# What's the big deal with hearing instrument classifiers?

Author
Donald Hayes, Ph.D.
Director, Clinical Research
Favorite sound: Blues guitar

unitron™ Love the experience

A Sonova brand

# Introduction

Modern hearing instruments all provide some degree of automatic program switching based on acoustic classification. The simplest of such devices have been available almost since the turn of the century – and it's hard to believe it's been 19 years since then! Have you ever considered how the classification system quietly influences hearing instrument performance? While there are still some individuals who wish to manually control their hearing instruments, most people would prefer to put them on and forget about them, allowing the hearing instruments to automatically adapt to their changing listening environments. That places a lot of responsibility on the precision of the classification of which that device is capable.

As digital hearing instruments have become more sophisticated, their performance has steadily improved. But so 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. It was our first classifier trained using artificial intelligence to distinguish between four different acoustic scenes: quiet listening, speech in noise, noise, and music.

With the introduction of our conversational classifier on the North platform, we became so confident in our ability to correctly classify seven different listening environments that we use the classifier output to drive Log It All, a Unitron industry first. While datalogging tells you what the hearing instrument is doing over time, Log It All tells you how much time the wearer spends in each of the seven listening environments – showing you an overview of the wearer's listening lifestyle, helping you to individualize the wearer's experience in each environment. However, for Log It All to be of value, we have to be certain that the classifier is accurately categorizing these listening environments.

Classification is even more important for a good user experience. You can perfectly set up parameters for each listening environment at the first fit, but if the classifier that drives the automatic program switching mis-detects the acoustic environment, none of that will matter. For example, if the classifier thinks the wearer is listening to music while he is actually having a conversation in a quiet setting, the hearing instrument performance will be substandard, because it is optimized for the wrong listening environment.

Consequently, precise classification is an absolutely critical component of success with modern hearing instruments. At Unitron, we wanted to know: do we get it right? Have we trained our classifier to accurately detect the true acoustic environments in which consumers spend their time?

To answer our questions, we undertook a benchmarking study of our conversational classifier at the University of South Florida with Dr. David Eddins and Dr. Erol Ozmeral.

# What classifiers do

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

However, not all classification schemes work the same way. What makes them unique is the philosophy of the engineers who create them. It is these philosophies that drive their choices about which aspects of a given acoustic environment distinguish it from all others. Consider this: two manufacturers' hearing instruments could be exposed to the same acoustic environment and classify it differently. Why does this happen? It's because the designers of the two systems assigned different weightings to the various aspects of that acoustic environment. Therefore, the devices were measuring different aspects of the environment and making different decisions about the values of what they detected. Thus, they can reach different conclusions about the acoustic environment itself.

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

- (Kates, 1995) described a system based on cluster analysis of envelope modulation and spectral features to classify background noises into eleven classes: apartment, babble, dinner, dishes, gaussian, printer, traffic, typing, male talker, siren, and ventilation.
- (Nordqvist & Leijon, 2004) used hidden Markov models to develop a robust classification system for hearing instruments containing three classes: speech in traffic noise, speech in babble, and clean speech. (Büchler, Allegro, Launer, & Dillier, 2005) classified clean speech, speech in noise, noise and music using multiple approaches. The authors explained many types of feature extraction and then compared six different classifiers of low to moderate complexity, required for HA use.
- (Büchler, Allegro, Launer, & Dillier (, 2005) classified; clean speech, speech in noise, noise and music using multiple approaches. The authors explained many types of feature extraction and then compared six different classifiers of low to moderate complexity, required for HA use.
- (Lamarche, Giguere, Gueaieb, Aboulnasr, & Othman, 2010) tested two systems: Minimum Distance and Bayesian classifiers. In each case, the classifier can adapt to the listeners unique environments and tune itself accordingly. They chose distinctive features that are good for distinguishing between Speech, Noise and Music environments, including Depth of amplitude modulations, Modulation frequency ranges (0 – 4 Hz & 4 – 16 Hz), and Temporal variance of the instantaneous frequency. They found that both methods worked well. But they did tend to merge classes differently when merging down to two classes from three.

While this list isn't exhaustive, it shows many of the approaches available to engineers and scientists who develop these algorithms. While the philosophies of hearing instrument companies are proprietary, it's still possible to compare these schemes to one another and to a gold standard to evaluate what different systems have to offer. To that end, we developed a benchmarking approach based on replicating real listening environments in a controlled and repeatable setting. The approach and some of the outcomes will be described in this paper.

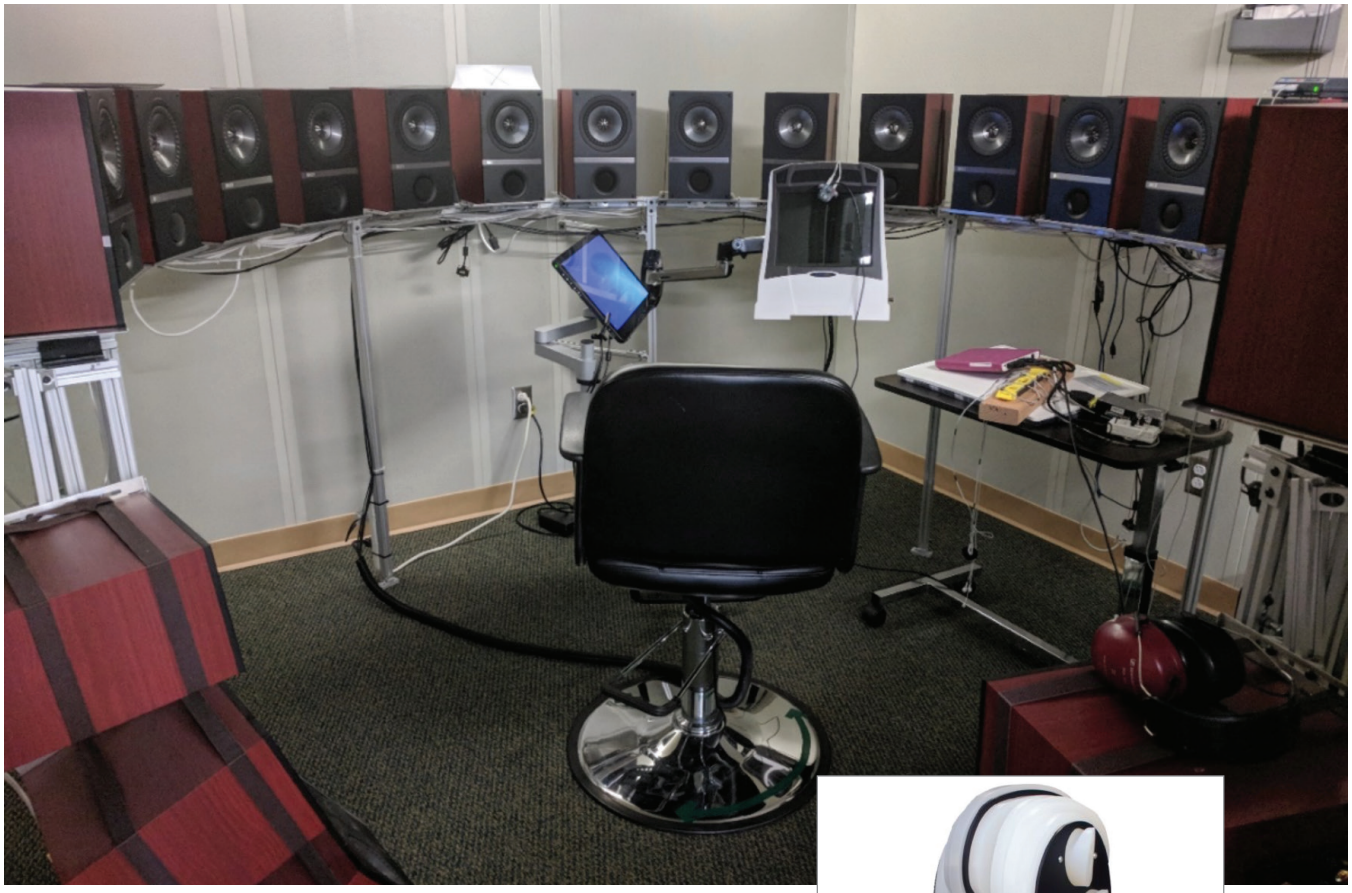
# The benchmarking approach

We chose to benchmark the classifiers by applying two types of comparisons. First, we compared all of the hearing instrument classifiers to a human gold standard. Second, we compared the classifier results for five manufacturers' hearing instruments to each other. Both approaches offer useful insights.

# The location

We conducted all of the measurements at the Auditory & Speech Sciences Laboratory at the University of South Florida. The sound room is shown in Figure 1.

**Figure 1**



The chair at the center of the room is surrounded by an array of 64 independently driven ear level loudspeakers. Though the room is a traditional sound treated testing chamber, plexiglass panels can be mounted on the walls and ceiling to create a more naturally reverberant environment. Human participants are seated in the chair at the center of the room while evaluating listening environments. We obtained hearing instrument data in sets of three devices at a time using a Klangfinder anthropomorphic system (Figure 2).

**Figure 2**

By replacing the human participants at the center of the room with the Klangfinder, it was possible to repeat all test conditions for all subjects and all hearing instruments in one location.

# The sound parkour

We began the measurement exercise by creating a sound parkour – a sort of acoustic obstacle course to put the classifiers through their paces. We defined the parkour in multiple dimensions, as shown along the header and the left column of Table 1. Each row of Table 1 describes the makeup of a single sound file that is two minutes in duration and represents a specific listening environment. This iteration of the parkour contains 26 listening environments (sound files). The simplest listening environment 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 SPL. There is almost no modulation and no temporal or spectral contrasts – just a soft, steady noise.

As you go down the table, the listening environments become more complex. For example, in the left column you will see that we added more talkers and several different types of background noise. We also experimented with different levels of music and background noise combined with speech in the very complex environments.

There is also a directional component to the speech, noise and music elements. As more speakers are added, their orientation relative to the front of the hearing instruments is updated to reflect where a speaker would normally stand or sit in that environment. This step incorporates any impact of directional processing. For example, note the orientation of the speakers – left, right and front –

**Table 1**

| 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 |
| | 3 | N/A | 0°, 300°, 60° | N/A | N/A | 55 dB |
| Quiet conversation with music | 1 | **Music** | 0° | **90°** | -3 | 55 dB |
| | 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 |
| | 3 | Subway | 270°, 0°, 90° | 0°, 90°, 180°, 270° | +10 | 70 dB |
| | 2 | Traffic | 270°, 90° | 0°, 90°, 180°, 270° | 0 | 70 dB |
| | 2 | Traffic | 270°, 90° | 0°, 90°, 180°, 270° | -10 | 70 dB |
| | 3 | Car | 270°, 0°, 90° | 0°, 90°, 180°, 270° | -10 | 70 dB |
| | 3 | Car | 270°, 0°, 90° | 0°, 90°, 180°, 270° | -15 | 80 dB |
| | 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 |
| | 3 | Traffic, **Music** | 300°, 0°, 60° | 0°, 90°, 180°, 270°, **90°** | -5, **-5** | 70 dB |
| | 3 | Traffic, **Music** | 300°, 0°, 60° | 0°, 90°, 180°, 270°, **90°** | -5, **+5** | 70 dB |
| | 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 |
| | 1 | 8T-Babble | 0° | 315°, 45°, 135°, 225° | 0 | 70 dB |
| | 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 |
| | 1 | 8T-Babble, **Music** | 0° | 315°, 45°, 135°, 225°, **90°** | 0, **-10** | 70 dB |
| | 1 | 10T-Babble, **Music** | 0° | 315°, 45°, 135°, 225°, 180°, **90°** | -5, **-10** | 75 dB |
| | 1 | 8T-Babble, **Music** | 0° | 315°, 45°, 135°, 225°, **90°** | 0, **0** | 70 dB |
| | 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 |
| | 2 | Sporting Event | 330°, 30° | 225°, 335° | +5 | 70 dB |

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. The directional component is also used for the noise and music in the sound files. Multiple iterations of the sound parkour have been used, of which Table 1 is a representative example.

Each sound file was looped for eight hours of continuous playback to each set of hearing instruments in the Klangfinder. There was no direct way to read the classifier probabilities from most of the devices. Instead, we relied on the datalogging results for eight hours' worth of a single file to determine how the classifier of each manufacturer logged that particular 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 sound file for eight hours/session was the most logical way to obtain stable classifier outcomes.

# What do actual classifier results look like?

Before looking at the results taken indirectly from five manufacturers' hearing instruments using datalogging output, it will be instructive to look at more detailed results from Unitron hearing instruments. It is possible for Drs. Eddins and Ozmeral to read out classifier probabilities from our hearing instrument instantaneously several times a second while they are being generated.

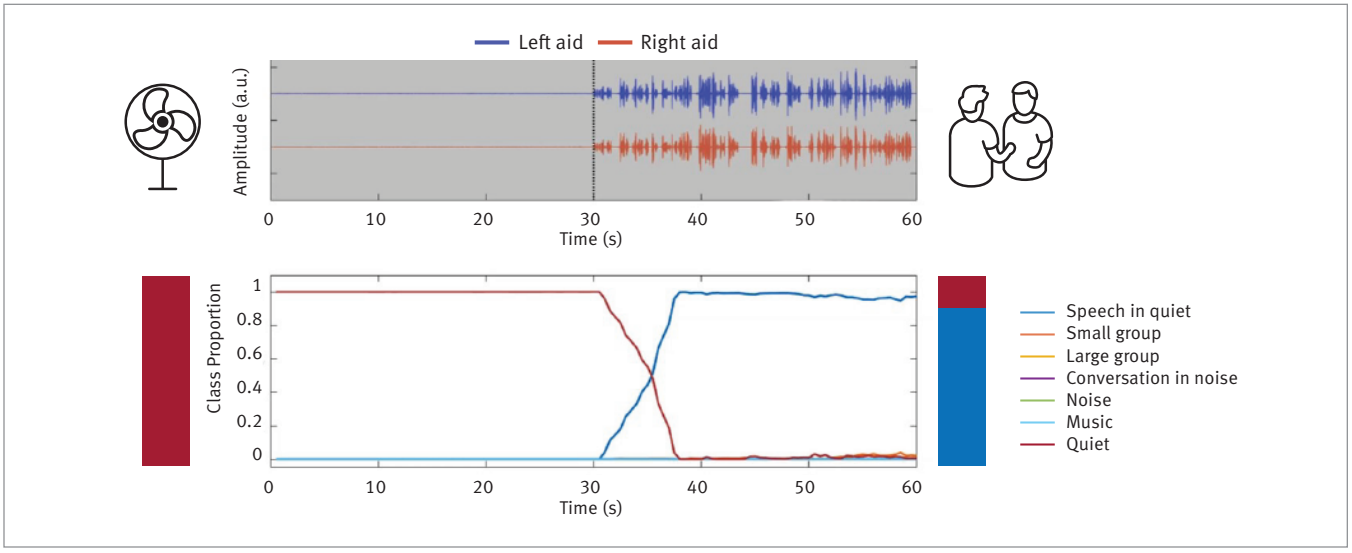
Figures 3 and 4 show actual classifier probabilities as determined by a pair of Unitron hearing instruments using this approach. The first case, Figure 3, shows 60 seconds worth of classifier probabilities for two very simple listening environments.

At the top of Figure 3 we can see 60 seconds of the original playback. The first half of this figure shows the final 30 seconds of the WAV file recording of the soft fan environment (the top row of Table 1). The second half shows the first 30 seconds of the two-minute recording of the quiet conversation WAV file with a single talker (second row of Table 1). These simple listening environments demonstrate how the classifier generates probabilities that almost exclusively represent a single acoustic 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. The first 30 seconds is a 100% probability (1 on the Class Probability axis) that this is a quiet listening environment. Given that it is a recording of a soft fan measured at only 40 dB SPL in a sound treated room, that classification is correct. The hearing instrument would spend these 30 seconds in the quiet listening environment of SoundNav.

At 30 seconds, the recording abruptly switches from the soft fan at 40 dB SPL to a single talker at 55 dB SPL. 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

**Figure 3**



one another at approximately 35 seconds. In this transition zone, SoundNav switches the hearing instrument from the quiet listening environment to the speech in quiet listening environment. The classifier actually detects the change almost immediately, but our developers made a conscious decision not to have the device react too quickly to every little change in the acoustic environment. Rapid changes could lead to reduced sound quality in dynamic listening environments as SoundNav attempts to keep up with all of the environmental fluctuations.

By 40 seconds and for the last 20 seconds of the recording, the probability of a speech in quiet listening environment is almost 100%.

The two vertical bars on the left and right of the classifier proportions section show the proportion of time spent in each of the seven possible listening environments for the pair of two-minute WAV files. The red bar on the left is the full two minutes of the soft fan WAV file, and the red and blue bar on the right show the proportion of time spent in each of the seven listening environments during the two minutes of speech in quiet WAV file. The slight red section represents the transition time at the beginning of the speech in quiet recording.

Figure 4 is an example of what happens in a more complex listening environment.

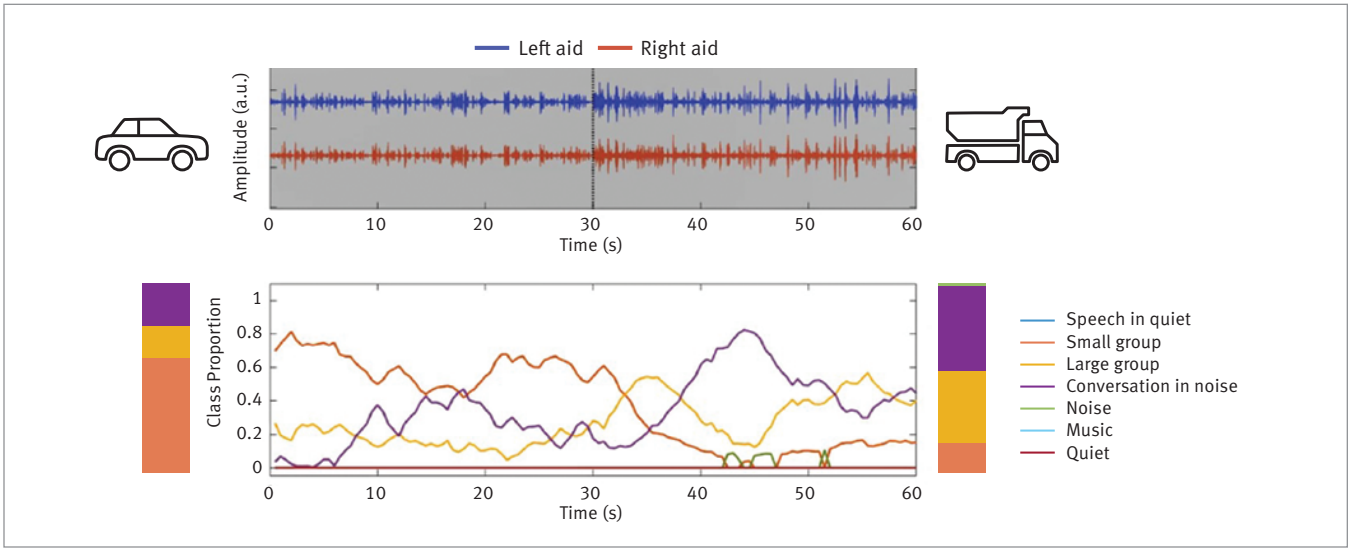
Here we can see the impact on the probabilities of two much more complex listening environments. In both cases, the listener is driving in the car along with three talkers. On the left side (the first 30 seconds) the car is much quieter with an overall level of about 70 dB and a -10 dB SNR (Signal to Noise Ratio). The overall levels are much more difficult in the second 30 seconds at an overall signal level of 80 dB with -15 dB SNR. These levels may look like nearly impossible SNRs for a hearing instrument wearer, but the car noise is distinctive in that almost all of the energy is in the very low frequencies (below 1000 Hz). As such, the SNRs look extreme, but almost all of the high frequency speech is clearly audible in both WAV files.

As the car changes speeds and the talkers start and stop, the classifier probabilities vary widely across a blend of three different listening environments. During the softer first 30 seconds, the highest probability is that of conversation in a small group, averaging 50% to 60%. As you 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 sound of the car noise becomes predominant. As the car speeds up, the classifier probabilities shift hard into the conversation in noise environment and 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 instruments from many manufacturers, it is clear that all of them would react similarly in both listening environments shown in Figure 3.

But what about the two environments in Figure 4? This is where philosophy plays a role. There is a lot going on in these listening environments and developers have to make some decisions about what to do. For exam-

**Figure 4**



ple, 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 overall level or SNR? The sound park-our is designed to look at all of those possibilities to tease out what relevant choices have been made.

# The gold standard

Table 1 lists sound files that represent several general listening environments that a hearing instrument wearer might encounter in real life. How did we know the files accurately represented the designated listening environments? We had 17 normal hearing listeners who defined for us what listening environments they thought were best represented by each sound file. Multiple answers were acceptable. The sound files were played back in randomized order for our listeners. They heard each sound file three times, and they described the environment for each iteration of every sound file. We then pooled all of their answers to compare to the hearing instrument classifiers.

In Figure 5 we see how the descriptions of our human listeners compared to the seven listening environments in our classifier:

**Figure 5**

| Young normals | Classifier |
|---|---|
| Quiet | Quiet |
| Speech in quiet | Quiet speech<br>Small group |
| Speech in noise | Large group<br>Speech in noise |
| Noise | Noise |
| Speech in music | |
| Music | Music |

Although there was some overlap in specific terminology, there were interesting differences in the interpretation of what those names meant. There were three names 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 our listeners and rarely exceeded 3% for any listening environment. For example, the fan sound file at the top of Table 1 was given a 100% probability of “quiet” by our classifier since the overall level was a mere 40 dB SPL, but our listeners called it “noise” 92% of the time. Interestingly, our listeners only gave us a probability of “noise” above 27% in just two other listening environments, both of which were quite loud. The really noisy sound files all contained speech and were therefore given the highest probabilities of “speech in noise” by our listeners. The same was true for the classifier, except it made a distinction on the basis of the type of noise, either multiple background talkers or engine noise such as trains, cars or traffic. Neither the listeners nor the classifier detected “music” very often, and only when it was much louder than everything else around it. But the listeners did offer a distinct category of “speech in music” mixed with “speech in noise” in seven environments where the classifier detected a “large group” (which they were, but the classifier ignored the music in favor of optimizing speech).

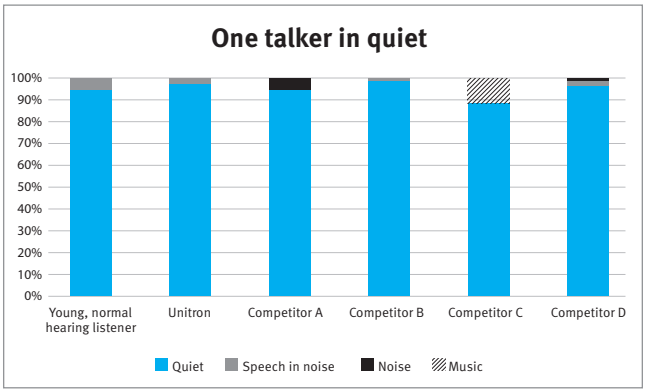
The main distinctions between the listeners and the classifier were not so much that they were detecting different things, but that they were prioritizing different aspects of the sound files or making slightly more precise distinctions in some cases. For example, one could easily argue that a soft fan 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 listening environments versus our young normal hearing listeners. This exercise isn't about who is right or who is wrong – rather, it's an opportunity to see how different classifiers compare. The results showed some hearing instruments are better at classification than others, and the different philosophies across companies tend to reveal themselves.

Let's start again with a simple example. Figure 6 shows how the young normal listeners and the five hearing instruments classified a single male talker from the front at 55 dB SPL.

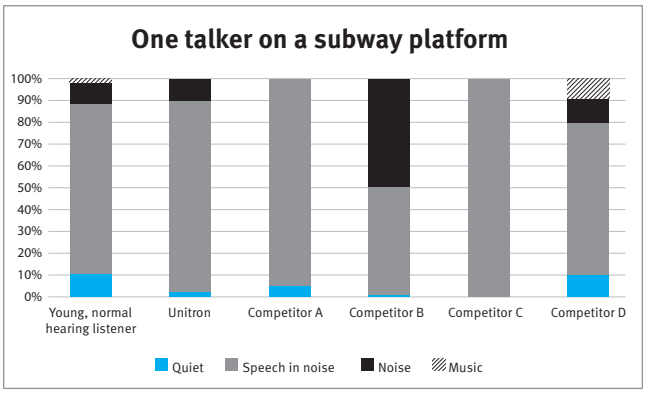
**Figure 6**




Different manufacturers have different classification schemes that use different names for the listening environments they classify. Using their descriptions of what each listening destination was intended for, we grouped the titles into four main categories: quiet, speech in noise, noise and music (as shown in the legend of Figure 6). These four general categories appear in all of the hearing instruments we tested under one name or another, but we used the generic names in our results to maintain the anonymity of the manufacturers and hearing instruments involved. Our normal listeners classified this sound file as quiet listening about 98% of the time. All five hearing instruments did the same.

Figure 7 is a bit more complex than Figure 6. There is once again a single talker directly in front of the listener, but the overall level of the sound file is now 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 7**

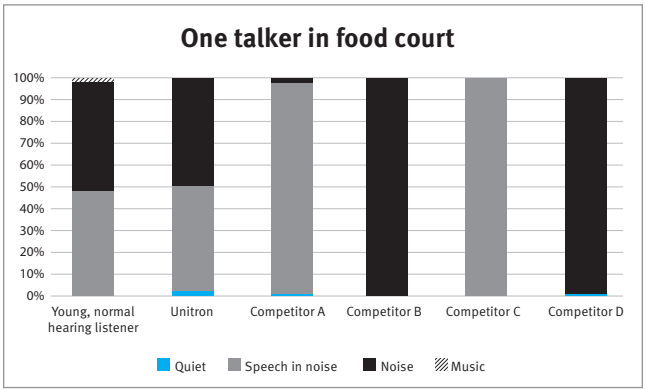



Our normal listeners classified this file as speech in noise about 83% of the time. They also said it was noise 4% of the time and quiet 10% of the time. Bearing in mind the level differences as trains came and went, it is fair to say that Unitron and Competitor D were closest to what the young normal hearing listeners told us. Competitor A was not far behind, however, Competitors B & C were very different.

This is where the differences in philosophy are first exposed. If we look at Competitor B, that instrument classified the environment as just noise about 50% of the time. It is clear that our normal listeners are reporting speech in noise fairly consistently. Therefore, the SNR must be reasonable most of the time. However, at 80 dB the overall level is quite high. Thus, we are inferring that Competitor B has a philosophy that is more sensitive to overall level than to SNR in this case, like the other four hearing instruments tested.

The background becomes even more complex in Figure 8. Here the listeners were evaluating a single talker from the front in a background of a food court at the mall near lunch time. The overall level was a bit lower at 70 dB at a 0 dB SNR. This is a complex background of dozens of people carrying on many conversations at once as well as the sound of the kitchens serving food and people walking by.

**Figure 8**



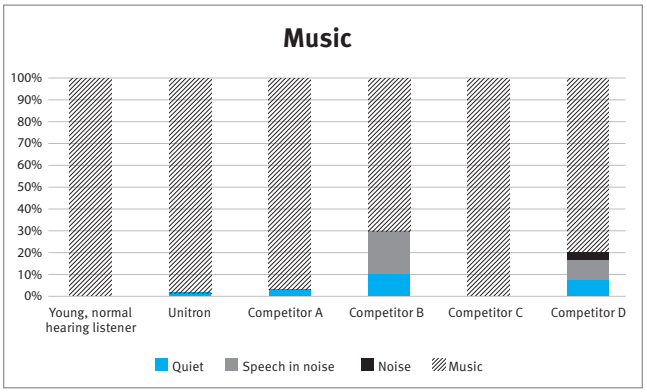

In this case, our normal hearing listeners report about 47% speech in noise and about 50% noise only. The other 3% was music. This time, the classifier results vary widely across manufacturers. While all classifiers offered some combination of speech in noise and noise, the percentages for Competitors A & C were completely the opposite of those for Competitors B & D.

This 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 instrument wearer "gives up" trying to follow the speech because the situation has become too difficult. Below the give up point, the listener will work to follow what is being said and report it as a speech in noise environment, expecting the hearing instrument to emphasize speech clarity. But once the give up point is crossed, the listener reports that it is too difficult to follow the speech or too loud to listen comfortably, and they would like the hearing instrument to emphasize comfort over clarity. Each classifier is built to make that decision at some point, and it's a purely acoustically driven decision. (Unless the listener switches to a manual program to override it.)

The first take away from Figure 8 is that Competitors A & C assume a higher give up point than Competitors B & D. Both Unitron and the normal listeners have indicated that this environment is pretty well right on the give up point line with a near 50/50 split between speech in noise and noise. This is perhaps the most striking example of philosophy impacting performance. Given that the give up point for different hearing-impaired people often varies widely, who is to say which of these companies will get it absolutely right for a particular listener?

The final example is for listening to music. In Figure 9, we see the results for music being played alone (with no other background sounds) at a level of 65 dB. This is not a high level for listening to music and doesn't replicate a live performance. Rather, it's closer to the level at which a hearing instrument wearer may listen to music while cooking or reading a book, but a bit louder than background music.

**Figure 9**




In this instance, the normal listeners, Unitron, Competitor A and Competitor C all indicated that this was essentially a pure music listening environment. Competitors B & D classified it differently at least 33% and 20% of the time respectively. The most common misclassification on this one was for speech in noise, and this is the one case where a clear and indefensible miss took place. Mistaking music for speech in noise is tantamount to setting up a hearing instrument for exactly the opposite type of performance you would prefer. It is generally accepted practice to set a music environment for broadband lightly processed reproduction. But speech in noise usually receives a heavy dose of directional microphones and noise cancelling designed among other things to reduce low frequency amplification. The music in this sound file was presented from 90 degrees azimuth and would have been effected by a directional microphone. To be fair, such a miss was not common for the five classifiers.

# Summary

Hearing instrument sound scene classification is a topic that gets precious little attention. Yet, it is one of the most important components of the instrument's architecture. Quietly 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, and heavily impact how a wearer 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 situations, almost all of the top hearing instruments 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 performance become obvious.

With SoundNav, a classifier trained using artificial intelligence, Unitron's results are highly consistent with those of our young normal hearing listeners.

# 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 lab at the University of South Florida.

# References

Büchler, M., Allegro, S., Launer, S., & Dillier, S. (2005). Sound classification in hearing aids inspired by auditory scene analysis. EURASIP Journal on Applied Signal Processing, 18, 2991–3002.

Kates, J. M. (1995). Classification of background noises for hearing-aid applications. J Acoust Soc Am, 97(1), 461-470.

Lamarche, L., Giguere, C., Gueaieb, W., Aboulnasr, T., & Othman, H. (2010). Adaptive environment classification system for hearing aids. J Acoust Soc Am, 127(5), 3124-3135. doi:10.1121/1.3365301

Nordqvist, P., & Leijon, A. (2004). An efficient robust sound classification algorithm for hearing aids. J Acoust Soc Am, 115(6), 3033-3041.

Unitron is a hearing solution company that believes people should feel really good about the entire hearing care experience, start to finish. Our ingenious products, technologies, services and programs offer a level of personalization you can't get anywhere else. Get ready to Love the experience.



1904-093-02

**unitron.com**

sonova
HEAR THE WORLD