Sound Scene and Spatial Speech Classification in Hearing Aids

Donald Hayes, Ph.D.
Director, Clinical Research

Do you ever think about what a modern hearing aid really is?
It is a system of interrelated parts working together in support of a philosophy.

**Two parts you may not think much about**

Benchmarking results on two of the components:

- Classification
  - Studies which highlight how it works and why it is important

- Spatial Detection
  - Studies that evaluate the benefits of spatial detection algorithms
  - What you can do when you know the direction of speech
Do you ever consider how automatic hearing aids classify acoustic environments?

We do!
Environmental Classification in Hearing Aids

**Philosophy:** The developers’ choice of “detectors” and the decision rules weighted by design philosophy

**Consequences:** Engaging, disengaging, and parameter value selection for signal processing features

**Importance:** The best possible signal processing is only as good as the accuracy of the classifier and quality of the signal processing decisions driven by the classifier results.

Why is this clinically relevant? It comes back to “Log It All”
The Global Listening Environment Study (GLES) took advantage of the Log It All feature to acoustically map the listening environments experienced by hearing aid wearers in 10 countries.

Clinics from ten countries participated

<table>
<thead>
<tr>
<th>Country</th>
<th>Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>134</td>
</tr>
<tr>
<td>Canada</td>
<td>87</td>
</tr>
<tr>
<td>France</td>
<td>78</td>
</tr>
<tr>
<td>Germany</td>
<td>87</td>
</tr>
<tr>
<td>Netherlands</td>
<td>141</td>
</tr>
<tr>
<td>New Zealand</td>
<td>140</td>
</tr>
<tr>
<td>Slovenia</td>
<td>58</td>
</tr>
<tr>
<td>South Africa</td>
<td>55</td>
</tr>
<tr>
<td>Spain</td>
<td>67</td>
</tr>
<tr>
<td>USA</td>
<td>208</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1055</strong></td>
</tr>
</tbody>
</table>
Greater variability within groups of people of the same age than there is across ages

Age is not a strong predictor of the time people spend in conversations
What we saw in the GLES

Average classifier results will be extremely similar for any large group of people regardless of:

- Age
- Country
- Gender
- Population density (rural, small town, big city)
- Hearing aid experience

Why is that?

Is it because:

1. On average everybody experiences roughly the same distribution of acoustic environments in their daily lives?
2. Our classifier sees the same distribution regardless of a person’s daily life?

Individual results would indicate it is the first one. But we must also rule out the second one.

We do that through a Benchmarking study.
Benchmarking Studies - USF

Multiyear project with the University of South Florida

- Principle Investigators
  - David Eddins, Ph.D.
  - Erol Ozmeral, Ph.D.
- Together with Sonova they have studied
  - Classifier Performance
  - Azimuth Detection Behavior
  - Take home comparison to leading competitor

Classifier performance

First we designed the sound parkour to simulate real world listening

Then we had 16 normal hearing listeners “classify” each sound file

Finally we had our hearing aids “classify” the same sound files
The sound parkour
(a walk through the acoustic park)

Defining a listening environment
<table>
<thead>
<tr>
<th>Condition</th>
<th>Short ID</th>
<th>Talkers</th>
<th>Background Noise</th>
<th>Talker Distribution°</th>
<th>Noise Distribution°</th>
<th>SNR</th>
<th>Overall Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiet Listening</td>
<td>Q-Fan</td>
<td>0</td>
<td>Fan noise</td>
<td>N/A</td>
<td>N/A</td>
<td>75 dB</td>
<td></td>
</tr>
<tr>
<td>Quiet Conversation</td>
<td>Q-1T</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>75 dB</td>
<td></td>
</tr>
<tr>
<td>Quiet Conversation with Music</td>
<td>Q-3T-M</td>
<td>1</td>
<td>Music</td>
<td>0°, 300°, 60°</td>
<td>N/A</td>
<td>75 dB</td>
<td></td>
</tr>
<tr>
<td>Small Group Conversation with noise</td>
<td>C-Tube1</td>
<td>2</td>
<td>Subway</td>
<td>270°, 0°, 90°</td>
<td>0°, 90°, 180°, 270°</td>
<td>10, -15, 70 dB</td>
<td></td>
</tr>
<tr>
<td>Small Group Conversation with Noise/Music</td>
<td>C-Tra1</td>
<td>3</td>
<td>Traffic</td>
<td>270°, 0°, 90°</td>
<td>0°, 90°, 180°, 270°</td>
<td>10, -10, 70 dB</td>
<td></td>
</tr>
<tr>
<td>Large Group Conversation</td>
<td>LG1</td>
<td>6T-Babble</td>
<td>0°</td>
<td>315°, 45°, 135°</td>
<td>5°</td>
<td>65 dB</td>
<td></td>
</tr>
<tr>
<td>Large Group Conversation with Music</td>
<td>LG2</td>
<td>1</td>
<td>10T-Babble</td>
<td>0°, 315°, 45°, 135°, 225°, 180°</td>
<td>5°</td>
<td>75 dB</td>
<td></td>
</tr>
<tr>
<td>TV</td>
<td>0</td>
<td>TV (CSI S01E01)</td>
<td>N/A</td>
<td>0°</td>
<td>N/A</td>
<td>70 dB</td>
<td></td>
</tr>
</tbody>
</table>
Transition Case 1
Quiet (Q-Fan) → Speech in Quiet (Q-1T)

Transition Case 3
Small Group with Music (Q-3T-M) → Small Group in Subway (3T-Tube)
Transition Case 4
Two Talkers in Traffic (2T-Traf) $\rightarrow$ Three Talkers in Car (3T-Car)

Mean Proportions

Transition Case 5
3 Talkers in Car-Avg (3T-Car-A) $\rightarrow$ Three Talkers in Car-Loud (3T-Car-L)

Mean Proportions
That is Classification under the microscope

Classifier Benchmarking

You can benchmark against either:

- A gold standard
- Similar devices

Gold Standard = 16 normal hearing listeners
Similar devices = 4 top competitor’s Premium hearing aids
Classifier performance

First we designed the sound parkour to simulate real world listening.

Then we had 16 normal hearing listeners “classify” each sound file.

Finally we had our hearing aids “classify” the same sound files.

Hearing aids versus human classification.
Basic classification

Soft Speech

Loud Speech

Complex classification

Difficult SNR

Favorable SNR
What does the benchmarking tell us?

Comparisons to normal listeners and competitors show:

1. In simple acoustic environments all manufacturers do a pretty good job of classification
2. The differences in philosophy appear when the environment becomes more complex
   a) Some are sensitive to overall level
   b) Some to SNR
   c) Some lock in on speech
3. Log It All results are valid because our classification correlates well with normal listeners, even in complex (blended) environments.

Now let us look at some real world data
Ecological Study

Ecological momentary assessment (EMA) involves repeated sampling of subjects' current behaviors and experiences in real time, in subjects' natural environments.

EMA aims to minimize recall bias, maximize ecological validity, and allow study of microprocesses that influence behavior in real-world contexts.


Ecological Study

1. EMA = Ecological Momentary Assessment

2. Joint effort between: Unitron/UWO under the ORF/MITACs
   • Unitron - Leonard Cornelisse
   • NCA - Susan Scollie, Danielle Glista
   • UWO Geography – Jason Gilliland, Tayyab Shah

3. They generated way more data than we have time for today
Environmental Classification in Hearing Aids

Method of evaluation:

- EMA study with UWO
- NCA/Geography Dept.
- 6 intrepid grad students took to the streets of London over 54 hours in 1 sec intervals.

- Wearing only:
  - Laptop
  - HA’s
  - NOAHLink
  - GPS
  - LENA
  - Cell Phones with Ratings

In other words:

Some basic measurements

1. Average over all level in the real world (55 hours of it anyway)

2. Average Signal to Noise Ratio

3. Left and right ear differences

In case you ever wondered
Distribution of overall levels experienced

79% of samples

1 second samples over 55 hours

SIGNAL LEVEL X RANGE (DB)

Average ear “level” difference

94% <= 3 dB Difference

SIGNAL LEVEL X DIFFERENCE (DB)
Distribution of SNRs x Level

Accounts for 96.6% of samples

Distribution of SNRs x percentage of time

SNR RANGES (DB)

SNR RANGES (DB)
What we learned about the classifier

1. Some useful but not surprising information

2. Some further information to consider for the future

Log-It-All: Primary Classes (52%)

Percentage of time class proportion exceeded 75%
Log-It-All: Primary (52%) & Mixed Classes (48%)

What does the EMA study tell us?

These results are based on one day (9 hours) of recordings from each of 6 people

1. Our participants were in acoustic environments that were clearly defined as primary classes based on our system about half the time.
2. The rest of the time the acoustic environments were complex enough to be classified as mixed (blended) classes.
3. Those proportions will vary depending on the philosophy and priorities set by each manufacturer.
How about Spatial Detection?

Two parts you may not think much about

Benchmarking results on two of the components

✓ Classifier
  ✓ Studies which highlight how it works and why it is important

• Spatial Detection
  • Studies that evaluate the benefits of spatial detection algorithms
  • What you can do when you know the direction of speech

unitron.
Benchmarking Studies - USF

Multiyear project with the University of South Florida

- Principle Investigators
  - David Eddins, Ph.D.
  - Erol Ozmeral, Ph.D.
- Together with Sonova they have studied
  - Classifier Performance
  - Azimuth Detection Behavior
  - Take home comparison to leading competitor

Azimuth detection in two parts (n=19)

Locating the direction of speech

- Presented short snips of speech and asked the listener to indicate its direction
  - Speech came from one of 8 directions
  - One snip every 4 seconds for 16 seconds

- Speech perception from different directions
- CNIT
  - Adaptive Speech in Noise test
  - Asked to repeat numbers from different directions and at different SNR’s
Locating the direction of speech

0 dB SNR

-6 dB SNR
Effect of amplification and directional enhancements on localization

Main effects of device ($p = 0.05$), snr ($p < .001$), and location ($p < .001$). No significant interactions.

Benefit relative to Unaided condition

Effect of device and location reaching significance ($p = 0.07$, $p = 0.08$, respectively). No other main effects or interactions.
Azimuth detection in two parts ($n=19$)

Locating the direction of speech

• Presented short snips of speech and asked the listener to indicate its direction
  • Speech came from one of 8 directions
  • One snip every 4 seconds for 16 seconds

Speech perception from different directions

• CNIT
  • Adaptive Speech in Noise test
  • Asked to repeat numbers from different directions and at different SNR’s

CNIT

Noise is presented from four speakers: 0°, 90°, 180° & 270° azimuth simultaneously

Running speech is presented from one of three directions (0°, 90° or 180°) at a time.
  The hearing aids detect the direction of the speech
  This focuses aids in the desired direction before the numbers test begins

Speech and noise levels are at a fixed at a +6 dB SNR

The listener is asked to ignore the words repeat a 3 integer sequence presented from the same azimuth as the speech.
The level of the integers is varied adaptively when the listener gets them right or wrong.
Benefit of amplification and directional enhancements on speech ID in noise

• In each case, directional enhancements were shown to have an advantage relative to when they were off.

• Benefit, however, relative to unaided was only significant (p = 0.05) at the Front.

(n=19)  ***p<.001; **p<.01; *p=0.07

Unaided hearing thresholds seem to matter

Listeners arbitrarily split in half by pure-tone average PTA (500, 1000, 2000 Hz) threshold.

• High HL (n = 15; PTA >=35 dB HL)
• Low HL (n = 15; PTA <35 dB HL)

High HL group received very little aided benefit (left panel).
Low HL group received large aided benefit at the Front and Side, and little benefit at the Back.
What does this say about azimuth detection?

1. People with hearing loss frequently have degraded localization abilities and adding hearing aids rarely improves that.
2. The largest benefits we found for localization were from the sides and people did better with more signal processing than unaided.
3. Regarding speech perception we saw significant benefits from more processing (Speech Pro on/off) from all three directions.
4. From the front there was a significant benefit of more processing over the unaided condition.