Environmental Classification in Hearing Aids

Donald Hayes, Ph.D.[1]

ABSTRACT

There are two parts to this article. The first is a general overview of how hearing aid classification works, including a comparison study of normal-hearing listeners and multiple manufacturers' hearing aids while listening to a sound parkour composed of a multitude of acoustic scenes. Most hearing aids applied nearly identical classification for simple listening environments. But differences began to appear across manufacturers' products when the listening environments became more complex. The second section reviews the results of a study of the acoustic ecology (listening environments) experienced by several cohorts of hearing aid users over a 4-month period. The percentages of time people spent in seven different listening environments were mapped. It was learned that they spent an average of 57% of their time in conversation and that age is not a good predictor of the amount of time spent in most listening environments. This is because, when grouped by age, there was little to no difference in the distribution of time spent in the seven listening environments, whereas there was tremendous variability within each age group.

KEYWORDS: classification, log it all, listening environments

Modern hearing aids provide some degree of automatic program switching based on acoustic classification. The simplest of such hearing aids have been available almost since the turn of the century—and it is hard to believe it has been 21 years since then. Have you ever considered how the classification system quietly influences hearing aid performance? While some individuals wish to control their hearing aids manually, there is evidence for the superiority of hearing aids that automatically adapt to their changing listening environments, thereby allowing users to put the hearing aids on and forget about them.[1] This expectation places a lot of responsibility on the precision of the classification system in their hearing aids.

[1]Director, Clinical Research, Unitron, Ontario, Canada.

Address for correspondence: Donald Hayes, Ph.D., Clinical Research, Unitron, 20 Beasley Drive, Kitchener, ON, Canada N2E 1Y6 (e-mail: don.hayes@unitron.com).

As digital hearing aids have become more sophisticated, their performance has steadily improved. So too has the complexity of the underlying acoustic classification schemes that make it all possible. With the launch of Indigo in 2005, Unitron introduced a new type of classification system. The classifier was unique because it was trained using artificial intelligence to distinguish between four distinct acoustic scenes: quiet listening, speech in noise, noise, and music (see the article by Fabry and Bhowmik in this issue for more discussion about the use of artificial intelligence to train classifiers). Modern hearing aids are so reliant upon their classifiers' precision that it is almost impossible to overstate the importance of this unseen actuator constantly working in the background.

With the introduction of the conversational classifier on the north platform in 2015, we became so confident in our ability to correctly classify seven different listening environments that we used the classifier output to drive a feature called "Log It All."[2] While datalogging records what the hearing aid is doing over time, Log It All records the amount of time spent in each of seven listening environments. By providing an overview of the user's listening lifestyles, the clinician can use Log It All to better individualize their experiences in each of those environments. However, for Log It All to be of value, we had to be certain that the classifier was capable of accurately categorizing these listening environments in the first place.

Classification is even more critical for a good user experience. The clinician can perfectly set up parameters for each listening environment at the initial fitting visit. But if the classifier that drives the automatic program switching miscategorizes the listening environment, none of that will matter. For example, suppose the classifier detects a music environment while the user is actually having a conversation in a quiet setting. In this case, the hearing aid performance will be substandard because it is optimized for the wrong listening environment.

Consequently, precise classification is an absolutely essential component of success with modern hearing aids. Upon developing a seven-environment classifier and a tool like Log It All that records its output, we needed to know: Does the classifier get it right? Has the classifier been trained to accurately detect the actual listening environments in which users spend their time? What follows is a description of a pair of investigations that were explicitly conducted to answer these two questions. Covering them in chronological order, we started with a field study of Log It All, which led us to an extensive laboratory study of the underlying classification that drives it.

WHAT CLASSIFIERS DO

Automatic classifiers sample the current listening environment and generate probabilities for each of the listening destinations available in the automatic program. The hearing aid will switch to the listening destination for which the highest probability is generated. It will switch again when the listening environment changes enough to trigger a higher probability for another listening destination in the classifier.

Not all classification schemes work the same way. What makes them unique is the philosophy of the engineers who create them. These philosophies drive their choices about which aspects of a given listening environment distinguish it from all others. Consider this, two manufacturers' hearing aids could be exposed to the same listening environment and classify it differently. Why might this happen? Contradictory classifications will occur to the extent that the developers of the two systems assign different weightings to the various aspects of that listening environment. If the two classifiers measure different aspects of the environment, they may make different decisions about the values of what they detected. Thus, they reach different conclusions about the listening environment itself.

For example, consider these representative approaches to acoustic classification in hearing aids:

- Described a system based on cluster analysis of envelope modulation and spectral features to classify background noises into 11 classes: apartment, babble, dinner, dishes, gaussian, printer, traffic, typing, male talker, siren, and ventilation.[3]
- Used hidden Markov models to develop a robust classification system for hearing aids containing three classes: speech in traffic noise, speech in babble, and clean speech.[4]

- Classified clean speech, speech in noise, noise, and music using multiple approaches.[5] The authors explained many feature extraction types and then compared six different classifiers of low to moderate complexity required for hearing aid use.
- Tested two systems: Minimum distance and Bayesian classifiers.[6] Each type of classifier can adapt to the user's unique listening environment and tune itself accordingly. They chose distinctive features that have been shown to reliably distinguish between speech, noise, and music environments. These features included the listening environment's depth of amplitude modulations, modulations in the frequency ranges (0–4 Hz and 4–16 Hz), and temporal variance of the instantaneous frequency. They found that both methods worked well. But the two approaches did tend to merge classes differently when merging down to two classes from three.

This list is not exhaustive, but it does show many of the approaches available to engineers and scientists who develop these algorithms. Most of the classification that is done in Unitron products is proprietary, as with all manufacturers. Many of these approaches have come a long way since the early studies. Modern hearing aids have several orders of magnitude more processing power allowing the developers to include many acoustic features in the detection phase of classification. Some features are detected in the time domain early on the signal pathway, others in the frequency domain after Fast Fourier transform. Some acoustic features are short term, on the order of milliseconds, whereas others tend to vary over much longer time frames. For example, temporal modulation depth and modulation rate may be detected in the time domain. Or the modulation of a given frequency band can be detected during frequency domain processing. The presence of speech can be detected on the basis of rapid switching of energy across frequency bands or by the movement of spectral peaks and valleys over time. Furthermore, now that hearing aids can communicate with one another in real time, it possible to estimate differences in signal-to-noise ratio (SNR) between the ears for improved directional performance in noise (see the articles by Derleth et al; Jespersen et al; and Andersen et al in this issue for examples of this processing). Unitron hearing aids currently detect over 40 distinct acoustic features to help with classification.

While the philosophies of hearing aid companies are proprietary, it is still possible to compare their classification systems to one another and to a gold standard to document what different systems have to offer. To that end, we developed a benchmarking approach based on replicating natural listening environments in a controlled and repeatable setting. The approach and some of the outcomes will be described in this article.

BENCHMARKING

To ascertain the precision of the classifier, a benchmarking study was conducted along with Dr. David Eddins and Dr. Erol Ozmeral at the University of South Florida. The classifiers were benchmarked according to two types of comparisons. First, the hearing aid classifiers were compared with a human gold standard. Second, the classifier results for five brands of hearing aids were compared with each other. Both approaches offer valuable insights.

Test Environment

All of the measurements were conducted at the Auditory and Speech Sciences Laboratory at the University of South Florida. The room is a traditional sound-treated testing chamber. The sound room is shown in Fig. 1.

Acoustic scenes that simulated real-world listening environments were created using an array of 64 independently driven ear-level loudspeakers that surrounded a chair at the center of the room. Normal-hearing listeners sat in the chair while they evaluated the listening environments. Hearing aid output for all test conditions was also recorded from the hearing aids using a Klangfinder positioned in the same place as the listeners.

A Klangfinder is an anthropomorphic system that can record the output from three pairs of hearing aids simultaneously (Fig. 2).
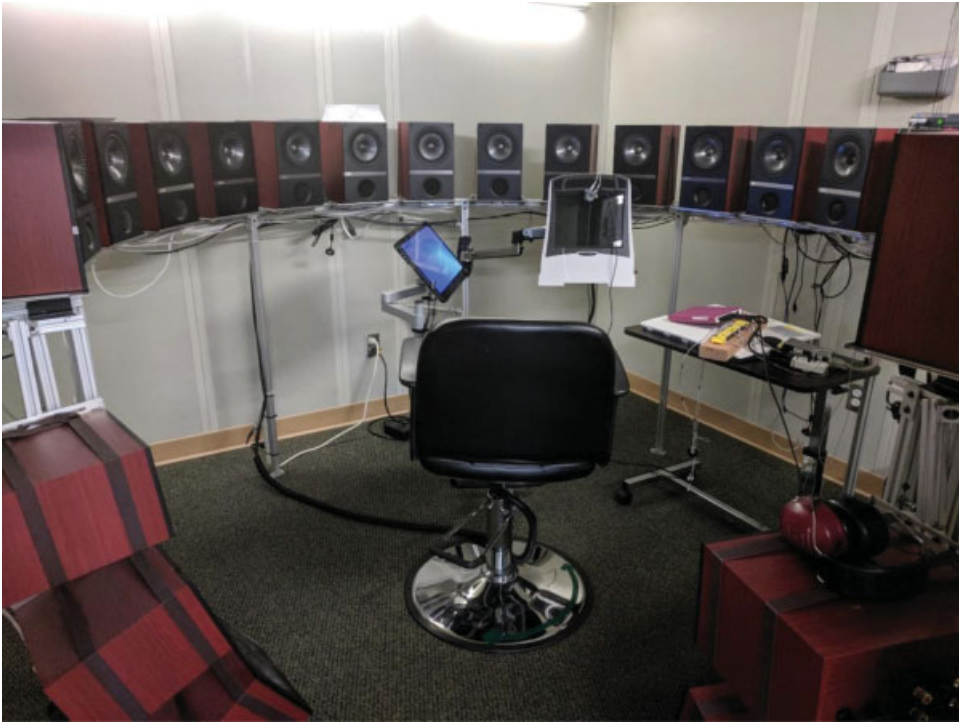
Figure 1 Sound room at the Auditory Speech Sciences Laboratory, University of South Florida. The normal-hearing participants sat in the chair at the center of the 64- speaker array. Alternatively, the Klangfinder was placed in the center of the room when the hearing aid classifiers were tested instead of the human participants.

The Sound Parkour

Using a sound parkour—an acoustic obstacle course of sorts—the classifiers were tested in 26 acoustic scenes that varied in difficulty. It is important to note that the parkour was not designed to exercise a single classifier to test its accuracy. If you consider all of the characteristics of different listening environments in a

Figure 2 Klangfinder acoustic manikin head. Three ear simulators are mounted on either side of the head to obtain measurements from three pairs of hearing aids at the same time.

multidimensional space, then the parkour should be designed to move about in that space, testing the classifiers in one or more dimensions at a time. The concept for the parkour began by including what were thought to be the most relevant aspects that define a range of listening environments and where they exist in multidimensional space: the number and location of the speakers; the type and direction of any noise and/or music; and the levels of the speech, noise, and music relative to one another.

As indicated by the header and the left column of Table 1, the parkour was defined along multiple dimensions: the number of talkers and their spatial distribution; the noise source(s) and their spatial distribution; the SNR; and overall level. Each row of Table 1 describes the makeup of a single sound file that

is 2 minutes in duration and represents a specific acoustic scene. The simplest acoustic scene is called quiet listening (in the top row). There is no speech, just the soft sound of a fan running steadily with an overall level of 40 dB sound pressure level (SPL); there is almost no modulation and no temporal or spectral contrasts— just a soft, steady noise.

As you go down the table, the complexity of the listening environments was increased by adding talkers and changing the background noise. The music and background noise levels were also manipulated to vary the types and complexity of the environments.

There was also a directional component to the speech, noise, and music elements in the sound parkour. As more talkers were added, their orientation relative to the front of the

Table 1   Description of the acoustic scenes which comprised the Parkour for this study

Condition	Talkers	Background Noise	Talker Distribution*	Noise Distribution*	SNR	Overall Level
Quiet Listening	0	Fan noise	N/A	0°, 90°, 180°, 270°	N/A	40 dB
Quiet Conversation	1	N/A	0°	N/A	N/A	55 dB
Quiet Conversation	3	N/A	0°, 300°, 60°	N/A	N/A	55 dB
Quiet Conversation with Music	1	Music	0°	90°	-3	55 dB
Quiet Conversation with Music	3	Music	0°, 300°, 60°	90°	-3	55 dB
Small Group Conversation with noise	3	Subway	270°, 0°, 90°	0°, 90°, 180°, 270°	+15	70 dB
Small Group Conversation with noise	3	Subway	270°, 0°, 90°	0°, 90°, 180°, 270°	+10	70 dB
Small Group Conversation with noise	2	Traffic	270°, 90°	0°, 90°, 180°, 270°	0	70 dB
Small Group Conversation with noise	2	Traffic	270°, 90°	0°, 90°, 180°, 270°	-10	70 dB
Small Group Conversation with noise	3	Car	270°, 0°, 90°	0°, 90°, 180°, 270°	-10	70 dB
Small Group Conversation with noise	3	Car	270°, 0°, 90°	0°, 90°, 180°, 270°	-15	80 dB
Small Group Conversation with noise	3	Food court	300°, 0°, 60°	0°, 90°, 180°, 270°	0	70 dB
Small Group Conversation with Noise/Music	3	Traffic, Music	300°, 0°, 60°	0°, 90°, 180°, 270°, 90°	0, -5	70 dB
Small Group Conversation with Noise/Music	3	Traffic, Music	300°, 0°, 60°	0°, 90°, 180°, 270°, 90°	-5, -5	70 dB
Small Group Conversation with Noise/Music	3	Traffic, Music	300°, 0°, 60°	0°, 90°, 180°, 270°, 90°	-5, +5	70 dB
Small Group Conversation with Noise/Music	3	Traffic, Music	300°, 0°, 60°	0°, 90°, 180°, 270°, 90°	-10, +15	75 dB
Large Group Conversation	1	6T-Babble	0°	315°, 45°, 135°	+5	65 dB
Large Group Conversation	1	8T-Babble	0°	315°, 45°, 135°, 225°	0	70 dB
Large Group Conversation	1	10T-Babble	0°	315°, 45°, 135°, 225°, 180°	-5	75 dB
Large Group Conversation with Music	1	6T-Babble, Music	0°	315°, 45°, 135°, 90°	+5, -10	65 dB
Large Group Conversation with Music	1	8T-Babble, Music	0°	315°, 45°, 135°, 225°, 90°	0, -10	70 dB
Large Group Conversation with Music	1	10T-Babble, Music	0°	315°, 45°, 135°, 225°, 180°, 90°	-5, -10	75 dB
Large Group Conversation with Music	1	8T-Babble, Music	0°	315°, 45°, 135°, 225°, 90°	0, 0	70 dB
Large Group Conversation with Music	1	10T-Babble, Music	0°	315°, 45°, 135°, 225°, 180°, 90°	-10, 0	75 dB
Television Viewing	0	TV (CSI S01E01)	N/A	0°	N/A	70 dB
Television Viewing	2	Sporting Event	330°, 30°	225°, 335°	+5	70 dB

Each row describes one acoustic scene. The simplest scenes are at the top of the table, and they increase in complexity as one progresses down to the bottom of the table. Each sound scene is defined by the information in the corresponding row moving from left to right: the number of talkers, the type of background noise when present, the orientation of the various talkers on the horizontal around the listener, the orientation of the noise sources around the listener, Signal-to-noise ratio (SNR), and overall level. In some acoustic scenes, there is also music present. When present, Music is listed as a background noise, its orientation was always 90 degrees relative to the listener with varied signal-to-music ratios (SMRs). Positive SMRs simulated acoustic scenes where the music was in the background and more negative SMRs for acoustic scenes where the music was primary.

hearing aids was updated to reflect where a talker would typically stand or sit in that environment. This step incorporates any impact of directional processing. For example, note the orientation of the talkers—left, right, and front—in the subway environment. This "talker distribution" is what you would experience on a subway platform in the London subway when sitting between two companions with another person in front of you carrying on a conversation. For the record, the subway sound was a recording made in the London Tube. The talkers were varying combinations of male and female voices recorded in a sound studio in Toronto while listening to noise under headphones. Similar logic was used when assigning directional components to the noise and music sound files. The traffic noise was recorded outside of the Unitron office in Kitchener, Ontario; the food court was from a nearby mall at lunchtime; and the TV sporting event was a German soccer match—I don't recall who won. Multiple iterations of the sound parkour have been used in other studies. Table 1 represents the version used for this one.

The sound files for a single acoustic scene were looped for 8 hours of continuous playback to each pair of hearing aids in the Klangfinder. There was no direct way to read the classifier probabilities from most of the hearing aids. Instead, datalogging results were used to determine how each manufacturer's classifier logged that particular 8-hour listening environment. Given that the datalogging of time spent in a given listening environment is most likely driven by classifier probabilities over time, looping a single acoustic scene for 8-hour session was the most logical way to obtain stable classifier outcomes. Each acoustic scene is represented by a given 2-minute sound file defined by the corresponding row of Table 1. The hearing aids were assessed in groups of three as the Klangfinder has three sets of ears. The acoustic output and signal processing of the hearing aids was not relevant to this experiment as the datalogging results were the dependent variable of interest. The hearing aids were set for minimal gain to reduce the potential for acoustic feedback.

WHAT DO ACTUAL CLASSIFIER RESULTS LOOK LIKE?

Before examining the datalogging results from the other manufacturers' hearing aids, it is instructive to look at more detailed results from the Unitron hearing aids. Instantaneous classifier probabilities from the Unitron hearing aid were written to a file several times a second as they were generated. Figs. 3 and 4 show actual classifier probabilities as determined by a pair of hearing aids using this approach. The first case, Fig. 3, shows 60 seconds worth of classifier probabilities for two very simple listening environments.

The top of Fig. 3 shows 60 seconds of the time waveforms that were recorded from the output of the left and right hearing aids. The first half of the waveforms was from the final 30 seconds of a soft fan at 40 dB SPL (quiet listening acoustic scene in the top row of Table 1). The second half of the waveforms was from the first 30 seconds of the Quiet Conversation with a single talker acoustic scene (second row of Table 1). These simple acoustic scenes demonstrate how the classifier generates probabilities that almost exclusively represent a single listening environment. The bottom center of the figure is time synched with the recordings and shows the distribution of probabilities for each of the seven possible listening environments in the Unitron classifier. For the first 30 seconds, the classifier indicated a 100% probability (one on the class–probability axis) for the quiet listening environment. Given that it is a recording of a soft fan measured at only 40 dB SPL in a sound-treated room, the classification is correct. The hearing aid would spend these 30 seconds in the quiet listening environment.

At 30 seconds, the recording abruptly switches from the soft fan to the single talker. From 30 seconds to approximately 37 seconds, the classifier probabilities are in transition. Note how the probability of speech in quiet immediately begins to rise as the probability of quiet listening drops. The two probabilities transect one another at approximately 35 seconds. In this transition zone, the hearing aid switches from the quiet listening environment to the speech in quiet environment. In reality, the classifier detects the change almost immediately. However, if the hearing aid rapidly changes
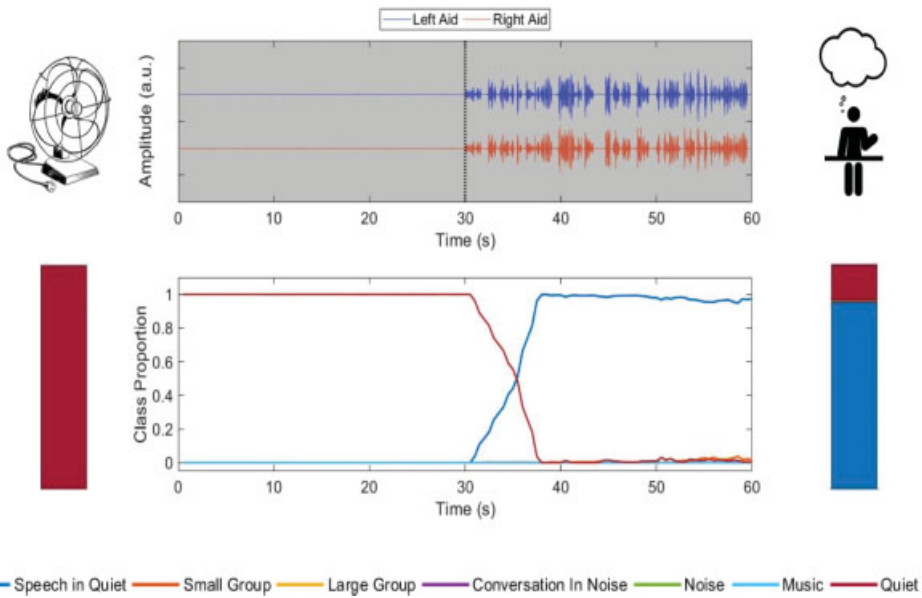
Figure 3 Simple case—**Top:** 60 seconds of acoustic recordings. The first 30 seconds consisted of a small fan inside a sound room at 40-dB SPL. The last 30 seconds consisted of a single speaker at 55 dB SPL as recorded through left and right hearing aids. **Bottom:** Classifier output probabilities for a seven-category classification system. All seven categories are shown in the legend by color code. As the probability of a given category rises from 0 to 1 (0–100%, respectively), the corresponding colored line rises as well. The top and bottom figures are synchronized in time.
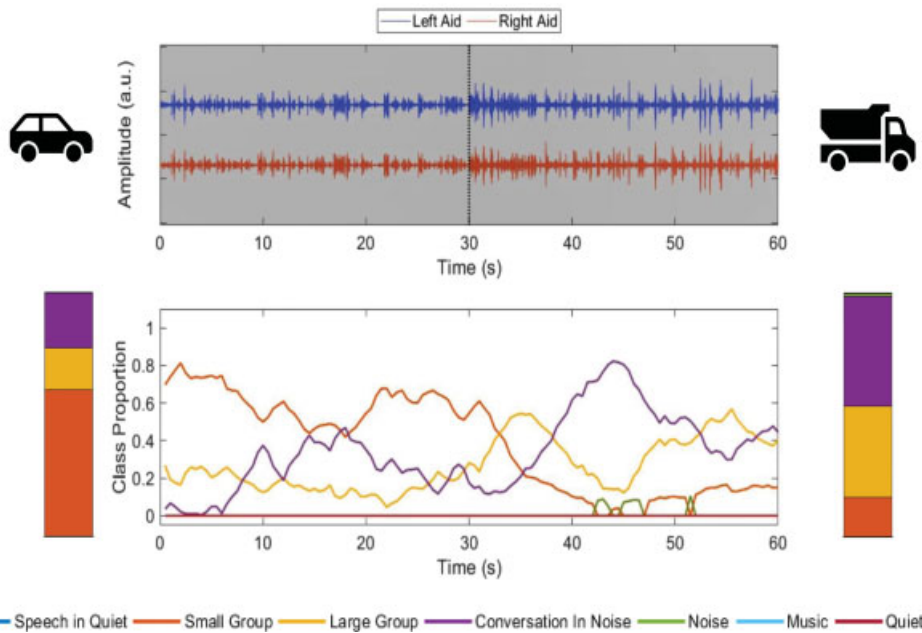
Figure 4 Complex case—**Top:** 60 seconds of acoustic recordings. The first 30 seconds consisted of three talkers in a car at 70-dB SPL overall and −10 dB SNR. The last 30 seconds was even more difficult. They consisted of three talkers in a car at 80 dB SPL overall and −15 dB SNR as recorded through left and right hearing aids. **Bottom:** Classifier output probabilities for a seven-category classification system. All seven categories are shown in the legend by color code. As the probability of a given category rises from 0 to 1 (0–100%, respectively), the corresponding colored line will rise as well. The top and bottom figures are synchronized in time.

between listening environments in response to every environmental fluctuation, sound quality will be negatively impacted in dynamic listening environments. Therefore, the developers made a conscious decision not to have the hearing aid react too quickly to every little change in the listening environment. By 40 seconds and for the last 20 seconds of the recording, the probability of a speech in quiet environment is almost 100%.

The entire stimulus in this example consisted of 2 minutes of the quiet listening acoustic scene (soft fan) followed by 2 minutes of the quiet conversation acoustic scene. The cumulative proportion of time each listening environment was classified during the first 2 minutes is shown by the vertical bar on the bottom-left of Fig. 3. As expected from the earlier discussion, the bar is entirely red, indicating that this was appropriately classified as a quiet listening environment. The cumulative proportion of time each listening environment was classified during the last 2 minutes is shown by the vertical bar on the bottom-right of Fig. 3. It is mostly blue, which appropriately corresponds to a speech in quiet environment. The slight red section represents the transition time when the stimulus switched to the Quiet Conversation acoustic scene. Fig. 4 is an example of what happens in a more complex listening environment.

Here we can see the impact on the probabilities of two much more complex listening environments. In both cases, the listener is driving in a car along with three talkers. For the first half of the time, the car is much quieter, with an overall level of approximately 70 dB SPL and a −10 dB SNR. For the second half of the time, the overall level is higher (80 dB SPL), and the SNR is much more difficult (−15 dB SNR). These levels may look like nearly impossible SNRs for a hearing aid user, but car noise is distinctive in that almost all of the energy is very low in frequency (below 1,000 Hz). As such, the overall SNRs look extreme, but the SNR in the high frequencies is much more favorable and the speech signal can be clearly seen in the time waveforms (compare the time waveforms in Fig. 4 to the second half of the time waveforms in Fig. 3).

As the car noise changes (simulated by changes in speed in this example) and as the

talkers start and stop, the classifier probabilities vary widely across a blend of different listening environments. In this example, three different listening environments are detected by the classifier. When the overall level is lower during the first 30 seconds, the highest probability is conversation in a small group, averaging 50 to 60%. As one might expect, conversation in noise is also detected, varying from 0 to 50%. Conversation in a large group has a smaller but still noticeable probability hovering around 15 to 20% throughout. Once the overall level goes up and the SNR gets worse, the car noise becomes predominant. As the car speeds up, the classifier probabilities shift hard into the conversation in noise environment, while conversation in a small group drops below 20%.

Take a moment to reflect on these two examples. The first one is easy. Having benchmarked hearing aids from many manufacturers, it is clear that each would react similarly in both listening environments shown in Fig. 3. But what about the two environments in Fig. 4? This is where the developers' philosophies play a role. A lot is going on in these listening environments, and developers have to decide how they want their algorithms to react. For example, what is more important: eliminating the car noise or enhancing the speech? At what point is the overall level too loud and not worth worrying about the speech? Is that decision based on the overall level, the SNR, or a combination of both? The sound parkour is designed to look at all of these possibilities and illuminate the relevant choices made during algorithm development.

THE GOLD STANDARD

Table 1 lists several general acoustic scenes that a hearing aid user might encounter in real-world listening environments. To evaluate the validity of the labels assigned to each acoustic scene, 20 normal-hearing listeners described the listening environments they thought best represented each acoustic scene. Multiple answers were considered acceptable. The sound files associated with each acoustic scene were played back in randomized order to each listener. Listeners heard each acoustic scene three times and described the environment for each

Table 2 An example of how the young normal-hearing participants (gold standard) classified the acoustic scenes compared with the Unitron classifier

Young Normals	Classifier
Quiet	Quiet
Speech in Quiet	Quiet Speech
	Small Group
Speech in Noise	Large Group
	Speech in Noise
Noise	Noise
Speech and Music	
Music	Music

The relationship is not exact, but it is very close.

repetition. The responses were pooled across listeners and compared with the hearing aid classifiers.

Table 2 compares the descriptions of the human listeners to the seven labels used by the classifier.

Although there was some overlap in specific terminology, there were interesting differences in interpretation for what those descriptors meant. There were three descriptors for listening environments used by both the listeners and the classifier: quiet, noise, and music. However, the interpretation of each term was often quite specific. Quiet was used very infrequently by the listeners and rarely exceeded 3% for any listening environment. For example, the soft fan acoustic scene at the top of Table 1 was given a 100% probability of quiet by the classifier since the overall level was a mere 40-dB SPL. But the listeners identified it as noise 92% of the time. Interestingly, the listeners' probability for noise was above 27% in just two other listening environments, both of which were quite loud. The very noisy acoustic scenes all contained speech and were therefore given the highest probabilities of speech in noise by the listeners. The same was true for the classifier, except it made a distinction on the basis of noise type: either multiple background talkers or engine noise such as trains, cars, or traffic. Neither the listeners nor the classifier identified the acoustic scenes as music very often. Both identified acoustic scenes as music only when the music was much louder than everything else around it. But the listeners offered a distinct category of speech in music mixed with speech in noise for seven acoustic scenes. In contrast, the classifier identified these as a large group, which they were, thereby ignoring the music in favor of optimizing the speech.

The main distinctions between the listeners and the classifier were not so much that they were detecting different things but that they were prioritizing different aspects of the acoustic scenes or were making slightly more precise distinctions in some cases. For example, one could easily argue that a soft fan at 40 dB SPL is both quiet and a noise. Both are correct interpretations of the same listening environment.

THE MULTIPRODUCT COMPARISON

The following results show how premium products from five manufacturers, including Unitron, classify several acoustic scenes versus our young normal-hearing listeners, who are referred to in Figs. 5, 6, 7 and 8 as our "gold standard." This exercise was not about who was right or who was wrong—rather, it was an opportunity to see how different classifiers compare. The results illuminate how the different philosophies across manufacturers reveal themselves by virtue of how they are grouped across brands. Unitron is called out by name because the accuracy of Unitron's classifier has relevance for the second study described after this one. In that study, we used Log It All to map the probability of people's experiences in different real-world listening environments. We could not establish the accuracy of Log It All results without also comparing the Unitron classifier to the perception of normal-hearing listeners under controlled acoustic scenes.

Let's start again with a simple example. Fig. 5 shows how the gold standard and the five hearing aids classified a single male talker from the front at 55 dB SPL.

Different manufacturers have different classification schemes that use different names for the listening environments they classify. Using their descriptions of what each listening category was intended for, the titles were grouped into four main categories: quiet, speech in noise, noise, and music (as shown in the legend of Fig. 5). These four general

categories appear in all of the hearing aids we tested under one name or another, but the generic names were used to maintain the anonymity of the manufacturers and hearing aids involved. The gold standard classified this acoustic scene as quiet listening approximately 98% of the time. All five hearing aids did essentially the same.

The acoustic scene represented in Fig. 6 is a bit more complex than that in Fig. 5. As before, there was a single talker directly in front of the listener, but the overall presentation level was 80 dB SPL with a nominal SNR of 0 dB. The background noise is a subway train in the London Tube, and the levels varied as trains arrived and departed.
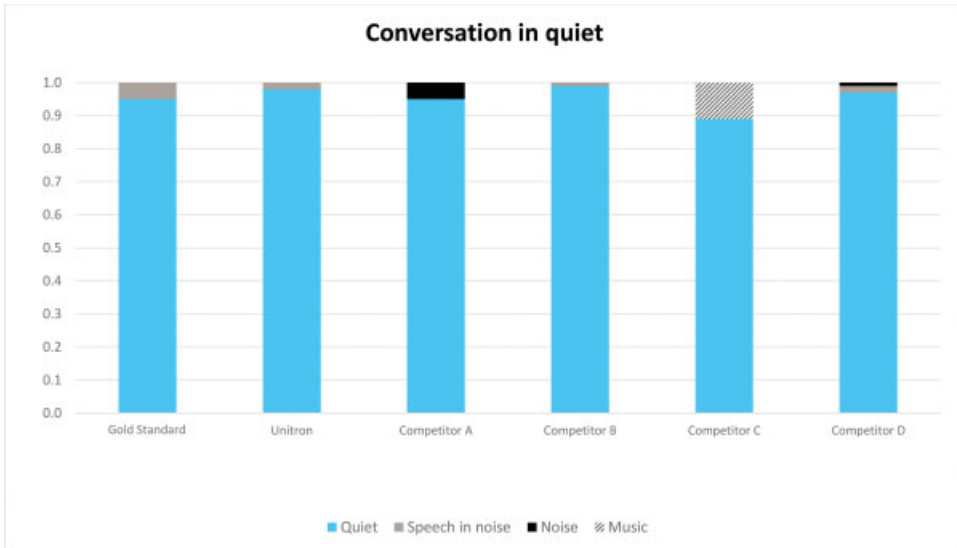
Figure 5 Gold standard based on 20 young normal-hearing listeners' subjective assessments of a quiet conversation in the parkour. This assessment is compared with the datalogging classification from five hearing aid manufacturers for the same quiet conversation.
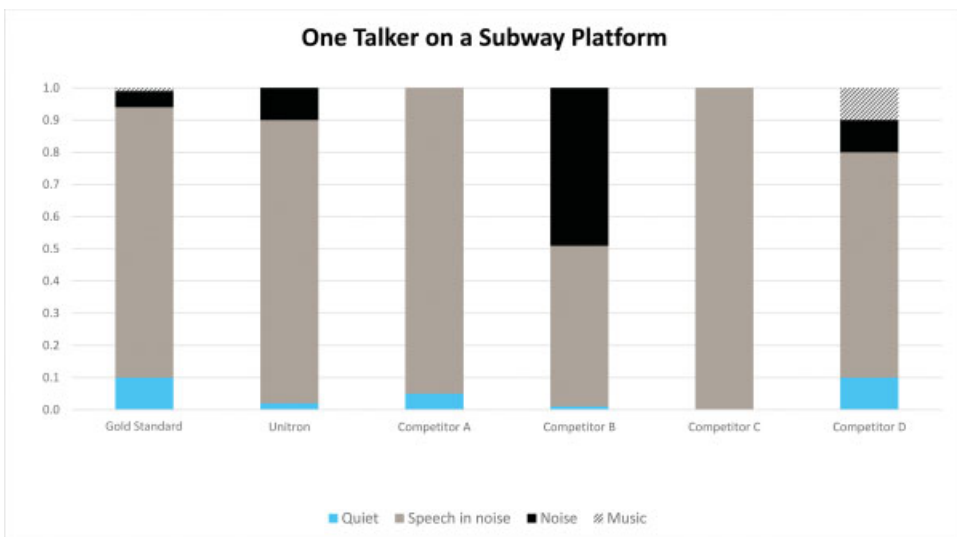

Figure 6 Gold standard plus five hearing aid manufacturers logged classification of a single talker on a subway platform in front of the listener in the London Tube.
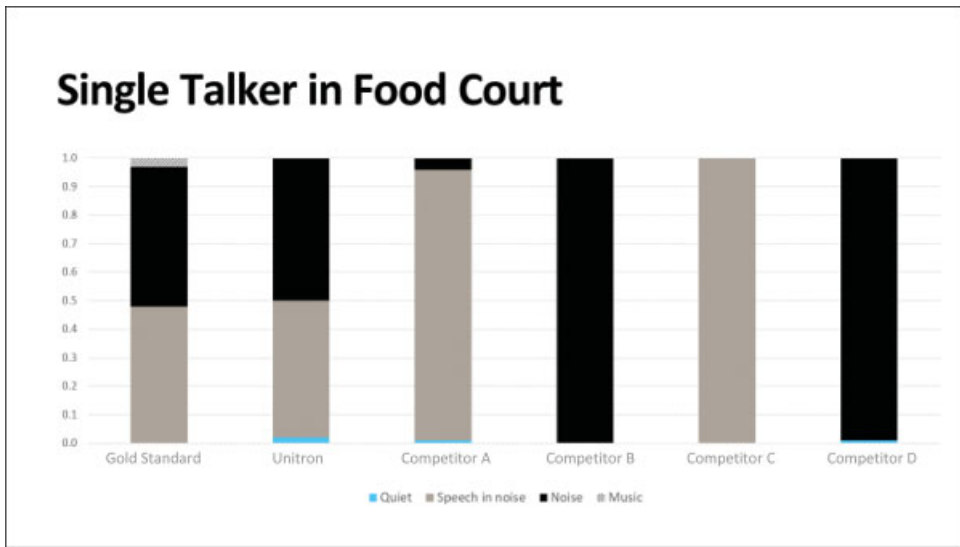
Figure 7 Gold standard plus five hearing aid manufacturers logged classification of a single talker in a food court at the mall during lunch hour.

The gold standard listeners classified this acoustic scene as speech in noise approximately 83% of the time. They also said it was noise 4% of the time and quiet 10% of the time. Bearing in mind the level differences as trains came and went, it is fair to say that Unitron and competitor D were closest to what the gold standard listeners reported. Competitor A was quite similar as well. However, competitors B and C were very different.

This example is where the differences in philosophy are first apparent. Competitor B classified the environment as just noise approximately 50% of the time, whereas competitor C classified it as speech in noise 100% of the time. It is clear that the gold standard listeners reported speech in noise relatively consistently. Therefore, the SNR must have been reasonable most of the time. However, at 80 dB SPL, the overall level is quite high. Thus, it is reasonable to infer that
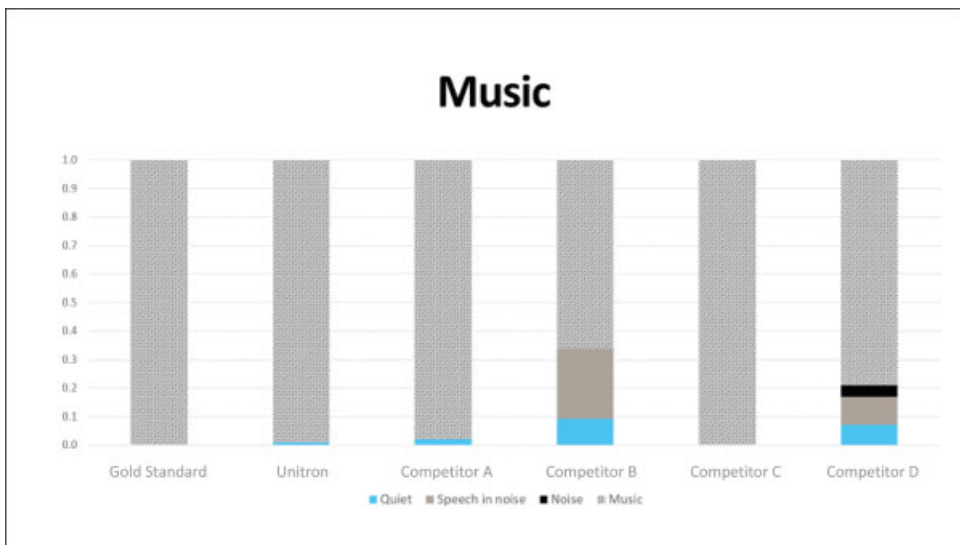

Figure 8 Gold standard plus five hearing aid manufacturers logged classification of music at 65 dB SPL.

competitor B has a philosophy that is more sensitive to overall level than to SNR in this example like the other four hearing aids tested. From this example, it seems that if there is speech and if there is also noise during the sampling interval (not necessarily at the same time), competitor C will exclusively classify it as speech in noise.

The background represented in Fig. 7 is even more complex. A single talker was presented from the front, and the background was a mall food court near lunchtime. This is a complex background with dozens of people carrying on many conversations at once as well as the sound of the kitchens serving food and people walking by. Compared with the previous example, the overall presentation level was lower at 70-dB SPL at a 0-dB SNR.

For this example, the gold standard listeners reported approximately 47% speech in noise and approximately 50% noise only. The other 3% was music. This time, the classifier results varied widely across manufacturers. While all classifiers offered some combination of speech in noise and noise, the percentages for competitors A and C were completely the opposite of those for competitors B and D.

These contrasting results may be the perfect example of philosophical differences in what Unitron Hearing Scientist Leonard Cornelisse calls the "give-up point." He defines the "give-up point" as the signal level and/ or SNR where the hearing aid user "gives up" trying to follow the speech because the situation has become too difficult. Below the give-up point, the user will work to follow what is being said and reports it as a speech in noise environment, expecting the hearing aid to emphasize speech clarity. But once the give-up point is crossed, the user reports that it is too difficult to follow the speech or too loud to listen comfortably, and they would like the hearing aid to emphasize comfort over clarity. Every classifier is built to make this decision at some point. It is a purely acoustically driven decision unless the user switches to a manual program to override it.

The first takeaway from Fig. 7 is that competitors A and C assume a higher give-up point than competitors B and D. With a near 50/50 split between speech in noise and noise,

both the Unitron classifier and the gold standard listeners indicated that this environment is very close to the give-up point. This acoustic scene is perhaps the most striking example of how a developer's philosophy will impact the classifier's behavior. Since the give-up point for different hearing aid users often varies widely, who is to say which of these classifiers will get it absolutely right for a particular user?

The final example represented by Fig. 8 is Music presented alone at 65 dB SPL with no other background sounds. This is not a high level for listening to music and it does not replicate a live performance. Instead, it is closer to the level at which a hearing aid user may listen to music while cooking or reading a book, but a little louder than background music.

In this example, the gold standard listeners, Unitron, competitor A, and competitor C all indicated that this was essentially a pure music environment. Competitors B and D classified it differently at least 33% and 20% of the time, respectively. The most common misclassification was for speech in noise, and this is the one example where a clear and indefensible "miss" occurred. Mistaking music for speech in noise is tantamount to setting up a hearing aid for exactly the opposite behavior than what a user would prefer. It is generally accepted practice to set a music environment for broadband lightly processed reproduction. But a speech in noise environment is often heavily processed by directional microphones and noise canceling, which, among other things, tends to reduce low-frequency amplification. To be fair, such a miss was not common for the five classifiers.

BENCHMARKING SUMMARIZED

Hearing aid sound scene classification is a topic that gets precious little attention. Yet, it is one of the most critical components of a hearing aid's architecture. Stealthily running in the background, classifiers make all of the decisions about which sets of processing parameters are the most valid in any given listening environment. Consequently, the classifier's accuracy in a particular environment may heavily impact how a user hears.

Classification decisions are based as much on philosophy as on acoustics. As such, not all classifiers are equal in all situations. Most of the

time, particularly in simple listening environments, almost all of the top hearing aids will converge on highly consistent outcomes that correspond with how a normal-hearing listener would classify the environment. But once the listening environment becomes more complex, the differences in philosophy and sometimes behavior become apparent.

In this benchmarking study, the Unitron classifier provided results that were consistent with a group of young normal-hearing (gold standard) listeners. This was an important finding for the following study that investigated the real-world listening environments of more than 1,000 hearing aid users from around the world.

GLOBAL LISTENING ENVIRONMENT STUDY

One of the most critical steps in any new product development is the proper validation of each new feature on actual hearing aid users in their own listening environments. Validation in this sense can be described as a test of the efficacy of the feature under ecologically valid conditions. In other words, it is important to confirm that the feature works properly and that it adds value for the hearing aid user in their listening environment. Upon establishing that the classifier output was representative of the perception of the gold standard normal-hearing listeners, the next logical question was, "What can be learned from Log It All"? There was a genuine fear that although we had a seven-category automatic program, most people would spend almost all of their time in one or two listening environments, such as quiet listening and one-on-one conversation.

Log It All provided the means by which to properly assess the types of listening environments where people spend their time in the real world. Log It All directly reads the output of the classifier and makes it available for later inspection. This was a milestone moment because, for the first time, the hearing aids themselves could directly and automatically monitor the users' acoustic ecology while they went about their daily lives. Using this new tool, we were able to answer the question, "What percentage of their time do people spend in each of seven different lis-tening environments?." The classifier provides moment-by-moment probabilities that the user is in one of the seven acoustic scenes based on the presence of speech, noise, or music.

- *Quiet*—very low noise floor and no speech, such as reading a book or working on a computer.
- *Conversation in quiet*—soft or average level speech, typically one on one with minimal noise.
- *Conversation in a small group*—a relatively easy listening environment with more than one talker and some background noise as in a quiet cafe or small family gathering.
- *Conversation in a crowd*—a large group scenario with significant noise due to multiple talkers.
- *Conversation in noise*—talking to someone in the presence of a steady noise such as a busy restaurant, kitchen, or public transportation.
- *Noise*—background noise is present and speech is absent; the noise could be as soft as a nearby air conditioner or as loud as would be expected near machinery.
- *Music*—music is not merely present (as in background music) but is the primary signal in the acoustic scene, such as a live concert or a radio station.

Data Collection

The objective of the Global Listening Environment Study (GLES) was to quickly collect as much data as possible about the acoustic ecology of as many and varied types of people as possible. All of the data were collected in the 4-month span between April 1, 2016, and July 28, 2016. The data consisted of information that clinicians had readily available either from their files or from the hearing aids themselves through the fitting software. Table 3 shows the 10 countries where data were collected and the number of useable records obtained in each country.

Data were collected from purchasers of new hearing aids after two full weeks of daily use. The information obtained was the user's (1) Log It All results, (2) age (by decade), (3) gender, (4) hours per day of hearing aid use, and (5) population density of the place where they resided. The population density values

Table 3 List of the 10 countries where data were collected from hearing aids for the Global Listening Environment Study and the number of useable records obtained from each country following the application of the 6 hours of use rule

Country	Records
Australia	134
Canada	87
France	78
Germany	87
Netherlands	141
New Zealand	140
Slovenia	58
South Africa	55
Spain	67
USA	208
Total	1055

were used directly to categorize the groups numerically but also by type of location: rural, small town, suburban, or urban dwellers. See Table 4 for the breakdown.

Data Analysis

Many interesting relationships can be mined from this dataset, for example: Do people of different ages spend more or less time in noise? Are there differences by gender? Are urban dwellers in more noisy environments than peo-

ple who live in more rural areas? For this article, the most relevant relationships will be addressed later. First, it is important to remember that the primary purpose was to obtain a proper sampling of as many listening environments as possible. For this purpose, it is important to know how long users actually wore their hearing aids.

The first question most clinicians would want to ask to someone to whom they just sold a set of hearing aids would be, "How much do you wear them?" This can be answered by looking at the datalogging.

Fig. 9 is a histogram of the hours of use for all participants from all countries. The large blip at 15 hours is due to the way the data were collected. It was mistakenly assumed that no first-time users would use their hearing aids for more than 15 hours a day. Hence, this was the highest value in the data entry form for clinicians to record datalogging usage, meaning that all times equal to or greater than 15 hours per day were entered as 15 hours. When averaged across all users, the mean usage time was 9.9 hours/day with a 3.5-hour standard deviation.

Ideally, the most representative sampling of the fullest range of listening environments should come from those who wore their hearing aids full time. Those who wore them for 1 or 2 hours a day almost certainly used them situationally. They have one or two specific listening situations where they have significant difficulty and wear their hearing aids only in those

Table 4 The type of data collected as part of each record

Gender	Age in Decades	Location	Population Density	Logged (Hrs/day)	Listening Environments
Male	20 - 29	Rural	0 - 1000	0	Conversation 1 on 1
Female	30 - 39	Small Town	1001 - 10,000	1	Conversation Small Group
	40 - 49	Suburban	10,001 - 25,000	2	Conversation Large Group
	50 - 59	Urban	25,001 - 100,000	3	Conversation in Noise
	60 - 69		100,001 - 500,000		Quiet Listening
	70 - 79		500,001 - 1,000,000	13	Noise only
	80 - 89		1,000,000 - 10,000,000	14	Music
	90+			15	

This consisted of information immediately available to the clinician as a part of routine record keeping or intuitively obvious details such as the population density near the clinic site. The data points included the hearing aid user's gender, their age by decade at the time of the fitting, a general nominal categorization of the population surrounding the clinic by density, numeric categorization of population density, data logging results of daily average use time, and percentage of time recorded by Log It All for each of the seven acoustic environments for which the classifier provides probabilities.
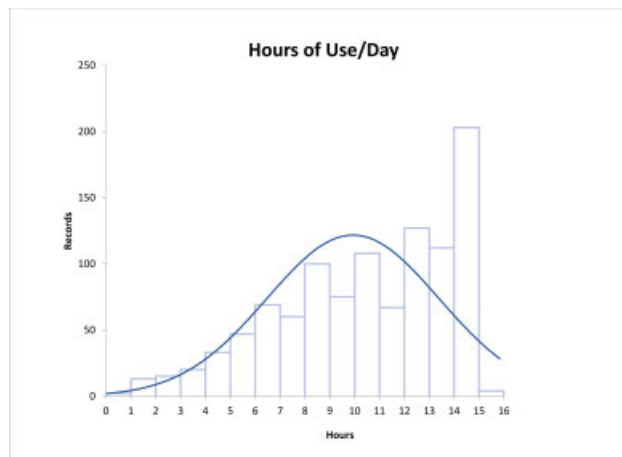
Figure 9 Count of the average hourly use logged per day during the first 2 weeks of hearing aid ownership. Each record is one user's average hourly use. Average full-day use was only tracked up to 15 hours. Anything longer than 15 hours was counted as 15 hours of average daily use.

situations. Such people might skew the distribution of listening environments unequally toward a given set of difficult situations, and therefore, their results were dropped entirely from the rest of the analysis. Since the mean wearing time of 9.9 hours minus 1 standard deviation (3.5 hours) is equal to 6.4 hours, anyone whose wearing time was less than 6 hours was dropped, leaving a total of 1,055 records (shown in Table 3) for the rest of the analyses.

As previously stated, a lot of data were generated and analyzed in many different ways. But some questions are much more clinically relevant than others. From here on, the results concerning the following five questions will be discussed.

1. How much time do hearing aid users spend in each of the seven listening environments?
2. How does the distribution of listening environments vary by age?
3. How much time is spent in conversational environments?
4. How much time is spent listening in noise?
5. How much time is spent listening in quiet?

HOW MUCH TIME DO HEARING AID USERS SPEND IN EACH LISTENING ENVIRONMENT?

Fig. 10 shows the median and interquartile ranges of time spent in each of the seven listening environments for all participants in the study. The four categories to the left represent the four listening environments recognized by the classifier as being conversational in nature. The three categories to the right are not conversational environments, meaning that if speech was present, it was not the predominant factor for classification. The cloud of symbols over each box and whisker plot show the individual usage times for participants. The "■" and "▲" symbols represent the outliers. There are a couple of interesting points worth discussing about this figure.

This pattern of use, where the greatest amount of time (highest median) is spent in quiet listening, followed by conversation in a small group, and continuing down to the smallest median category—music—is very stable no matter how the data are analyzed. This is shown in further detail in Figs. 11 and 12. It is the predominant pattern throughout the remainder of this discussion. However, even though this pattern repeats for group results, no matter how the data are parsed, it is not representative of all individuals. Evidence that there is considerable individual variability in this dataset can easily be seen in the variance of the individual results, which is substantial. Also worth mentioning is that the median time spent by hearing aid users in each of the seven listening environments is at least 5%.

HOW DOES THE DISTRIBUTION OF LISTENING ENVIRONMENTS VARY BY AGE?

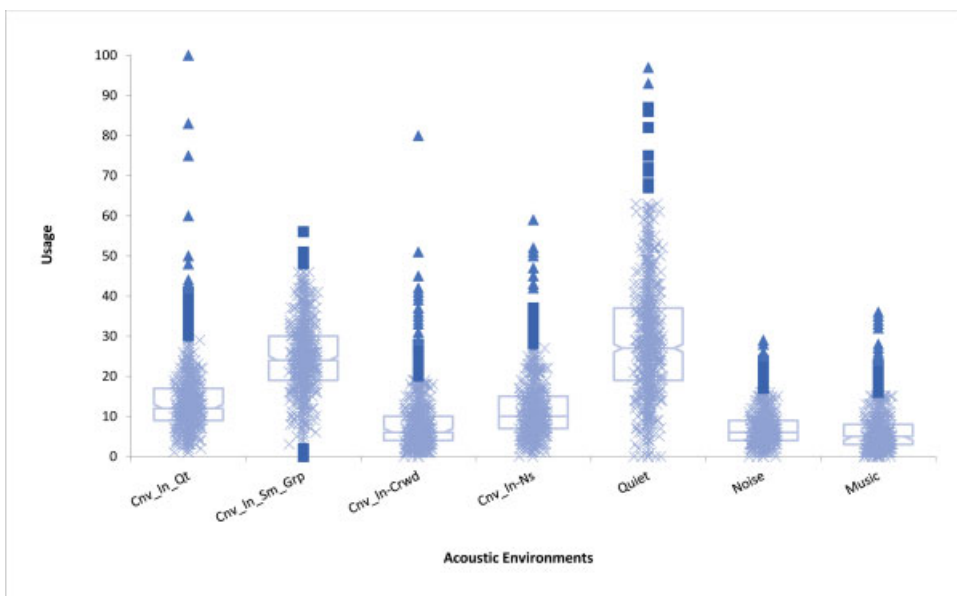At this point, one may ask: "Sure, when you average everybody together, they will cover all

Figure 10   Log It All records the proportions of seven listening environments from the classifier. Shown is a box and whisker plot of time spent in each of the seven listening environments for all participants with 6 or more hours of average daily use. The center line of each box is the median percentage of time for that listening environment and the edges are the interquartile ranges from the 25th percentile at the bottom to the 75th percentile at the top. The "X" markers are individual results, and the "■" and "▲" represent outliers.
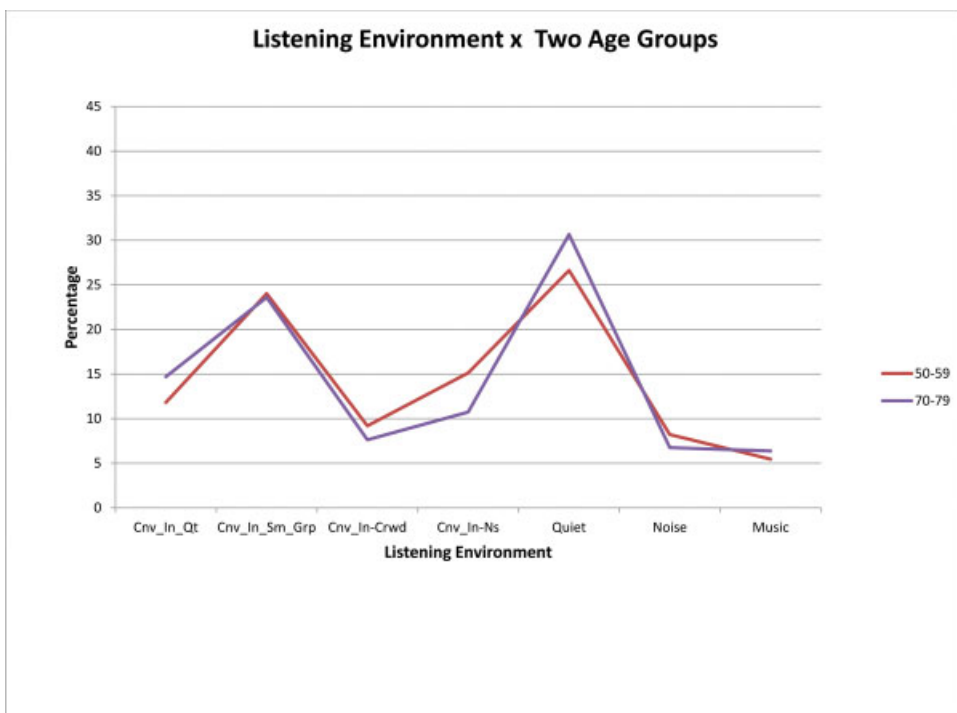
Figure 11   The mean percentage of time spent in each of the seven listening environments for two age cohorts, 50- to 59-year-olds and 70- to 79-year-olds.
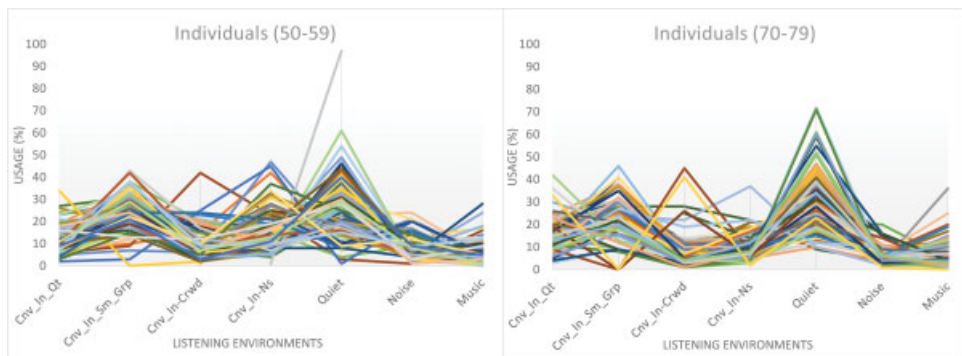
Figure 12 Individual Log It All percentages for each individual in the two age cohorts, 50- to 59-year-olds and 70- to 79-year-olds. The individual plots demonstrate the significant variability even within cohorts of similar age.

seven listening environments. But is it not the case that younger people in the workforce spend a lot more time in conversation and noise than retired 80-year-olds who just watch game shows and bake cookies at home all day?" It is a cheeky question. But is it grounded in fact?

To examine how listening habits varied by age, the results were analyzed by age using 10-year cohorts. For simplicity, Figs. 11 and 12 show a direct comparison of just two cohorts, 50 to 59 years of age and 70 to 79 years of age. First, Fig. 11 shows the mean percentage of time spent in each listening environment for both cohorts. These two cohorts were chosen because the 50 to 59 group is most likely predominated by people who are still working, whereas the 70 to 79 group is likely mostly retired. The age difference between any two people in each group is between 11 and 30 years. Furthermore, the take-home message would be the same for any two sets of cohorts. But these two, in particular, are among the two largest in the sample.

It is evident that there are no notably large differences in the two distributions of mean results. One small set of differences is that the older group averages 4% more time in quiet listening, which is offset by 4.5% less time in conversation in noise relative to the younger group. But in general, the overlap in these two lines is not at all what one might expect. Fig. 12 provides a complete picture of the individual results.

Fig. 12 shows the Log It All results for every individual in both groups from Fig. 11.

The data from these two cohorts make it clear that there is a considerable amount of individual variability even within a given age group. It also brings home two very important points about these hearing aid users.

1. Across groups—average results are highly consistent.
2. Within groups—individual results are highly inconsistent.

Therefore, age is not a good predictor of a given individual's acoustic ecology.

HOW MUCH TIME IS SPENT IN CONVERSATIONAL ENVIRONMENTS?

One of the most important motivators for a hearing aid purchase is the desire for improved performance in conversation. Or at least to better understand speech. Thus, it is reasonable to ask the question, "How much time do people spend listening to speech or otherwise engaged in conversation?" To help answer this question, Fig. 13 shows the sum of the percentages for the four conversational listening environments and the sum of the percentages for the three nonconversational listening environments.

HOW MUCH TIME IS SPENT LISTENING IN NOISE?

One of the key complaints among most hearing aid users is the inability to properly communicate in background noise. Log It All uses classifier probabilities to track three types of
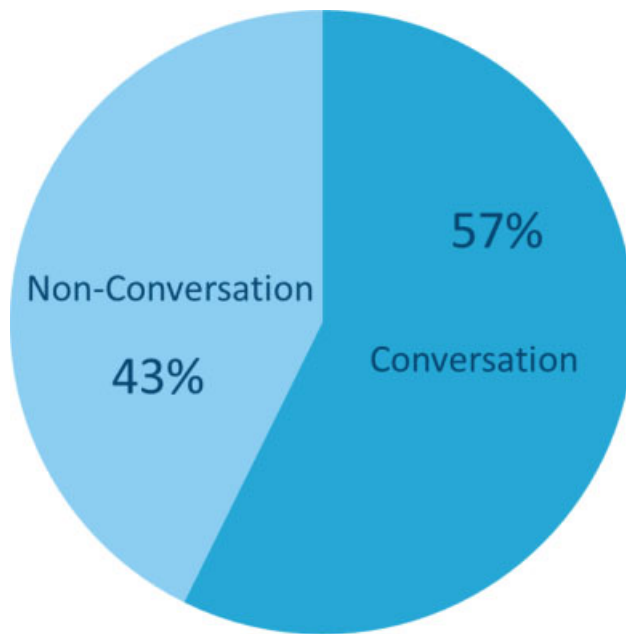
Figure 13 The percentage of time the hearing aids classified the acoustic environments as being either conversational or nonconversational in nature.

noise environments: conversation in a crowd, conversation in noise, and exclusively noise. Crowd noise is when the background is predominated by other talkers. Conversation in noise can be whenever the background is predominantly a type of noise that is not speech, such as riding in a car, working in a busy kitchen, or working on a construction site. Exclusive noise is when there is no speech present or the SNR is so unfavorable that speech understanding is likely to be nonexistent. The only rule is noise in the absence of speech. Unlike the other two categories, this category has no level dependency. Even a quiet fan will fall into this category if there is no speech. Hence, it is not likely something that most hearing aid users would complain about.

When looking at time spent in noise, there is a clear age effect. As stated previously, the variability within an age cohort far exceeds the differences between cohorts. Nonetheless, it is clear from Fig. 14 that the older people in this study spent less time in the full range of noise backgrounds than did younger people.

Given the description above, the "listen in noise" bars in Fig. 14 may be slightly misleading. The dark blue bars show the mean ±1 standard deviation of time spent conversing in a crowd for each cohort by age. The middle blue bars show the mean ± 1 standard deviation for conversing in nonspeech noise. The light blue bars represent the summation of all three noise environments added together; so, they have larger standard deviations. Thus, the light blue bars show the percentage of time a given cohort spent in all types of noise. There is a clear trend in Fig. 14 where the total time in noise drops from approximately 37% at 50 to 59 years of age down to only 21% in the 90+ age range. This is one of the more notable age-related findings from this study. In general, people spend less time on average in noise as they age. But even at 90+ years of age, individuals are still in noisy environments as much as 21% or one-fifth of the time. Having said this, it is worth noting that for all but the oldest group, the standard deviations for each bar are larger than the mean differences across ages. This indicates that there is still a lot of variability in individual acoustic ecologies.

HOW MUCH TIME IS SPENT LISTENING IN QUIET?

Needless to say, there is a flip side to listening in noise. Some people purchase hearing aids because they inhabit complex and challenging
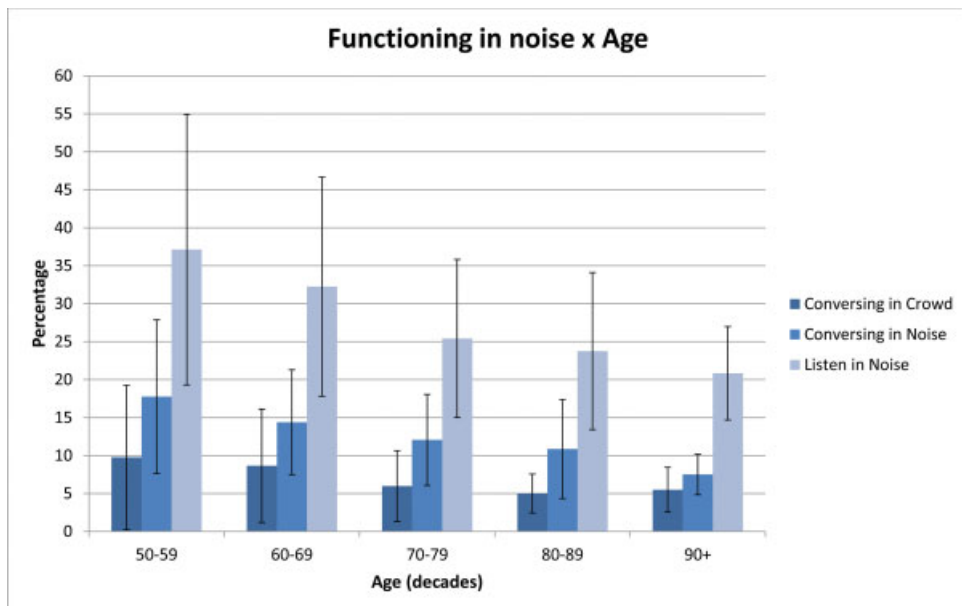
Figure 14 The mean percentage of time spent in various acoustic environments classified as containing noise ± 1 standard deviation, broken down by age cohorts in decades. The darkest (left) bars show the percentage of time spent in acoustic environments where there is conversation and the background noise consisted of other people talking. The medium (middle) bars are the percentages for environments containing speech where the background noise is nonspeech, such as public transportation or machines. The lightest bars (right) show the percentage of time spent in all acoustic environments containing noise when combined. Therefore, the percentages of the first two bars (left and middle) are part of the sum total of these bars.

listening environments. Others struggle to hear the soft speech of their spouse or grandchildren. It is beneficial to know which category a given individual falls into when choosing the appropriate technology. For the purpose of the "quiet listening" analysis, the percentages of three different listening environments were grouped together. Fig. 15 shows the sum of the three environments: one-on-one conversation, small group conversation, and quiet listening where there is no speech. These three classifications all require a reasonably quiet listening environment.

It should come as no surprise that Fig. 15 looks a little like the inverse of Fig. 14. In this analysis, the mean percentage of time in quiet was lowest for the youngest cohort at 38.3% and highest for the oldest group at 51.5%. Once again, it is important to remember that although the average percentage of time in quiet listening environments trends upward with age, the standard deviation about the mean for every age range is considerably larger than the average difference across ages.

Thus, it stands to reason that although there is a general trend for people to spend less time in noise and more time in quiet as they age, the rule cannot be applied uniformly to specific individuals: There are 90-year-olds who spend as little as 30% of their time in quiet listening environments and 50-year-olds who spend as little as 20% in noise. Those people may not be in the plurality for their ages, nor are they outliers.

GLOBAL LISTENING ENVIRONMENT STUDY SUMMARIZED

The purpose of the GLES was to get a sense of the percentage of time people spend in different listening environments. The results revealed a lot of useful information. It demonstrated how much time average hearing aid users spend in different listening environments. It was found that, on average, users spent most of their time, roughly 57%, in listening environments where conversation was predominant. It was also learned that the average distribution of time spent in
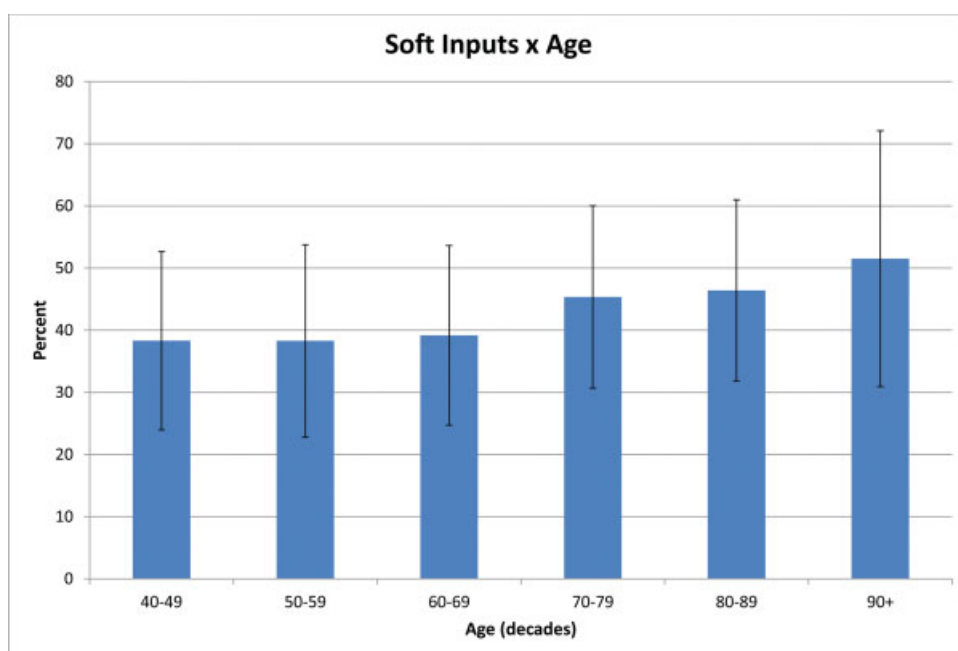
Figure 15 Mean percentage of time ± 1 standard deviation spent in soft- or low-level acoustic environments broken down by age cohorts in decades.

different environments was highly consistent regardless of almost every other factor that was examined. However, that consistency at the group level belied substantial individual variability in each user's listening experiences.

Going into this study, there was some thought that older people who were less active in the workforce might spend less time in more challenging listening environments. But the average data indicated that, regardless of a person's age, differences in the percentage of time people spent in various listening environments are much smaller than expected.

CONFLICT OF INTEREST

The author is a salaried employee of Unitron, which is a brand of the Sonova Group. All features and hearing aids specifically named are Unitron products and features.

ACKNOWLEDGMENTS

I would like to acknowledge the contributions of Dr. Ozmeral and Dr. Eddins who worked closely with us to develop the sound parkour and undertake the data collection in their laboratory at the University of South Florida for the classifier study.

REFERENCES

1. Rakita L, Jones C. Performance and preference of an automatic hearing aid system in real-world listening environments. Hearing Review 2015;(December): 28–34
2. Howard T Log It All. Reference #027–5896– 02.2015:1–6
3. Kates JM. Classification of background noises for hearing-aid applications. J Acoust Soc Am 1995;97 (01):461–470
4. Nordqvist P, Leijon A. An efficient robust sound classification algorithm for hearing aids. J Acoust Soc Am 2004;115(06):3033–3041
5. Büchler M, Allegro S, Launer S, Dillier N. Sound classification in hearing aids inspired by auditory scene analysis. EURASIP J Appl Signal Process 2005;18:2991–3002
6. Lamarche L, Giguère C, Gueaieb W, Aboulnasr T, Othman H. Adaptive environment classification system for hearing aids. J Acoust Soc Am 2010; 127(05):3124–3135