according to each frequency's importance for speech perception (ANSI, 1997) and summed across frequencies. The resulting number was used in the PCA-MLR analysis. Several collapsing criteria were initially considered of which the following two were finally used as possible predictors of outcomes:

1. The number of L_2 levels for which DPOAEs were present ($DPOAE_N$). At each test frequency, this variable could be equal to any integer number between 0 and 8, depending on the number of L_2 levels for which DPOAEs were present. This variable likely reflects the input level range where cochlear compression exists because DPOAEs are generated by cochlear nonlinearity. The variable disregards the DPOAE response amplitudes because the response amplitudes are highly variable both in listeners with normal hearing and hearing loss (Gorga et al., 1997).
2. Total DPOAE pressure ($DPOAE_{\mu Pa}$, in dB SPL). At each test frequency, this variable was calculated as the sum of the DPOAE amplitudes (in μPa) recorded at the eight L_2 levels and converted back into decibels (Reavis et al., 2011). This variable is an overall

measure of the total DPOAE response amplitude. The emphasis is on the larger responses because the summation of responses across L_2 levels is done in μPa and hence tends to disregards small responses typical of low L_2 levels.

Indicators of Temporal Representation and Processing Capacity

Binaural masking level difference. The BMLD (in dB) is a measure of the improvement in the detection threshold of a pure tone signal masked by noise when the tone is anti-correlated (180° out-of-phase) in the two ears relative to that for an interaurally correlated pure tone masked by noise. The BMLD involves binaural hearing abilities. We nonetheless decided to include it under the temporal processing category because performance in this test not only requires being able to process the time variation of the interaural statistics of the tone and the noise (e.g., Durlach, 1963) but also reflects the fidelity of coding temporal information at each ear.

A dynamic version of the BMLD measurement was used in the current study. In the dynamic test, the BMLD is the improvement in detection threshold of a tone in noise whose interaural correlation varies in time relative to that of a tone in noise that is correlated at the two ears and whose interaural correlation is unchanging in time. The dynamic test was used because dynamic BMLD appears to be correlated with self-reported auditory disability (Gatehouse & Akeroyd, 2006). Specifically, dynamic BMLD appears related to four items in the speech section of the SSQ questionnaire (Gatehouse & Noble, 2004): "speech in noise," "speech in speech contexts," "multiple speech-stream processing and switching," and "identification of sounds and objects," which suggests that BMLD might predict speech-in-noise intelligibility.

Our BMLD experiment was almost identical to that of Gatehouse and Akeroyd (2006). We measured the binaural masked threshold of a 500 Hz pure tone embedded in a binaural noise with a bandwidth from 0.25 to 1 kHz. The tone had a duration of 30 ms including 15-ms cosine-squared onset and offset ramps. The duration of the noise was 500 ms with 20-ms onset and offset ramps. The tone was centered in time with the noise. The level of the noise was fixed either 30 dB above the absolute threshold for the noise or at 50 dB/Hz, whichever was higher. Two binaural masked conditions were measured. In the first condition, both the tone and the masker noise had the same phase at the two ears. In the second condition, the tone was presented with opposite phase between ears and the noises presented to the two ears had a correlation coefficient between the two ears that varied from -1 to 1 with a rate of 2 Hz. The dynamic BMLD was calculated as the difference of the tone threshold between the two conditions. The tone level at threshold was measured using a two-interval, two-alternative forced-choice adaptive procedure with feedback. The initial tone level was set sufficiently high that the listener always could hear the tone. Tone level was then changed according to a two-down, one-up adaptive procedure to estimate the 71% point on the psychometric function (Levitt, 1971). An initial step size of 6 dB was applied, which was decreased to 2 dB after three reversals. The adaptive procedure continued until a total of 12 reversals in tone level were measured. Threshold was calculated as the mean tone level at the last 10 reversals. A measurement was discarded if the standard deviation of the last 10 reversals exceeded 6 dB. Three threshold estimates were obtained in this way, and their mean was taken as the threshold. If the standard deviation of these three measurements exceeded 6 dB, one or more additional threshold estimates were obtained and included in the mean.

Prior to the BMLD task, the absolute threshold of the noise was measured using a two-alternative forced-choice procedure in which the level of the noise was varied in successive trials according to an adaptive two-down, one-up rule to estimate the 71%-correct point in the psychometric function (Levitt, 1971).

Forward masking recovery rate. The rate of recovery from forward masking (FMRR) is a measure of the ability to hear a signal preceded by an intense masker sound. It was included as a possible predictor of hearing-aid outcomes because (a) fast recovery is thought to be related with the ability to perceive weak phonemes preceded by more intense phonemes in running speech (Gregan, Nelson, & Oxenham, 2013) and (b) it is thought to reflect neural recovery from previous stimulation (Oxenham, 2001).

Here, FMRR was assessed by measuring the level of a tonal masker required to just mask a brief fixed-level tonal probe as a function of the time interval between the masker offset and the probe onset. The slope of function relating masker level against masker-probe time interval (in units of dB/ms) was regarded as the recovery rate. In general, the recovery rate can depend on cochlear mechanical compression and on the postcochlear recovery rate (Nelson et al., 2001). To minimize the confounding effects of compression, the probe frequency was set above 2 kHz, and masker frequency was set an octave (or more) below the probe frequency, a condition where cochlear responses are assumed to be linear (i.e., free from cochlear mechanical compression; Lopez-Poveda et al., 2003). Although forward masking recovery was measured for a single probe frequency, there is evidence that the recovery rate is approximately independent of probe frequency (Lopez-Poveda et al., 2003; Pérez-González et al., 2014).

The procedure for measuring FMRR and the obtained values have been published and analyzed elsewhere (Johannesen et al., 2014) and so the reader is referred to the relevant study for detailed information.

Frequency-modulation detection threshold. The FMDT is a measure of the ability to detect changes in the sound waveform. This variable was included as a candidate predictor of hearing-aid outcomes because the reduced speech intelligibility of listeners with hearing loss appears to be associated with their (in)ability to use the information conveyed in the rapid temporal changes of speech sounds, known as temporal fine structure (Lorenzi et al., 2006). FMDTs are thought to provide a measure of this ability as they presumably depend on the quality with which frequencies are coded in the phase locking of auditory nerve discharges or on the ability of a listener to discriminate frequencies based on such a code (Moore & Sek, 1996). Furthermore, FMDTs are correlated with unaided SRTs in a two-talker masker condition (Strelcyk & Dau, 2009).

Here, the FMDT is defined as the minimum detectable excursion in frequency for a pure tone carrier of 1500 Hz modulated in frequency at a rate of 2 Hz. The level of the tone was set 30 dB above the absolute threshold for the carrier tone (30 dB sensation level). The procedure for measuring FMDT and the obtained values have been published and analyzed elsewhere and so the reader is referred to the relevant study for detailed information (Johannesen et al., 2016).

Temporal integration. TI is a measure of the improvement in the detection threshold of a sound as the sound duration increases. This measure was included as a possible predictor for two reasons. First, TI is reduced for hearing-impaired listeners, and the reduction appears unrelated to cochlear mechanical dysfunction (Plack & Skeels, 2007). Second, steeper-than-normal threshold-duration functions could be indicative of disrupted auditory nerve activity (Zeng, Kong, Michalewski, & Starr, 2005) and of primary deafferentation (Marmel, Rodríguez-Mendoza, & Lopez-Poveda, 2015), both of which might affect speech perception particularly in noise backgrounds (Lopez-Poveda, 2014).

Here, TI is defined as the difference between absolute detection thresholds for short (10 or 30 ms) and long (210 ms) pure tones. TI was measured for pure tones with frequencies 0.5, 1, 2, 4, and 6 kHz. At 0.5 kHz, the short tone duration was 30 ms with 15-ms cosine-squared onset and offset ramps, and the long duration tone was 210 ms with 10-ms ramps. At all other frequencies, the tones had 10 and 210 ms durations and 5-ms onset and offset ramps.

The procedure for measuring absolute thresholds and the obtained values have previously been published in Johannesen et al. (2014). Those values were reused here to calculate the TI predictor.

Hearing-Aid Fitting Procedures

An experienced audiologist fitted participants with two hearing aids using Starkey's Inspire^{TM} (version 7.0.1602) fitting software and following standard clinical procedures. In the first fitting session, amplification was set according to the National Acoustic Laboratories' nonlinear fitting procedure, version 1 (NAL-NL1; Byrne et al., 2001). A probe microphone placed in the user ear was used to make sure that the target NAL-NL1 amplification was reached. Some users, particularly first-time hearing-aid users, complained that sound levels were too loud with the prescribed settings. When that happened, an attempt was made to address the user's complaint by reducing the overall output volume in the hearing aids. The aim was to reduce the loudness for soft or medium or loud sounds without changing the compression prescribed by the NAL-NL1 rule. If this was not sufficient, then the amplification for soft or loud sounds was adjusted independently, thus modifying the compression prescribed by NAL-NL1. Whenever amplification settings differed from NAL-NL1, users were called back after two weeks and a second attempt was made to readjust hearing-aids settings to make them closer to the NAL-NL1 target. Users were asked to wear their hearing aids regularly for one month. After that time, all participants were called in for fine tuning of their hearing aids. At this time, their specific complaints or requests were addressed by adjusting the amplification settings without restrictions. Participants were asked to wear their hearing aids regularly for a second month before outcomes were measured. During the second month, participants were allowed to have their amplification revised at any moment.

Hearing-Aid Settings Used as Predictors

The REIG was measured for the final hearing-aid settings and for the two ears of each participant but only the REIG of the test ear was used as a predictor. The REIG was measured for an International Speech Test Signal (Holube, Fredelake, Vlaming, & Kollmeier, 2010), for three input levels of 50, 65, and 80 dB SPL, and for frequencies of 0.5, 1, 2, 4, and 6 kHz. With this information, the following predictors were obtained (Table 6).

1. The REIG (in dB) for an input level of 50 dB SPL ($\text{REIG}_{50\text{dB}}$).
2. The REIG for an input level of 65 dB SPL ($\text{REIG}_{65\text{dB}}$).

3. The REIG for an input level of 80 dB SPL ($\mathrm{REIG}_{80\mathrm{dB}}$).
4. The RECE (in dB/dB) applied by the hearing aid at low input levels: $\mathrm{RECE}_{\mathrm{LO}} = 1-(\mathrm{REIG}_{50\mathrm{dB}} - \mathrm{REIG}_{65\mathrm{dB}})/15$.
5. The RECE applied by the hearing aid at high input levels: $\mathrm{RECE}_{\mathrm{HI}} = 1-(\mathrm{REIG}_{65\mathrm{dB}} - \mathrm{REIG}_{80\mathrm{dB}})/15$.

The $\mathrm{REIG}_{50\mathrm{dB}}$ and $\mathrm{REIG}_{65\mathrm{dB}}$ were included as predictors because they can affect audibility and reduced audibility decreases speech-in-noise intelligibility (e.g., Peters et al., 1998). The $\mathrm{RECE}_{\mathrm{LO}}$ and $\mathrm{RECE}_{\mathrm{HI}}$ were included as predictors because compression might distort the amplified speech presented to the hearing-aid user and distortion can reduce speech intelligibility (Bode & Kasten, 1971). A second reason was that $\mathrm{RECE}_{\mathrm{LO}}$ and $\mathrm{RECE}_{\mathrm{HI}}$ may interact with residual cochlear compression (BMCE) or cochlear amplification loss, and the interaction may influence speech intelligibility. For example, a hearing-aid user with substantial residual cochlear compression (low BMCE) in combination with high hearing-aid compression (low RECE) may show poorer aided speech intelligibility than (the more typical) user with low residual compression (high BMCE) and high HA compression (low RECE).

Reading Span Test

Akeroyd (2008) reviewed a number of studies linking cognition and speech reception in noise. He concluded that of the many cognitive tests that have been used to investigate such a link, tests of working memory have, mostly, given significant results. He further concluded that "the reading-span test (Daneman & Carpenter, 1980), which is often taken as a measure of working memory capacity, is most effective and predicts performance on a wide range of cognitive tasks." For this reason, the Spanish version (Elosúa, Gutiérrez, García Madruga, Luque, & Gárate, 1996) of the reading-span test was applied here, and the score in this test was used as a potential predictor of hearing-aid outcomes.

The procedure and scoring were virtually identical as designed by Daneman and Carpenter (1980). The test included five levels. In the first level, two sentences were presented to the participant in a computer screen. The participant was asked to read aloud each of the two sentences. After reading the last sentence, the participant was asked to recall the last word in each sentence. The score was 1 when the words and the order were remembered correctly and 0 otherwise. For each successive level, the number of sentences increased by one. For Levels 2 to 5, the score was 1 when the words were remembered correctly, 2 if the words were also remembered in the right order, and 0 otherwise. Three runs were conducted at all levels. The score for each level was the sum of the scores across the three runs times the number of sentences at that level. The test continued by increasing the level until the participant scored 0 in a level or Level 5 was completed. The total score was the sum of the scores across the five levels. Larger scores indicate better working memory capacity.

Speech Intelligibility Index in Quiet

The NAL-NL1 hearing-aid prescription compensates at most for half of the threshold elevation (Byrne et al., 2001). In addition, some participants might have requested less hearing-aid gain in their fine tuning sessions to achieve listening comfort (see Humes, 2002). As a result, the low-intensity portions of the amplified speech spectrum might have been below the participant's absolute thresholds, thus reducing intelligibility (e.g., Peters et al., 1998; Woods et al., 2013). We attempted to predict the potential effects of reduced audibility on aided speech-in-noise intelligibility by using the SII (ANSI, 1997).

The conventional SII indicates the proportion of the speech spectrum that is above the absolute threshold *and* above the background noise (ANSI, 1997). The same would be true for an aided speech-in-noise situation, except that the speech and the noise would be both amplified by the hearing aid before being presented to the listener. Here, however, the SII was calculated disregarding the background noise. In other words, the SII informed of the proportion of the amplified speech spectrum that was above absolute threshold. The rationale behind this approach is that if the full speech spectrum were audible, then performance deficits in a masker background would be due to the presence of the masker rather than reduced audibility and would thus reflect supra-threshold deficits. Therefore, the present SII calculations did *not* take into account the background noise but only the absolute thresholds of the listeners, the speech spectrum, and the REIGs. The resulting SII will be referred to as the (aided) $\mathrm{SII}_{\mathrm{Q}}$ to emphasize that it is the SII had there been a quiet background. If audibility were reduced, then intelligibility would be reduced because part of the speech spectrum would still be inaudible due to the combination of the listener's absolute threshold and hearing-aid gain. In other words, $\mathrm{SII}_{\mathrm{Q}}$ would be negatively correlated with outcomes. The $\mathrm{SII}_{\mathrm{Q}}$ was calculated for the aided binaural listening situation assuming that the listener used the better ear in each frequency band. In all other aspects, our SII calculations conformed to ANSI (1997) for 1/3 octave bands. Our approach was identical to that of Peters et al. (1998).

Questionnaires

Three questionnaires were administered:

1. The COSI (Dillon, James, & Ginis, 1997). The COSI is a short open-form clinically feasible questionnaire aimed at assessing the hearing improvements provided by the hearing-aid treatment and is tailored to the deficits considered important for each individual. It was administered at the time of entering the study to identify the hearing aspects considered important by each participant and after the 2-month hearing-aid trial period to ask participants' about their improvement in those specific aspects obtained from using hearing aids. Larger COSI scores indicate greater improvements.
2. The SSQ questionnaire (Gatehouse & Noble, 2004). The SSQ questionnaire is a large closed-form scientific-oriented questionnaire and hence more suited to across subject comparisons. The questionnaire was administered at the time of entering the study to assess the participants' subjective (baseline or unaided) hearing abilities. Scores ranged from 0 to 10 with higher scores indicating better hearing abilities (Gatehouse & Noble, 2004). The questionnaire was administered again after the 2-month hearing-aid trial period to assess the participants' subjective improvement in hearing abilities. The scale for post-treatment (SSQ_B) scores ranged from -5 to $+5$, with negative or positive values indicating decreased or increased hearing ability (Jensen, Akeroyd, Noble, & Naylor, 2009).
3. For completeness, the IOI-HA questionnaire was also administered (Cox et al., 2000, Cox, Stephens, & Kramer, 2002). The IOI-HA is intended to be a minimum core set of questions for assessing hearing-aid user satisfaction and is available in multiple languages, including Castilian Spanish. Larger scores in the IOI-HA questionnaire indicate greater benefits from using hearing aids.

We stress that the COSI, IOI-HA, and posttreatment SSQ_B questionnaires measured *improvement* (or benefit) from treatment. Pretreatment SSQ questionnaire scores for each of the three questionnaire sections (SSQ) were used as potential predictors of outcomes.

Aided Speech-in-Noise Reception Thresholds

The SRT_N was used as an indicator of speech intelligibility in noisy environments when wearing hearing aids. That is, the SRT_N was used as a measure of hearing-aid outcome. The SRT_N was measured using the Castilian-Spanish version (Huarte, 2008) of the hearing-in-noise test (Nilsson, Soli, & Sullivan, 1994). Sentences uttered by a male speaker were presented to the hearing-aid user in the presence of a masker sound. The masker consisted of two talkers (one male and one female) played in reverse (time-reversed two-talker masker). This masker was made as explained by Johannesen et al. (2016) and was used because it has similar temporal and spectral properties as forward speech—and was thus expected to have the same energetic masking properties as speech—but without semantic information that may contribute to informational masking (e.g., Hornsby & Ricketts, 2007). Participants wore their two hearing aids during SRT_N measurements.

The target sentences and the masker were played simultaneously via a loudspeaker placed at eye level 1 m away in front of the listener (quasi free-field conditions). Stimuli were filtered to minimize the effects of room acoustics on the stimulus spectrum. The filter was such that a noise with a spectrum equal to the long-term spectrum of speech emitted by the loudspeaker would have the long-term average spectrum of speech when recorded with a microphone placed at the position of the listener's head (Byrne et al., 1994).

To measure an SRT, the speech level was held constant at 65 dB SPL and the masker level was varied adaptively to find the signal-to-noise ratio (SNR) at which the hearing-aid user recognized 50% of the presented sentences. The initial SNR step was 4 dB and was reduced to 2 dB after five reversals. Each trial had a total of 20 reversals. The SRT_N (i.e., the SNR at 50% correct sentence identification) was calculated as the mean SNR for the last 15 sentences. The procedure was otherwise identical as for the original hearing-in-noise test (Nilsson et al., 1994). The SRT_N was measured three times, and the mean was noted as the final result.

Acoustic Stimuli and Apparatus

For all psychoacoustic measurements, stimuli were digitally generated or stored as digital files with a sampling rate of 44100 Hz. They were digital-to-analog converted using an RME Fireface 400 sound card with a 24-bit resolution and were played through either Sennheiser HD-580 headphones in the case of measurements of predictor variables or a loudspeaker (Behringer model Truth B2031A) in the case of the SRT_N measurements. Participants sat in a double-wall sound attenuation booth during data collection.

Statistical Procedures

All statistical analyses were carried out using Matlab (version 2008b, The Mathworks).

Pairwise Correlations. Simple pairwise correlations between measured variables and outcomes (Table 8) were investigated using Pearson's correlation coefficient. The significance of the obtained values was calculated using a two-tailed Student's *t* test.

Principal Component Analysis. PCA was conducted with the aim of reducing redundancy across predictors due to their possible colinearity. For example, variables commonly affected by audiometric thresholds (e.g., $\mathrm{SRT_Q}$, $\mathrm{DPOAE_N}$, $\mathrm{HL_{OHC}}$, etc.) were expected to covary and be possibly redundant. The PCA was conducted on 19 on the 28 measured variables (i.e., on the variables listed in Table 8 except BMLD; see Methods section). The variables were standardized (z-transformed) prior to the PCA.

PCs are typically retained for further analysis when their eigenvalues (of the transformation matrix) are larger than 1. We identified six components that met this criterion (PC1 to PC6; Table 10). However, it is strongly recommended to try several sizes for the retained components in order to assess the robustness of the interpretation of the results (Abdi & Williams, 2010). Furthermore, when the identified components are to be used as predictors of an outcome measure, there is no logical reason why the outcome measure should not be closely tied to the least important component (Cox, 1968). For these reasons, here, we explored retaining up to nine components to minimize the risk of missing significant outcome predictors but only report the solution that was most stable in terms of loadings. The seven retained components (Table 9) were rotated orthogonally using the *varimax* criterion to facilitate their interpretation.

Stepwise MLR Models. It was assumed that each outcome measure (or dependent variable) could be expressed as the sum of multiple independent variables, each multiplied by a coefficient (linear combination). The coefficients were optimized for the linear combination of predictors to explain the largest possible amount of variance in the dependent variable. The procedure is known as MLR. As explained in the Methods section, MLR models were obtained using either the identified PCs or the measured predictors with the higher loadings in the components. In either case, the MLR analysis was conducted in a stepwise fashion to minimize the impact of colinearity between predictors as well as to assess their relative importance for outcome.

In each step of the stepwise procedure, the most significant new variable was added to the model. Sometimes, one or more previously included variables became nonsignificant after adding a new variable (those being colinear with the newly included variable). In this case, the originally included variables were excluded from the model when their probability for a significant contribution increased above $p > .10$. Here, we report variables whose contribution to the predicted variance was significant at the $p \leqslant .05$ level. The variable selection process ended when additional variables did not add significantly to the predicted variance and none of the included variables needed to be excluded from the model. This procedure is commonly known as stepwise MLR with forward selection. The final models omit colinear variables and, most importantly, inform about the relative importance of the various predictors. As is common practice, the variance explained by the models was adjusted for the number of variables used in the model (Theil & Goldberger, 1961).

Distributions of Predictors and Outcomes

As explained earlier, the values used here for some predictors have been reported in earlier studies (Johannesen et al., 2014, 2016; Pérez-González, Johannesen, & Lopez-Poveda, 2014). The values for other predictors, however, are reported here for the first time. This section summarizes the range of values for the latter and compares them with those reported by independent studies.

Table A1 shows the 5, 25, 50, 75, and 95 percentiles for the numerical predictors and outcome measures. Most of the predictors were measured for all 68 participants. Others, however, could not be measured for all participants.

We could not measure BMLDs for 14 participants because their hearing losses were so large at 500 Hz (the probe frequency used to measure BMLDs) that the sound levels involved in measuring their BMLDs would exceed the safety output limit of our system (105 dB SPL). For the 54 participants for whom we obtained a BMLD, the mean BMLD was 1.8 dB (standard deviation = 3 dB). For comparison, Gatehouse and Akeroyd (2006) reported a mean BMLD of 2.8 dB (standard deviation of 4.5 dB). The slightly smaller BMLD for the present participants is likely due to their having higher hearing loss at 500 Hz than the participants used by Gatehouse and Akeroyd (2006).

The present FMDTs were reported in Johannesen et al. (2016). They were in the range 0.7 to 2 (in units of $\log_{10}$[Hz]), thus similar to the range of values reported by Strelcyk and Dau (2009; 0.7–1.7, when converted to the present units). The participants in the study of Strelcyk and Dau (2009) had almost normal audiometric thresholds at frequencies $\leqslant 1$ kHz while the present listeners had typically greater hearing losses over that frequency range (Figure 1), something that might explain the higher upper limit in the present FMDTs.

The present FMRR values ranged from 0.05 to 0.27 dB/ms. These values were similar to those reported

Table A1. Values of the 5th, 25th, 50th, 75th, and 95th Percentiles for Numerical Predictors and Outcome Measures.

Predictor or outcome measure	5%	25%	50%	75%	95%	*p*	*N*
Predictor							
Age (years)	38	54	61	74	81	.400	68
PTT (dB HL)	35	44	52	56	63	.090	68
Air-bone gap 9(dB)	−4	−1	1	3	9	.310	68
ULL (dB HL)	88	100	109	115	120	.041*	68
SRT$_{Q}$ (dB HL)	34	42	49	57	68	.410	68
HL$_{OHC}$ (dB)	16	25	29	33	38	.250	67
HL$_{IHC}$ (dB)	9	12	15	17	25	.031*	67
BMCE (dB/dB)	0.12	0.26	0.38	0.52	0.69	.043*	67
DPOAE$_{N}$	0.2	0.6	1.0	1.9	3.2	$2.1 \cdot 10^{-5}$*	68
DPOAE$_{\mu Pa}$ (dB SPL)	−4	−1	1	4	10	.350	65
BMLD (dB)	−2.8	0.13	1.9	3.6	7.2	.390	54
FMDT (log10[Hz])	0.77	1.12	1.3	1.52	1.88	.260	68
FMRR (dB/ms)	0.052	0.092	0.12	0.15	0.27	.024*	67
TI (dB)	1.4	2.8	3.7	4.5	5.9	.650*	68
REIG$_{50dB}$ (dB)	7	13	17	22	28	.380	68
REIG$_{65dB}$ (dB)	4	10	13	18	24	.410	68
REIG$_{80dB}$ (dB)	1	6	9	13	18	.460	68
RECE$_{LO}$ (dB/dB)	0.56	0.68	0.75	0.82	0.87	.650	68
RECE$_{HI}$ (dB/dB)	0.54	0.64	0.72	0.79	0.89	.960	68
RSpan	4	6	12	18	32	$1.64 \cdot 10^{-9}$*	68
SSQ-speech	3.2	4.2	5.0	5.9	7.3	.470	68
SSQ-spatial	4.3	6.2	7.2	8.4	9.7	.840	68
SSQ-qualities	5.8	6.9	7.9	8.9	9.4	$2.9 \cdot 10^{-6}$*	68
SII$_{Q}$	0.43	0.58	0.69	0.76	0.85	.320	68
Outcome measures							
SRT$_{N}$ (dB SNR)	−0.88	0.80	2.25	3.90	7.03	.240	68
SSQ$_{B}$-speech	−0.07	1.33	2.46	3.21	4.04	.207	68
SSQ$_{B}$-spatial	−0.17	0.85	2.38	3.41	4.72	.590	68
SSQ$_{B}$-qualities	−0.19	0.54	1.83	3.14	4.48	.191	68
COSI	2.32	3.67	4.12	4.67	5.00	.212	68
IOI-HA	3.10	3.64	4.14	4.50	4.86	.735	68

Note. The probability (*p*) of the corresponding distribution not being Gaussian (two-tailed, chi-squared test for goodness-of-fit) is shown, and the number of participants for whom each predictor was measured. PTT = pure tone threshold; ULL = uncomfortable loudness levels; SRT = speech reception threshold; HL$_{OHC}$ = outer hair cell loss; HL$_{IHC}$ = inner hair cell loss; BMCE = basilar membrane compression exponent; DPOAE = distortion product otoacoustic emission; BMLD = binaural-masking level difference; FMDT = frequency-modulation detection threshold; FMRR = forward-masking recovery rate; TI = temporal integration; REIG = real-ear insertion gain; SSQ = speech, spatial, and qualities; SII$_{Q}$ = speech intelligibility index in quiet; COSI = client-oriented scale of improvement; IOI-HA = international outcome inventory for hearing aids.

*$p < .05$.

in several other studies (e.g., Lopez-Poveda et al., 2003; Lopez-Poveda, Plack, Meddis, & Blanco, 2005) but slightly lower than the 0.1 to 0.5 dB/ms range reported by Gregan et al. (2013). Recovery from forward masking is slower for masker levels above than below 92 dB SPL (Wojtczak & Oxenham, 2009). This, however, is unlikely to explain the difference between present values and the values reported in Gregan et al. (2013), because the two studies used similar masker level ranges. The reason for the difference is uncertain.

The present participants had a mean score in the reading span test of 13.6 (standard deviation = 9.2). These values are slightly lower than those (mean = 15.1, standard deviation = 8.5) reported by Elosúa et al. (1996) for the Spanish version of the test probably because the present participants were older (age range: 25–82 years) than the participants tested by Elosúa et al. (age range: 12–16 years).

The mean pretreatment SSQ-questionnaire scores were 5.8 (speech), 7.1 (spatial), 7.8 (qualities), and the corresponding standard deviations 1.3, 1.6, and 1.3, respectively. Gatehouse and Noble (2004) did not report mean scores for each section but for the individual items in each section. The mean scores across items for each section can, however, be calculated from their Table 1 and were around 1.5 lower than the present ones while the present standard deviations for each section of the SSQ were around half of those calculated from Table 1 of Gatehouse and Noble (2004).

As for outcome measures, the present aided SRT_N were measured having participants wear their two hearing aids with final settings. For the present participants, the mean SRT_N was 2.6 dB SNR (standard deviation = 2.7 dB), thus slightly higher than the mean −1 dB SNR (standard deviation = 3 dB) reported by Festen and Plomp (1990). The small discrepancy in range across studies may be due to the different type of amplification employed across studies: We used the level-dependent amplification provided by the hearing aids while Festen and Plomp (1990) used level-independent amplification prescribed by the National Acoustics Laboratory Revised (NAL-R) fitting rule (Byrne & Dillon, 1986). It may also be due to our using insufficient amplification for some participants; indeed, some of the present participants requested less amplification than prescribed by the NAL-NL1 rule to achieve comfortable loudness while the participants used by Festen and Plomp (1990) were given the NAL-R amplification. It may also be due to differences in the intrinsic difficulty of the speech material used.

The present SSQ_B scores represent mean reported *improvement* in speech, spatial, and quality aspects of hearing. The scores reported by Gatehouse and Noble (2004) represented absolute hearing ability for each item in the questionnaire for groups of unaided, unilaterally aided, and bilaterally aided listeners. Therefore, the scores in the two studies may not be compared directly. Improvement scores, however, can be calculated from Table 2 in Gatehouse and Noble (2004) by subtracting the scores of the bilaterally aided and unaided groups. The difference in question was 2.5, 1.8, and 1.4 for the three SSQ sections of the questionnaire, respectively, with corresponding standard deviations of 1.0, 1.4, and 0.9. These mean difference SSQ scores may be thought of as the benefit of aided listening, and hence may be compared with the present SSQ_B results (Table A1). The comparison reveals that the present benefit of aided listening is slightly larger than that reported by Gatehouse and Nobel (2004) while the standard deviations were similar. Jensen et al. (2009) also used the SSQ questionnaire to assess hearing-aid benefit and reported slightly larger scores than the present SSQ_B scores. Overall, the present SSQ_B scores seem to be typical.

Acknowledgments

The authors thank William S. Woods, Jason Galster, and Olaf Strelcyk for useful suggestions and Almudena Eustaquio-Martín for technical support. The authors also thank the editor and two anonymous reviewers for their excellent reviews.

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 Starkey Ltd. (USA), by Junta de Castilla y León, MINECO (refs. BFU2012-39544-C02 and BFU2015-65376-P), and European Regional Development Funds.

Notes

1. Without Bonferroni correction for multiple comparisons, the critical R for statistical significance at $p \leqslant .05$ is equal to .24; with correction for 20 comparisons (as in Table 8), the critical R is .36.
2. Loadings are analogous to pairwise correlation coefficients between the corresponding component and predictor. In other words, the square of the loadings indicates the predictor variance explained by the component (Abdi & Williams, 2010).

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*, 433–459. doi:10.1002/wics.101.

Abrams, H. B., & Kihm, J. (2015). An introduction to MarkeTrak IX: A new baseline for the hearing aid market. *Hearing Review*, *22*, 16.

AAO-HNS (1993). Academy Responses to the FDA Request for Comment on Hearing Aid Regulations, American Academy of Otolaryngology—Head and Neck Surgery. *Bulletin*, *16–17*, 26–28.

Akeroyd, M. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*, S53–S71.

American National Standards Institute (ANSI). (1997). *S3.5 methods for calculation of the speech intelligibility index*. New York, NY: Author.

ANSI. (2004). *S3.6 specification for audiometers*. New York, NY: Author.

Bode, D. L., & Kasten, R. N. (1971). Hearing aid distortion and consonant identification. *Journal of Speech and Hearing Research*, *14*, 323–331.

Boothroyd, A., Springer, N., Smith, L., & Schulman, J. (1988). Amplitude compression and profound hearing loss. *Journal of Speech and Hearing Research*, *31*, 362–376.

British Society of Audiology (2011). Recommended procedure. Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. *British Society of Audiology*, *2014*, 1–32. Available at: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf.

Brooks, D. N. (1989). The effect of attitude on benefit obtained from hearing aids. *British Journal of Audiology*, *23*, 3–11.

Buus, S., & Florentine, M. (2001). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. *Journal of the Association for Research in Otolaryngology*, *3*, 120–139.

Byrne, D., & Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, *7*, 257–265.

Byrne, D., Dillon, H., Ching, T., Katsch, R., & Keidser, G. (2001). NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *Journal of the American Academy of Audiology*, *12*, 37–51.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., . . . Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America*, *96*, 2108–2120.

Cahana-Amitay, D., Spiro, A., Sayers, J. T., Oveis, A. C., Higby, E., Ojo, E. A., . . . Obler, L. K. (2016). How older adults use cognition in sentence-final word recognition. *Aging, Neuropsychology, and Cognition. A Journal of Normal and Dysfunctional Development*, *23*, 418–444.

Cárdenas, M. R., & Marrero, V. (1994). *Cuaderno de logoaudiometría*. Madrid, Spain: Universidad Nacional de Educación a Distancia.

CHABA (1988). Speech understanding and aging. *Journal of the Acoustical Society of America*, *83*, 859–895.

Cox, D. R. (1968). Notes on Some Aspects of regression Analysis. *Journal of the Royal Statistical Society*, *131*, 265–279.

Cox, R. M., Alexander, G. C., & Gray, G. A. (2007). Personality, hearing problems, and amplification characteristics: Contributions to self-report hearing aid outcomes. *Ear and Hearing*, *28*, 141–162.

Cox, R. M., Hyde, M., Gatehouse, S., Noble, W., Dillon, H., Bentler, R., . . . Hallberg, L. (2000). Optimal outcome measures, research priorities and international cooperation. *Ear and Hearing*, *21*, 106S–115S.

Cox, R. M., Stephens, D., & Kramer, S. (2002). Translations of the international outcome inventory for hearing aids (IOI-HA). *International Journal of Audiology*, *41*, 3–26.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik, & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York, NY: Plenum.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Dillon, H., James, A., & Ginis, J. (1997). The Client Oriented Scale of Improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids. *Journal of the American Academy of Audiology*, *8*, 27–43.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *Journal to the Acoustical Society of America*, *35*, 1206–1218.

Edwards, B. (2003). Hearing aids and hearing impairment. In S. Greenberg, W. A. Ainsworth, A. N. Popper, & R. R. Fay (Eds.), *Speech processing in the auditory system* (vol. 18, pp. 339–421). New York, NY: Springer.

Elosúa, M. R., Gutiérrez, F., García Madruga, J. A., Luque, J. L., & Gárate, M. (1996). Adaptación española del "Reading Span Test" de Daneman y Carpenter. *Psicothema*, *8*, 383–395.

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

Füllgrabe, C., Moore, B. C. J., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, *6*, 347.

Füllgrabe, C., & Rosen, S. (2016a). Investigating the role of working memory in speech-in-noise identification for listeners with normal hearing. *Advances in Experimental Medicine and Biology*, *894*, 29–36. doi:10.1007/978-3-319-25474-6_4.

Füllgrabe, C., & Rosen, S. (2016b). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing. *Frontiers in Psychology*, *7*, 1268.

Gatehouse, S., & Akeroyd, M. (2006). Two-eared listening in dynamic situations. *International Journal of Audiology*, *45*, S120–S124.

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, *43*, 85–89.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., & Peters, J. O. (1997). From aboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss. *Ear and Hearing*, *18*, 440–455.

Gregan, M. J., Nelson, P. B., & Oxenham, A. J. (2013). Behavioral measures of cochlear compression and temporal resolution as predictors of speech masking release in hearing-impaired listeners. *Journal of the Acoustical Society of America*, *134*, 2895–2912.

Hansen, M. (2002). Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear and Hearing*, *23*, 369–380.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, *49*, 891–903.

Hopkins, K., & Moore, B. C. J. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *Journal of the Acoustical Society of America*, *130*, 334–349.

Hornsby, B. W. Y., & Ricketts, T. A. (2007). Directional benefit in the presence of speech and speech like maskers. *Journal of the American Academy of Audiology*, *18*, 5–16.

Huarte, A. (2008). The Castilian Spanish hearing in noise test. *International Journal of Audiology*, *47*, 369–370.

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *Journal of the Acoustical Society of America*, *112*, 1112–1132.

Jensen, N. S., Akeroyd, M. A., Noble, W., & Naylor, G. (2009). The Speech, Spatial and Qualities of Hearing scale (SSQ) as a benefit measure. *NCRAR conference on The Ear-Brain System: Approaches to the Study and Treatment of Hearing Loss*, Portland, OR.

Johannesen, P. T., & Lopez-Poveda, E. A. (2008). Cochlear nonlinearity in normal-hearing subjects as inferred psychophysically and from distortion-product otoacoustic emissions. *Journal of the Acoustical Society of America*, *124*, 2149–2163.

Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., & Lopez-Poveda, E. A. (2016). The influence of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of audible speech in noise by hearing-impaired listeners. *Trends in Hearing*, *20*, 1–14.

Johannesen, P. T., Pérez-González, P., & Lopez-Poveda, E. A. (2014). Across-frequency behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss. *Frontiers in Neuroscience*, *8*, 214.

Lewsen, B. J., & Cashman, M. (1997). Hearing aids and assistive listening devices in long-term care. *Journal of Speech-Language Pathology and Audiology*, *21*, 149–152.

Liberman, M. C., & Dodds, L. W. (1984). Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves. *Hearing Research*, *16*, 55–74.

Keidser, G., Dillon, H., Carter, L., & O'Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends in Amplification*, *16*, 211–223.

Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, *1*, e24.

Keidser, G., O'Brien, A., Carter, L., McLelland, M., & Yeend, I. (2008). Variation in preferred gain with experience for hearing aid users. *International Journal of Audiology*, *47*, 621–635.

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *Journal of the Acoustical Society of America*, *64*, 1386–1391.

Kim, S., Frisina, R. D., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2006). Effect of age on binaural speech intelligibility in normal hearing adults. *Speech Communication*, *48*, 591–597.

Kochkin, S. (2002). MarkeTrak VI: Consumers rate improvements sought in hearing instruments. *Hearing Review*, *9*, 18–22.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *Journal of Neuroscience*, *29*, 14077–14085.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Löhler, J., Akcicek, B., Wollenberg, B., Schönweiler, R., Verges, L., Langer, Ch., . . . Ernst, A. (2015). Results in using the Freiburger monosyllabic speech test in noise without and with hearing aids. *European Archives of Otorhinolarygology*, *272*, 2135–2142.

Lopez-Poveda, E. A. (2014). Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech. *Frontiers in Neuroscience*, *8*, 348.

Lopez-Poveda, E. A., & Barrios, P. (2013). Perception of stochastically undersampled sound waveforms: A model of auditory deafferentation. *Frontiers in Neuroscience*, *7*, 124.

Lopez-Poveda, E. A., & Johannesen, P. T. (2012). Behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss. *Journal of the Association for Research in Otolaryngology*, *13*, 485–504.

Lopez-Poveda, E. A., Plack, C. J., & Meddis, R. (2003). Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing. *Journal of the Acoustical Society of America*, *113*, 951–960.

Lopez-Poveda, E. A., Plack, C. J., Meddis, R., & Blanco, J. L. (2005). Cochlear compression in listeners with moderate sensorineural hearing loss. *Hearing Research*, *205*, 172–183.

Lorenzi, C., Gilbert, G., Carn, H., & Moore, B. C. J. (2006). Speech perception problems on the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences USA*, *103*, 18866–18869.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, *42*, S49–S58.

Lunner, T., Rudner, M., & Rönnberg, J. (2009). Cognition and hearing aids. *Scandinavian Journal of Psychology*, *50*, 395–403.

Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology*, *18*, 604–617.

Marmel, F., Rodríguez-Mendoza, M. A., & Lopez-Poveda, E. A. (2015). Stochastic undersampling steepens auditory threshold/duration functions: Implications for understanding auditory deafferentation and aging. *Frontiers in Aging Neuroscience*, *7*, 63.

Marozeau, J., & Florentine, M. (2007). Loudness growth in individual listeners with hearing losses: A review. *Journal of the Acoustical Society of America*, *122*, EL81–EL87.

Marriage, J. E., Moore, B. C. J., Stone, M. A., & Baer, T. (2005). Effects of three amplification strategies on speech perception by children with severe and profound hearing Loss. *Ear and Hearing*, *26*, 35–47.

Mills, D. M., Feeney, M. P., & Gates, G. A. (2010). Evaluation of cochlear hearing disorders: Normative distortion product otoacoustic emission measurements. *Ear and Hearing*, *28*, 778–792.

Moore, B. C.J. (2014). Development and current status of the "Cambridge" loudness models. *Trends in Hearing*, *18*, 1–19.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2010). Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task. *Journal of the Acoustical Society of America*, *128*, 360–371.

Moore, B. C. J., & Glasberg, B. R. (1997). A model of loudness perception applied to cochlear hearing loss. *Auditory Neuroscience*, *3*, 289–311.

Moore, B. C. J., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, *188*, 70–88.

Moore, B. C. J., Glasberg, B. R., & Stone, M. A. (2010). Development of a new method for deriving initial fitting for hearing aids with multichannel compression: CAMEQ2-HF. *International Journal of Audiology*, *49*, 216–227.

Moore, B. C. J., & Sek, A. (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *Journal of the Acoustical Society of America*, *100*, 2320–2331.

Moore, B. C. J., Stainsby, T. H., Alcántara, J. I., & Kühnel, V. (2004). The effect on speech intelligibility of varying compression time constants in a digital hearing aid. *International Journal of Audiology*, *43*, 339–409.

Müller, J., & Janssen, T. (2004). Similarity in loudness and distortion product otoacoustic emission input/output functions: Implications for an objective hearing aid adjustment. *Journal of the Acoustical Society of America*, *115*, 3081–3091.

Neely, S. T., Johnson, T. A., & Gorga, M. P. (2005). Distortion-product otoacoustic emission measured with continuously varying stimulus level. *Journal of the Acoustical Society of America*, *117*, 1248–1259.

Nelson, D. A., Schroder, A. C., & Wojtczak, M. (2001). A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *110*, 2045–2064.

Newman, C. W., Wharton, J. A., Shivapuja, P. G., & Jacobson, G. P. (1994). Relationships among psychoacoustic judgments, speech understanding ability and self-perceived handicap in tinnitus subjects. *Audiology*, *33*, 47–60.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*, 1085–1099.

Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife*, *5*, e16747.

Oxenham, A. J. (2001). Forward masking: Adaptation or integration? *Journal of the Acoustical Society of America*, *109*, 732–741.

Palmer, A. D., Newsom, J. T., & Rooks, K. S. (2016). How does difficulty communicating affect the social relationships of older adults? An exploration using data from a national study. *Journal of Communication Disorders*, *62*, 131–146.

Perez, E., McCormack, A., & Edmonds, B. A. (2014). Sensitivity to temporal fine structure and hearing-aid outcomes in older adults. *Frontiers in Neurosience*, *8*, 7. doi:10.3389/fnins.2014.00007.

Pérez-González, P., Johannesen, P. T., & Lopez-Poveda, E. A. (2014). Forward-masking recovery and the assumptions of the temporal masking curve method of inferring cochlear compression. *Trends in Hearing*, *18*, 1–14.

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of the Acoustical Society of America*, *103*, 577–587.

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research*, *223*, 114–121.

Plack, C. J., & Skeels, V. (2007). Temporal integration and compression near absolute threshold in normal and impaired ears. *Journal of the Acoustical Society of America*, *122*, 2236–2244.

Plack, C. J., Drga, V., & Lopez-Poveda, E. A. (2004). Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss. *Journal of the Acoustical Society of America*, *115*, 1684–1695.

Reavis, K. M., McMillan, G., Austin, D., Gallun, F., Fausti, S. A., Gordon, J. S., . . . Konrad-Martin, D. (2011). Distortion-product otoacoustic emission test performance for ototoxicity monitoring. *Ear and Hearing*, *32*, 61–74.

Robles, L., & Ruggero, M. A. (2001). Mechanics of thec mammalian cochlea. *Physiological Review*, *81*, 1305–1352.

Ruggero, M. A., Rich, N. C., & Recio, A. (1996). The effect of intense acoustic stimulation on basilar-membrane vibrations. *Auditory Neuroscience*, *2*, 329–345.

Ruy, I. S., Ahn, J. H., Lim, H. W., Joo, K. Y., & Chung, J. W. (2012). Evaluation of masking effects on speech perception in patients with unilateral chronic tinnitus using the hearing in noise test. *Otology and Neurotology*, *33*, 1472–1476.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., . . . Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, *9*, 159–197.

Searchfield, G. D., Kaur, M., & Martin, W. H. (2010). Hearing aids as an adjunct to counseling: Tinnitus patients who choose amplification do better than those who don't. *International Journal of Audiology*, *49*, 574–579.

Shera, C. A., & Guinan, J. J. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs. *Journal of the Acoustical Society of America*, *105*, 782–798.

Sommers, M. S. (1997). Speech perception in older adults: The importance of speech-specific cognitive abilities. *Journal of the American Geriatrics Society*, *45*, 633–637.

Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., & Kates, J. M. (2015). Working memory and intelligibility of hearing-aid processed speech. *Frontiers in Psychology*, *6*, 256.

Strelcyk, O., & Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *Journal of the Acoustical Society of America*, *125*, 3328–3345.

Theil, H., & Goldberger, A. S. (1961). On pure and mixed statistical estimation in economics. *International Economic Review*, *2*, 65–78.

van Toor, T., & Verschuure, H. (2002). Effects of high-frequency emphasis and compression time constants on speech intelligibility in noise. *International Journal of Audiology*, *41*, 379–394.

Wojtczak, M., & Oxenham, A. J. (2009). Pitfalls in behavioral estimates of basilar-membrane compression in humans. *Journal of the Acoustical Society of America*, *125*, 270–281.

Woods, W. S., Kalluri, S., Pentony, S., & Nooraei, N. (2013). Predicting the effect of hearing loss and audibility on amplified speech reception in a multi-talker listening scenario. *Journal of the Acoustical Society of America*, *133*, 4268–4278.

World Health Organization. (2015). Deafness and hearing loss. Fact sheet n° 300. Retrieved from http://www.who.int/mediacentre/factsheets/fs300/en/

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., & Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *Journal of Neurophysiology*, *93*, 3050–3063.