The Collaboration between Hearing Aid Users and Artificial Intelligence to Optimize Sound

Laura Winther Balling, Ph.D.,[1] Lasse Lohilahti Mølgaard, Ph.D.,[2] Oliver Townend, B.Sc.,[1] and Jens Brehm Bagger Nielsen, Ph.D.[2]

ABSTRACT

Hearing aid gain and signal processing are based on assumptions about the average user in the average listening environment, but problems may arise when the individual hearing aid user differs from these assumptions in general or specific ways. This article describes how an artificial intelligence (AI) mechanism that operates continuously on input from the user may alleviate such problems by using a type of machine learning known as Bayesian optimization. The basic AI mechanism is described, and studies showing its effects both in the laboratory and in the field are summarized. A crucial fact about the use of this AI is that it generates large amounts of user data that serve as input for scientific understanding as well as for the development of hearing aids and hearing care. Analyses of users' listening environments based on these data show the distribution of activities and intentions in situations where hearing is challenging. Finally, this article demonstrates how further AI-based analyses of the data can drive development.

KEYWORDS: hearing aid classification, artificial intelligence, listening intention, data-driven hearing care

Over the years, hearing aids have been getting increasingly advanced, going from analog to digital signal processing, from linear to nonlinear gain prescriptions, and from a single general processing scheme for all listening environments to sound classification schemes that adjust the processing to the specific listening environment. Nonetheless, a fundamental problem remains that hearing aids are designed for the average ear of the average user, and they are designed using assumptions about what is typical in a given listening environment. Even the individually adjusted, customized fitting performed by a qualified clinician cannot

---

[1]Widex, Lynge, Denmark; [2]WS Audiology, Lynge, Denmark.

Address for correspondence: Laura Winther Balling, Ph.D., Widex, Nymøllevej 6, 3540 Lynge, Denmark (e-mail: laba@widex.com).

account for every real-life situation that the individual hearing aid user experiences.

Problems associated with the average hearing solution may be both general and specific. The most serious are, of course, the general problems that arise when the individual hearing aid user deviates from the assumptions on which the hearing aid processing, gain rationale, and fitting are based. The hearing aid user is likely to be dissatisfied with how they hear, which may cause them to use their hearing aids less or not at all, thereby leading them to experience all of the negative impacts of such behavior.[1] The clinician, in turn, may struggle to help the hearing aid user and spend valuable time on repeated fine-tunings that do not necessarily solve the problems.

One general problem is that preferred gain levels vary substantially between individuals,[2] which is obscured by the fact that gain targets are shown during the fitting as exact levels, not ranges. This holds for both the proprietary fitting rationales provided by hearing aid manufacturers and the generic fitting rationales such as DSL[3] and NAL-NL2.[4] Keidser and Dillon[2] summarized preferred gain levels for 189 adult hearing aid users. Even disregarding the most extreme cases, which may be considered outliers, the average preferred gain for the individual varies by approximately 15 dB. This indicates that, for a substantial number of hearing aid users, the fitting rationale will not provide a level of gain that is in accordance with their preferences. Smeds et al[5,6] showed similarly large ranges of loudness preference, though with some variation depending on the input sound.

Another important parameter to consider is loudness discomfort levels, which also show substantial variation. Summarizing studies including 710 ears, Bentler and Cooley[7] showed loudness discomfort levels ranging from below 70 to above 130 dB SPL for hearing losses up to 80 dB HL. For hearing losses above 80 dB HL, the range gets smaller as one would expect, but it still spans approximately 30 dB SPL.

For both preferred loudness and loudness discomfort, the hearing aids may be adjusted appropriately by a knowledgeable clinician, but this puts a significant burden on the clinician, and it is not an option for all users. Alternatively, equipping hearing aid users with a volume control may help alleviate the problem, but this

still falls short of solving it because preferences vary across frequencies and sound levels.

In addition to these general problems, the hearing aid user may also experience more local problems, arising in specific situations for the individual user whose hearing aids are generally well-fitted. If this happens often enough, problems that start as specific may lead to more general dissatisfaction. Contributing to this type of problem is the fact that hearing aids are generally fitted in a hearing clinic's relatively sterile acoustic environment, which is different from the environmental acoustics experienced in real-life listening environments outside the clinic.

Specific problems may arise when the listening intention assumed by the hearing aid in its automatic adjustments does not match the user's actual intention. In the extreme, the same user may be in the same listening environment but with entirely different intentions. For a hearing aid user who listens to music in the company of other people, the sound classification settling on a "music" sound class may be appropriate and enhance the enjoyment of music. But in other cases, if the user's intention is interacting with one or more conversation partners, this setting may be annoying. Sitting on a park bench near a playground, a hearing aid user may wish to exclude ambient noise entirely to concentrate on work or on having a conversation, they may want to read or relax while being aware of the surroundings, or they may need to focus on the sounds from the playground to monitor a child. The hearing aid classification will be based on assumptions—built into the hearing aid during its development—of what is typical or average in the specific listening environment; therefore, deviation from this is potentially problematic. See the article by Hayes in this issue for more information about the development of environmental classification systems in hearing aids.

In summary, both general and specific problems arise from the fact that hearing aids are designed for the average user in average listening environments, and from the fact that a hearing aid's classification of the listening environment does not always match the user's intent. A knowledgeable clinician may solve some of these problems, but this requires that the hearing aid user is able to explain their listening experiences and preferences, and that the clinician is able to translate these explanations into appropriate settings, both of which are

difficult tasks.[8] Volume controls, separate adjustments of left and right hearing aids, and multichannel equalizers also aim to alleviate these problems, but they do not necessarily solve them, and they require the user to interact with increasingly complex controls. This challenge is the motivation behind using artificial intelligence (AI) to optimize sound, as described next.

USING ARTIFICIAL INTELLIGENCE TO OPTIMIZE SOUND

A key insight about sound preferences is that hearing aid users are able to choose which of two sound settings they prefer much more systematically than they are able to describe in words which sound they prefer. This insight lies behind the literature on the use of A-B comparisons for the fine-tuning of hearing aids,[9–11] and it is at the core of the Widex SoundSense Learn (SSL) solution, which uses AI to optimize sound. SSL is based on the user indicating their degree of preference between two sound settings, A and B, in their smartphone app (see the interface shown in Fig. 1c). SSL uses machine learning to translate the user's degree of preference into a manipulation of hearing aid
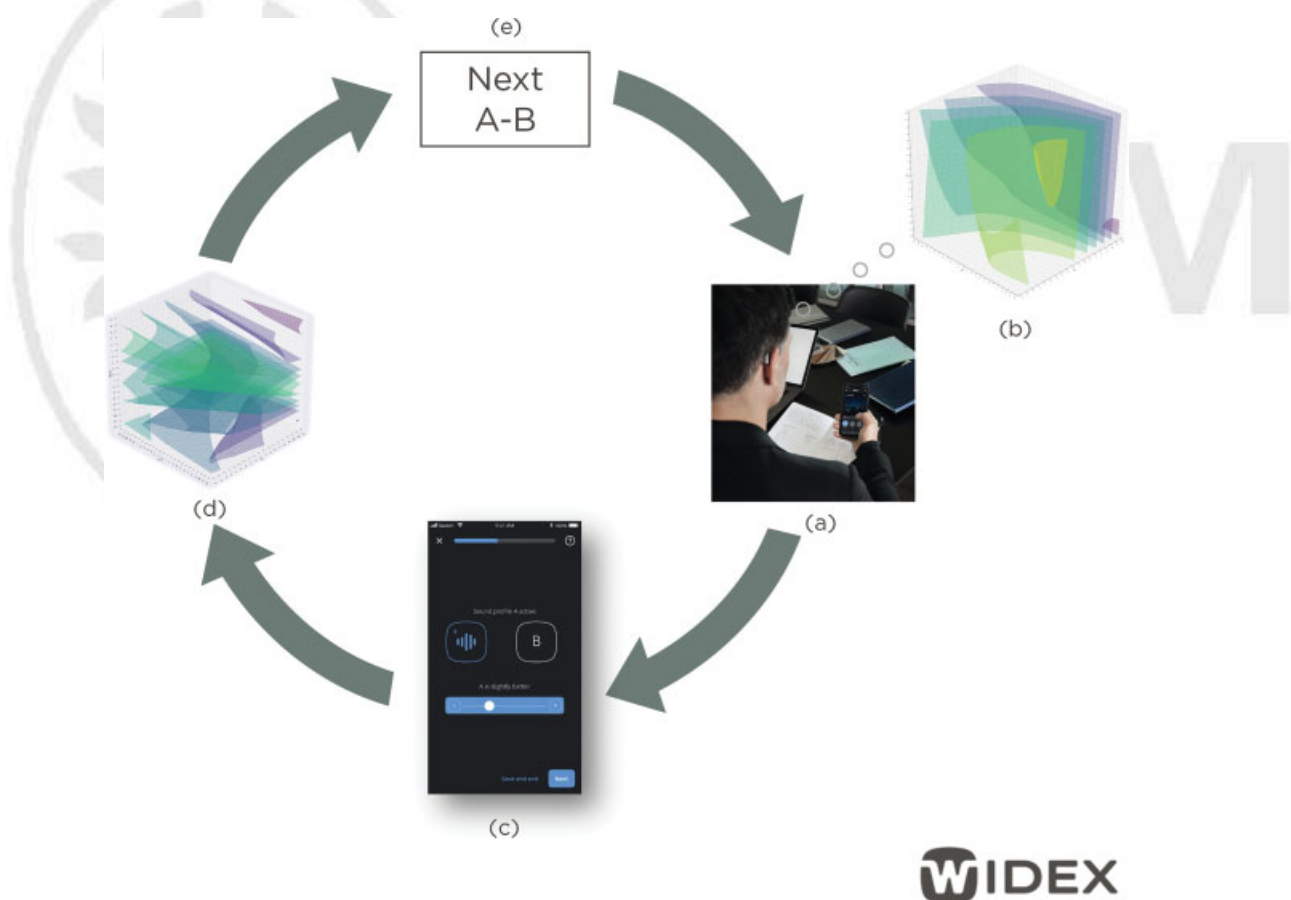

Figure 1 A conceptual overview of the steps of the SoundSense Learn (SSL) optimization process. The user is in a listening environment where the sound is not optimal for them (**a**). Their internal preference function for this listening environment is shown in (**b**) with gain in the three frequency channels illustrated as three dimensions. The colors (online only) represent preferences, going from yellow for the best match to the user's preference, over green and blue, to purple for the worst match. The first step in the process is the user indicating their activity and intention in the moment (not shown), after which they perform an initial A-B comparison (**c**), from which the Gaussian Process model creates a first estimate of the user's internal preference function (**d**). In turn, this estimation is used to create the next two settings to be compared by the user (**e**), at which stage the process returns to (**c**) where the user again indicates their degree of preference. SSL aims to match the settings in the hearing aid shown in (**d**) as closely as possible to the user's preference function shown in (**b**). The process stops when the user is satisfied with the settings, or when the model has converged, so that further improvements to the sound are not expected.

gain to optimize the sound for the given user in the given listening environment. The parameters manipulated are gain in three different frequency channels: bass (0.1–0.7 kHz), middle (0.6–3.6 kHz), and treble (2.2–10 kHz). For the bass channel, the gain may be turned up or down by 12 dB relative to the general setting. For the middle and treble channels, the gain may be turned down by 12 dB and up by 6 dB.

Machine learning, AI, and related terms are widely used across various fields, including hearing science and hearing aid marketing (see the articles by Andersen et al and by Fabry and Bhowmik in this issue for additional discussion on these topics). It is beyond the scope of this article to review and explain these concepts in detail, but briefly "machine learning" in SSL refers to the use of a specific mathematical model known as a Gaussian process[12] together with information theory[13] to generate personalized settings based on user input, as explained in the following. This is AI because a computer performs a task in real life and in the moment that a human would otherwise be required to do. In this case, it would require a clinician to manually manipulate the relevant parameters based on verbal feedback from the user. Crucial to the AI system, the human user remains an integrated part of the process, providing input that is used to update the machine-learning model and, in turn, fine-tune the hearing aid parameters.

An overview of the SSL process is provided in Fig. 1. The process assumes that the user's preference for sounds as a function of the hearing aid settings is encoded in an internal preference function within the user's mind, which is illustrated in Fig. 1b. When two sound settings are compared, the internal preference function determines the user's degree of preference. A user's internal preference function cannot be directly observed but is estimated by SSL (Fig. 1d), based on their responses to a series of A-B comparisons.

Active Learning and Bayesian Optimization

The machine learning in SSL utilizes active learning, also known as sequential experimental design. Active learning entails that the machine-learning model is continuously updated based on input from the user. The active learning approach

is instrumental in cases where a model only has access to a limited number of assessments and where assessments are expensive. In SSL, the assessments are A-B comparisons, which are considered expensive because large numbers of assessments cost time and effort for a large number of users. The active learning is implemented in a Bayesian optimization[14,15] framework, using a nonparametric Bayesian Gaussian process (GP)[12] model for pairwise comparisons. This allows SSL to model user preferences even though they are inherently "noisy." This noise arises because human preferences are not always stable and because the preferences are formed in listening environments that fluctuate.

Active learning aims either to learn a full function or to find the optimum of an unknown function. For adjusting a hearing aid, learning a full function would correspond to modeling the user's preferences for all possible settings, including bad settings. In contrast, finding the optimum corresponds to learning *just enough* information to determine the hearing aid setting that optimizes the user's preference. SSL does the latter because its aim is precisely to provide individualized best settings. This has the advantage that, although it is fundamentally a more difficult AI problem, it requires less effort from the user.

In SSL, the active-learning paradigm uses the knowledge captured by the Bayesian model based on previous A-B comparisons to select the next A-B samples for the user to compare, with the goal of maximizing the information yield provided by the user's response. Several criteria exist for estimating this,[16] of which SSL uses expected improvement (EI). In general, EI measures the amount of improvement a sample (in this case, an A-B comparison) can potentially yield on the maximal value of the function. Initially, the model does not know the user's preferences; so, the first set of hearing aid settings is chosen at random. After each response from the user, the GP model is updated (trained), and the EI criterion is evaluated to find a new A-B comparison that will best inform the model about where to find the optimal setting for the user. Importantly, the user's response is a degree-of-preference rating, which provides more information to the modeling process per user input than a forced-choice

paradigm. Nonetheless, it remains a relatively simple perceptual task.

The SSL mechanism has the advantage of allowing users to indicate preference without having to describe it in words, which in turn would have to be translated into parameter settings by the clinician. In addition, the use of AI allows SSL to cover a much wider parameter space than human comparisons alone could possibly do. In the extreme, the combination of 13 steps for each of three handles gives 2,197 possible settings; if these were to be exhaustively compared, more than 2.4 million A-B comparisons would be necessary. Using AI, by contrast, optimal settings should be reached in maximally 24 A-B comparisons, which is just 0.001% of the total possible comparisons.

Use of Soundsense Learn

The SSL settings are available immediately in the listening environment that prompted the user to initiate the SSL process, but they may also be used more generally if the user chooses to save them as a personal program. In this way, it is up to the user to decide whether they consider their preferences to be local to the listening environment or to be more globally relevant.

SSL is designed to allow the user to address hearing problems in the moment, without having to consult their clinician. However, the clinician and the dialogue between hearing aid user and clinician remain crucial for a successful hearing aid fitting. To leverage the insights generated by the hearing aid user's interaction with SSL, the Widex fitting software includes a functionality called "Real Life Insights", which allows the clinician to see (1) the personal programs that the user has created (using SSL or other manipulations), (2) what settings were reached, and (3) how much each program has been used. Since the SSL data in themselves are completely anonymous, this functionality requires the user to consent to having their SSL data shared with their clinician through a secure cloud server. This enables the clinician to observe any systematic trends in the programs that the user has created and used frequently; if these are sufficiently systematic, it may make sense to implement them as more universal adjustments to gain. Even if no

general changes are warranted, information about the personal programs provides a good starting point for the important dialogue between the clinician and the hearing aid user.

Studies of User Experiences

An important part of the development of SSL included studies on how hearing aid users use and experience the feature. Development tests[17] showed that eight out of ten participants could use an early version of SSL to reach settings that were significantly preferred in blind comparisons with the baseline settings of the hearing aids. For the remaining two participants, the SSL prototype did not reach convergence, indicating that their preferences were not systematic and that there was no preference function for SSL to maximize. So, while SSL may optimize sound for most users, it may not be a helpful tool for some users.

These initial tests focused on the sound quality of music using blind comparisons in a laboratory setting. Two additional blind comparison studies were conducted in which the scope was broadened to consider three different sound attributes: basic audio quality, listening comfort, and speech clarity.[18,19] During the first session of these studies, participants listened via headphones to a range of recordings of different acoustic scenes that were presented to hearing aids mounted on a KEMAR. For each scene, they used SSL to create a personalized program. Recordings were then made using these programs, and, in a second session, participants rated their personalized programs in comparison to their baseline setting. The results for both studies showed the strongest effects for basic audio quality, where the SSL settings in the personalized programs were rated significantly higher than the baseline setting. There was also a preference for the SSL settings for listening comfort, but this was significant only in the first study.[19] Finally, there was no significant difference between the SSL settings and the baseline settings for speech clarity.

As interesting as they are, these blind laboratory studies remain somewhat removed from the real-life use of hearing aids that SSL was designed for. Therefore, questions on the use of SSL were included in a large-scale survey of hearing aid satisfaction,[20] where 118 experienced hearing aid users wore Widex hearing aids in their

daily lives and rated their satisfaction with different aspects of use. Of the 118 participants in the overall hearing aid survey, 53 indicated that they had used SSL. In other words, not everyone felt the need or motivation to adjust their hearing aids, which is also what one would expect given a well-fitted modern hearing aid, but a substantial number did try out the functionality. Of the participants who did try it out, more than 70% had experienced an improvement in at least one listening environment, while almost 80% would recommend the feature to others. These are, of course, rather general questions that do not provide any detail on the listening environments in which users experienced an improvement. Instead, this question may be addressed from a different perspective by exploring the data generated when SSL is used.

LEARNING FROM DATA

A key characteristic of SSL is that, although the preference assessments and gain calculations run on the user's smartphone, the preference assessments are also stored anonymously on a cloud computer. This means that SSL generates large amounts of data, including information on (1) the settings and amount of use of the programs created, (2) the activities and intentions indicated by users when SSL is used, and (3) the settings compared and the associated degree of preference. These data support both specific improvements to the SSL algorithms and a general understanding of the problems and preferences of hearing aid users in real life. With observations from thousands of users, these data are on a scale that is usually not seen in hearing research studies, which typically rely on smaller samples of users.[21–23] Moreover, although there may be a selection bias because individuals who consent to share their data are not representative of all SSL users, or of all hearing aid users, these biases are likely to be smaller than for traditional laboratory studies, which are associated with more participant effort and more restrictions in terms of accessibility and inclusion criteria.

Data for Development

One example of how data have been used to improve the SSL process itself is *model conver-*

gence—the point at which the expected improvement (EI) from additional A-B comparisons is so low that the model has sufficient information about the user's preference. Without data from a very large number of users, in the original version of SSL, it was expected that the progress toward convergence would be linear with a maximum of 20 comparisons. In contrast, later versions of SSL operate with a more refined convergence measure based on the preference assessments from users of the original version who agreed to share their anonymous data in the cloud. Convergence calculations are based on how much EI remains across the entire input space (i.e., all combinations of gain settings in the three frequency channels) for different iterations of the A-B comparisons across users. Based on this, the degree of convergence can be calculated in the moment for the individual user so that there are generally fewer A-B comparisons than if SSL simply stopped after a fixed number of comparisons. This means that, although current SSL versions are capped at 24 comparisons, convergence is often achieved much earlier, and the comparisons can stop.

Settings, Listening Environments, and Intentions

The discussion now shifts from the specific tuning of the machine-learning algorithms to a more general understanding based on an interrogation of the data. The first step in this process is to explore the variation in the final settings reached by the users. Fig. 2 shows a sample of 20,000 programs created and saved in the year 2020. Each program is represented by a dot which indicates its gain settings for the bass, middle, and treble frequency channels. This is an important validation of SSL's utility because the individual programs (dots) are distributed all over the cube, which indicates that there is a wide range of individual preferences in individual listening environments rather than general group preferences. The only pattern that can be observed directly is some concentration of programs around the edges of the cube, indicating gain that was maximally changed in one or more of the frequency channels.

Another way these data are informative is in an analysis of the distribution of activities
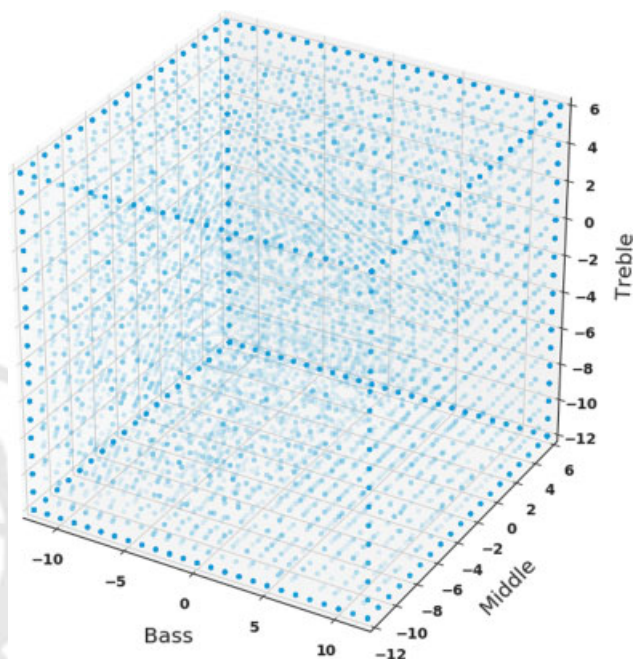
Figure 2 Sample of 20,000 SSL programs with their settings in the bass, middle, and treble frequency channels as shown on the different axes. Each dot represents a program, with darker colors indicating overlapping programs.

and listening intentions indicated by hearing aid users because these provide some detail about the listening environments where hearing aid users are not entirely satisfied and so choose SSL to remediate their dissatisfaction. For this analysis, a sample of 31,772 programs created between March 2020 and February 2021 is considered with a focus on the activities and intentions that users indicate before they start SSL comparisons to create a personalized program. Fig. 3 shows the distributions of activity and intention tags (categories selected by the user that correspond to their listening environment and listening goal, respectively) for this sample. These distributions should be generalized with some caution since the programs were created during periods of extensive COVID-19 lockdowns in much of the world, which does affect SSL program creation.[24] However, there are still general trends of interest in these data.
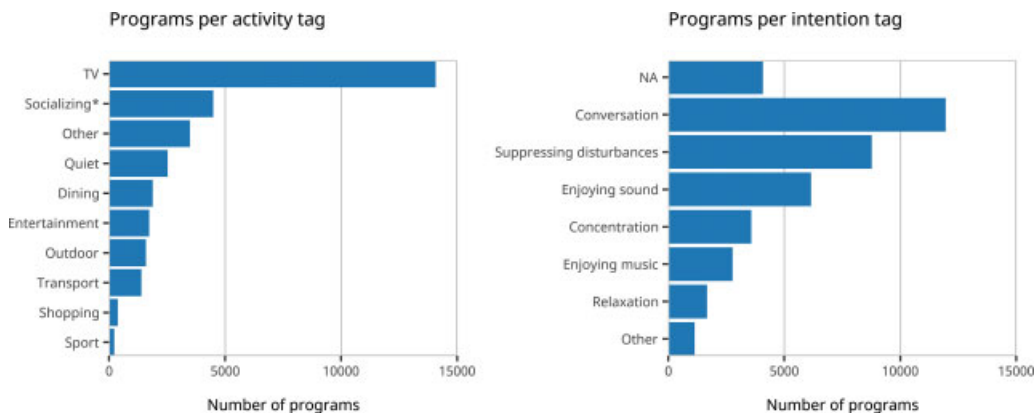

Figure 3 Distribution of SSL programs in terms of the activities (1 per program, left) and intentions (0–2 per program, right) indicated.

Starting with the activity tags, watching TV is the overwhelmingly most frequent activity indicated by users, a fact which is probably driven by two different circumstances. The first is that watching TV is a very frequent activity, especially for the older age groups[25] who are also more likely to wear hearing aids. The second is that a setting where the user watches TV is likely to be a relatively easy one to create SSL programs for, in contrast to a one-to-one conversation, where it is likely to be more difficult to systematically complete the series of A-B comparisons required for program creation.

Nonetheless, the second most frequent activity is "Socializing*," with the asterisk indicating that it is not the activity tag from the app, but a collection of three different activity tags—"Socializing," "Party," and "Family Gathering"—which are all likely to involve some level of to-be-attended speech along with noise that includes competing speech. It is noteworthy that programs for these activities are relatively frequent, even though it is likely more challenging to create programs for these types of interactive listening environments. On the other hand, these listening environments are ones in which hearing is notoriously difficult,[26,27] which may motivate users to create personal programs despite the attentional resources they need to invest in performing the A-B comparisons. This trade-off between the desire to improve sound and the potential challenge of creating programs may also help explain the relative rareness of programs for "Sport" and "Shopping," where creating programs is likely more challenging and the need for improvements likely lower. However, the low numbers for these activities may also be COVID-19 related.

Turning to the intentions indicated by users, shown in the right panel of Fig. 3, the most frequent intentions are "Conversation" and "Suppressing disturbances." The co-occurrence of these two intentions represents cases where SSL was activated to address problems with speech in noise, which shows up in more than 3,000 unique programs. Another major type of intention is enjoyment: "enjoying sound" and "enjoying music" are frequently used and may be understood together as cases where SSL was used to improve sound quality.

When considering what these distributions indicate about hearing aid users' listening environments, it is important to remember that SSL is used when their hearing experiences are suboptimal. This means that these data do not fully represent hearing aid users' typical listening environments. Instead, they represent those listening environments where there is some degree of dissatisfaction. This does not make the data less interesting; in fact, one could argue that these situations, where hearing aid settings are suboptimal for the individual user, are the most interesting for both clinicians and hearing aid developers. However, it remains crucial to remember, as discussed earlier, that some listening environments, such as watching TV, are more conducive to creating SSL programs than others, such as socializing.

An interesting supplement to this view of hearing aid users' everyday lives is found in the article by Hayes in this issue, where Fig. 10 shows the distribution of everyday environments as detected by the classifier in the hearing aid. The data in that article are informative about the listening environment detected by the hearing aids, which may not always correspond to the user's experience, but do have the advantage of being collected independently of how the user perceives the situation. By contrast, in the SSL data, the user's intention and experience are central, with a focus on those cases where the sound is not entirely satisfactory. This means that the SSL data add a layer of information that is crucial to understanding hearing in real life. But it also introduces a potential bias because the distribution of intentions and activities depend partially on the relative ease of using SSL in the individual listening environments. A further difference is that the current data were collected in the pandemic-affected world of 2020 and 2021, while the data reported by Hayes were collected before the pandemic.

Activity-Dependent Clustering of Degree-of-Preference Ratings

In addition to the analysis of the final SSL settings discussed in the previous section, an alternative approach to the data is an analysis of

all the paired comparisons and degree-of-preference ratings that users give while creating SSL programs. Such an approach to the data represents an alternative view of SSL as a preference sensor, a mechanism that maps users' preferences, as well as a preference optimizer that provides personalized settings for the individual user.[28] Both views are possible because of the central role of the internal preference function as outlined earlier. Analyzing all paired comparisons has the advantage that it provides much richer information than if only the final settings were considered. However, it also presents the challenge that there is too much data and too many complex patterns for simple descriptive analyses of the type presented earlier. Instead, machine-learning techniques are employed again, this time to analyze the data.

To explore systematicities in users' preferences for specific activities, the analyses focus on the activity tags that hearing aid users select at the outset of creating an SSL program. To keep the analyses computationally manageable, a statistically representative subset of A-B comparisons was drawn among the entire set of A-B comparisons for each activity. A subset of 32,000 A-B comparisons was found to be fully representative for the entire set of A-B comparisons while being computationally feasible.

For each activity tag, the subset of A-B comparisons and degree-of-preference ratings were analyzed in three steps. First, a GP model (see earlier) was trained for each situation, thereby assigning probabilities to all possible preference functions based on the users' degrees of preference. Some preference functions are associated with large probabilities, while other preference functions are associated with lower probabilities. Second, based on these assigned probabilities, 10,000 preference functions were sampled. Consequently, preference functions with higher assigned probabilities would be more frequent in the samples than preference functions with lower assigned probabilities. Third, for all the 10,000 sampled preference functions, the gain setting that maximized the preference function was picked. This resulted in a set of 10,000 settings representing programs that the model predicts to be optimal for different users, based on the probabilities derived from the degree-of-preference ratings.

These may be thought of as ideal programs representing the preferences expressed in the full set of A-B comparisons conducted by users for the given activity. The advantage of considering these 10,000 predicted programs is that they represent both mean and variance across many users' preference functions, whereas considering 10,000 individual programs would only allow us to consider final settings. In other words, considering the predicted programs allows us to go beyond the individual user's optimal settings and consider the preference functions more broadly.

The predicted programs were distributed much more systematically in the three-dimensional space than the programs shown in Fig. 2. They are shown for four different activities in Fig. 4, clustered using a density-based clustering algorithm.[29] This algorithm estimates the number of clusters that are necessary and sufficient to describe the data for each activity. It essentially helps summarize the user preferences in a meaningful way, predicting how likely each of them would be.

The outcome of these analyses was a small number of clusters for each activity. Each cluster includes a proportion of the predicted programs for that activity. Table 1 shows the four most frequent clusters per activity. The clusters labeled in Table 1 as clusters 1 through 4 do not represent the same settings across activities; in fact, the settings tend to differ between activities, as discussed later. Instead, the proportions for cluster 1 simply describe how frequent the most frequent cluster is for the given activity, and similarly for clusters 2 through 4. This analysis provides information about how systematic the predicted programs—and by implication, users' preferences—are for each activity. At one extreme, all predicted programs constitute one cluster for transport. In contrast, for watching TV, the predicted programs are distributed over more clusters, with the top four accounting for 74% of programs. The remaining activities fall between these two, with relatively good coverage by the two most frequent clusters of around 70% or more of the predicted programs.

Together with considering *how many* of the predicted programs are represented by a given cluster, it is also interesting to analyze the
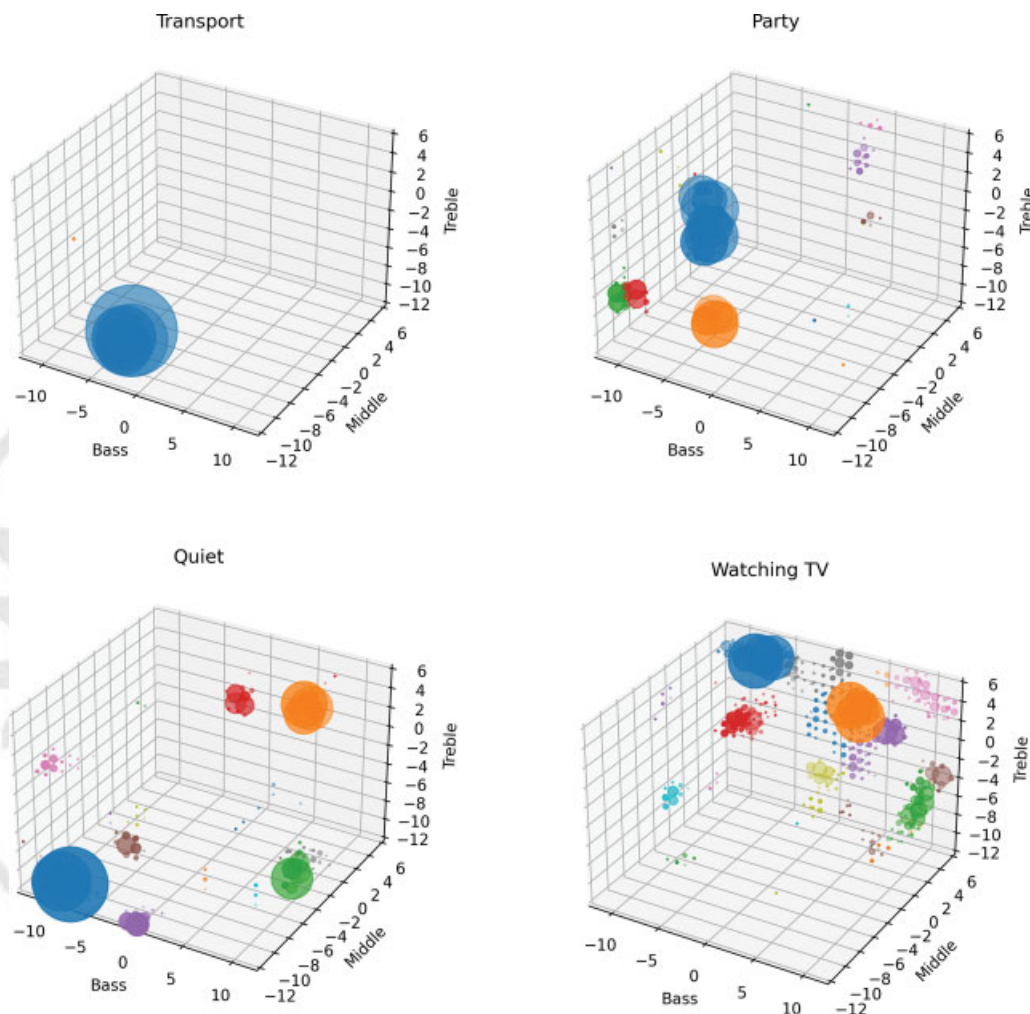
Figure 4 Clusters of predicted programs for four activity tags: transport, party, quiet, and watching TV. Different colors correspond to different clusters, with the size corresponding to the number of predicted programs represented by a given cluster.

Table 1 The Percentage of Predicted Programs Covered by the Four Strongest Clusters for each Activity

Activity	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Dining	57.9	14.9	8.2	5.0
Entertainment	61.6	29.2	6.5	1.3
Family gathering	84.7	6.3	2.8	1.8
Outdoor	54.8	14.8	9.4	8.8
Party	68.4	23.3	3.6	3.2
Quiet	47.9	26.7	12.9	4.3
Shopping	79.2	7.3	4.4	2.8
Socializing	52.4	26.6	8.8	4.5
Transport	100	–	–	–
Watching TV	37.4	14.8	14.0	7.9

settings of the three gain parameters for each cluster. In Fig. 4, four different activities that represent different levels of systematicity and interpretability are considered. First, as indicated earlier, transport represents an extremely systematic case, with a single cluster characterized by a general turning down of volume, though less for the bass than the middle and treble channels. In addition to being systematic, this is also relatively clearly interpretable as the users' attempts to reduce the transport noise.

A second activity where the clusters are audiologically interpretable is the "Party" tag. Here, the most frequently occurring cluster (with 68%) turns down the bass and treble but keeps the middle fixed or a little turned up; this could represent an attempt to reduce noise while preserving speech intelligibility. The second-most frequent cluster (23%) instead represents a general turning down of volume, though more so for the higher frequencies. The "Quiet" listening environment is described by two fairly frequent patterns: (1) all frequencies are turned down, likely representing a desire for more quiet, and (2) the bass and treble are turned up, likely representing a desire for more awareness of the surroundings. Finally, "Watching TV" includes many different settings, although most include increased gain in the middle, which is important for speech. All these interpretations of settings are educated guesses about users' preferences, but one should not overlook the fact that preferences may be driven by many factors that are specific to individual users and individual listening environments.

Some activities are more difficult to interpret, including a relatively well-defined activity like "Watching TV" and a fuzzier label like "Entertainment." In fact, a criticism that could be raised against the SSL activity tags is that they are not all well-defined and may not all be easily identifiable and distinguishable for SSL users. Another example of this is the potential overlaps between "Socializing," "Party," and "Family Gathering" (which were grouped together in Fig. 3). This indicates that a revision of the activity tags may be needed to ensure that they are easy to distinguish and map onto user experiences.

In addition to showing interesting patterns of preference in different listening environments, these cluster-based analyses also have a more direct application in a recent update of the SSL functionality under the label "My Sound" in the Widex MOMENT app. The update retains the SSL functionality described earlier (Fig. 1). It also introduces a faster option for personalizing sound where the user chooses between two sound settings sampled from the best clusters for the activity that the user has indicated. In this way, the preferences of previous users produce recommendations for current users. Importantly, the analyses are dynamic, which means that if preferences shift—for instance, as a result of changed defaults in the hearing aids—the recommendations can also be updated. Given the large amount of data the clusters are based on, one would expect preferences to be relatively stable, but having the ability to update remains important. For example, we recently observed that the global pandemic resulted in a shift in the distribution of activity tags.[24]

DISCUSSION

This article has focused on the development and use of AI features to optimize sound. In this section, the discussion shifts to connections between the different aspects and findings from SSL and their implications for the future of AI in hearing aids. An important point to emphasize upfront is the human hearing aid user's role in the Widex AI solution: In contrast to sound classification mechanisms that may also be based on AI[30] (see also the article by Andersen et al in this issue), the machine-learning algorithms in SSL operate continuously on the input from the individual user so that the sound settings come about through the interaction between the user and the application. This also means that the SSL data are informative about both the application and the user's hearing needs; these aspects must be considered together.

Another user-focused aspect to consider is the effect that the availability of an adjustment mechanism like SSL has on a hearing aid user, potentially empowering the user with a larger sense of control over their hearing. One may

hypothesize that such a sense of empowerment may *in itself* improve hearing aid satisfaction, *in addition* to any concrete improvements in specific listening environments. To what extent such general effects apply is an interesting question for future research. SSL is different from other self-adjustment methods, such as the "Goldilocks" method[31,32] and the ear machine algorithm,[33] among other things by being intended explicitly for repeated in-the-moment use, which is likely to engender a more general sense of empowerment.

Empowerment and motivation may also contribute to the observed differences between laboratory studies and large-scale usage data. While the laboratory studies did not show a significant effect on perceived speech clarity, the conversation activity tag was indicated as an intention for thousands of SSL programs in the sample analyzed, both alone and in combination with the "Suppress disturbances" intention. The fact that the users saved these programs indicates that the programs help them in speech-in-noise environments. Likely, this finding is partly driven by the fact that SSL use in real life is motivated by a desire to improve the sound, while in the laboratory, participants adjusted the sound whether they felt the need to or not. The adjustments give an experienced benefit for some participants in the laboratory studies, but not for all, which may be related to whether they felt a need to improve the sound. Another difference is that participants were asked to focus exclusively on speech clarity in the artificial laboratory setup, which is a narrower focus than what is likely in real-life speech-in-noise environments. One may, for example, speculate that the experience of a more natural or more comfortable sound in a real-life listening environment may free up cognitive resources for focusing on understanding speech.

In addition to the positive effects of empowerment and motivation, an important area of future study is the listening environments in which hearing aid users re-use the programs they have created and their satisfaction with the programs in these potentially varying environments. Key questions to address are how specific the preferences expressed in SSL programs are for a given listening environment and to what extent changes may meaningfully be applied more generally in data-based hearing care.

Although big data of the kind analyzed in this article represent a wide variety of individual needs and experiences, it is important to remember that analyses like those reported here still focus on the average and the typical. This means that, in addition to data-driven general solutions, which are likely to become increasingly important in hearing care as in other fields, optimization for specific listening environments, based on a collaboration between the individual user and the AI, remains crucial.

CONFLICTS OF INTEREST
All the authors are employees of Widex and WS Audiology. No other conflicts of interest exist.

REFERENCES

1. Ferguson MA, Kitterick PT, Chong LY, Edmondson-Jones M, Barker F, Hoare DJ. Hearing aids for mild to moderate hearing loss in adults. (Review) Cochrane Database Syst Rev 2017;9(09):CD012023

2. Keidser G, Dillon H. What's new in prescriptive fittings down under? In: Palmer CV, Seewald R (eds.). Hearing Care for Adults. Stäfa, Switzerland: Phonak AG; 2006

3. Scollie S, Seewald R, Cornelisse L et al. The desired sensation level multistage input/output algorithm. Trends Amplif 2005;9(04):159–197

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

5. Smeds K, Keidser G, Zakis J et al. Preferred overall loudness. II: Listening through hearing aids in field and laboratory tests. Int J Audiol 2006;45(01): 12–25

6. Smeds K, Keidser G, Zakis J et al. Preferred overall loudness. I: Sound field presentation in the laboratory. Int J Audiol 2006;45(01):2–11

7. Bentler RA, Cooley LJ. An examination of several characteristics that affect the prediction of OSPL90 in hearing aids. Ear Hear 2001;22(01):58–64

8. Caswell-Midwinter B, Whitmer WM. The perceptual limitations of troubleshooting hearing-aids based on patients' descriptions. Int J Audiol 2021; 60(06):427–437

9. Kuk FK. Paired comparisons as a fine-tuning tool in hearing aid fittings. In: Valente M (ed.). Strategies for Selecting and Verifying Hearing Aid Fittings. New York: Thieme Medical Publishing; 1992:125–150

10. Kuk FK, Lau C. Comparison of preferred frequency gain settings obtained with category rating and modified simplex procedure. J Am Acad Audiol 1996;7(05):322–331

11. Amlani AM, Schafer EC. Application of paired-comparison methods to hearing AIDS. Trends Amplif 2009;13(04):241–259

12. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press; 2006

13. MacKay DJC. Information Theory, Inference and Learning Algorithms. Cambridge, UK: Cambridge University Press; 2003

14. Mockus J. On Bayesian methods for seeking the extremum. In: Proceedings of the IFIP Technical Conference; July 1974: 400–404

15. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. J Glob Optim 1998;13(04):455–492

16. Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: a review of Bayesian optimization. Proc IEEE 2015; 104(01):148–175

17. Nielsen JBB, Nielsen J, Larsen J. Perception-based personalization of hearing aids using gaussian processes and active learning. IEEE/ACM Trans Audio Speech Lang Process 2014;23(01):162–173

18. Jensen NS, Hau O, Nielsen JBB, Nielsen TB, Legarth SV. Perceptual Effects of Adjusting Hearing-Aid Gain by Means of a Machine-Learning Approach Based on Individual User Preference. Trends Hear 2019;23:1–23

19. Townend O, Nielsen JB, Balslev D. SoundSense Learn—listening intention and machine learning. Hear Rev 2018;25(06):28–31

20. Balling LW, Townend O, Switalski W. Real-life hearing aid benefit with Widex EVOKE. Hear Rev 2019;26(03):30–36

21. Jensen NS, Hau O, Lelic D, Herrlin P, Wolters F, Smeds K. Evaluation of auditory reality and hearing aids using an ecological momentary assessment (EMA) approach. Proc Int Congr Acoust 2019; 2019(September):6545–6552

22. Schinkel-Bielefeld N, Kunz P, Zutz A, Buder B. Evaluation of hearing aids in everyday life using ecological momentary assessment: What situations are we missing? Am J Audiol 2020;29 (3S):591–609

23. Holube I, von Gablenz P, Kowalk U, Bitzer J. Assessment of acoustical properties and subjective perception in everyday life. In: Proceedings of the 23rd International Congress on Acoustics, Aachen; 2019:7639

24. Balling LW, Townend O, Mølgaard LL, Jespersen CB, Switalski W. AI-driven insights from AI-driven data. Hear Rev 2021;28(01):27–29

25. The Nielsen Total Audience Report. August 2020. Available from https://www.nielsen.com/us/en/insights/report/2020/the-nielsen-total-audience-report-august-2020/, last accessed August 9, 2021

26. Picou EM. MarkeTrak 10 (MT10) survey results demonstrate high satisfaction with and benefits from hearing aids. Semin Hear 2020;41(01):21–36

27. Bronkhorst AW. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acta Acust 2000;86 (01):117/128

28. Nielsen JBB, Mølgaard LL, Jensen NS, Balling LW. Machine-learning based hearing-aid personalization: preference optimizer, preference sensor or both? In: 7th International Symposium on Auditory and Audiological Research. 2019:40

29. Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 2002; 2002:603–619

30. Brændgaard M. An Introduction to MoreSound Intelligence. Tech Paper 2020. Available from https://wdh01.azureedge.net/-/media/oticon/main/pdf/master/whitepaper/69674uk_tech_paper_moresound_intelligence.pdf?la=en&rev=3F19&hash=6A1037B3951F262345E45C4A725D3CC7, last accessed August 9, 2021

31. Boothroyd A, Mackersie C. A "goldilocks" approach to hearing-aid self-fitting: user interactions. Am J Audiol 2017;26(3S):430–435

32. Mackersie C, Boothroyd A, Lithgow A. A "Goldilocks" approach to hearing aid self-fitting: ear-canal output and speech intelligibility index. Ear Hear 2019;40(01):107–115

33. Nelson PB, Perry TT, Gregan M, VanTasell D. Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. Trends Hear 2018; 22:2331216518798264