

Perceptual Effects of Adjusting Hearing-Aid Gain by Means of a Machine-Learning Approach Based on Individual User Preference

Trends in Hearing
Volume 23: 1–23
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216519847413
journals.sagepub.com/home/tia



Niels Søgaard Jensen¹ , Ole Hau¹, Jens Brehm Bagger Nielsen¹, Thor Bundgaard Nielsen², and Søren Vase Legarth²

Abstract

This study investigated a method to adjust hearing-aid gain by use of a machine-learning algorithm that estimates the optimal setting of gain parameters based on user preference indicated in an iterative paired-comparison procedure. Twenty hearing-impaired participants completed this procedure for 12 different sound scenarios. During the adjustment procedure, their task was to indicate a preference based on one of three sound attributes: Basic Audio Quality, Listening Comfort, or Speech Clarity. In a double-blind comparison of recordings of the processed scenarios, and using the same attributes as criteria, the adjusted gain settings were subsequently compared with two prescribed settings of the same hearing aid (with and without activation of an automatic sound-classification system). The results showed that the adjustment method provided a general improvement of Basic Audio Quality, an improvement of Listening Comfort in a traffic-noise scenario but not in three scenarios with speech babble, and no significant improvement of Speech Clarity. A large variation in gain adjustments was observed across participants, both among those who did benefit and among those who did not benefit from the adjustment. There was no clear connection between the gain adjustments and the perceived benefit, which indicates that the preferred gain settings for a given sound scenario and a given listening intention are highly individual and difficult to predict.

Keywords

hearing aids, self-adjustment, listening intention, personalization, sound quality

Date received: 9 November 2018; revised: 29 March 2019; accepted: 3 April 2019

Introduction

The initial fitting of modern hearing aids is normally done by setting the hearing-aid gain according to a given fitting rationale that prescribes the gain as a function of frequency and input level. The individual prescription of gain is determined predominantly by the hearing-aid user's hearing loss as measured by an audiogram. A variety of different generic fitting rationales have been suggested, for example, the NAL-NL2 rationale (Keidser, Dillon, Flax, Ching, & Brewer, 2011), the DSL v5.0 rationale (Scollie et al., 2005), and the CAM2 rationale, also known as CAMEQ2-HF (Moore, Glasberg, & Stone, 2010). Furthermore, most hearing-aid manufacturers have developed their own proprietary fitting rationales dedicated to their own hearing aids. The hearing-aid fitter will typically be offered a choice in the

fitting software between the proprietary rationale and one or more of the generic rationales.

The different fitting rationales are developed based on different compensation philosophies (e.g., targeting loudness normalization or speech-intelligibility maximization) and assumptions about listening needs, and they often provide quite different gain prescriptions (Keidser, Brew, & Peck, 2003). Furthermore, the various rationales have been updated over time as new knowledge has been gathered and as new hearing-aid

¹Widex A/S, Lyngø, Denmark

²SenseLab, FORCE Technology, Hørsholm, Denmark

Corresponding author:

Niels Søgaard Jensen, Widex A/S, Nymøllevej 6 DK-3540 Lyngø, Denmark.

Email: nije@widex.com



technology has been developed (Smeds et al., 2015). Thus, the question about which rationale is the “best” and should be selected for a given hearing-aid user is not trivial.

Some studies have included within-subject comparisons of different generic rationales (e.g., Ching et al., 2013; Marriage, Moore, & Alcántara, 2004; Moore, Alcántara, & Marriage, 2001; Moore & Sek, 2013). Although these studies found an overall trend toward benefit of or preference for some of the rationales over others, the benefit or preference varied across participants, experimental conditions, and outcome domains. This is very much in line with other studies that have shown how the user preferences for gain-frequency responses and compression-parameter settings for hearing aids depend on the sound environment that the user is exposed to (e.g., Keidser, 1995; Keidser et al., 2005; Keidser, Dillon, & Byrne, 1995; Moore, Füllgrabe, & Stone, 2011; Preminger, Neuman, Bakke, Walters, & Levitt, 2000; Sabin, Hardies, Marrone, & Dhar, 2011). Furthermore, an important observation is that while there may be some general group trends in terms of what type of gain-frequency response and compression settings are preferred, there is also an individual variation in preference that cannot be easily explained.

These findings suggest that while a given—generic or proprietary—fitting rationale may work very well for some users some of the time, it is very difficult to develop a rationale that works well for all users in all types of situations. In certain situations, the user may need something different from what the rationale prescribes. This could obviously be a matter of personal preference since different persons may have different wishes and needs, for example, regarding the frequency shaping and the compression characteristics, but it could also be caused by a mismatch between the listening intention assumed by the fitting rationale and the actual listening intention of the user (Korzepa et al., 2018). The listening intention varies within and between situations and listeners (Wolters, Smeds, Schmidt, Christensen, & Norup, 2016), and predicting the intention based only on the acoustic signal picked up by a microphone is not always possible.

In addition to the input-dependent gain-frequency response prescribed by the fitting rationale, most modern hearing aids include various types of adaptive signal processing to address specific needs in certain sound environments. Features such as noise reduction and adaptive directional microphones will change the gain depending on the characteristics of the sound environment in order to meet the assumed needs of the user in specific environments (e.g., Bentler, 2005; Chung, 2004; Desjardins, 2016; Neher, Wagener, & Fischer, 2016). Furthermore, sound-classification systems have been developed to classify different sound environments into

a number of predefined categories that are associated with different settings of the hearing-aid parameters (e.g., Alexandre, Cuadra, Rosa, & Lopez-Ferreras, 2007; Büchler, Allegro, Launer, & Dillier, 2005; Nordqvist & Leijon, 2004). However, even though sound-classification algorithms have shown promising results in terms of correct classification (e.g., Nordqvist & Leijon, 2004) and, accordingly, are implemented in many modern hearing aids, they still do not address the situation where the individual user has a special preference that differs from the norm or where the user’s listening intention differs from what the algorithm assumes.

The obvious response to the problem of unaddressed individual needs is to provide more personalized solutions where gain and other hearing-aid parameters are set according to individual preferences rather than group preferences. This is basically the purpose of fine-tuning a hearing aid, which is an integrated part of a normal hearing-aid fitting, and which has been shown to be important for a successful overall patient outcome (Kochkin et al., 2010). In this part of the fitting, the user may inform the fitter about perceptual experiences with the hearing aids, and the fitter may use this information to adjust the hearing-aid parameters to improve the individual user’s experience (Anderson, Arehart, & Souza, 2018). The fine-tuning process may also include giving the user access to additional programs in the hearing aid. These additional programs may be adjusted to fit the individual needs in specific situations (Keidser, 1995).

For the fine-tuning process to be successful, the user must be able to express his or her listening experiences, and the fitter must be able to interpret this expression and translate it to an appropriate adjustment of hearing-aid parameters (Elberling & Hansen, 1999). It obviously also requires that the user actually connects with the fitter, either physically in the clinic or online. The cases where these prerequisites are not met are addressed by the research on, and development of, self-fitting hearing aids where hearing-aid users themselves get access to adjusting their own hearing-aid settings, without the involvement of a hearing-aid fitter (e.g., Convery, Keidser, Dillon, & Hartley, 2011; Convery, Keidser, Hickson, & Meyer, 2018; Keidser & Convery, 2016). In self-fitting hearing aids, the option to make self-adjustments is one of the basic processes needed (Keidser & Convery, 2016). The idea of self-fitting hearing aids has received increasing attention in recent years in parallel with the emergence of new product categories such as smartphone hearing apps, hearables, Personal Sound Amplification Products, and the upcoming Over-The-Counter product category in the United States (American Academy of Audiology, 2018).

A variety of different approaches to self-adjustment have been suggested. In the “Goldilocks” procedure

(Boothroyd & Mackersie, 2017; Mackersie, Boothroyd, & Lithgow, 2019), users have access to three controls adjusting the low-frequency (LF) cut (“Fullness”), the overall amplitude (“Loudness”), and the high-frequency boost (“Crispness”), respectively. Based on the reported studies, the authors found the method to be a “speedy, reliable and feasible alternative to, or supplement to, conventional fitting procedures.” Nelson, Perry, Gregan, and VanTasell (2018) reported a study on a commercially available self-fitting approach involving two user controls, “loudness” (adjusting gain, compression and limiting) and “fine tuning” (adjusting the tilt of the frequency response). One noticeable finding in that study was a very large between-subject variation in the final self-adjusted gain settings, which further supports the aforementioned findings on variations in individual preference. The variation could not be explained by any of the listener factors (age, gender, duration of hearing-aid use, hearing threshold, acceptable noise level and real-ear characteristics) included in a subsequent analysis reported by Perry, Nelson, and Van Tasell (2019). Convery, Keidser, Seeto, and McLelland (2017) investigated a commercially available self-fitting hearing aid where fine-tuning of the device could be performed using a three-band equalizer. They reported that even though the majority of participants could complete the fine-tuning process, a number of errors made by the participants were observed. Dreschler, Keidser, Convery, and Dillon (2008) compared four different sets of controls for self-adjustment of hearing aids, where each set included two or three controls (adjusting volume, tone balance or gain in three frequency bands). They concluded that the gain settings obtained with all four sets of controls were reproducible and not strongly depending on the combination of control keys, whereas it was found that the final setting strongly depended on the baseline gain setting (i.e., the starting point for the adjustment).

One assumption underlying the self-adjustment methods is that the user understands the parameter controls and operates them in a meaningful way. The user may not always have this awareness but rather have the feeling of wanting something else without being able to state what it is. In that case, operation of gain controls may lead to a trial-and-error process with a high risk of failure. An alternative is to optimize the parameter setting based on the user’s preferences. This has been exploited in methods based on use of paired comparisons (Amlani & Schafer, 2009), where the user only has to indicate their preference between two settings determined by an underlying algorithm. One approach that has been included in various studies is the modified simplex method (e.g., F. K. Kuk & Pape, 1992, 1993; Neuman, Levitt, Mills, & Schwander, 1987; Preminger et al., 2000). The modified simplex method prescribes a systematic iterative approach to

finding the optimal setting of a set of hearing-aid parameters, and the studies mentioned show that the method indeed has the potential of doing so. However, the modified simplex method has not gained popularity in clinical or commercial contexts. One reason for this could be the rather large number of comparisons and thereby the amount of time needed to actually find the optimal setting of parameters, not least when multiple hearing-aid parameters are involved. The problem of long convergence times was also experienced in studies where genetic algorithms inspired by evolution theory were used to adjust hearing-aid processing based on user-preference input (Baskent, Eiler, & Edwards, 2007; Durant, Wakefield, Van Tasell, & Rickert, 2004).

In recent years, new ways of tailoring hearing-aid processing to individual needs based on direct input from the user or on registration of user behavior have been suggested (e.g., Aldaz, Puria, & Leifer, 2016; Johansen et al., 2017; Korzepa et al., 2018). The increased processing power of hearing aids and, not least, the option to connect hearing aids to a smartphone to integrate the processing power of the smartphone in the entire hearing solution have allowed for more advanced and computationally demanding technologies like machine learning to be applied. This, for instance, makes self-adjustment of hearing aids based on the paired-comparison approach applicable in practice. Making the paired-comparison approach sufficiently fast to allow hearing-aid users to use it to adjust their own hearing aids while being in daily-life situations was indeed one of the main goals for the Interactive Hearing Aid Personalization System (IHAPS) suggested by Nielsen, Nielsen, and Larsen (2015), see also Nielsen, Nielsen, Jensen, and Larsen (2013) and Nielsen (2015).

The IHAPS uses machine learning and active learning to drive a sequence of paired comparisons. In each comparison, the user assesses the degree of preference between two different settings of hearing-aid parameter values. With the goal of finding an optimal setting in as few comparisons as possible, IHAPS iteratively determines the two settings of the next paired comparison based on what was learned from the paired comparisons already made by the user. Partly inspired by an adaptive-personalization approach (Heskes & Vries, 2005), this is done by continuously learning and updating a nonparametric machine-learning model of a hypothetical (unobserved) internal response function (IRF) that describes the user’s preference as a function of the hearing-aid parameter values. The final suggestion for the optimal setting, that is, the parameter values that provide the maximum preference, is determined as the maximum of the estimated IRF. From a machine-learning perspective, the key characteristics of the IHAPS are the use of Gaussian-process priors (Rasmussen & Williams, 2006) for the nonparametric modelling of the IRF, and the

resulting nonparametric Bayesian framework that enables the system to tolerate realistic amounts of randomness associated with human assessment while having a sufficiently flexible model of the IRF. For a more detailed description of the machine-learning approach, see Nielsen et al. (2015).

Running the IHAPS involves completing the following three steps in a looped manner: (a) Calculation of two hearing-aid parameter settings that the hearing-aid user should compare using a paired A-B comparison approach, (b) User assessment of the degree-of-preference between the two settings, and (c) Updating the user-specific IRF given the results of all past assessments. The user's indication of a degree of preference separates the method from most previous paired-comparison methods, which only involve indication of preference for one or the other setting. The more detailed information available in the degree of preference is utilized by the machine-learning algorithm to allow a faster convergence toward the optimal setting. The results reported by Nielsen et al. (2015), obtained by using the IHAPS to adjust hearing-aid gain in two or four frequency bands, indicate that the method in most cases was able to converge reliably toward the maximum of the IRF, and that improvements in user satisfaction provided by the hearing aid could be obtained. A basic requirement for the method to be successful was that the user was consistent when indicating the preference in the paired comparisons, as inconsistency reduced the efficiency of the machine-learning algorithm. This requirement obviously also applies to other methods based on the paired-comparison approach.

A method very similar to IHAPS has been implemented in a commercially available hearing aid, the Widex EVOKE, under the commercial name SoundSense Learn (SSL). SSL is administered by an app, which connects the user's smartphone to the hearing aids. The functionality of SSL is based directly on the IHAPS system described earlier. The parameters affected by the current version of the SSL method include static gain in three frequency bands that cover the entire frequency range of the hearing aid. Thus, the dynamic performance of the adaptive signal-processing systems in the hearing aid (e.g., compression and noise reduction) is not affected by the SSL procedure. The obvious consequence is that SSL only searches for a performance optimum within the parameter space defined by the three static gain parameters, whereas it is unable to find a global optimum within the parameter space defined by all parameters in the hearing-aid processing. The SSL method is intended to be part of a hearing aid, which has already been fitted using a traditional fitting approach where a fitter prescribes an initial setting and fine-tunes the setting of the hearing aid. SSL comes into play when the user experiences a specific daily-life situation in which his or her listening intention is not

addressed optimally by the hearing aids. As indicated in the description of the IHAPS system, the aim of SSL is to guide the user rapidly toward a more acceptable setting of the gain parameters by examining degrees of preference for judiciously chosen alternatives. The final adjusted setting may be saved in the app and reactivated the next time the user is in a similar situation.

In the study described in this article, some of the perceptual effects of carrying out the SSL procedure were investigated using the commercially available version of the system. The design of the study was to some extent inspired by a prior study performed in Widex's own lab (Townend, Nielsen, & Balslev, 2018). For example, the two studies both involved placing the test hearing aids on an acoustic manikin during the individual adjustment where participants listened to the output via headphones. The following assessment of hearing-aid settings was then made using recordings (made in the ears of the manikin) of sound processed by the hearing aids. This allowed for a double-blind test design and a smooth execution of the comparison of different hearing-aid settings. However, Townend et al. (2018) used a more simple setup when it came to the recording and the reproduction of sound scenarios included in the testing. In their study, they found that the SSL method, on a group level, could improve sound quality and listening comfort, while no effect on speech intelligibility was found.

This study focused on three perceptual sound-quality attributes: Basic Audio Quality, Listening Comfort, and Speech Clarity. Although the two former attributes were comparable to the Sound Quality and Listening Comfort attributes used by Townend et al. (2018), the Speech Clarity attribute was different from the Speech Intelligibility attribute used in that study. The change was made based on the expectation that changing the gain parameters affected by the SSL procedure would have the potential to provide a positive change in perceived speech clarity, whereas an improvement in perceived speech intelligibility (on a group level) seemed much less likely, as indicated by the findings by Townend et al. (2018).

The basic research question in this study was: What is the self-assessed benefit of user-driven adjustment of amplification when compared with threshold-based prescriptions with or without supplementation by input-dependent algorithms? In practice, this was done by comparing a gain setting obtained with the SSL procedure with two prescribed settings of the same hearing aid (based on the audiogram only), one with the hearing aid's sound-classification system turned on and one with the system turned off. A second purpose of this study was to look for possible systematic trends in the gain adjustments and to investigate possible connections between the gain adjustments and the perceptual effects of the adjustments.

Table 1. Overall Contents of the Study Protocol.

| Visit | Activity |
|----------------|--|
| 1 | Information and signing of consent form Familiarization to sound scenarios Instruction and training in self-adjustment procedure Self-adjustment of test hearing aids in each of 12 sound scenarios |
| Between visits | Recording of each sound scenario processed through test hearing aids in three different settings |
| 2 | Instruction in rating procedure Subjective rating of each recording, that is, each combination of sound scenario and hearing-aid setting |

Methods

The experimental part of the study was conducted at SenseLab, FORCE Technology in Hørsholm, Denmark. Ethical clearance for conducting the study was obtained from the Research Ethics Committees of the Capital Region of Denmark (case no. 16038586).

Overview of Protocol

The study protocol included two visits to the lab for each participant. The overall contents of the protocol are listed in Table 1. The visits for a given participant took place on two different days, with each visit taking around 2 hr to complete. The different activities listed in the table are described in more detail in the remainder of this section.

Participants

Twenty participants (12 men and 8 women) with sensorineural hearing loss participated in the study. The number of participants was based on the study by Townend et al. (2018) who used a similar experimental method and found significant main effects with $N=19$. The average age was 72 years (range: 54–83 years). The average audiogram is shown in Figure 1. The pure tone average hearing loss averaged across 0.5, 1, 2, and 4 kHz in both ears was 44 dB HL (range: 26–59 dB HL). All participants were recruited from SenseLab's pool of test participants. They were all experienced hearing-aid users and were fitted bilaterally with hearing aids of various brands and models. They all had prior experience as participants in listening tests. Prior to participation, they received information about the study and signed a consent form.

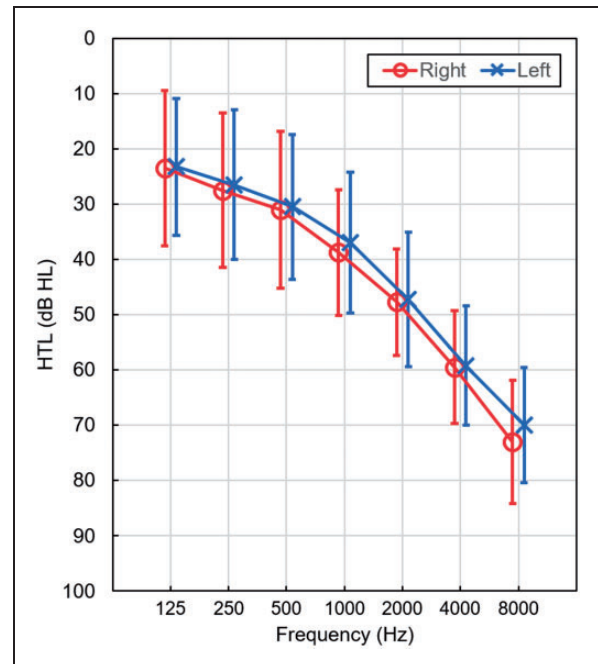


Figure 1. Mean left and right hearing threshold levels (HTLs) for the 20 participants. The audiograms have been slightly displaced for clarity. Error bars indicate ± 1 standard deviation.

Hearing Aids

Widex EVOKE F2 440 RIC (Receiver-In-Canal) hearing aids were used as test hearing aids. Each hearing aid was equipped with a type M receiver and a closed silicone ear mould shaped to fit the ear canal of the acoustic manikin (G.R.A.S. KEMAR 45BB), on which the hearing aid was mounted during use.

Three different settings of the hearing aids were used in the study. The first setting was a reference (REF) setting in which the hearing aids were programmed according to the proprietary Widex Fitting Rationale that prescribes the nonlinear gain of the hearing aid. As explained by Schmidt (2018), one design criterion of the Widex Fitting Rationale is that it should prescribe a long-term aided response similar to that prescribed by the NAL-NL1 algorithm (Byrne, Dillon, Ching, Katsch, & Keidser, 2001) when listening to speech at a normal conversation level. The hearing aids were programmed using the audiogram-based fitting option in the Widex fitting software, “Compass GPS” (version 3.0.142). In the REF setting, the automatic sound classification system used in the EVOKE hearing aid was turned off using a custom Matlab script, while all other features were in their default settings as prescribed by the fitting software. The default settings included the dynamic parameters of the dual speed compression system used in the hearing aid. This system combines a slow compressor that adjusts the gain according to the overall input

signal level and the hearing loss of the user, and a fast compressor that adjusts the gain based on fast changes and the modulation characteristics of the input signal. For the slow compressor, the attack and release times are approximately 1.5 s and 17 s, while for the fast compressor, the time constants are approximately 12 ms and 130 ms. The time constants do not depend on listener characteristics like, for example, hearing loss or age. For a more detailed explanation of the compression system, see F. Kuk, Slugocki, Korhonen, Seper, and Hau (2018).

The second setting, called Universal (UNI) in accordance with the terminology used in the fitting software, corresponds to normal use of the hearing aid, that is, with the sound-classification system activated. The basic gain setting in UNI is identical to the REF setting but, depending on the hearing aid's automatic classification of the sound environment (into one of 11 predefined classes), one or more of the processing parameters will be modified according to the sound class detected. The modifications involve both static and dynamic gain parameters in the hearing aid where the dynamic parameters determine the function of compression and noise reduction. Some of the sound classes only differ by their dynamic-parameter settings, which mean that they have the same long-term (static) gain for a given input signal. In this study, the hearing aids were used in 12 different sound scenarios (see below), which activated seven of the 11 sound classes in the hearing aid. Five of the sound classes were activated by two of the sound scenarios each, while the last two sound classes were activated by one sound scenario each. To eliminate the risk of switching between sound classes during trials and to avoid having to wait for the switching between sound classes to occur, the automatic classification was switched off and the hearing aids were fixed in the parameter setting for the sound class associated with a given sound scenario. This was done using a custom hearing-aid programming script.

The third setting (SSL) was based on the use of the SSL self-adjustment method, which was described in the introduction, and which is available in the test hearing aid and run using the accompanying EVOKE smartphone app (available for free download). During normal use of the hearing aids, the app would be installed on the user's smartphone. In this study, the app was run on an iPad to ease the handling in the lab. The adjustment procedure was completed for each of the 12 scenarios, and a unique SSL setting was obtained for each scenario. The setup used for reproduction of the scenarios and for listening through the hearing aids (which were mounted on a KEMAR) is described later in this section.

Although the machine-learning algorithms used by SSL were very much in accordance with the descriptions

provided by Nielsen et al. (2015), the graphical user interface was modified substantially in the commercially available version, and thus, in this study. A snapshot of the user interface used in the study is shown in Figure 2.

When running the SSL procedure, the participant had to switch between two settings, A and B, using the two buttons on the screen. The task was then to indicate the degree-of-preference between the two settings on a scale going from "A is much better" to "B is much better" using a slider (see Figure 2). In case of no preference between the settings, the slider should be placed on the middle of the scale. When a preference had been indicated, the participant pressed the "Next" button. The algorithm then calculated two new settings to be compared, and the looping process continued. Nielsen et al. (2015) indicate that convergence toward the maximum of the IRF normally was obtained within (at most)

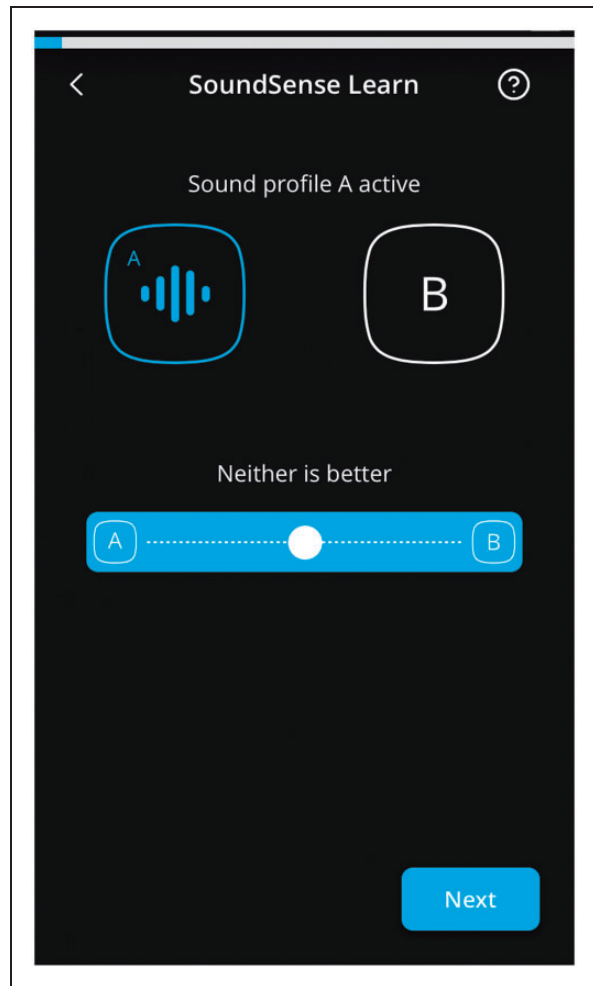


Figure 2. The graphical user interface for the SoundSense Learn app function. The A and B buttons are used to switch between two different settings of the hearing aids. The preference is indicated using the slider.

20 iterations. During normal use of the SSL procedure, the user can stop at any time, but in this study, all SSL adjustments were based on 20 iterations, partly to give the underlying machine-learning algorithm the best possible conditions for determining the optimal setting and partly to make the test conditions equal for all participants. After an SSL adjustment had been completed, the setting was saved as a separate program in the app, allowing for later recall.

The parameters adjusted by the commercially available SSL procedure used in this study include gain in three frequency bands, Low, Mid, and High, covering the entire frequency range of the hearing aid. The Low band includes bands 1 to 6 (0.1–0.7 kHz) of the hearing aid's 15-band filter bank, the Mid band includes bands 5 to 12 (0.6–3.6 kHz), while the High band includes bands 11 to 15 (2.2–10 kHz). To ensure a smooth frequency response, applying a gain change in one of the three SSL bands would have a reduced effect in the two filter-bank bands shared with an adjacent SSL band. The result of the SSL adjustment was accordingly a possible (static) gain change in each of the three bands, relative to the gain in the baseline setting, which always was the UNI setting. Thus, the setting of the remaining hearing-aid parameters in the SSL setting, including all the dynamic compression and noise-reduction parameters, was identical to the UNI setting.

Some limitations on the possible gain change have been applied in the SSL procedure. In the Low band, gain can be changed in both directions by up to 12 dB (in 2-dB steps). In the Mid and High bands, gain can also be decreased by 12 dB (in 2-dB steps), but only increased by 6 dB (in 1-dB steps). The latter limit is applied in order to avoid problems with acoustic feedback, which could be the result of an excessive increase in gain in those frequency regions.

Sound Scenarios and Sound Attributes

Twelve different sound scenarios were used for the SSL adjustments of the hearing aids, and subsequently, for comparison of the adjusted hearing-aid setting (SSL) with the two prescribed settings of the test hearing aids (REF and UNI). The sound scenarios were divided into three groups of four scenarios. The four scenarios within a group were selected to allow the participants to focus on a specific sound attribute during the SSL adjustment as well as during the following assessment. The three attributes were Basic Audio Quality, Listening Comfort, and Speech Clarity.

Besides oral instructions, the participants received short written descriptions of the attributes, which were intended to increase orthogonality of the attributes as well as help define meaning. The descriptions were available during the familiarization procedure (in which

participants were introduced to the different sound scenarios in a virtual reality setup), the SSL adjustment, and the assessment. The descriptions were in Danish. Here, a (nonvalidated) translation into English is provided.

Basic Audio Quality: “Your subjective assessment of the sound quality”. *Listening Comfort:* “A pleasant and effortless listening experience in the given context. A high level of listening comfort means that you hear the surroundings and the people around you with an appropriate balance between them. When judging the listening comfort, you should imagine that you are in the situation shown in the virtual-reality goggles and on the picture in the instructions.” *Speech Clarity:* “The speech appears to be clear in comparison with the background noise. A high level of clarity of speech means that all parts of the speech are perceived as being precise and distinct.”

Table 2 provides a short description of the 12 sound scenarios, grouped according to the sound attribute to be evaluated for each scenario. Furthermore, the table includes the average A-weighted sound pressure level of each scenario (measured at the position of the KEMAR) and the name of the hearing-aid sound class that was activated by the scenario.

As seen in Table 2, all four Basic Audio Quality scenarios were music signals. This choice was made because the Basic Audio Quality of music typically is fairly easy to assess. The Listening Comfort scenarios did not include distinct speech, but three of the four scenarios included speech babble. It should be noticed that none of the Listening Comfort scenarios was presented at levels higher than 70 dBA, and it was not the expectation (nor the intention) that the SSL adjustments focusing on Listening Comfort should be made out of a wish to just reduce loudness. The Speech Clarity scenarios were all characterized by including speech in a background of noise at varying speech-to-noise levels. Again, no very loud scenarios were included. The frequency responses of the 12 scenarios, averaged across the entire signal, are shown in Figure 3.

Recording and Reproduction of Sound Scenarios

The recording of each sound scenario included surround audio quality as well as 360 video and was made on location using a SoundField ST350 Portable Microphone and a GoPro OMNI camera. The scenarios were reproduced in a listening room fulfilling the ITU-R BS.1116-3 recommendations (International Telecommunication Union, 2015), using a standard 22.2 setup established with 22 Genelec loudspeakers (models 8050A, 8330A, 8320A, 8020C). The two subwoofers were not used in this study. The loudspeakers were positioned in three horizontal layers: 3 loudspeakers at -25° elevation at 0° and $\pm 45^\circ$ azimuth; 10 loudspeakers at 0° elevation

Table 2. The 12 Sound Scenarios Used in the Study Grouped According to the Three Sound Attributes.

| Attribute | No. | Description | Level (dBA) | Sound class |
|---------------------|-----|-----------------------------------|-------------|--------------------|
| Basic Audio Quality | 1 | R&B music (Ray Charles) | 70 | Contemporary music |
| | 2 | Pop music (Shubidua) | 70 | Contemporary music |
| | 3 | Classical music (Debussy) | 70 | Classical music |
| | 4 | Classical music (Mozart) | 70 | Classical music |
| Listening Comfort | 5 | Traffic noise, no speech | 70 | Urban (no speech) |
| | 6 | Café noise, including babble | 63 | Party (no speech) |
| | 7 | Canteen noise, including babble | 70 | Party (no speech) |
| | 8 | Outdoor setting, including babble | 70 | Urban (speech) |
| Speech Clarity | 9 | Family-dinner conversation | 63 | Social |
| | 10 | Woman talking in canteen | 70 | Party (speech) |
| | 11 | Man talking in café | 60 | Party (speech) |
| | 12 | Couple talking in the street | 56 | Urban (speech) |

Note. The table includes the presentation sound pressure level (at KEMAR's position) and the sound class activated in the hearing aid (in the UNI and SSL settings).

at 0°, ±30°, ±60°, ±90°, ±135°, and 180° azimuth; 8 loudspeakers at +40° elevation at 0°, ±45°, ±90°, ±135°, and 180° azimuth; and 1 loudspeaker at +90° elevation (just above the KEMAR). The distance from the top loudspeaker to the position corresponding to the point between KEMAR's ears (without KEMAR being present) was 145 cm, while the distance was 206 cm for the remaining 21 loudspeakers. The shorter distance to the top loudspeaker was compensated for by delaying the signal to that loudspeaker.

The playback was administered by a PC equipped with Audition CC2017. Signals from the PC were routed to the loudspeakers through an RME MADiface XT audio interface, an RME M-32 DA converter, and a TC Electronic EQ Station equalizer.

The reproduction setup is illustrated in Figure 4, which shows the setup used for the SSL adjustments and the subsequent recording of scenarios processed by the different hearing-aid settings. For reasons of simplicity, only the middle ring of loudspeakers is shown in the figure.

Familiarization Procedure

The visual input, provided through virtual-reality goggles, was used to familiarize the participants with the sound scenarios and the corresponding listening intentions used during the optimization and later assessment of the sound attributes. Due to the nature of the tasks where participants had to look at a screen, the tasks could not be completed while wearing the goggles, which were therefore only used for the familiarization procedure. The procedure was completed with the participant sitting in the reference position in the listening room, wearing a pair of HP Windows Mixed Reality

VR1000-100nn virtual-reality goggles and listening to the sound presented from the loudspeakers. For the Comfort and Clarity scenarios, the videos showed the actual real-life environment where the sound was recorded. For the Quality scenarios, the videos showed a living room where the prerecorded music was reproduced from two loudspeakers. During the familiarization procedure, the experimenter was present in the listening room, instructing the participant and providing information about the specific attribute that the participant should focus on in each scenario.

SSL Adjustment Procedure

To conduct a double-blind comparison of the hearing-aid settings, we decided to use an approach where the hearing aids were positioned on the ears of an acoustic manikin (KEMAR) in the listening room during the adjustment procedure, rather than on the ears of the participant. That allowed for subsequent recordings (made in the ears of the KEMAR) of all combinations of hearing-aid settings and scenarios, and these recordings could then be presented via headphones in a comparative assessment of the different hearing-aid settings.

During the adjustment procedure, the output of the hearing aids was picked up by the KEMAR microphones and presented to the participants (located in a neighboring room) via a pair of Sennheiser HD650 headphones. The signal to the headphones was routed through an equalizer (Mackie Quad EQ) to compensate for the acoustic effects of the KEMAR ear canal and the headphone response, and a Sonifex RB-HD3 headphone amplifier, see Figure 4.

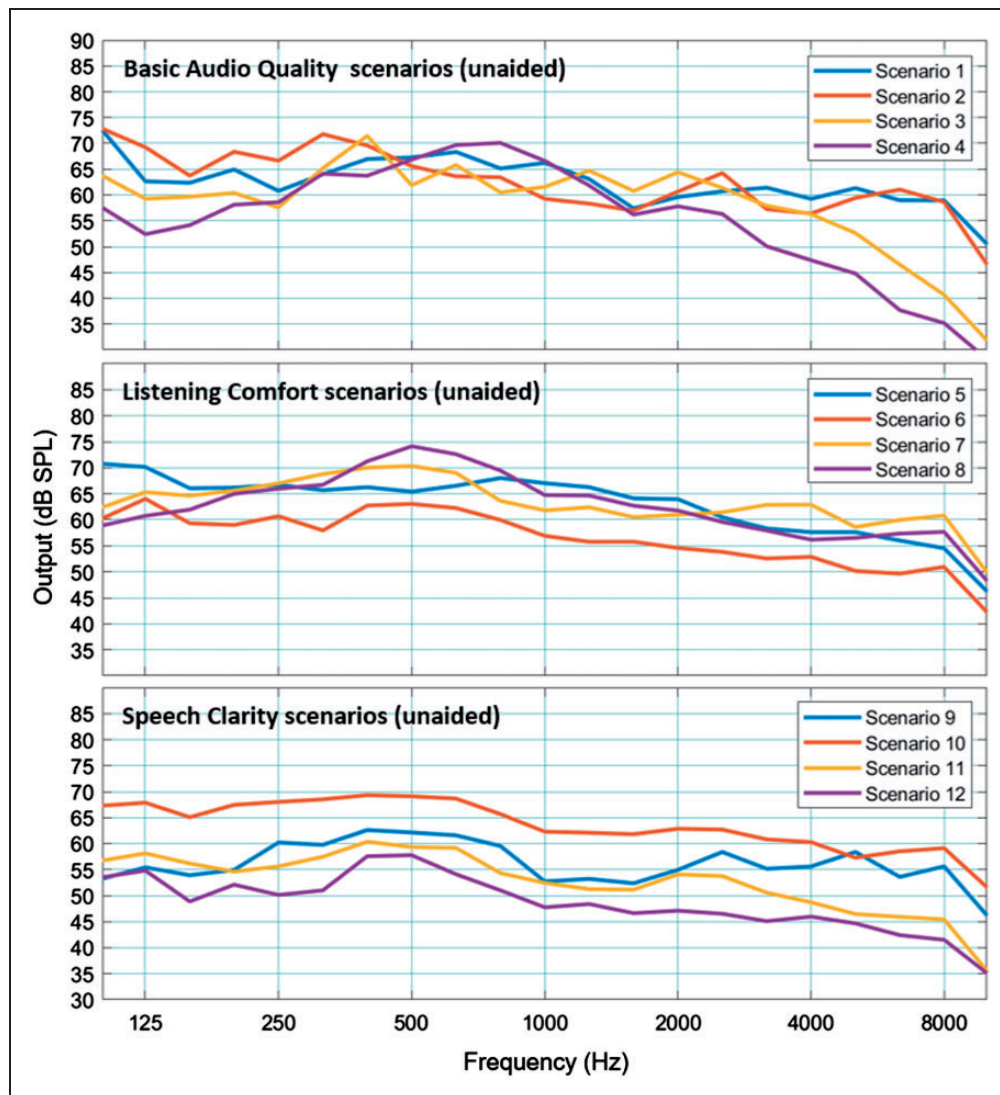


Figure 3. Frequency responses for the Basic Audio Quality scenarios (upper panel), the Listening Comfort scenarios (middle panel), and the Speech Clarity scenarios (lower panel), based on recordings made in the open ear of KEMAR. The frequency responses show 1/3 octave SPL values averaged across both ears of KEMAR.

Prior to each SSL adjustment, the hearing aids were programmed in the UNI setting corresponding to the given sound scenario (see earlier). The programming was done using a separate fitting PC with the fitting software (including custom scripts for hearing-aid programming) installed. Wireless connection to the hearing aids was established using a Widex USBLink neckloop (connected to the PC via an USB cable), which was mounted around the neck of the KEMAR. The hearing aids were connected to the iPad used for the SSL adjustment using a Bluetooth connection, and the hearing aids thus had to be paired with the iPad prior to the adjustment.

The SSL adjustment procedure was initiated in the app. Prior to the actual adjustment, each subject completed a training session consisting of seven A-B comparisons in order to get familiar with the user interface

and the task. After completing the training, the SSL adjustment was completed for each of the 12 scenarios. The sound was turned on prior to activating the app to allow the adaptive processing in the hearing aids to settle. During the adjustment, the participant had access to a sheet with a still picture of the scenario (taken from the video) to allow them to recollect the situation from the familiarization procedure. Furthermore, they had access to the description of the attribute which they were supposed to focus on (and optimize) during the adjustment procedure.

When the final SSL setting for a given scenario had been determined after 20 iterations, the setting was saved as a separate program in the app. When determining the scenario presentation order, the scenarios were grouped according to the three attributes so that the participants

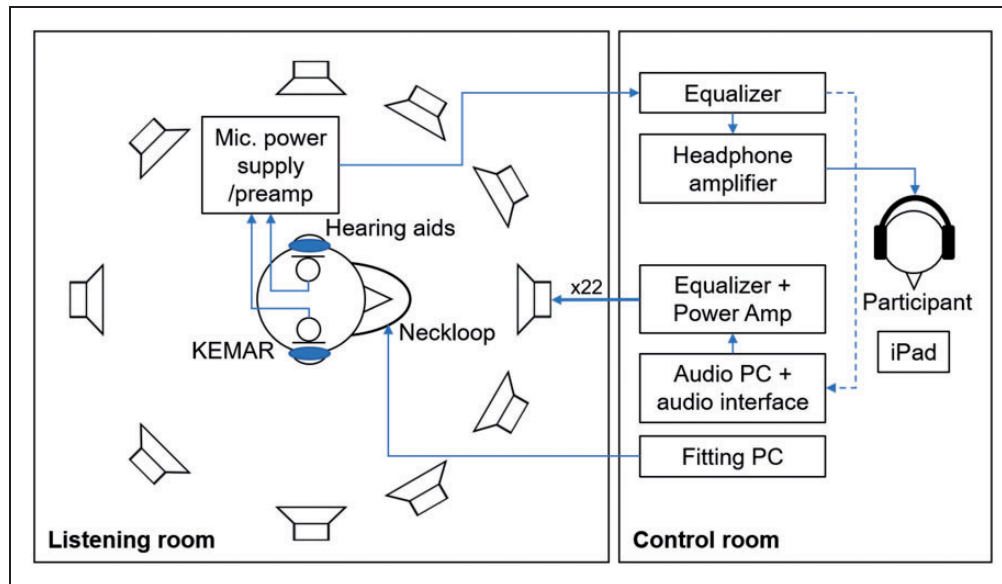


Figure 4. The setup used for SSL adjustments and the subsequent recordings of processed sound scenarios. The addition needed to establish the recording setup is indicated by the dotted connection between equalizer and Audio PC in the control room. The figure only shows the middle horizontal ring of loudspeakers. The hearing aids were connected wirelessly with neckloop and iPad.

would not have to change their focus (from one attribute to another) between scenarios. The order of the three attributes was counterbalanced across participants. Within each group, the presentation order of the four scenarios was randomized.

Recording of test samples

When all 12 SSL adjustments had been made, the participant was sent home. For each sound scenario, the hearing aids (still mounted on KEMAR) were then programmed in the three settings: REF, UNI, and SSL. When programmed in a given setting, each of the 12 sound scenarios was reproduced by the loudspeaker setup, and a recording of the signal captured by the KEMAR microphones was made at 48 kHz/24 bit. Again, an equalizer was used to compensate for the KEMAR ear canal and headphone response. Thus, for each subject a total of 3 (settings) \times 12 (scenarios) = 36 recordings were made. For each recording, the signal was on for 30 s before the actual recording was made in order for the adaptive features in the hearing aid to settle according to the scenario. The duration of the recordings was between 20 and 30 s (equal to the length of the sound sample). Following the recording, all recordings were cut and time-aligned to enable direct crossfade during the following assessment procedure.

Assessment Procedure

At a second visit to the lab, the participants completed a comparison of the three hearing-aid settings in the

same 12 sound scenarios used for the SSL adjustments. The comparison was administered by the SenseLabOnline software (developed by SenseLab), run on a PC, which allowed for a double-blind comparison where neither participant nor experimenter was aware of the identity of the settings that were compared at any given time. A balanced design was used where the scenarios were grouped in blocks of four according to the attribute (for the same reasons as the blocking used during the adjustments), and where the order of the blocks was counterbalanced across participants, while the presentation order of the four scenarios was randomized.

The participants listened to the recorded scenarios via Sennheiser HDA650 headphones, where the signal from the PC was routed through a FiiO Andes headphone amplifier. The participants had access to the same descriptions of scenarios and corresponding attributes that were used during the SSL adjustments. During the assessment, the participants were seated in front of a screen displaying a Danish version of the user interface shown in Figure 5. The participants' task was to listen to all three settings, always shown in random order on the screen, and rate each of the settings with respect to the relevant attribute (Basic Audio Quality, Listening Comfort or Speech Clarity). The assessment of each setting was made by placing a slider on a continuous rating scale with five labels (English translations shown in Figure 5). The rating of each setting (the slider position on the scale) was transformed to a rating between 0 and 100. The participants could switch between the three settings as they wanted, and they had the option to

zoom in on a selected part of the sample (however, this last option was rarely used).

When the ratings of the settings had been made in all 12 scenarios, the entire program was repeated (with the attributes and scenarios presented in another order). Thus, the test design included two ratings of all combinations of hearing-aid setting and sound scenario, that is,

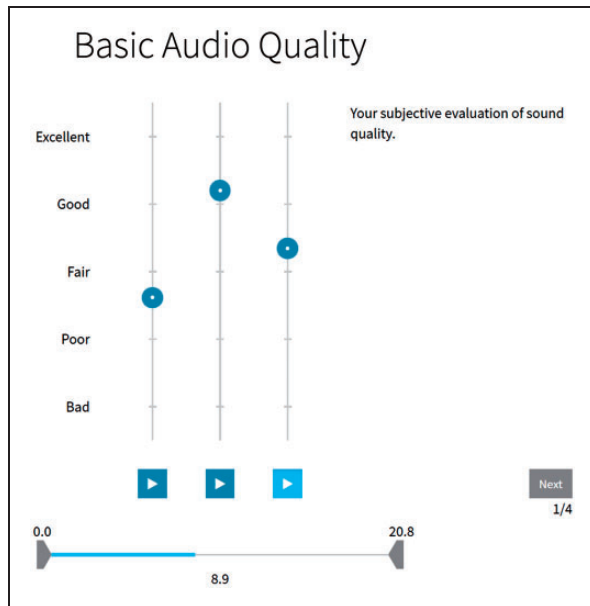


Figure 5. The graphical user interface used for rating the three settings in a given sound scenario. The participant switched between the three settings using the buttons below the scale, and the ratings were made by positioning the sliders on the scales.

a total of 3 (settings) × 12 (scenarios) × 2 (repetitions) = 72 ratings made by each participant.

Results

All data analysis was performed using the Tibco Statistica software (version 13).

Rating Data

All 20 participants completed the entire protocol. Thus, all combinations of hearing-aid setting and sound scenario were rated twice by all participants, and there were no missing data. Box plots showing the distributions of ratings of the hearing-aid settings are shown in Figure 6 for each of the three attributes. For each combination of attribute and hearing-aid setting, each participant contributed eight data points (two ratings of four scenarios), and each boxplot is therefore based on 160 data points. The mean ratings across all 160 data points are indicated with squares in Figure 6.

Interquartile plots of residuals indicated that the assumption about normally distributed residuals was fulfilled, and the rating data for each of the three attributes were analyzed using a general linear mixed model analysis of variance (ANOVA). The same structure was used for each of the three models, with Rating (0–100) being the dependent variable, Participant (1–20) as a random effect, and Setting (REF, UNI, and SSL), Scenario (1–4) and Repetition (1–2) as fixed effects. All second-order interactions, but no third and higher order interactions, were included in the models. The outcome of each of the three analyses is summarized in Table 3.

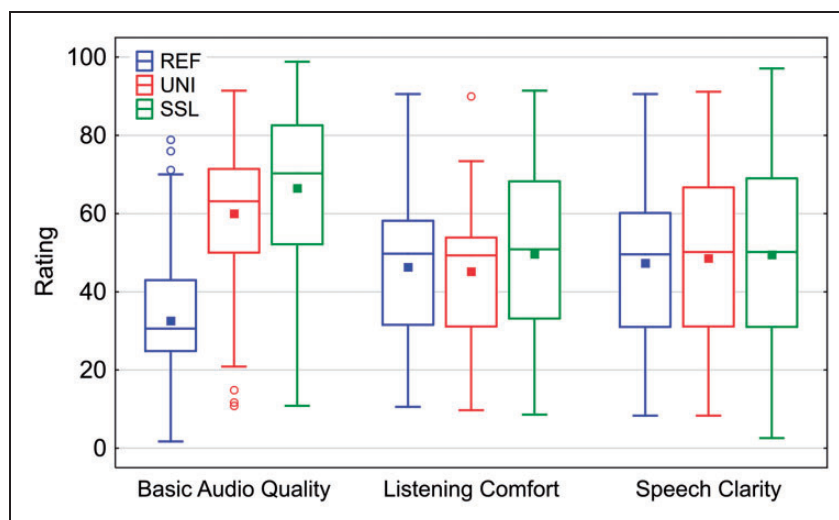


Figure 6. Boxplots of ratings of each of the three settings, REF, UNI, and SSL, on each of the three attributes. The boxes indicate quartiles, while whiskers indicate minimum or maximum of nonoutliers. Mean values are shown as squares, while outliers (outlier coefficient 1.5) are marked as circles. REF = Reference; UNI = Universal; SSL = SoundSense Learn.

Table 3. *df* and *F*- and *p* Values for the Effects Included in the Mixed-Model ANOVAs for Each of Three Attributes, BAQ, COM, and CLA.

| Effect | <i>df</i> | BAQ | | COM | | CLA | |
|--------------------------|-----------|----------|-----------|----------|-----------|----------|-----------|
| | | <i>F</i> | <i>p</i> | <i>F</i> | <i>p</i> | <i>F</i> | <i>p</i> |
| Participant | 19 | 1.74 | .068 | 2.57 | .0060* | 2.09 | .019* |
| Setting | 2 | 40.05 | <.00001* | 1.23 | .30 | .36 | <i>ns</i> |
| Scenario | 3 | 4.76 | .0050* | .12 | <i>ns</i> | 28.52 | <.00001* |
| Repetition | 1 | 1.93 | .18 | 1.25 | .28 | 1.4 | .25 |
| Participant × Setting | 38 | 9.14 | <.00001* | 3.78 | <.00001* | 2.57 | <.00001* |
| Participant × Scenario | 57 | 1.52 | .013* | 1.21 | .16 | 2.70 | <.00001* |
| Participant × Repetition | 19 | 1.23 | .23 | 1.94 | .011* | 1.16 | .29 |
| Setting × Scenario | 6 | 1.95 | .073 | 3.13 | .0054* | .89 | <i>ns</i> |
| Setting × Repetition | 2 | .31 | <i>ns</i> | 2.03 | .13 | .28 | <i>ns</i> |
| Scenario × Repetition | 3 | .82 | <i>ns</i> | 1.61 | .19 | 1.60 | .19 |
| Error | 329 | — | — | — | — | — | — |

Note. Significant effects ($p < .05$) are indicated by an asterisk (*). *df* = degrees of freedom; ANOVAs = analyses of variance; BAQ = Basic Audio Quality; COM = Listening Comfort; CLA = Speech Clarity.

The mean rating of Basic Audio Quality was 32.6 for the REF setting, 60.0 for the UNI setting, and 66.5 for the SSL setting. Thus, turning on the sound classifier offered a substantial Quality improvement of 27.9 scale points, and applying the SSL procedure offered a further Quality improvement of 6.5 scale points. The statistical analysis showed that the main effect of Setting was significant ($p < .0001$, see Table 3). A Tukey's HSD post hoc test showed significant differences ($p < .05$) between all pairs of the three mean values.

Averaged across the three settings, the mean Quality ratings of the four sound scenarios were 54.2 (Scenario 1), 50.3 (Scenario 2), 51.0 (Scenario 3), and 56.6 (Scenario 4), and the ANOVA showed a significant main effect of Scenario ($p = .005$, see Table 3). A Tukey's HSD post hoc test on the Scenario effect showed significant differences ($p < .05$) between Scenarios 2 and 4 and between Scenarios 3 and 4, while none of the other pairwise differences was significant. Thus, the main effect was mainly driven by Scenario 4 (Mozart) that was rated higher than two out the other three scenarios.

Although the main effect of Participant failed to reach the .05 level of significance in the analysis of the Quality ratings, it should be noted that Participant interacted significantly with both Setting and Scenario (see Table 3). Thus, not surprisingly, the effect of Setting and Scenario varied between participants, which contributes to the rather large spread in ratings. The Participant × Setting interaction is illustrated in the upper panel of Figure 7 where the ratings of each setting (averaged across four scenarios and two repetitions) are plotted for each participant (numbered 1–20). It is quite evident that while the overall trend (i.e., REF < UNI < SSL) is found in the data for many of the participants, some very different

individual patterns are also observed. It is, for example, noteworthy that two participants (10 and 20) rate SSL substantially lower than UNI. It is also noteworthy that one participant (3) rated all three settings rather poorly (mean ratings around 20–25).

Turning to the analysis of the Listening Comfort data, the mean rating of Comfort was 46.3 for the REF setting, 45.2 for the UNI setting, and 49.6 for the SSL setting. The statistical analysis showed that the main effect of Setting was not significant ($p = .30$, see Table 3), and accordingly, the pairwise differences between settings were not tested. However, the significant interaction between Setting and Scenario ($p = .0054$, see Table 3) indicates that the effect of setting varied across the scenarios. This is illustrated in Figure 8 that shows the mean Comfort ratings for all combinations of settings and scenarios. Although the differences between settings are quite small for Scenarios 6 to 8, there is a somewhat clearer trend in Scenario 5 where the mean ratings are 41.2 for the REF setting, 43.8 for the UNI setting, and 54.1 for the SSL setting. Thus, turning on the sound classifier offered a mean Comfort benefit of 2.6 scale points, while the SSL adjustment offered an additional Comfort benefit of 10.3 scale points. A Tukey's HSD post hoc test on the Setting × Scenario effect showed that the differences between SSL and the two other settings observed for Scenario 5 were the only two significant ($p < .05$) pairwise differences in the Comfort ratings. Thus, while a general Comfort benefit of SSL was not observed, SSL provided a benefit in this specific scenario. Scenario 5 was a traffic-noise signal, and it differed from the three other Comfort scenarios by not including speech babble. This could indicate that the task of increasing Listening Comfort was easier to solve when speech babble was not present.

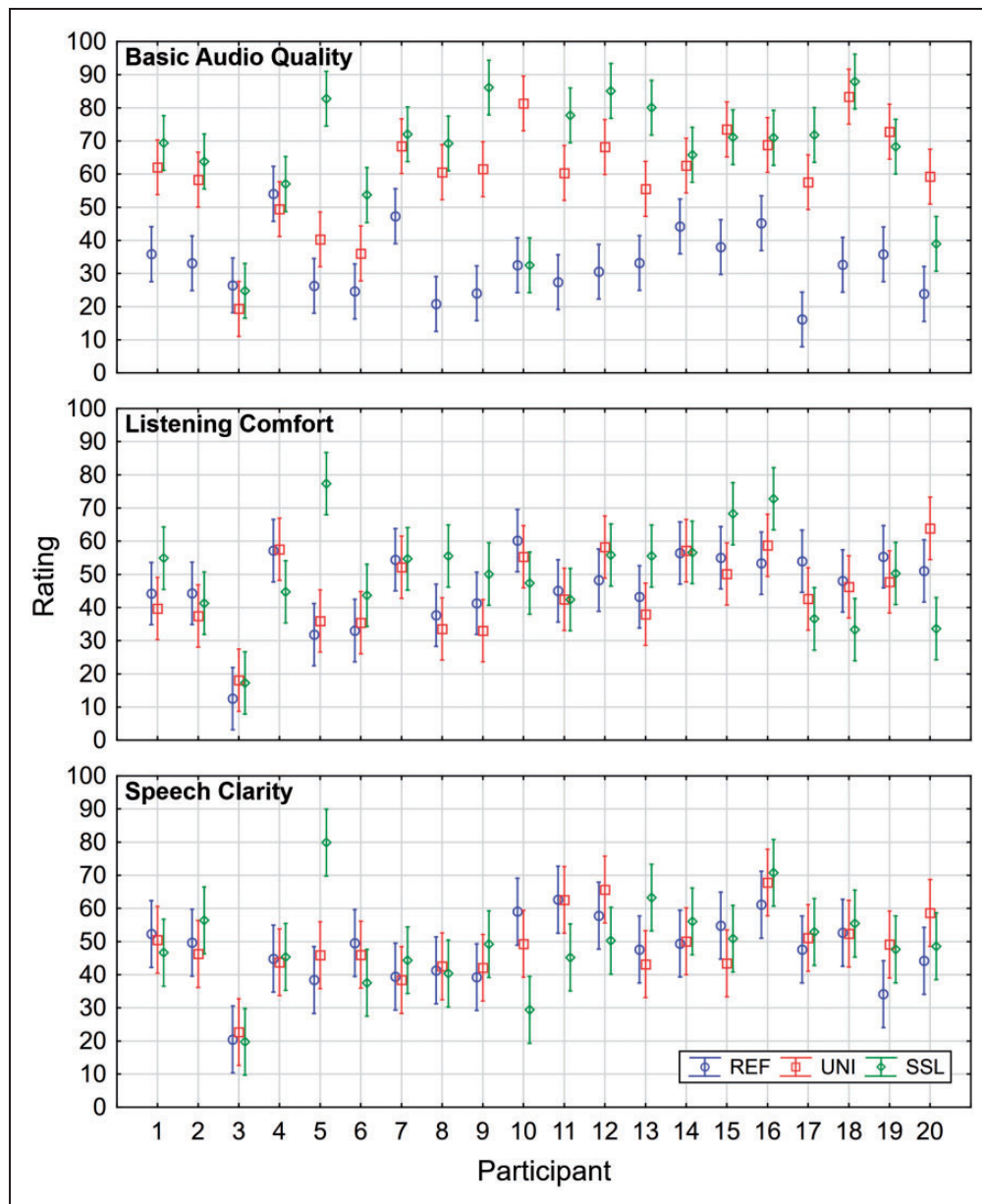


Figure 7. Individual ratings of Basic Audio Quality (upper panel), Listening Comfort (middle panel), and Speech Clarity (lower panel), averaged across four scenarios and two repetitions, for each of the three hearing-aid settings, and for each of the 20 participants. Error bars indicate 95% confidence intervals. REF = Reference; UNI = Universal; SSL = SoundSense Learn.

In contrast to the analysis of the Quality ratings, the Participant main effect was significant in the Comfort analysis. Furthermore, the interaction between Participant and Setting was significant, which is illustrated in the middle panel in Figure 7. Some of the trends observed for the Quality ratings were also seen in the Comfort ratings. Thus, there were subjects showing rather large preferences for SSL, but also some who gave lower ratings to SSL (including the two participants, 10 and 20, who also rated SSL lower in the Quality domain). It should also be noted that the

participant (3) who provided very low Quality ratings did the same in the rating of Comfort.

Finally, the analysis of the Speech Clarity data showed that the mean rating of Clarity was 47.3 for the REF setting, 48.6 for the UNI setting, and 49.5 for the SSL setting. The statistical analysis showed that the main effect of Setting was not significant (see Table 3), and accordingly, the pairwise differences between settings were not tested. As for the other two attributes, a significant interaction between participant and setting was observed, which is illustrated in the lower panel of

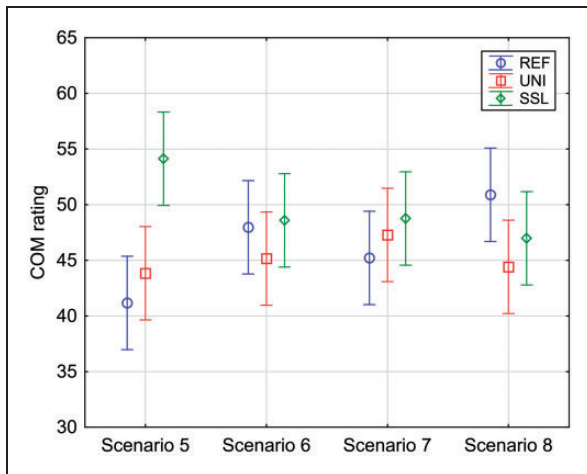


Figure 8. Mean Listening Comfort ratings (across 20 participants and two repetitions) for each combination of setting and scenario. Error bars indicate 95% confidence intervals. REF = Reference; UNI = Universal; SSL = SoundSense Learn; COM = Listening Comfort.

Figure 7. Some of the same observations that were made for the two other panels in the same figure can be made for this one. That is, both cases of relatively high and relatively low SSL ratings were observed, and the participant (3) with low ratings of Quality and Comfort also provided low ratings of Clarity.

Averaged across the three settings, the mean Clarity ratings of the four sound scenarios were 62.0 (Scenario 9), 43.9 (Scenario 10), 52.9 (Scenario 11), and 35.1 (Scenario 12), and the ANOVA showed a highly significant main effect of Scenario ($p < .0001$, see Table 3). However, there was no significant interaction between Setting and Scenario, indicating that the effects of the settings did not vary across the rather different scenarios.

For all three attributes, the general trend was that the ratings increased slightly from the first to the second rating of the same condition. This could indicate the presence of a training effect. However, the Repetition effect was not significant in any of the three cases (see Table 3). Accordingly, this effect will not be taken into account in the discussion of the results.

Gain Adjustments

The effects of the SSL adjustments on the long-term average output frequency response of the hearing aids, measured in the ear of KEMAR and averaged across participants, for each of the 12 sound scenarios are shown in Figure 9. The SSL effect corresponds to the difference between the SSL and UNI frequency responses. Furthermore, each panel includes the frequency response of the REF setting, allowing a direct comparison of the mean frequency response of all three settings.

A general observation across sound scenarios is that the frequency responses for the three settings are quite close to each other (within 5–6 dB), but as it will be shown later, there was a substantial individual variation in the adjustments. The only major deviations are the REF responses in the Quality scenarios where substantially lower gain is observed at low and mid frequencies compared with the other settings (see the left panels in Figure 9). The main explanation for this is that the adaptive processing (besides the sound-classification system) was applied in the REF setting. Since all the music samples used for the Quality scenarios activated the noise-reduction system in the hearing aid, a rather substantial gain reduction was effectuated in the REF setting. When the sound-classification system is turned on and music is detected by the system, the noise reduction will be deactivated, and gain at low and mid frequencies will be substantially increased as seen by the difference between the REF and UNI frequency responses.

It should be noted that for some of the scenarios, the REF and UNI settings have the same mean frequency response. In these cases, the given sound class activated in the UNI setting does not include a change of specific static gain parameters, but it would involve changes of other hearing-aid parameters (e.g., compression time constants) that have audible effects without affecting the long-term output frequency response shown in Figure 9. The lack of frequency-response differences and the fact that no extreme differences in compression-parameter settings appear across the sound classes may have made it more difficult for some of the participants to hear and relate to the differences between these REF and UNI settings during the Comfort and Clarity rating tasks. This could have contributed to the observation in Figure 7 that the individual mean ratings of REF and UNI were close to each other for many participants. However, it should be noted that other participants did rate UNI and REF differently on the Comfort and Clarity attributes, indicating that the differences in processing between settings provided a perceivable difference.

To explore the SSL adjustments made by the participants, the change in gain relative to the baseline (i.e., the UNI setting) in the three frequency bands was read directly from the app. The app indicated the adjustment in steps (of 1 or 2 dB, depending on the band), which were converted here to dB. All the individual gain adjustment values in the three bands are plotted in Figure 10 for each of the 12 scenarios, with red for the Low, green for the Mid, and blue for the High frequency band. The scale for gain adjustment from –12 dB to +12 dB is shown on the left y -axis; note that it is only in the Low band that gain increases between +6 dB and +12 dB are possible.

The panels in Figure 10 also include the perceptual effects of the adjustment, as measured by the ratings

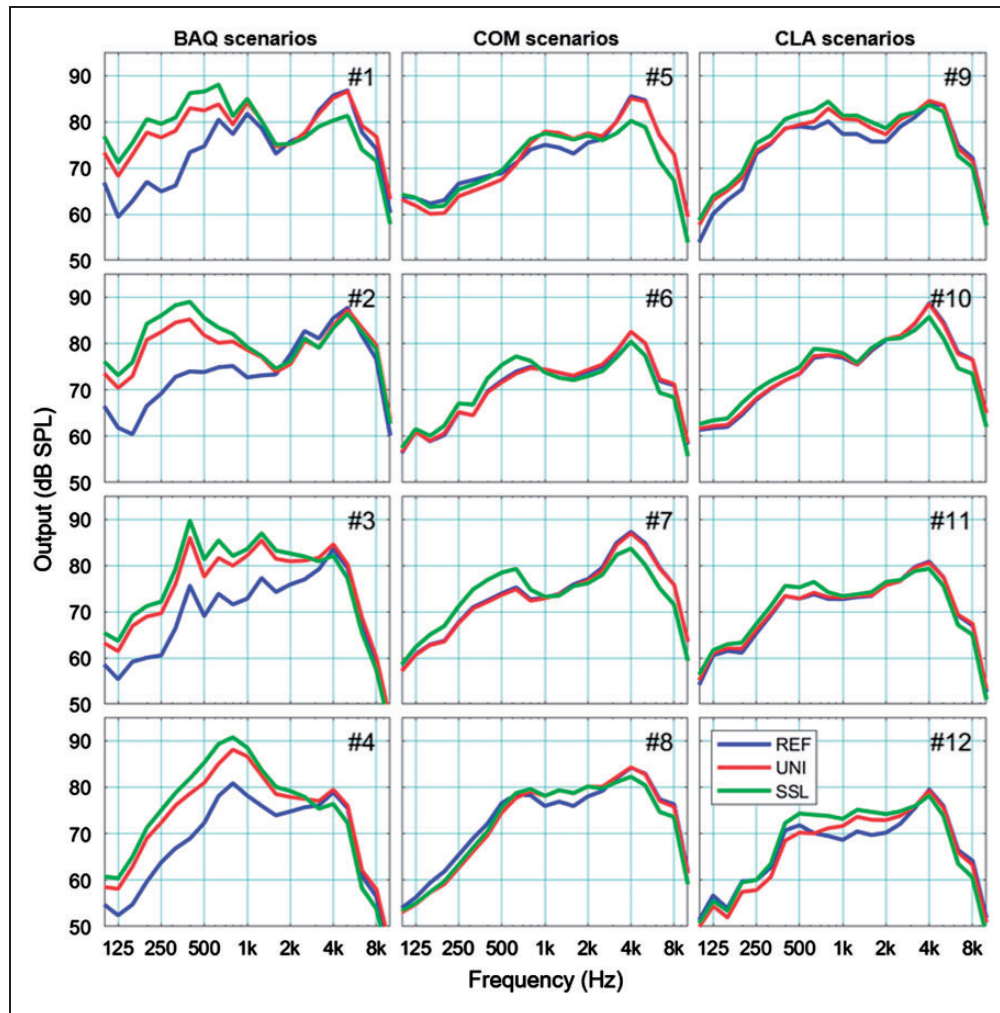


Figure 9. Long-term average output frequency responses for the three hearing-aid settings in the four Basic Audio Quality scenarios (left panels), the four Listening Comfort scenarios (middle panels), and the four Speech Clarity scenarios (right panels), measured in the ear of KEMAR. Each frequency response shows 1/3 octave SPL values averaged across both ears of KEMAR. BAQ = Basic Audio Quality; COM = Listening Comfort; CLA = Speech Clarity; REF = Reference; UNI = Universal; SSL = SoundSense Learn.

reported earlier, in the following way: For each SSL adjustment, the “SSL benefit” was calculated as the difference between the SSL rating and the UNI rating, averaged across the two repetitions. In each of the 12 panels in Figure 10, the participants (shown on the x -axis) have been rank-ordered according to the magnitude of their SSL benefit, which also is shown in the plot with open circles (right y -axis). In each panel, a vertical dotted line is drawn between the participants with negative SSL benefits (i.e., SSL disadvantages) and those with positive SSL benefits. For the gain adjustments in each of the three frequency bands, linear regression lines have been added to show the trend in the gain adjustments with respect to the SSL-benefit rank order.

Various noteworthy observations may be made based on the data presented in Figure 10. One is that the entire range of gain adjustments (-12 to $+12$ dB for the Low

band and -12 to $+6$ dB for the Mid and High bands) was used for all three bands in almost all the scenarios. At the same time, there are some obvious overall trends in the gain adjustments. The mean gain adjustment (across participants) in each band corresponds to the midpoint of the regression line drawn for that band, and the general trends were an increase of gain in the Low band and a decrease of gain in the High band, with the gain adjustment the Mid band being somewhere in the middle. The mean gain adjustments for the four Quality scenarios were 5.8 dB, 1.1 dB, and -2.9 dB for the Low, Mid, and High band, respectively. For the Comfort scenarios, the similar adjustments were 5.0 dB, -0.4 dB, and -3.3 dB, and for the Clarity scenarios, they were 5.0 dB, 1.5 dB, and -1.5 dB. Thus, across attributes the most pronounced mean adjustment away from the baseline (the UNI setting) was an increase of the gain in the Low

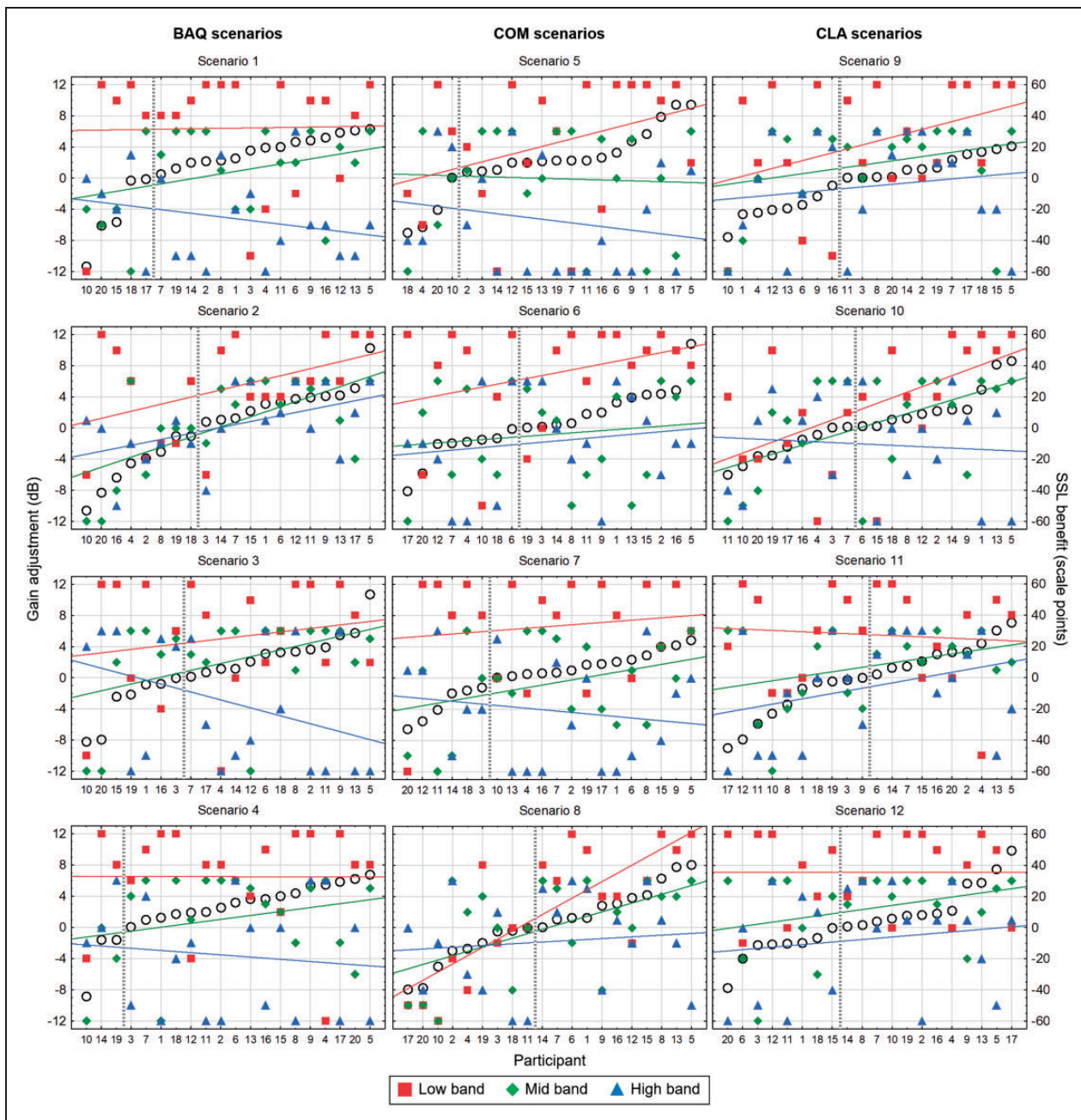


Figure 10. Individual SSL gain adjustments away from the UNI gain setting (left y-axis) in the Low band (red), Mid band (green), and High band (blue) for each of the 12 sound scenarios. In each panel, the participants (on the x-axis) have been rank ordered according to the magnitude of their SSL benefit (i.e., the difference between ratings of SSL and UNI). The SSL benefit (right y-axis) is indicated by black open circles. Linear regression lines have been added to indicate overall trends in gain adjustment as functions of SSL-benefit ranking. The vertical dotted line in each panel separates participants with positive and negative SSL benefits. The participant numbers on the x-axis correspond to those shown in Figure 7. BAQ = Basic Audio Quality; COM = Listening Comfort; CLA = Speech Clarity; SSL = SoundSense Learn.

band. The panels in Figure 10 also clearly show that the majority of the gain adjustments in the Low band (the red squares) were positive. However, it should be noticed that cases of no or negative adjustments in the Low band occurred, also among those participants who benefited from SSL (as indicated by a positive SSL benefit score).

In all 12 sound scenarios, there are several cases of gain adjustments reaching either the lower limit (-12 dB) or the upper limit ($+12$ dB in the Low band, $+6$ dB in the Mid and High bands) of the adjustment range. Thus, the average gain adjustments reported earlier have been affected by these floor or ceiling effects, and it seems

likely that even larger adjustments would have occurred if the method had allowed it.

The regression lines included in the plots do not indicate a consistent pattern when it comes to possible connections between the SSL adjustments made and the SSL benefit obtained. In all cases, it should obviously be noted that the variation of the adjustments around the regression lines was quite substantial (and even limited by the floor or ceiling effects mentioned earlier). To investigate the relationships between gain adjustment and SSL benefit, a Spearman rank correlation coefficient was calculated for all three bands in all 12 scenarios. After applying a Bonferroni correction to account for multiple correlations, it was only the correlation between the Low band adjustment and the SSL benefit in Scenario 8 that came out as significant (Spearman $R = .76$, $p < .01$). Although the plot for Scenario 8 indeed shows a rather clear association between increase in Low gain and SSL benefit, it also includes individual cases showing the opposite pattern. Thus, the overall observation is that it is difficult to predict the resulting SSL benefits based on the gain adjustments. This underlines the individual nature of the adjustments made by the participants.

Discussion

A main research question in this study was whether the SSL adjustment procedure could provide perceptual benefits for the participants, as indicated by their ratings of the three settings, and in particular by the difference between the ratings of the SSL and UNI settings. The results reported earlier indicate that the answer depends on the sound scenario as well as the sound attribute that the participants were instructed to optimize during the SSL adjustment and subsequently focus on when comparing the hearing-aid settings. In the Basic Sound Quality domain, a significant benefit of the SSL was observed compared with the two other settings. The results were more mixed in the Listening Comfort domain, where an SSL benefit was only observed in one of the four scenarios. In the Speech Clarity domain, benefits on a group level were not observed in any of the scenarios. However, for all three domains, there were differences in individual performance across participants, where SSL benefits were observed for some participants, while others experienced disadvantages.

The analysis of the Basic Audio Quality rating data showed a significant main effect of setting and significant differences between all three settings (REF, UNI, and SSL). Thus, the results indicated a benefit of turning on the sound classification system (in the UNI setting) and a further benefit of making the SSL adjustment. It should be noted that the mean rating of REF across the four Quality scenarios (32.6) was quite low, being

just above the “poor” label on the rating scale, see Figure 5. The obvious explanation for this is the rather substantial low- and midfrequency gain reduction caused by the noise reduction when the sound-classification system is turned off (as seen in Figure 9). The overall findings for the Quality attribute correspond quite well with the results reported by Townend et al. (2018), who used a similar experimental approach, but implemented in another setup and using different sound scenarios, to investigate the SSL procedure. Using a general Sound Quality attribute, they reported a mean SSL benefit of 8.9 scale points when comparing with a setting similar to the UNI setting used in this study (and using the same type of rating scale), as compared with the mean SSL benefit of 6.5 scale points found in this study.

When participants had the task of optimizing Listening Comfort in the SSL adjustments, they only succeeded in making a significant mean improvement in one out of the four scenarios. The scenario was traffic noise (No. 5), which, in contrast to the three other Comfort scenarios, did not include speech. As already speculated, the fact that speech babble was part of the three other scenarios may have contributed to making the adjustment task more difficult by adding confusion in terms of what to focus on. This was supported by spontaneous comments made by some participants who said that they found it “impolite to listen in on private conversations” when they listened to the three Comfort scenarios including speech babble. This reaction could have introduced inconsistency in their paired comparisons, which would reduce the effectiveness of the SSL procedure in terms of finding the correct maximum of the IRF (Nielsen et al., 2015). Townend et al. (2018) also included Listening Comfort in their investigation and found a significant SSL benefit of almost 20 scale points across three different sound scenarios. However, they also found the strongest effect for a traffic-noise scenario and smaller effects when speech babble was part of the scenario, which is consistent with the current results.

In the discussion of the SSL effects on Listening Comfort, it should be noted that none of the four scenarios included in this study was very loud. Three of the scenarios were presented at a level of 70 dBA, and the level of the fourth (No. 6, café noise) was 63 dBA. The motivation for this choice was that the SSL adjustments should not be aimed at just reducing gain (and thereby the loudness) to avoid uncomfortably loud levels. This was also reflected in the definition of the Listening Comfort attribute given to the participants as part of the instructions, where they were told to focus on the ability to “hear the surroundings and the people around you with an appropriate balance between them.” The analysis of the gain adjustments indeed showed that the participants did not just turn the gain

down to improve comfort. In fact, on average they increased the gain in the Low band, while only minor mean gain reductions were observed in the Mid and High bands. Especially for one of the Comfort scenarios (No. 8, outdoor setting), it is interesting to notice that the participants who reduced gain the most (in particular at low frequencies) during the SSL adjustment were the ones who perceived the lowest SSL benefit, see Figure 10. The definition of Listening Comfort deviated somewhat from the definition used by Townend et al. (2018). The latter focused more on the ability to tolerate being in a given sound scenario for a long time, and it could be speculated that a higher focus on tolerance urged participants to prefer a larger gain reduction. Thus, the difference in definitions may have contributed to the difference in SSL benefits between the studies.

For Speech Clarity, there was no significant difference between the three settings across the four scenarios. Among the three attributes included in the study, it was the one where the application of the SSL procedure seemed least likely to provide a benefit. Townend et al. (2018) included (perceived) speech intelligibility as an attribute in their study and found no SSL benefit for that domain. That finding was not that surprising, given the fact that they compared the SSL setting with a setting (similar to the UNI setting used in the present study), which in fact is designed to optimize speech intelligibility for the “average user.” Since the SSL procedure only adjusts gain in three rather wide frequency bands, it is quite difficult (if not impossible) to significantly improve speech intelligibility if the baseline is fairly close to the optimal gain setting, which it is expected to be in many cases. This was the reason for not using speech intelligibility as an attribute in this study. Instead, it was decided to include the sound quality-related attribute Speech Clarity, where it as expected to be more likely that gain adjustments in three bands could have a perceptible positive effect, even if speech intelligibility as such was not affected. However, that did not seem to be the case. One explanation could of course be that speech clarity was already optimized by the UNI setting, but another explanation could be that it, in practice, was difficult for the participants to separate speech clarity from perceived speech intelligibility. To actually improve clarity (or speech intelligibility) in cases where the baseline gain setting is already optimized, it would probably be necessary to give the SSL procedure access to other hearing-aid parameters, for example, compression time constants or directionality parameters. The SSL-benefit differences across attributes correspond well with spontaneous comments from some of the participants who indicated that it was easier to relate to the Quality attribute, and thereby easier to make the SSL adjustments and the subsequent ratings, than it was for the Comfort and Clarity attributes.

Across all three sound attributes, there was large individual variation in both the ratings and in the observed SSL benefits, as well as in the gain adjustments made. Some individual cases have already been highlighted when reporting the results earlier. Compared with the remaining participants, one participant (3) did, quite consistently, provide very low ratings on all three sound attributes. From a pure data-distribution perspective, this participant acted as an outlier, but since no known technical issues occurred during neither adjustment nor assessment, since no other explanation for the low ratings could be found, and since the participant in all other aspects performed the task according to the instructions, the participant’s data were not excluded from the analysis. However, it remains a fact that this specific subject drags the between-subject variance up and all mean ratings down.

Other special cases included two participants (10 and 20) who experienced an SSL disadvantage on all three attributes. Although the magnitude of the disadvantage varied across attributes, it was in some cases quite substantial. The most likely explanation is that these two participants were unable to perform the adjustment task in a consistent manner, which means that the machine-learning algorithm did not converge toward the maximum of the IRF (Nielsen et al., 2015). Such cases suggest the need for a confidence indicator that can provide information about whether convergence has occurred during the optimization procedure. Such a convergence indicator was not available at the time of this study, but it has subsequently been developed and added to the commercially available SSL procedure. In the study, more training in the adjustment task could perhaps have made a difference, but it could also be speculated that the task is too cognitively demanding for some people. Participants in this study were not screened for cognitive abilities, but it should be noted that participant 10 (83 years old) was the oldest of all participants in the study, whereas participant 20 (71 years old) was closer to the mean age of the participants. From a clinical perspective, these results indicate that there may be users who will have difficulties obtaining a benefit from self-adjustment procedures like the SSL procedure. But it should be noted that the remaining 18 participants obtained an average SSL benefit on at least one of the attributes, and six of them obtained a benefit for all three attributes. This indicates, on the other hand, that there may be many users who will be able to benefit from the method, at least in some situations and for some listening intentions.

In a discussion of the variation in the observed SSL benefits, it also needs to be reiterated that the SSL procedure only affected static hearing-aid gain, while dynamic parameters of the hearing-aid processing were not included. Ignoring the dynamic aspects obviously

means that the SSL procedure only searched for a preference optimum within a limited part of the global parameter space provided by the hearing aid. Previous research has shown that the dynamic performance of compression indeed has an effect on the perceived benefit, and that different compression settings may be preferred for different outcome domains (Gatehouse, Naylor, & Elberling, 2006a). Furthermore, part of the variation in benefit across compression systems with different dynamic characteristics may be explained by personal factors such as cognitive capacity and auditory ecology of the listener (Gatehouse, Naylor, & Elberling, 2006b). Including dynamic parameters in the SSL procedure would perhaps have provided other patterns in the SSL-benefit variation and taking other personal factors into account could perhaps have explained some of this variation. As suggested previously, including other (dynamic) processing parameters could perhaps also have improved the capability of the SSL procedure to provide benefits on speech clarity.

In the design of the study, we assumed that the listening intentions dictated by the three sound attributes would make sense to the participants in all the 12 sound scenarios. However, the participants were not asked about what their intention would have been in the different scenarios if a given listening intention had not been imposed upon them. If a discrepancy existed between the “enforced” and the “personal” listening intention, this could in principle have decreased the observed benefit of SSL. In this respect, it could be argued that the study design actually disregarded one of the main reasons for developing this type of self-adjustment method, namely, the ability to address listening intentions different from those suggested by the sound scenario.

The most striking trend in the gain adjustments was an increase in LF gain, which came out of the SSL adjustments made by many (but not all) of the participants. Other studies involving various self-adjustment methods have also found a preference toward increasing LF gain compared with a baseline determined by a generic fitting rationale (Preminger et al., 2000, Nelson et al., 2018). The mean increase of LF gain could obviously be seen as an indication of insufficient LF gain in the UNI setting and thereby in the underlying fitting rationale. However, the experimental setup needs to be taken into account in this discussion, most importantly that the hearing aids were mounted on a KEMAR in the lab, and not on the ears of the participant in the real world, during the SSL adjustments. One important consequence of this was that the participant’s own voice was not captured and amplified in the experimental setup. The lack of own voice could have affected the LF gain preference (as also suggested by Nelson et al., 2018). Had own voice (and possible issues related to it) been included in the setup, it could have changed the mean

preference toward a lower LF gain setting, closer to the prescribed UNI setting.

Although the KEMAR ear acoustics were compensated for in the recording and headphone reproduction of the scenarios processed by the three settings, no compensation for the individual variation in ear acoustics was introduced. As a consequence, audible differences between what the participants actually listened to and what they would have listened to if they had worn the hearing aids on their own ears may have occurred. Thus, in principle, it could be argued that some of the participants who obtained an SSL benefit perhaps would have performed better with the prescribed settings (REF and UNI) if the hearing aids had been fitted on their own ears, and that a smaller (or no) SSL benefit would have been obtained in that situation. However, there is no reason to believe that the variance caused by the fitting on KEMAR is of a much larger magnitude than the between-user variance in the (not unusual) case where a clinical fitting is based on the audiogram only with no further verification of the fitting (e.g., via real-ear measurements).

In this study, the baseline (starting point) for the SSL adjustment was the UNI setting where gain was prescribed by the Widex proprietary fitting rationale, and where the adaptive processing, including the sound-classification system, was turned on. Thus, the actual gain provided by the UNI setting varied across scenarios. This makes it difficult, and in some cases even misleading, to compare the gain adjustments across scenarios. The approach was chosen to reflect the actual use of the system in real life where the adaptive features indeed would be turned on. The adjustments were not performed using other settings as baseline, and the question about whether the final SSL setting depends on the baseline setting can therefore not be assessed in this study. Some studies where participants had self-adjusted hearing-aid gain parameters using different types of user controls (directly affecting the gain) have shown that the choice of baseline has affected the final setting (Dreschler et al., 2008; Keidser, Dillon, & Convery, 2008). However, F. K. Kuk and Lau (1995) did not observe such an effect in a study where they used the modified simplex method to optimize the gain setting, and where they concluded that changes in the baseline did not affect the final setting but only the time required to reach it. The difference between findings may be explained by the type of user interaction. It may make a difference whether the user has to determine the gain adjustments by operating a number of controls, or whether the adjustments are made based on sound preference (in a paired-comparison procedure). This could suggest that the setting determined by the SSL procedure is more likely to be independent of the baseline (which indeed has been the intention in the development of the

machine-learning procedure), but this remains to be verified experimentally.

When considering the adjustments made, it is important to remember that the mean gain changes are based on individual gain changes of much larger magnitudes than the mean, and that the individual gain changes went in both positive and negative directions. These observations correspond well with the large spread in gain changes observed in the study by Nelson et al. (2018). If focusing on only the participants who experienced an SSL benefit, that is, the data points on the right side of the vertical lines in the panels in Figure 10, there seems to be no clear trend in the adjustments. Considering only the six participants (2, 5, 7, 9, 13, and 16) who obtained an SSL benefit for all three attributes, the mean difference (averaged across the six participants) between the maximum and the minimum gain adjustment made for the four Quality scenarios was 8.3 dB in the Low band, 5.3 dB in the Mid band, and 11.8 dB in the High band. For the Comfort scenarios, the corresponding values were 12.7 dB, 8.8 dB, and 12.7 dB, and for the Clarity scenarios, they were 8.8 dB, 12.7 dB, and 10.3 dB. Thus, on average, the gain adjustments (within the same band and for the same sound attribute) spanned a range of 9.6 dB. The need to make these adjustments can accordingly not be addressed by a generic gain change in the fitting rationale. This is in line with the findings by Perry et al. (2019) who found it “unlikely” that a modification of the NAL-NL2 rationale, based on the reported self-adjustments, would result in the preferred gain for many users. From a clinical perspective, these findings suggest that individual users may require individual and quite different solutions in order to improve the setting of their hearing aid in a given situation where a certain listening intention is to be supported. In the discussion of these data (and other data from the study), it is important to remember that they were obtained in a specific test setup and with three specific hearing-aid settings being compared. Thus, generalization of the results to other self-adjustment approaches should be made with caution.

This study was designed to allow for a double-blind comparison of hearing-aid settings and thereby an assessment of the perceptual effects of the SSL procedure. However, this meant that the study did not take into account any feelings of psychological ownership that may be associated with being actively involved in the fitting of one's own hearing aid (Convery et al., 2011). This feeling of being empowered may amplify the satisfaction associated with a pure perceptual benefit, and from a clinical perspective, it may thereby impact the entire experience of using a self-adjustment procedure in real life. More research needs to be done to investigate the ownership effects associated with the use of the SSL method and other self-adjustment methods.

Conclusions

The results showed a significant benefit of the SSL self-adjustment method on Basic Audio Quality. An SSL benefit for Listening Comfort only appeared for traffic noise, while no benefit for this attribute was observed for sound scenarios including speech babble. No SSL benefit for Speech Clarity was observed. No clear relationship between individual gain adjustments and SSL benefit was found. From a clinical perspective, the results of the study indicate that user-driven adjustment of amplification based on a machine-learning approach has the potential to provide benefits for some (but not necessarily all) users, most likely in the Basic Audio Quality domain. The results also indicate that different users have different amplification preferences, within the same sound scenario and for a given listening intention, that are impossible to address with the gain prescribed by one given fitting rationale based on the audiogram.

Acknowledgements

The authors wish to thank Jesper Ramsgaard for his contributions to the planning of the study, and Laura Winther Balling, Francis Kuk, and Martin Dahlquist for their comments during preparation of the manuscript. The authors also thank Michael Stone and two anonymous reviewers for their helpful comments to an earlier version of the manuscript.


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: The study received financial support from Widex A/S.

ORCID iD

Niels Sogaard Jensen  <http://orcid.org/0000-0002-6647-3607>

References

- American Academy of Audiology. (2018). *The audiologist's guide to hearing aids, PSAPs, hearables and OTC devices*. Retrieved from American Academy of Audiology: https://www.audiology.org/sites/default/files/publications/resources/20180130_AuD_Guide_OTC.pdf.
- Aldaz, G., Puria, S., & Leifer, L. J. (2016). Smartphone-based system for learning and inferring hearing aid settings. *Journal of the American Academy of Audiology*, 27(9), 732–749. doi:10.3766/jaaa.15099.
- Alexandre, E., Cuadra, L., Rosa, M., & Lopez-Ferreras, F. (2007). Feature selection for sound classification in hearing aids through restricted search driven by genetic algorithms.

- IEEE Transactions on Audio, Speech, and Language Processing*, 15(8), 2249–2256. doi:10.1109/TASL.2007.905139.
- Amlani, A. M., & Schafer, E. C. (2009). Application of paired-comparison methods to hearing aids. *Trends in Amplification*, 13(4), 241–259. doi:10.1177/1084713809352908.
- Anderson, M. C., Arehart, K. H., & Souza, P. E. (2018). Survey of current practice in the fitting and fine-tuning of common signal-processing features in hearing aids for adults. *Journal of the American Academy of Audiology*, 29(2), 118–124. doi:10.3766/jaaa.16107.
- Baskent, D., Eiler, C. L., & Edwards, B. (2007). Using genetic algorithms with subjective input from human subjects: Implications for fitting hearing aids and cochlear implants. *Ear and Hearing*, 28(3), 370–380. doi:10.1097/AUD.0b013e318047935e.
- Bentler, R. A. (2005). Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence. *Journal of the American Academy of Audiology*, 16(7), 473–484. doi:10.3766/jaaa.16.7.7.
- Boothroyd, A., & Mackersie, C. (2017). A “Goldilocks” Approach to Hearing-Aid Self-Fitting: User Interactions. *American Journal of Audiology*, 26(3 Suppl), 430–435. doi:10.1044/2017_AJA-16-0125.
- Büchler, M., Allegro, S., Launer, S., & Dillier, N. (2005). Sound classification in hearing aids inspired by auditory scene analysis. *EURASIP Journal on Advances in Signal Processing*, 2005(18), 2991–3002. doi:10.1155/ASP.2005.2991.
- 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(1), 37–51.
- Ching, T. Y. C., Johnson, E. E., Hou, S., Dillon, H., Zhang, V., Burns, L., . . . Flynn, C. (2013). A comparison of NAL and DSL prescriptive methods for paediatric hearing-aid fitting: Predicted speech intelligibility and loudness. *International Journal of Audiology*, 52, S29–S38. doi:10.3109/14992027.2013.765041.
- Chung, K. (2004). Challenges and recent developments in hearing aids: Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends in Amplification*, 8(3), 83–124. doi:10.1177/108471380400800302.
- Convery, E., Keidser, G., Dillon, H., & Hartley, L. (2011). A self-fitting hearing aid: Need and concept. *Trends in Amplification*, 15(4), 157–166. doi:10.1177/1084713811427707.
- Convery, E., Keidser, G., Hickson, L., & Meyer, C. (2018). Factors associated with successful setup of a self-fitting hearing aid and the need for personalized support. *Ear and Hearing*. Advance online publication. doi:10.1097/aud.0000000000000663.
- Convery, E., Keidser, G., Seeto, M., & McLelland, M. (2017). Evaluation of the self-fitting process with a commercially available hearing aid. *Journal of the American Academy of Audiology*, 28(2), 109–118. doi:10.3766/jaaa.15076.
- Desjardins, J. L. (2016). The effects of hearing aid directional microphone and noise reduction processing on listening effort in older adults with hearing loss. *Journal of the American Academy of Audiology*, 27(1), 29–41. doi:10.3766/jaaa.15030.
- Dreschler, W. A., Keidser, G., Convery, E., & Dillon, H. (2008). Client-based adjustments of hearing aid gain: The effect of different control configurations. *Ear and Hearing*, 29(2), 214–227. doi:10.1097/AUD.0b013e31816453a6.
- Durant, E. A., Wakefield, G. H., Van Tasell, D. J., & Rickert, M. E. (2004). Efficient perceptual tuning of hearing aids with genetic algorithms. *IEEE Transactions on Speech and Audio Processing*, 12(2), 144–155. doi:10.1109/TSA.2003.822640.
- Elberling, C., & Hansen, K. V. (1999). *Hearing instruments: Interaction with user preference*. In A. N. Rasmussen, P. A. Osterhammel, T. Andersen, & T. Poulsen (Eds.), *Auditory models and non-linear hearing instruments. Proceedings of the 18th Danavox Symposium* (pp. 341–357). Copenhagen, Denmark: Holmens Trykkeri.
- Gatehouse, S., Naylor, G., & Elberling, C. (2006a). Linear and nonlinear hearing aid fittings—1. Patterns of benefit. *International Journal of Audiology*, 45(3), 130–152. doi:10.1080/14992020500429518.
- Gatehouse, S., Naylor, G., & Elberling, C. (2006b). Linear and nonlinear hearing aid fittings—2. Patterns of candidature. *International Journal of Audiology*, 45(3), 153–171. doi:10.1080/14992020500429484.
- Heskes, T., & Vries, B. d. (2005). *Incremental utility elicitation for adaptive personalization*. Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence, Brussels, Belgium (pp. 127–134). Retrieved from <https://repository.uibn.ru.nl/bitstream/handle/2066/32628/32628.pdf>.
- International Telecommunication Union. (2015). *Recommendation ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems*. Retrieved from https://www.itu.int/dms_pubrec/itu-r/rec/b/BS-REC-BS.1116-3-201502-1!!PDF-E.pdf.
- Johansen, B., Petersen, M. K., Korzepa, M. J., Larsen, J., Pontoppidan, N. H., . . . Larsen, J. E. (2017). Personalizing the fitting of hearing aids by learning contextual preferences from internet of things data. *Computers*, 7(1), 1–21. doi:10.3390/computers7010001.
- Keidser, G. (1995). The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids. *Ear and Hearing*, 16(6), 575–586.
- Keidser, G., Brew, C., Brewer, S., Dillon, H., Grant, F., & Storey, L. (2005). The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input. *International Journal of Audiology*, 44(11), 656–670. doi:10.1080/14992020500266803.
- Keidser, G., Brew, C., & Peck, A. (2003). Proprietary fitting algorithms compared with one another and with generic formulas. *The Hearing Journal*, 56(3), 28–32. Retrieved from https://journals.lww.com/thehearingjournal/Fulltext/2003/03000/Proprietary_fitting_algorithms_compared_with_one.5.aspx.
- Keidser, G., & Convery, E. (2016). Self-fitting hearing aids: Status quo and future predictions. *Trends in Hearing*, 20, 1–15. doi:10.1177/2331216516643284.

- Keidser, G., Dillon, H., & Byrne, D. (1995). Candidates for multiple frequency response characteristics. *Ear and Hearing, 16*(6), 562–574.
- Keidser, G., Dillon, H., & Convery, E. (2008). The effect of the base line response on self-adjustments of hearing aid gain. *The Journal of the Acoustical Society of America, 124*(3), 1668–1681. doi:10.1121/1.2951500.
- Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research, 1*(e24), 88–90. doi:10.4081/audiores.2011.e24.
- Kochkin, S., Beck, D. L., Christensen, L. A., Compton-Conley, C., Fligor, B. J., Kricos, P. B., ... Turner, R. G. (2010). MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success. *Hearing Review, 17*(4), 12–34. Retrieved from <http://www.hearingreview.com/2010/04/marketrak-viii-the-impact-of-the-hearing-healthcare-professional-on-hearing-aid-user-success/>.
- Korzepa, M. J., Johansen, B., Petersen, M. K., Larsen, J., Larsen, J. E., & Pontoppidan, N. H. (2018). *Modeling user intents as context in smartphone-connected hearing aids*. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (pp. 151–155). New York, NY: ACM. doi:10.1145/3213586.322621.
- Kuk, F., Slugocki, C., Korhonen, P., Seper, E., & Hau, O. (2018). Evaluation of the efficacy of a dual variable speed compressor over a single fixed speed compressor. *Journal of the American Academy of Audiology*. Advance online publication. doi:10.3766/jaaa.17127.
- Kuk, F. K., & Lau, C.-C. (1995). Effect of initial setting on convergence to optimal hearing aid setting using a simplex method. *British Journal of Audiology, 29*(5), 263–269. doi:10.3109/03005369509076741.
- Kuk, F. K., & Pape, N. M. (1992). The reliability of a modified simplex procedure in hearing aid frequency-response selection. *Journal of Speech, Language, and Hearing Research, 35*(2), 418–429. doi:10.1044/jshr.3502.418.
- Kuk, F. K., & Pape, N. M. C. (1993). Relative satisfaction for frequency responses selected with a simplex procedure in different listening conditions. *Journal of Speech, Language, and Hearing Research, 36*(1), 168–177. doi:10.1044/jshr.3601.168.10.1044/jshr.3601.168.
- Mackersie, C., Boothroyd, A., & Lithgow, A. (2019). A “Goldilocks” approach to hearing aid self-fitting: Ear-canal output and speech intelligibility index. *Ear and Hearing, 40*(1), 107–115. doi:10.1097/AUD.00000000000000617.
- Marriage, J., Moore, B. C., & Alcántara, J. I. (2004). Comparison of three procedures for initial fitting of compression hearing aids. III. Inexperienced versus experienced users. *International Journal of Audiology, 43*(4), 198–210. doi:10.1080/14992020400050028.
- Moore, B. C., Alcántara, J., & Marriage, J. (2001). Comparison of three procedures for initial fitting of compression hearing aids. I. Experienced users, fitted bilaterally. *British Journal of Audiology, 35*(6), 339–353. doi:10.1080/00305364.2001.11745252.
- Moore, B. C., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing, 32*(5), 556–568. doi:10.1097/AUD.0b013e31820b5f4c.
- Moore, B. C., Glasberg, B. R., & Stone, M. A. (2010). Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *International Journal of Audiology, 49*(3), 216–227. doi:10.3109/14992020903296746.
- Moore, B. C., & Sek, A. (2013). Comparison of the CAM2 and NAL-NL2 hearing aid fitting methods. *Ear and Hearing, 34*(1), 83–95. doi:10.1097/AUD.0b013e3182650adf.
- Neher, T., Wagener, K. C., & Fischer, R.-L. (2016). Directional processing and noise reduction in hearing aids: Individual and situational influences on preferred setting. *Journal of the American Academy of Audiology, 27*(8), 628–646. doi:10.3766/jaaa.15062.
- Nelson, P. B., Perry, T. T., Gregan, M., & VanTasell, D. (2018). Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. *Trends in Hearing, 22*, 1–13. doi:10.1177/2331216518798264.
- Neuman, A. C., Levitt, H., Mills, R., & Schwander, T. (1987). An evaluation of three adaptive hearing aid selection strategies. *The Journal of the Acoustical Society of America, 82*(6), 1967–1976. doi:10.1121/1.395641.
- Nielsen, J. B. (2015). *Systems for personalization of hearing instruments: A machine learning approach* (Doctoral dissertation, DTU Compute PHD-2014; No. 325). Technical University of Denmark, Kgs. Lyngby, Denmark. Retrieved from http://orbit.dtu.dk/files/108304925/phd325_Nielsen_JB.pdf.
- Nielsen, J. B., Nielsen, J., Jensen, B. S., & Larsen, J. (2013). *Hearing aid personalization*. Paper presented at Proceedings of the 3rd NIPS Workshop on Machine Learning and Interpretation in Neuroimaging, Lake Tahoe, NV. Retrieved from http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6716/pdf/imm6716.pdf.
- Nielsen, J. B., Nielsen, J., & Larsen, J. (2015). Perception-based personalization of hearing aids using Gaussian processes and active learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23*(1), 162–173. doi:10.1109/taslp.2014.2377581.
- Nordqvist, P., & Leijon, A. (2004). An efficient robust sound classification algorithm for hearing aids. *The Journal of the Acoustical Society of America, 115*(6), 3033–3041. doi:10.1121/1.1710877.
- Perry, T. T., Nelson, P. B., & Van Tasell, D. J. (2019). Listener factors explain little variability in self-adjusted hearing aid gain. *Trends in Hearing, 23*, 1–17. doi:10.1177/2331216519837124.
- Preminger, J. E., Neuman, A. C., Bakke, M. H., Walters, D., & Levitt, H. (2000). An examination of the practicality of the simplex procedure. *Ear and Hearing, 21*(3), 177–193.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.
- Sabin, A. T., Hardies, L., Marrone, N., & Dhar, S. (2011). Weighting function-based mapping of descriptors to

- frequency-gain curves in listeners with hearing loss. *Ear and Hearing*, 32(3), 399–409. doi:10.1097/AUD.0b013e318202b7ca.
- Schmidt, E. (2018). The Widex fitting rationale for EVOKE. *WidexPress*, 38. Retrieved from <http://webfiles.widex.com/WebFiles/9%20502%204866%20001%2001.pdf>.
- Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Larnagaray, D., . . . Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, 9(4), 159–197. doi:10.1177/108471380500900403.
- Smeds, K., Dahlquist, M., Paludan-Müller, C., Larsson, J., Hertzman, S., Båsjö, S. (2015). Proprietary hearing aid gain prescriptions: Changes over time. *Hearing Review*, 22(5), 16–22. Retrieved from <http://www.hearingreview.com/2015/04/proprietary-hearing-aid-gain-prescriptions-changes-time/>.
- Townend, O., Nielsen, J. B., & Balslev, D. (2018). SoundSense learn—Listening intention and machine learning. *Hearing Review*, 25(6), 28–31. Retrieved from <http://www.hearingreview.com/2018/05/real-life-applications-machine-learning-hearing-aids-2/>.
- Wolters, F., Smeds, K., Schmidt, E., Christensen, E. K., & Norup, C. (2016). Common sound scenarios: A context-driven categorization of everyday sound environments for application in hearing-device research. *Journal of the American Academy of Audiology*, 27(7), 527–540. doi:10.3766/jaaa.15105.