


A Comparison of Environment Classification Among Premium Hearing Instruments

Trends in Hearing
Volume 24: 1–14
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216520980968
journals.sagepub.com/home/tia


Anusha Yellamsetty, Erol J. Ozmeral, Robert A. Budinsky, and David A. Eddins 

Abstract

Hearing aids classify acoustic environments into multiple, generic classes for the purposes of guiding signal processing. Information about environmental classification is made available to the clinician for fitting, counseling, and troubleshooting purposes. The goal of this study was to better inform scientists and clinicians about the nature of that information by comparing the classification schemes among five premium hearing instruments in a wide range of acoustic scenes including those that vary in signal-to-noise ratio and overall level (dB SPL). Twenty-eight acoustic scenes representing various prototypical environments were presented to five premium devices mounted on an acoustic manikin. Classification measures were recorded from the brand-specific fitting software then recategorized to generic labels to conceal the device company, including (a) Speech in Quiet, (b) Speech in Noise, (c) Noise, and (d) Music. Twelve normal-hearing listeners also classified each scene. The results revealed a variety of similarities and differences among the five devices and the human subjects. Where some devices were highly dependent on input overall level, others were influenced markedly by signal-to-noise ratio. Differences between human and hearing aid classification were evident for several speech and music scenes. Environmental classification is the heart of the signal processing strategy for any given device, providing key input to subsequent decision-making. Comprehensive assessment of environmental classification is essential when considering the cost of signal processing errors, the potential impact for typical wearers, and the information that is available for use by clinicians. The magnitude of differences among devices is remarkable and to be noted.

Keywords

Hearing aids, environment classification, auditory scene analysis, speech in noise, product comparison

Received 1 May 2020; Revised 20 November 2020; accepted 24 November 2020

Introduction

In most listening situations, a dynamic mixture of sounds from multiple sound sources simultaneously reaches our ears. Despite the fact that the elements in this mixture are typically interleaved and overlapped in time and frequency, the auditory system is able to accurately parse and group different patterns of the sound sources in terms of timing, space, and frequency into a coherent sound stream through a process known as auditory scene analysis (Bregman, 1990; Büchler et al., 2005). This phenomenon of parsing, grouping, and streaming is one of the theoretical bases for the classification system in hearing aids. In an analogous process, modern hearing aids automatically classify the incoming

acoustic mixture into one or more of a larger set of sound scene categories. Because everyday situations present a mixture of speech sounds, musical sounds, environmental sounds, and low-level or quiet moments, the hearing aid is tasked with continuous classification of the listener's current acoustic environment.

Department of Communication Sciences and Disorders, University of South Florida, Tampa, Florida, United States

Corresponding author:

David A. Eddins, Communication Sciences and Disorders, University of South Florida, 4202 E. Fowler Ave, PCD 1017, Tampa, FL 33620, United States.

Email: deddins@usf.edu



The automatic classification is based on many different acoustic analyses over various time scales, some of which are categorical and some of which are scalar. Environmental classification is perhaps the most important function of a modern hearing aid that relies on current environmental factors to guide decision-making regarding automatic signal processing involving simple and advanced digital signal processing (DSP) features.

Decisions regarding the nature and methods of signal processing are based, in part, on an initial and ongoing classification of the current acoustic scene. The resulting classification is used to populate the datalogging feature of the fitting software for most hearing aids and was available for each of the devices evaluated in the current investigation. Such datalogging information can be used by a clinician to better understand the nature of the listening environments encountered by a given patient. Clinicians also can use this information in patient counseling, as a troubleshooting tool, or as the basis for device adjustment or accessory recommendations. While information from the datalogging feature within the fitting software is the only data readily available to clinicians, it should be noted that investigation of hearing aid classifiers based on this datalogging data is cursory rather than comprehensive. The datalogging feature does not reveal the dynamics of the classification output and does not reveal if, when, or how the hearing aid uses level detectors, estimates signal-to-noise ratio (SNR), estimates the presence of wind noise or feedback, or compares information across aids. For example, others have shown that output SNR can vary significantly among multiple devices for the same speech-in-noise environment (Miller et al., 2017). Nevertheless, the datalogging feature is what is made available to the clinician, and thus what forms the basis for clinical decision-making. The present study is a detailed evaluation of the environment classifiers, as measured by company-specific datalogging, from five major hearing aid companies and is presented in conjunction with listener judgments.

Advanced DSP features are intended to adapt the corresponding signal processing to a scene class to improve listener experience. In other words, the choice of which DSP feature(s) to engage and the strength of engagement requires some knowledge of the types of stimuli present in the acoustic environment (i.e., environment classification). In most cases, the critical information regarding classification and subsequent decision-making is not widely distributed or known, as the rules governing such processing and the consequences of those rules are typically proprietary and technical in nature. From a broader clinical perspective, a given hearing aid is created with a certain design philosophy that includes the nature of the level-dependent gain-

frequency model, the dynamics of that model in terms of amplitude compression, the engagement of other signal processing features, and interactions among these. In many ways, the initial step in implementing the design philosophy begins with environmental classification, a process that typically is not well understood outside of the design team. This investigation represents an initial attempt to gain an understanding of similarities and differences among the environmental classification processes employed in the premium products of five major hearing aid companies. It involves collecting information about environmental classification that technically is available to the clinician, though the data collection methods are generally prohibitive for the average clinician or clinician-scientist. In collecting and analyzing such data, we highlight several key acoustic features that influence such classification including overall level, SNR, stimulus source number, and stimulus source type, from which some of the underlying philosophical differences can be inferred.

Early sound classification algorithms were developed based on subjective judgment, such as listening-environment preference (Elberling, 1999; Fedtke et al., 1991). Based on a library of relevant sounds and different kinds of background competition, multiple amplification schemes were developed by identifying different hearing aid characteristics for these desired listening conditions (Kates, 1995; Keidser, 1995, 1996). Consequently, current hearing aids can be conceptualized as providing several different “programs” with each program tailored to a particular class of sound environments and/or to particular user preferences. With advances in automatic processing, however, the concept of distinct programs is giving way rapidly to dynamic arrays of individual signal processing features and sets of features that may be engaged or disengaged synchronously, individually, or by degree based on classification of the acoustic environment and other real-time monitoring such as sound pressure level (SPL) and SNR.

Today, environment classifiers available in premium hearing aids possess a fixed number of environment classes (as many as nine; e.g., speech-in-quiet, quiet, speech-in-noise, noise alone, music, etc.). Each device classifier is pre-trained on a known set of audio files using computational algorithms that learn which sound features are best associated with each class. The algorithms often follow standard approaches like a Bayes classifier (Lamarche et al., 2010; Ostendorf et al., 1998), neural networks (Freeman, 2007; Park & Lee, 2020; Zhao et al., 2018), or Hidden Markov models (Dong et al., 2007; Freeman, 2007; Nordqvist & Leijon, 2004). Training data are deconstructed into spectro-temporal acoustic features as they would be in real-time in the device, ranging from simple (e.g., overall level or level within frequency channel) to complex feature sets including those based on perceptual models of human hearing

(e.g., modulation frequency and depth; mel-frequency cepstral coefficients, etc.; Ravindran et al., 2005). For example, complex scenes with speech are often classified based on their spectral profile and temporal envelope (Chen et al., 2014; Feldbusch, 1998; Kates, 1995), their statistical amplitude distribution (Wagener et al., 2008), or their characteristic temporal and/or spectral modulation frequencies (Nordqvist & Leijon, 2004; Ostendorf et al., 1998).

Classifiers exist physically as software stored and running on a microchip that creates a set of weighting functions that have dimensions specific to a company's desired number of classes and the acoustic features associated with those classes. As shown in Figure 1, at the earliest input stage (e.g., after the microphones), the classifier extracts the acoustic features of the incoming signal before applying weighting functions that project to a class or classes. Depending on the weight matrix, the system switches between the classes in postprocessing or blends class-dependent postprocessing. The resultant class or classes then affect decision rules for hearing aid features, such as directional microphone strategy, amplitude compression/expansion, and adaptive noise reduction. It is important to note that though the system performs these computations in real-time, companies often apply some temporal rules to avoid frequent DSP changes which could lead to adverse listening experiences. The pace at which a device may change environments is company-specific, may vary widely across companies, and is virtually unknown to and unknowable by the clinician (Mueller et al., 2014).

Every hearing aid manufacturer has engineers who design their unique classification schemes. There are potentially hundreds of acoustic parameters that could positively or negatively influence the quality of each of the classification schemes they design. But the devices themselves have limited physical resources to implement detection and actuation on the basis of all of those acoustic parameters. Thus, choices must be made, and limits set as to what is most important in the acoustic milieu for the purpose of their scheme. Those choices are based on the company's and the engineer's philosophy of what is going to be most efficacious for the listener in the

widest range of listening environments. It is that bespoke philosophy that determines how the device will classify and ultimately accommodate each listening environment in which it is worn. Hence, there is simply no way that all hearing aid classifiers are created equal. Understanding the philosophy of each company, therefore, should be a contributing factor when prescribing hearing aids according to the listener's individual lifestyle, abilities, desires, and needs. The present study was designed to assess the behavior of the classifiers of five different hearing aid companies using a broad array of acoustic environments. Because each company has its own class labeling, the results of the measurements given here were transformed to four major classes in order to compare across companies. In each case, great effort was taken to line up equivalent classes across companies based on what each uniquely named company class was intended to be used. Finer subclass divisions of each company might reveal the uniqueness and philosophical disparity for each company, but these granular points were not the focus of the study nor was the intention to single out any one specific company.

Previous reports on hearing aid classifiers have described individual methods of classification or have compared various types of classification tools (Abe et al., 2011; Büchler et al., 2005). Development of classification algorithms involves a balance between identifying and defining some number of relevant acoustic environments and the ability of classification procedures to do so accurately and efficiently. Information about the relevant environments and the frequency with which typical hearing aid wearers are in those environments has been obtained by self-report (Keidser, 2009; Walden et al., 2004), acoustic recordings and subsequent off-line analyses (Wagener et al., 2008), synchronized acoustic recordings and self-report (Wu & Bentler, 2012), and via datalogging features that catalog the classifier results over time during real-world hearing aid use (Humes et al., 2018; Taylor & Hayes, 2015). It is interesting and reassuring that each of these methods, with their relative advantages and disadvantages, has converged on very similar information. To summarize the results of the investigations cited earlier, the environmental descriptors

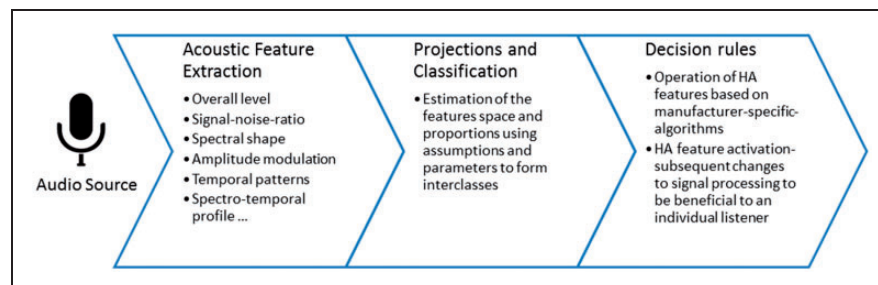


Figure 1. Schematic Diagram of an Automatic Scene Classification Algorithm. HA = hearing aid.

were equated, and the data were averaged. This average provides an approximation to the proportion of time that hearing aid wearers (sampled with a clear elderly age bias) are in “quiet” (~28%), “noise alone” (~23%), and speech plus noise (~29%). Interestingly, despite the fact that the chief complaint of a person with hearing loss is difficulty listening to speech in background noise (Beck & Le Goff, 2018; Nabelek et al., 1991; Wu & Bentler, 2012), and the fact that hearing aid wearers are most dissatisfied with the performance of their devices when they are in the same environment (Nabelek et al., 1991; Plyler et al., 2019; Turan et al., 2019; Walden et al., 2004), actual wearers are only in such environments a fairly small proportion of the time. Their complaints, however, are primary factors driving motivation to seek hearing aids (Olsen et al., 2014; Takahashi et al., 2007; Turan et al., 2019) and satisfaction with hearing aids (Huber et al., 2018; Kochkin, 2005; Korhonen et al., 2017; Picou, 2020; Taylor & Hayes, 2015; Wong et al., 2003; Wu et al., 2019). Thus, manufacturers continue to focus on development and refinement of signal cleaning strategies to mitigate the effects of background noise, in addition to refining the classification strategies used to govern the signal cleaning strategies.

Few investigations, however, have directly compared the results of the classification process across hearing aid companies under comparable circumstances. The study reported by Groth and Cui (2017) did just that and included two main components. The first was human subject evaluation of selected acoustic environments, for which interrater agreement was high and judgments appeared to represent accurate descriptors. The second involved assessment of hearing device classification of the same selected acoustic environments, as coded by the datalogging feature associated with each company’s fitting software. For the latter, accuracy was defined as agreement between device and human subject classification. For the quiet, speech, and steady noise environments, the classification performance was highly consistent among the devices from six different companies. For the speech babble and noise scenes, five of the six devices had similar classification performance. One device (from different companies) in each of those scenes had a fairly high proportion of unexpected classification results. As the scene complexity increased by combining turn-taking conversational speech with one of four different “noisy” backgrounds, more substantial differences among the devices were revealed. Each of those four scenes was considered to be speech in noise by the human subjects, though the proportion of time the six devices classified those scenes as speech in noise ranged from 98% to 41% with an average across devices of about 67% of the time. It would be even more interesting to know how the speech-in-noise scenes were classified when unexpected classes occurred. Likewise, there

was some variability in the accuracy of classification when faced with music as the primary or secondary source in the acoustic scene. Overall, the study revealed a fairly high degree of parity among the classifier results for relatively simple or unitary environments and more diverse results for the speech in noise and music scenes. To further challenge classifier performance, specific scenes could also vary systematically in overall level and in signal-to-background ratio. It is likely that many manufacturers use in their classification schemes estimates of overall level and signal-to-noise ratio as part of their comprehensive analyses. This would be especially interesting in the case of music, as one could imagine background music emerging as the primary signal of interest (or a distraction) as the music-to-background ratio increases from negative to positive. Furthermore, the classifier for a given device and given scene will always weigh the possible categories to a sum of 100%. Thus, when a classification result is unexpected, or when there is an ambiguous scene, it is important to consider the proportions assigned to each possible class. For these reasons, the present investigation includes a wide array of acoustic scenes with systematic changes in overall level and signal-to-background ratios in a design that is somewhat similar to that described by Groth and Cui (2017) but that presents a more detailed analysis of the classifier data.

Methods

Acoustic Scenes

Acoustic scenes were developed by mixing different speech and nonspeech sounds chosen from an in-house library of audio files of various durations. Original files were digitized at 44.1 kHz sampling rate and stored in separate mono 16-bit .WAV format. Speech passage recordings were drawn from both male (74 s) and female (54 s) talkers, and nonspeech sounds included music (214 s), a subway platform (177 s), a food court (238 s), a playing-card hall (240 s), and 10-talker babble (70 s). The speech sounds are reproduced on the “Phonak Sound CD 2” (D41-0508-02/0501/CU) distributed by Phonak AG. The music passage “My Babe Just Cares for Me” is distributed by FreeSound.org, as was the recording from the London tube subway sound. All other sounds were recorded and mastered at Unitron. Speech always was presented from a loudspeaker located at 0° relative to the head. The subway, food-court, and card hall each included four unique audio channels that were presented from four loudspeakers spatially separated by 90° (45°, 135°, 225°, and 315°). The 10-talker babble was a single recording presented diffusely from six loudspeakers (45°, 90°, 135°, 225°, 270°, and 315°). Music was a single track from a jazz artist (Ella

Fitzgerald, “My Baby Just Cares for Me”) presented in stereo from loudspeakers at 45° and 315°.

Eight primary acoustic scenes, 80 min in duration, were derived from looping the audio files. Among the primary scenes, 28 conditions were created by varying overall level (L Scenes) or the SNR (S Scenes). For subsets in which overall level varied, SNR was fixed. Likewise, for subsets for which SNR varied, overall level was fixed. Figure 2 (top row) shows the long-term average spectra for each of the isolated audio files (left panel) and primary scenes (L Scenes [middle panel] and S Scenes [right panel]) with 0-dB SNR and normalized to an RMS of 1 (magnitude is given in dB full scale). The temporal modulation index (Gallun & Souza, 2008; Krause & Braida, 2004) was computed for octave bands centered at 0.5, 1, and 4 kHz. The bottom row of Figure 2 provides the modulation index for the 1 kHz octave band, which indicates the dominant modulation rate and relative depth across scenes in that frequency region. Table 1 describes the individual conditions by their description of prominent audio (e.g., speech, music, and noise), overall level, SNR,

and modulation index (at .5, 1, and 4 kHz octave bands). Although not an exhaustive arrangement of acoustic scenes, all the SNR and level values were chosen to be within the typical range of realistic if not challenging listening environments. The music track was chosen because it contained both voiced and instrumental audio, and also because listeners have previously been shown to be more sensitive to differences in hearing aids when listening to jazz (Vaisberg et al., 2017).

Hearing Aid Classification

The acoustic scenes were presented in the free field using a 24-channel speaker (KEF Q100) array with subwoofer (KEF KUBE-I) in a sound-attenuating booth (Acoustic Systems RE-245). Digital-to-analog conversion was handled by a MOTU 24ao routed to three 8-channel power amplifiers (Ashley ne8250). To improve test efficiency, up to three pairs of hearing aids were evaluated simultaneously using a Klangfinder Twinface (Klangspektrum) head and ear simulator positioned in the center of the speaker array (41 in. radius) and adjusted such that the

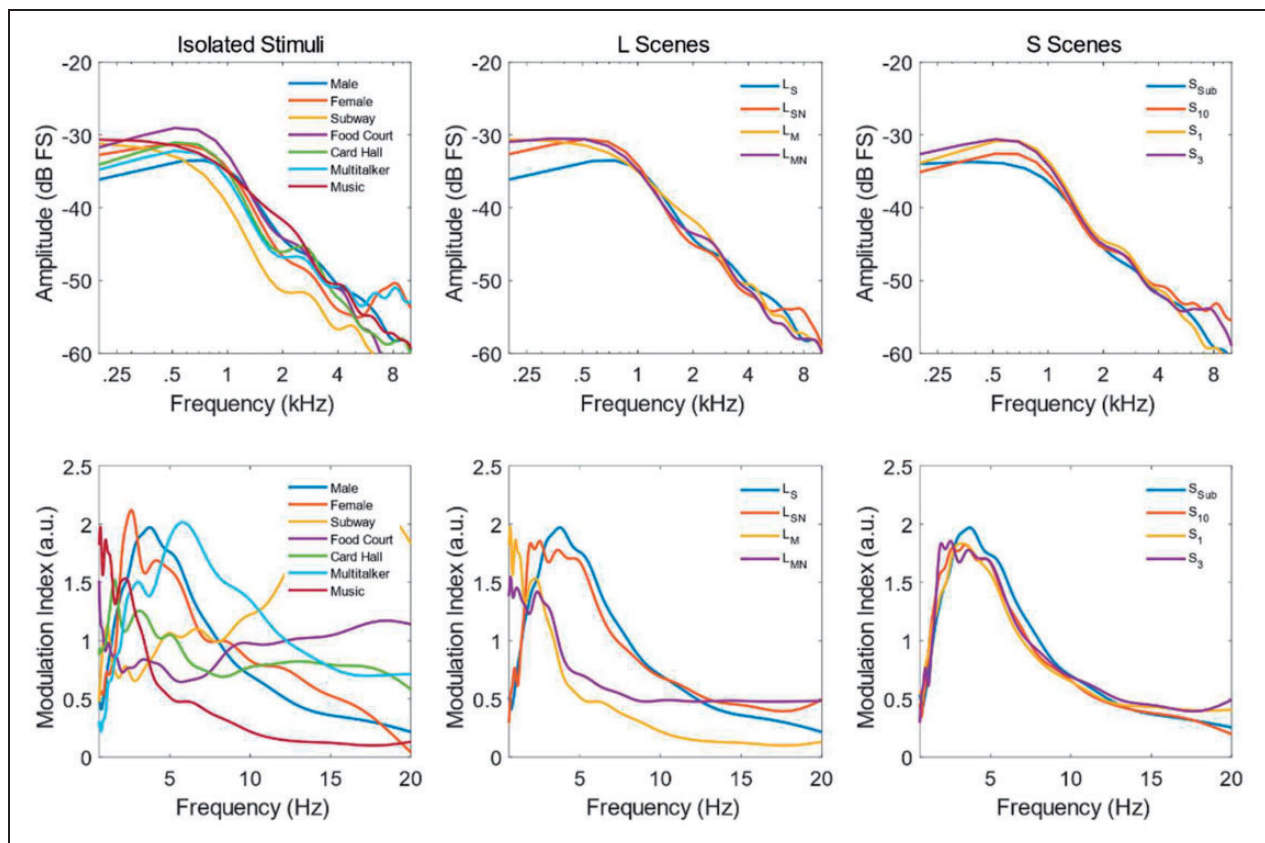


Figure 2. Upper row: The power spectrum (normalized RMS in dB full scale) of the isolate audio files (left panel), the primary scenes that varied overall level (L Scenes; middle panel), and those that varied the signal-to-noise ratio (SNR) between speech and background noise (S Scenes; right panel). Bottom row: The modulation index (MI) for the 1000-Hz one-octave frequency band for each of the Isolated Stimuli (left panel), L Scenes (middle panel), and S Scenes (right panel). Abbreviations in reference symbol subscripts: S = Speech; SN = Speech in Noise; M = Music; MN = Music in Noise; Sub = Subway Noise. FS = full scale.

Table 1. Acoustic Scenes.

Category	Reference symbol ^a	Audio scene description	Overall level (dB SPL)	Signal-to-noise ratio (dB)	Peak modulation rate (Hz) at 0.5 kHz	Peak modulation rate (Hz) at 1 kHz	Peak modulation rate (Hz) at 4 kHz
Level (L)	L _S	1 male	55, 70, or 85	n/a	5.1	5.4	3.5
	L _{SN}	1 male, 1 female in food court	55, 70, or 85	5	4.1	3.2	4.9
	L _M	Jazz music	55, 65, or 80	n/a	2.0	1.0	2.5
	L _{MN}	Jazz music in card hall	55, 65, or 80	0	17.1	2.2	17.1
SNR (S)	S _{Sub}	1 male in subway	80	-10, -5, 0, 5 or 10	7.2	3.7	3.6
	S ₁₀	1 male in 10-talker babble	70	-10, -5, 0, 5 or 10	5.7	5.4	3.5
	S ₁	1 male in food court	80	0, 5, 10	6.6	3.0	3.6
	S ₃	1 male, 1 female in food court	80	0, 5, 10	4.1	3.2	4.9

^aAbbreviations in reference symbol subscripts: S = Speech; SN = Speech in Noise; M = Music; MN = Music in Noise; Sub = Subway Noise.

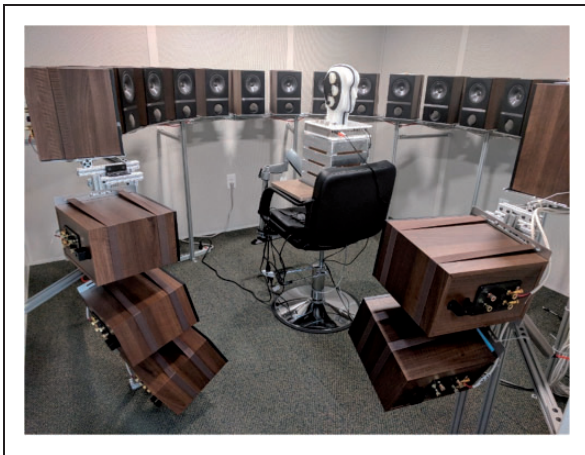


Figure 3. A Klangfinder Twinface Head and Ear Simulator Placed at the Center of the 24-Speaker Array.

center of the aperture of the middle ear canal was level with the center of the dual concentric cone drivers in the KEF loudspeakers (Figure 3).

To evaluate classification of acoustic scenes by hearing aids, audio scenes were presented for the full 80 min. This duration was chosen based on pilot data which revealed that some of the chosen devices required more than an hour but less than 80 min to reliably populate the data logging feature in the commercial fitting software. The premium models (launched in 2017) from each of five major hearing aid companies (subsequently labeled as A, B, C, D, and E) were selected for comparison. All devices had a miniature behind-the-ear, receiver-in-the-ear form factor (RITE or RIC), and were powered by size 312 zinc-air batteries. Device programming involved choosing the company “first fit” option and using the corresponding default settings. It is expected that other

fitting strategies would not impact classification performance, but this was not evaluated specifically. This was done via the company-specific clinical fitting software via the HiPro2 (Otometrics) USB-interface using the same generic mild-sloping hearing loss audiogram. The data-logging feature of each company was enabled and re-initiated prior to each audio scene presentation, and the placement of hearing aids on the Klangfinder Twinface was counterbalanced by company per condition to avoid possible effects of elevation differences.

The number of environment classes and corresponding company-supplied descriptions of the prototypical classes varied across the companies. The technical aspects of defining the environment and the analysis methods of the acoustic scenes are proprietary to each company. Therefore, to keep the analysis consistent across the devices, data-log classes were remapped to the following four generic classes – Speech in Quiet, Speech in Noise, Noise and Music. Re-mapping of classes was done based on a review of DSP features for a given class across all five devices using the information available publicly. This allowed for more direct comparisons per sound class among devices and the subjective classification from human subjects. However, it should be noted that by removing the granularity of the classifiers specific to each company, it is possible that the following observations will be considered too general and not capture the full breadth of each company’s classification philosophy. Nevertheless, the present design was chosen to provide the fairest cross-company comparisons.

Human Listener Judgment

Twelve young normal-hearing adults (age $M \pm SD$: 24 ± 2.25 years; 8 females, 4 males) participated in the environment judgment task. All had normal hearing

thresholds (i.e., <20 dB HL) at octave frequencies between 250 and 8000 Hz and reported no history of neurological disorders. Each provided written informed consent following procedures approved by the University of South Florida institutional review board.

To evaluate classification of acoustic scenes by normal-hearing listeners, two-channel audio files were recorded using microphones (1/2 in. B&K model 4134 condenser mic) mounted in Zwislocki ear simulators (B&K model DB100) in a KEMAR acoustic manikin (Knowles Electronics, Chicago, IL) and connected to a preamplifier (B&K model 2966), amplified with a G.R. A.S. model 12AA conditioner, and routed to the Motu 24ao audio interface that sampled the stimuli at 44.1 kHz. The sounds were equalized digitally for playback over Sennheiser Precision 580 headphones in a single-walled sound attenuating booth. Each recorded audio file was 2 min in duration. Processed audio files are provided in supplemental material of this report.

Listeners were presented with the same 28 conditions used in the device tests. Each self-paced test consisted of three trials per scene (pseudo-randomized). Listeners were instructed to identify the sound scene by using a maximum of four out of six key phrases: (1) “listening to speech in quiet,” (2) “listening to speech in noise,” (3) “listening to music,” (4) “mostly quiet,” (5) “mostly noise,” and (6) “mostly music.” The key phrases were designed to probe perceived foreground. To directly compare to the device tests, choices (1) and (4) were combined and (3) and (6) were combined to leave four generic classes as in the device tests earlier. Subjective classification was tallied for each sound scene and tested for inter- and intrasubject reliability. Intersubject variability was evaluated using an intraclass correlation coefficient (ICC) based on absolute agreement and a two-way mixed model in SPSS (Bland & Altman, 1999). Values greater than 0.9 indicate excellent reliability, values between 0.75 and 0.9 indicate good reliability, values between 0.5 and 0.75 indicate moderate reliability, and values less than 0.5 indicate poor reliability (Koo & Li, 2016).

Results

Inter- and Intrasubject Response Reliability for Audio Scenes

The intersubject reliability was inferred from two measures of ICC: (a) across all class judgments and (b) within an audio scene type. First, among the four different possible class judgments, listeners were considered to have excellent reliability for judging speech-in-noise (ICC = 0.97), noise (0.96), music (0.91), and quiet (0.89). Second, within each of the audio scenes, listeners judged scene classes with excellent reliability for seven of

the eight audio scene types (ICC between 0.83 and 0.99). Listeners’ judgments of the speech in a subway background (S_{Sub}) were only moderately reliable (0.69). Collectively, these results indicate that the designed audio scenes could be reliably classified by human listeners based on the four generic classes.

Level Change

A total of 12 audio scenes varied in overall level with fixed SNR (L Scenes). Among these were four types of scene: speech alone (L_S), speech in noise (L_N), music alone (L_M), and music in noise (L_{MN}). Within each scene type, there were three overall levels as indicated in the second column-from-the-left in Figure 4 (55, 70, and 85 dB SPL; also see Table 1). In Figure 4, each row is an audio scene, with the listener judgments (S) and device classifier outputs (A to E) represented by columns. Columns are grouped by the generic environment classes: Speech in Quiet, Speech in Noise, Noise, and Music. Each cell in the table indicates a percentage and a corresponding shade of gray as indicated in the color bar. For a given device (A to E) in a single row, the four corresponding cells sum to 100%. These are the percentages extracted from the company-specific data logging. For example, in the first row (L_S 55), Device A classified the audio scene proportionally as 92.8% Speech in Quiet and 7.2% Speech in Noise over the course of the 80-min presentation.

Speech-Dominant Scenes. In the first scene type (L_S), a male and female turn-taking conversation in quiet, subjects and devices mostly agreed in their assessment; specifically, a high percentage of the scene was classified as Speech in Quiet by subjects and by devices A, B, and E, independent of the change in level. That is, in the absence of background distractors, the other three classes did not register to a high degree. On the other hand, for levels at or above 70 dB SPL, Devices C and D transitioned from Speech in Quiet class to Speech in Noise class as the overall level increased. These results indicate that, in this type of acoustic scene, Devices C and D invoke level-sensitive algorithms for distinguishing among the Speech in Quiet and Speech in Noise classes, whereas the other devices and human judges did not weight level strongly in decision-making over this 30-dB range.

The second scene type (L_{SN}) included a food court background at 5-dB SNR. Whereas subjects and most devices classified this scene as a Speech in Noise scenario, independent of overall level, there were some nuances among the classifiers. Device C was consistently at 100% Speech in Noise for all levels. Devices A and E increased the proportion of the Speech in Quiet class as overall level increased, possibly indicating that the positive

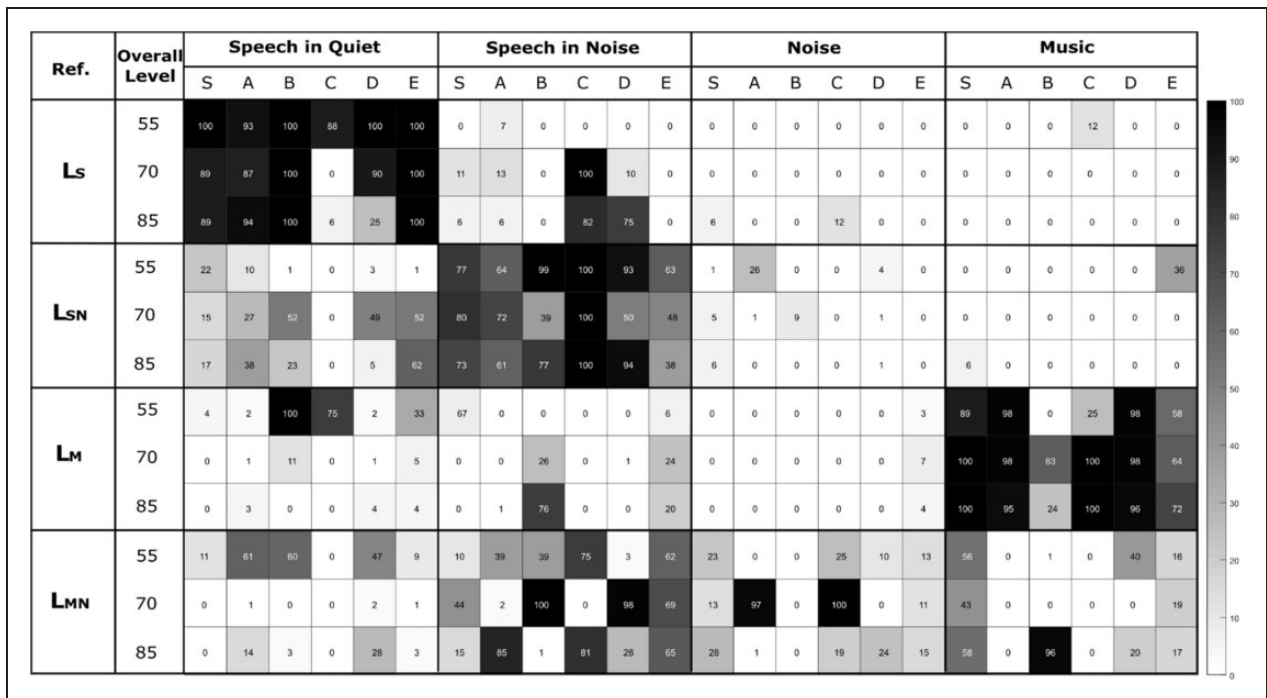


Figure 4. Environmental Classification of Audio Scenes With Varying Overall Levels. Each column represents classification by human subjects (S) and the five premium hearing aids (A-E). Each cell indicates the measured proportion of four generic classes: Speech in Quiet, Speech in Noise, Noise, and Music.

SNR (5 dB) interacted with overall level for these classifiers. Finally, Devices B and D performed in a more idiosyncratic fashion: at the lowest and highest overall levels, the classifier output was mostly in Speech in Noise (B: 99% and 77%; D: 93% and 94%, respectively), yet the intermediate level led to a split between Speech in Quiet (B: 52%; D: 49%) and Speech in Noise (B: 39%; D: 50%). Across devices, the proportion of speech-in-noise classification for speech-in-noise scenes ranged from 38 to 100%, a value strikingly like the 41 to 98% range reported by Groth and Cui (2017).

Music-Dominant Scenes. Because all modern premium devices include a classifier destination for music environments, the third and fourth scene types tested the likelihood of each classifier selecting the Music class at varying levels either in quiet (L_M) or in background noise (L_{MN}) consisting of card hall noise at an SNR of 0 dB. In quiet (L_M), human subjects judged the scene to be Music with greater than 90% proportion for each of the overall levels. In noise (L_{MN}), however, the proportion of Music judgments was considerably lower (between 43% and 58%), with the remaining percentage assigned mostly to the Speech in Noise or Noise classes. The distribution of weights among those classes appears to be level dependent, with Music and Noise having relatively higher weight for 55 and 85 dB and Speech in Noise having a higher weight for the 70 dB level.

Device classification in music was idiosyncratic, but detailed analysis of the cells in Figure 4 support logical inferences for each device. For the quiet (L_M) scene, Devices A and D mirrored the subject judgments in percentage classified as Music. Devices C, E, and B were progressively more level-dependent in their classification of Music. For device C, at 55 dB SPL, classification was 25% Music and 75% Speech in Quiet whereas at 70- and 85-dB SPL classification was 100% Music. For device E, classification percentage gradually increased from 58% to 72% with increasing level. At 55 dB SPL, the remaining percentage was attributed to Speech in Quiet while at 70- and 85-dB SPL the remaining percentage was attributed to Speech in Noise. For device B, classification at 55 dB SPL was 100% Speech in Quiet. At 70 dB SPL, classification was 63% music, 26% Speech in Noise, and 11% Speech in Quiet. At 85 dB SPL, classification was 24% Music and 76% Speech in Noise.

When music was presented in noise (L_{MN}), no devices mirrored the human subject judgments. The classification by devices A, B, C, and D clearly were level dependent with A and C progressing from Speech in Quiet dominant (device A) or Speech in Noise dominant (device B) at 55 dB SPL to Noise dominant at 70 dB SPL and Speech in Noise dominant at 85 dB SPL. Devices B and D progressed from Speech in Quiet dominant at 55 dB SPL to Speech in Noise dominant at 70 dB SPL and either Music dominant (device B) or evenly

distributed (device D) among the four classes at 85 dB SPL. Device E stood out as not being level dependent. At each level, the classification proportions remained approximately 65% Speech in Noise, about 15% Noise, and about 17% Music. While the aforementioned data reveal level and SNR dependencies when classifying music in quiet or mixed with background sounds, the variability among devices is consistent with the fixed-level, music alone data from Groth and Cui (2017).

SNR Change

The aforementioned level-change conditions showed that some device classifiers shifted destinations depending on the overall level whereas others were not affected as much by level changes. The presumption was that, like the human listeners in general, device classifiers that were mostly level-invariant would instead show effects from changing SNR. Figure 5 shows a similar heat map as in Figure 4. For this analysis, there were 16 audio scenes encompassing 4 scene types with varying SNRs (S Scenes; also see Table 1): a single talker with background subway noise (S_{Sub}), a single talker with 10-talker background (S_{10}), a single talker in a food court background (S_1), and three talkers in a food court background (S_3). Background stimuli were chosen to provide a variety of speech and nonspeech with their inherent spectral and modulation differences. For example, the subway background was a low-frequency, steady

background with minimal or no fluctuations in the spectrum with level changes as trains arrived and departed, and the 10-talker stimulus contained greater fluctuation at spectral regions common to speech.

Low-Frequency, Steady Background. The first scene (S_{Sub}) varied in SNR from -10 dB to 10 dB with an overall level held constant at 80 dB SPL. Subjects showed a consistent effect of SNR, shifting from a classification of Noise combined with Speech in Noise at negative SNRs to mostly Speech in Noise at SNRs greater than or equal to 0 dB. Because their judgments, shown in Figure 4, were largely level independent, this is consistent with our earlier presumption that decisions likely would be made based on level or SNR but not both. Device D followed a similar trend, and though Devices B and E also mirrored this trend; these devices also tended to classify a large proportion in Speech in Quiet at the highest SNR. Device A showed a different effect of SNR such that negative SNRs were classified as Speech in Noise, 0-dB SNR was classified mostly as Noise, and positive SNRs were again mostly Speech in Noise. Finally, Device C was 100% Speech in Noise for all SNRs, not only for this scene type, but also all other scene types. Leaving Device C aside for a moment, the other devices each showed some dependence on SNR at this relatively high overall level confirming expected effects of SNR.

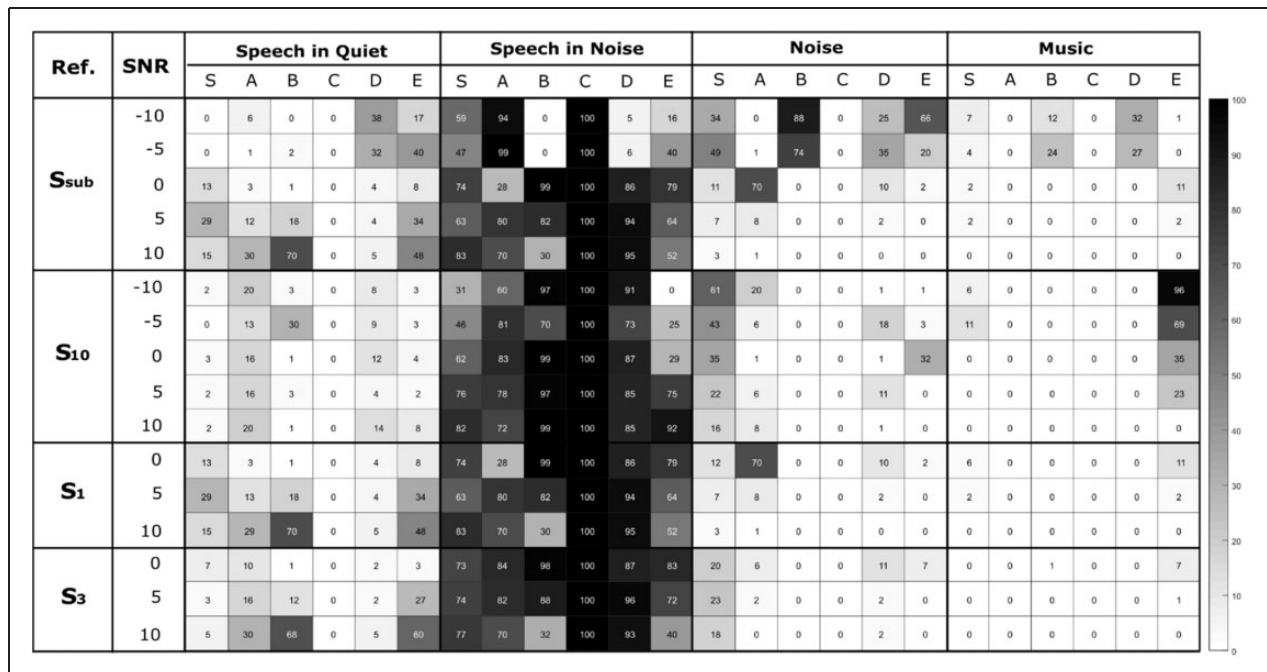


Figure 5. Environmental Classification of Audio Scenes With Varying Signal-To-Noise Ratios (SNR). Each column represents classification by human subjects (S) and the five premium hearing aids (A-E). Each cell indicates the measured proportion of four generic classes: Speech in Quiet, Speech in Noise, Noise, and Music.

Multitalker Background Babble. The next scene (S_{10}) also varied in SNR from -10 dB to 10 dB but with a constant level of 70 dB SPL. Subjects transitioned from predominantly Noise class judgments to Speech in Noise as SNR increased. No device classified the scene in a comparable way to the human subjects. Rather, Devices B, C, and D were consistently in the Speech in Noise destination independent of SNR, and though Device A was like these other devices, its classifier also dedicated a minor proportion to the Speech in Quiet class. The more surprising result was seen in Device E at low SNRs, the Music class was evoked to a high degree, even though no actual music was in the background noise. Without deeper knowledge of the company's classification system, including its acoustic feature analysis, it is challenging to infer the full implications of this result.

One Versus Three Speech Sources. The remaining two scene types (S_1 and S_3) tested three SNRs (0 , 5 , and 10 dB) and were fixed at 80 dB SPL. The addition of talkers in S_3 did not have considerable effects on subject or device classification. In general, classification was primarily Speech in Noise with some exceptions. Specifically, Device A tended to classify more as Noise at 0 -dB SNR with three talkers, and Device B shifted to Speech in Quiet destination at the high SNR as was seen in the S_{Sub} scene.

Discussion

Hearing aid classifiers were shown to be remarkably different from each other and, in some cases, from human subject judgments, despite all having some sensitivity to level, SNR, or both. Although the resulting heterogeneity may not be too surprising given the variety of acoustic analyses available to and methodologies employed by the different companies as well as their specific classification philosophies, the stark contrasts, however, could have a substantial impact on hearing-aid users and dispensing clinicians.

Effect of Overall Level and SNR on Environment Classification

The present results indicate that at low levels, all devices can reliably classify speech in quiet, but in two cases, increased level led to speech-in-noise classification (C and D). In contrast, the devices performed more idiosyncratically when speech was presented with a food court background. Whereas one device (C) had no dependence on overall level and consistently classified as Speech in Noise, the classification by other devices varied between Speech in Quiet and Speech in Noise in a manner dependent on overall level. Groth and Cui (2017) previously evaluated the classification accuracy of six different hearing aids in various scenes, including speech alone

or speech in noise, and reported that accuracy was poorer for most devices in noise backgrounds relative to human judgments. In that study, the "café" background was especially challenging, likely due to the presence of speech in the scene, and it is probable that the present food court scene with some distinct speech proved just as challenging for some of the classifiers. The fact that most devices showed a nuanced approach for dynamic backgrounds in both studies, however, may also indicate that for this type of background, a wider range of DSP features are available to the listener and processing is more dynamic over the course of the 80-min presentation.

When comparing device classification for scenes with music, again, the results show that for isolated scenes, like speech (L_S) or music (L_M) alone, the devices are mostly consistent with each other and in agreement with the human listeners. Device B was a unique case, which at the lowest level, classified the chosen jazz sample as Speech, at the medium level favored Music, and at the highest level favored Speech in Noise. Music has a wider range of levels and spectro-temporal characteristics compared to speech, which places greater demands on hearing aid circuits and algorithms for producing acceptable sound quality (Chasin & Russo, 2004). It is not surprising therefore that Music classification has been a more recent innovation in the industry, though recent studies have shown that music alone can be reliably classified (Büchler et al., 2005; Gil-Pita et al., 2015; Groth & Cui, 2017). For hybrid sounds containing both music and noise, however, there are known challenges for environmental classifiers (Büchler et al., 2005), just as there were for scenes with both speech and noise. In the present study, adding background noise to music was shown to steer device classifiers to a variety of classes that depended on overall level for four of the five devices. From human judgments, even at the highest overall level, music was considered the listening foreground for roughly 50% of the scene, which was not well-matched by the devices.

The second set of audio scenes was designed to measure the effects of SNR on environment classification. Except for Device E classifying Music somewhat erroneously in the S_{10} scene, most devices and the human subjects tended to classify these various speech-in-noise scenes as Speech in Noise. The Device C classifier was undeterred by SNR changes, classifying all scenes as Speech in Noise, likely due to the relatively high overall levels of the scenes. The other devices showed some variability in their classification approach, but the results of this analysis primarily show that both subjects and devices generally agree. One explanation for this is at these high levels, there is often a perceptual roll off (Dirks et al., 1977; Hannley & Jerger, 1981; Jerger & Jerger, 1971) in which the perceptual benefit of digital hearing

aids also decreases as input levels increase (Kuk et al., 2015). A large number of the acoustic scenes were presented at relatively high overall levels, which may have limited any chance to differentiate devices in the SNR-change conditions.

Understanding Classification Accuracy

Several studies on environment classification algorithms have shown that these tools can be extremely accurate, with up to 98% validation accuracy when learning the four primary classes (speech, speech in noise, music, and noise; e.g., Lamarche et al., 2010; Ravindran et al., 2005). The definition of accuracy can vary among investigations, however, and when considering actual device classifiers we must also consider the goals of the device under different scenarios. For example, Groth and Cui (2017) reported device classification accuracy as it related to human judgments, and they reported a wide range of accuracy across devices and listening conditions. The results of the present study demonstrate that the output of advanced hearing aid classifiers is often in contrast to human judgments, but again, this should not necessarily be an indication that devices were inaccurate in their classification. Rather, it may be the case that classification in these conditions has been intentionally biased to one class or another to support the desired DSP feature engagement and other adaptive processes according to the company's overall amplification philosophy (Hayes & Eddins, 2008). Thus, the present study reports only the percentage of time in which each classification was chosen for each scene condition.

Among the more surprising results of the present study were the variety of responses to music in noise at 0-dB SNR. The human listeners judged that this scene contained music in the foreground, but certainly recognized that noise and speech were also fittingly present. The devices, on the other hand, mostly avoided the music class, opting instead for either Noise or Speech in Noise. This distinction in understanding accuracy is important because even when music is present for example, a general bias towards the Speech in Noise and/or Noise classifications could be intended to engage decision rules for noise reduction to take precedence in order to preserve comfort and sound quality. This is also a good example of a case where the philosophy of the classifier and the perception of the listener may or may not be at odds. In natural environments, the appropriate hearing aid processing likely depends on the intent of the listener. The problem for the classifier is that the DSP features in use for speech clarity in noise and for improved sound quality for music are almost perfectly at odds with one another (Chasin & Russo, 2004). Speech clarity in noise leads to reductions in input levels and increased signal processing with directional

microphones, noise cancellers, and speech enhancement. But improved sound quality for music typically requires less heavy-handed signal processing including omnidirectional microphones, linear gain characteristics, and a removal of noise cancelling (Arehart et al., 2011; Croghan et al., 2014). Thus, the differences between what the listeners heard in this experiment and how the classifier responded nicely highlights the quandaries inherent in classification philosophy. In retrospect, it is also possible that the levels used for the audio scenes were not comparable to realistic scenarios from which the classifiers were trained. Although Smeds et al. (2015) observed natural music settings to range from 0 to 15 dB SNR, the average listening level was not greater than 70 dB SPL. In the present study, two of the three music scenes were at or above 70 dB SPL.

The current evaluation of environmental classifiers illustrates the importance of differences among devices and highlights differences in evaluation criteria. Rather than focusing on accuracy relative to human judgments, here the focus was on consistency, both within and among devices, in various sound-scenes. While extensive, this investigation certainly was not exhaustive. We chose to have listeners with normal hearing evaluate the scenes, as an internal reference and a means for comparison to previous investigations. We did not include listeners with hearing loss, though it would be of great interest to know how such listeners differ in their classifications relative to normal hearing, and whether such listeners classify sound scenes differently unaided versus aided. We developed a broad set of acoustic scenes, but did not cover all of the common environment scenes that listeners may encounter (Smeds et al., 2015; Wolters et al., 2016). We did not specifically consider reverberation that may have been inherent in the original recordings, and we did not manipulate reverberation with room treatments or audio processing methods. In the current set of sound scenes, all conditions were rather static, whereas moving sound sources and varying sound source distances might also impact environmental classification in important and potentially company-specific ways. Finally, and perhaps most profound in impact, we nor others have specifically investigated the performance of environmental classification relative to listener intent. At present, such classifiers still operate exclusively in the acoustic dimension without knowledge of listener intent or current focus of attention. Thus, there is a substantial likelihood of a mismatch between what the aid determines to be the prominent or important signal versus the signal to which the listener would like to attend. The present study did not measure listener intent and therefore cannot assess the accuracy of the devices in this way, but future work should consider accuracy of classifiers not only relative to foreground but also listener intent. Future innovations in classification technology

will undoubtedly seek to leverage human interfacing to incorporate the individual's unique perspective (Carlile et al., 2017), and the environment classifier will continue to be the bridge between the acoustic world and the DSP decision rules.

Conclusions

It is clear that premium hearing aid classification varies significantly among brands, and this presumably drives very different DSP-feature engagement subsequent to classification. This result itself connotes little valence if it can be assumed that individual companies understand the dependencies of their classifiers and take that into account when driving feature engagement via environment. The choice in feature activations and the subsequent changes to signal processing should always aim for some benefit to the individual listener. The idiosyncratic patterns revealed in some devices, and the sheer variability across devices, however, is more concerning from a clinician's perspective. Because the attributes of classifiers are not exposed to hearing health-care professionals to the same extent as the signal processing features they control, the importance of classification is often overlooked. The present data indicate that an individual clinician likely needs to know more information than they are provided to use classification data, as revealed via the datalogging feature in fitting software, for counseling, troubleshooting, and in making decisions about fitting adjustments. Presumably, knowledge of company DSP philosophy could better inform the clinician when prescribing premium hearing aids over base-level devices (Cox et al., 2016; Johnson et al., 2016). Therefore, understanding that there are differences and commonalities among companies may help the clinician give their best judgment in accordance with the patient's needs.

Acknowledgments

The authors would like to thank Don Hayes for consultation in the development of the acoustic scenes and the background on environment classification.


Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported, in part, by a multiyear project funded by Sonova, Inc. (to D. A. E.).

ORCID iD

David A. Eddins  <https://orcid.org/0000-0003-3716-2436>

Supplemental Material

Supplemental material for this article is available online.

References

- Abe, K., Sakaue, H., Okuno, T., & Terada, K. (2011). Sound classification for hearing aids using time-frequency images. *Proceedings of 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing* (pp. 719-724). <https://doi.org/10.1109/PACRIM.2011.6032982>.
- Arehart, K. H., Kates, J. M., & Anderson, M. C. (2011). Effects of noise, nonlinear processing, and linear filtering on perceived music quality. *International Journal of Audiology*, *50*(3), 177-190. <https://doi.org/10.3109/14992027.2010.539273>
- Beck, D., & Le Goff, N. (2018). Contemporary hearing aid amplification: Issues and outcomes in 2018. *Journal of Otolaryngology - ENT Research*, *10*(1), 00303. <https://doi.org/10.15406/joentr.2018.10.00303>.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135-160. <https://doi.org/10.1177/096228029900800204>
- Bregman, A. S. (1990). *Auditory scene analysis*. MIT Press.
- 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), 387845. <https://doi.org/10.1155/ASP.2005.2991>
- Carlile, S., Ciccarella, G., Cockburn, J., Diedesch, A. C., Finnegan, M. K., Hafter, E., ... Sagers, J. E. (2017). Listening into 2030 workshop: An experiment in envisioning the future of hearing and communication science. *Trends in Hearing*, *21*, 1-11. <https://doi.org/10.1177/2331216517737684>
- Chasin, M., & Russo, F. A. (2004). Hearing aids and music. *Trends in Amplification*, *8*(2), 35-47. <https://doi.org/10.1177/108471380400800202>
- Chen, J., Wang, Y., & Wang, D. (2014). A feature study for classification-based speech separation at low signal-to-noise ratios. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(12), 1993-2002. <https://doi.org/10.1109/TASLP.2014.2359159>
- Cox, R. M., Johnson, J. A., & Xu, J. (2016). Impact of hearing aid technology on outcomes in daily life I: The patients' perspective. *Ear and hearing*, *37*(4), e224-237. <https://doi.org/10.1097/AUD.0000000000000277>
- Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and hearing*, *35*(5), E170-E184. <https://doi.org/10.1097/AUD.0000000000000056>
- Dirks, D. D., Kamm, C., Bower, D., & Betsworth, A. (1977). Use of performance-intensity functions for diagnosis. *Journal of Speech and Hearing Disorders*, *42*(3), 408-415. <https://doi.org/10.1044/jshd.4203.408>
- Dong, R., Hermann, D., Cornu, E., & Chau, E. (2007). Low-power implementation of an HMM-based sound environment classification algorithm for hearing aid application.

- Proceedings of the 2007 15th European Signal Processing Conference* (pp. 1635–1638). <https://doi.org/10.1.1.436.2846>
- Elberling, C. (1999). Loudness scaling revisited. *Journal of the American Academy of Audiology*, 10(5), 248–260.
- Fedtke, T., Fuder, G., Hamann, D., & Haubold, J. (1991). Natürliche klangbilder. *Neue Technologien in der Hörgeräte-Akustik-Herausforderung an die Audiologie. [Natural sounds. New technologies in hearing-acoustics-challenge to audiology]* Materialsammlung vom 5. Multidisziplinären Kolloquium der GEERS-Stiftung am, 12, 116–136.
- Feldbusch, F. (1998). Geräuscherkennung mittels Neuronaler Netze [Noise detection using neural networks]. *Zeitschrift für Audiologie*, 1(1998), 30–36.
- Freeman, C. (2007). Audio environment classification for hearing aids using artificial neural networks with windowed input. *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing* 183–188.
- Gallun, F., & Souza, P. (2008). Exploring the role of the modulation spectrum in phoneme recognition. *Ear and Hearing*, 29(5), 800–813. <https://doi.org/10.1097/AUD.0b013e31817e73ef>
- Gil-Pita, R., Ayllón, D., Ranilla, J., Llerena-Aguilar, C., & Díaz, I. (2015). A computationally efficient sound environment classifier for hearing aids. *IEEE Transactions on Biomedical Engineering*, 62(10), 2358–2368. <https://doi.org/10.1109/TBME.2015.2427452>
- Groth, J., & Cui, T. (2017). How accurate are environmental classifiers in hearing aids? www.audiologyonline.com
- Hannley, M., & Jerger, J. (1981). PB rollover and the acoustic reflex. *Audiology*, 20(3), 251–258. <https://doi.org/10.3109/00206098109072699>
- Hayes, D., & Eddins, D. A. (2008). Comparison of three premium products: Does design philosophy matter? *Hearing Review*, 15(13), 24–30. www.hearingreview.com
- Huber, R., Bisitz, T., Gerkmann, T., Kiessling, J., Meister, H., & Kollmeier, B. (2018). Comparison of single-microphone noise reduction schemes: Can hearing impaired listeners tell the difference? *International Journal of Audiology*, 57(sup3), S55–S61. <https://doi.org/10.1080/14992027.2017.1279758>
- Humes, L. E., Rogers, S. E., Main, A. K., & Kinney, D. L. (2018). The acoustic environments in which older adults wear their hearing aids: Insights from datalogging sound environment classification. *American Journal of Audiology*, 27(4), 594–603. https://doi.org/10.1044/2018_AJA-18-0061
- Jerger, J., & Jerger, S. (1971). Diagnostic significance of PB word functions. *Archives of Otolaryngology*, 93(6), 573–580. <https://doi.org/10.1001/archotol.1971.00770060875006>
- Johnson, J. A., Xu, J., & Cox, R. M. (2016). Impact of hearing aid technology on outcomes in daily life II: Speech understanding and listening effort. *Ear and Hearing*, 37(5), 529–540. <https://doi.org/10.1097/AUD.0000000000000327>
- Kates, J. M. (1995). Classification of background noises for hearing-aid applications. *The Journal of the Acoustical Society of America*, 97(1), 461–470. <https://doi.org/10.1121/1.412274>
- Keidser, G. (1995). The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids. *Ear and Hearing*, 16(6), 575–586. <https://doi.org/10.1097/00003446-199512000-00004>
- Keidser, G. (1996). Selecting different amplification for different listening conditions. *Journal of the American Academy of Audiology*, 7, 92–104.
- Keidser, G. (2009). Many factors are involved in optimizing environmentally adaptive hearing aids. *The Hearing Journal*, 62(1), 26–28. <https://doi.org/10.1097/01.HJ.0000344340.83852.7e>
- Kochkin, S. (2005). MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age. *The Hearing Journal*, 58(9), 30–39. <https://doi.org/10.1097/01.HJ.0000286545.33961.e7>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Korhonen, P., Kuk, F., Seper, E., Morkebjerg, M., & Roikjer, M. (2017). Evaluation of a wind noise attenuation algorithm on subjective annoyance and speech-in-wind performance. *Journal of the American Academy of Audiology*, 28(1), 46–57. <https://doi.org/10.3766/jaaa.15135>
- Krause, J. C., & Braid, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, 115(1), 362–378. <https://doi.org/10.1121/1.1635842>
- Kuk, F., Lau, C. C., Korhonen, P., & Crose, B. (2015). Speech intelligibility benefits of hearing aids at various input levels. *Journal of the American Academy of Audiology*, 26(3), 275–288. <https://doi.org/10.3766/jaaa.26.3.7>
- Lamarche, L., Giguere, C., Gueaieb, W., Aboulnasr, T., & Othman, H. (2010). Adaptive environment classification system for hearing aids. *Journal of the Acoustical Society of America*, 127(5), 3124–3135. <https://doi.org/10.1121/1.3365301>
- Miller, C. W., Bentler, R. A., Wu, Y. H., Lewis, J., & Tremblay, K. (2017). Output signal-to-noise ratio and speech perception in noise: Effects of algorithm. *International Journal of Audiology*, 56(8), 568–579. <https://doi.org/10.1080/14992027.2017.1305128>
- Mueller, H. G., Ricketts, T. A., & Bentler, R. (2014). *Modern hearing aids: Pre-fitting testing and selection considerations*. Plural Publishing.
- Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Tolerant of background noises: Relationship with patterns of hearing aid use by elderly persons. *Journal of Speech, Language, and Hearing Research*, 34(3), 679–685. <https://doi.org/10.1044/jshr.3403.679>
- 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. <https://doi.org/10.1121/1.1710877>
- Olsen, S. Ø., Brännström, K. J., Adams, E., Gordon-Hickey, S., Moore, R., Morlas, H., . . . Altman, D. (2014). Does the acceptable noise level predict hearing-aid use? *International Journal of Audiology*, 53(1), 2–20. <https://doi.org/10.3109/14992027.2013.839887>
- Ostendorf, M., Hohmann, V., & Kollmeier, B. (1998). Klassifikation von akustischen Signalen basierend auf der Analyse von Modulationsspektren zur Anwendung in digitalen Hoergeraeten [Classification of acoustical signals

- based on the analysis of modulation spectra for the application in digital hearing devices]. *Fortschritte der Akustik*, 24, 402–404.
- Park, G., & Lee, S. (2020). Environmental noise classification using convolutional neural networks with input transform for hearing aids. *International Journal of Environmental Research and Public Health*, 17(7). <https://doi.org/10.3390/ijerph17072270>
- Picou, E. M. (2020). MarkeTrak 10 (MT10) survey results demonstrate high satisfaction with and benefits from hearing aids. *Seminars in Hearing*, 41(1), 21–36. <https://doi.org/10.1055/s-0040-1701243>
- Plyler, P. N., Tardy, B., & Hedrick, M. (2019). The effects of nonlinear frequency compression and digital noise reduction on word recognition and satisfaction ratings in noise in adult hearing aid users. *Journal of the American Academy of Audiology*, 30(2), 103–114. <https://doi.org/10.3766/jaaa.17068>
- Ravindran, S., Schlemmer, K., & Anderson, D. V. (2005). A physiologically inspired method for audio classification. *EURASIP Journal on Applied Signal Processing*, 2005(9), 1374–1381. <https://doi.org/10.1155/asp.2005.1374>
- Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenes. *Journal of the American Academy of Audiology*, 26(2), 183–196. <https://doi.org/10.3766/jaaa.26.2.7>
- Takahashi, G., Martinez, C. D., Beamer, S., Bridges, J., Noffsinger, D., Sugiura, K., ... Williams, D. W. (2007). Subjective measures of hearing aid benefit and satisfaction in the NIDCD/VA follow-up study. *Journal of the American Academy of Audiology*, 18(4), 323–349. <https://doi.org/10.3766/jaaa.18.4.6>
- Taylor, B., & Hayes, D. (2015). Does current hearing aid technology meet the needs of healthy aging? *Hearing Review*, 22(2), 22–26. www.hearingreview.com
- Turan, S., Unsal, S., & Kurtaran, H. (2019). Satisfaction assessment with Abbreviated Profile of Hearing Aid Benefit (APHAB) questionnaire on people using hearing aid having Real Ear Measurement (REM) eligibility. *The International Tinnitus Journal*, 23(2), 97–102. <https://doi.org/10.5935/0946-5448.20190017>
- Vaisberg, J. M., Folkeard, P., Parsa, V., Froehlich, M., Littmann, V., Macpherson, E. A., & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline* (article 20872). www.audiologyonline.com.
- Wagener, K. C., Hansen, M., & Ludvigsen, C. (2008). Recording and classification of the acoustic environment of hearing aid users. *Journal of the American Academy of Audiology*, 19(4), 348–370. <https://doi.org/10.3766/jaaa.19.4.7>
- Walden, B. E., Surr, R. K., Cord, M. T., & Dyrland, O. (2004). Predicting hearing aid microphone preference in everyday listening. *Journal of the American Academy of Audiology*, 15(5), 365–396. <https://doi.org/10.3766/jaaa.15.5.4>
- 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. <https://doi.org/10.3766/jaaa.15105>
- Wong, L. L., Hickson, L., & McPherson, B. (2003). Hearing aid satisfaction: What does research from the past 20 years say? *Trends in Amplification*, 7(4), 117–161. <https://doi.org/10.1177/108471380300700402>
- Wu, Y. H., & Bentler, R. A. (2012). Do older adults have social lifestyles that place fewer demands on hearing? *Journal of the American Academy of Audiology*, 23(9), 697–711. <https://doi.org/10.3766/jaaa.23.9.4>
- Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., DeVries, S., & Oleson, J. (2019). Efficacy and effectiveness of advanced hearing aid directional and noise reduction technologies for older adults with mild to moderate hearing loss. *Ear and Hearing*, 40(4), 805–822. <https://doi.org/10.1097/AUD.0000000000000672>
- Zhao, Y., Wang, D., Johnson, E. M., & Healy, E. W. (2018). A deep learning based segregation algorithm to increase speech intelligibility for hearing-impaired listeners in reverberant-noisy conditions. *Journal of the Acoustical Society of America*, 144(3), 1627. <https://doi.org/10.1121/1.5055562>