Human Detection and Localization of Sounds in Complex Environments

W.M. Hartmann
Physics - Astronomy
Michigan State University

QRTV, UN/ECE/WP-29
Washington, DC
4 May, 2010
Goals: GRB-51-18, QRTV

- Phases II-IV. Determine human and technical factors relevant to improving the safety of pedestrians in the presence of quiet road vehicles.

  e.g.:
  - Detection
  - Localization

  These are psychoacoustical topics
Psychoacoustics Goals

• Mathematical models of hearing processes.
• Consistent with animal physiology.
• Consistent with human imaging and encephalography.
• Explain data from listening experiments in which listeners respond to sounds.
Experiment

People from off the street

Irreproducible results
Psychoacoustical Methods

- Controlled environment
- Focused attention
- Trained listeners
- Long experiments
- Specialized protocol.
Experimental protocol

Two-interval forced choice

One of these intervals has a signal in it. The listener knows that in advance.

Interval 1

Interval 2
Experimental Goals Met

• Reproducible data
• Find the best possible performance for the human organism.
• Successfully test models.

But

• Relevant to the blind pedestrian on the street corner?
Single interval/ No interval

- Maybe there’s a signal…. Maybe not.

<table>
<thead>
<tr>
<th>Signal present:</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIT</td>
<td>MISS</td>
<td></td>
</tr>
<tr>
<td>FALSE</td>
<td>CORRECT</td>
<td></td>
</tr>
<tr>
<td>ALARM</td>
<td>REJECTION</td>
<td></td>
</tr>
</tbody>
</table>
Hit Rate + Miss Rate = 100
False Alarm + Correct Rejection = 100
Slope about 1.0

Listener in unbiased experiment
Miss rate about 0.
Is psychoacoustics of any use at all in this context?

• P, “It cannot be heard.”
• Conclusion: It can’t be heard.

• P, “It’s easily heard.”
• Conclusion: Maybe it can be heard.

• There’s more:
Active cochlea
Test

Conversational Speech

Inaudible

Cal
Detection

- In typical environments, background noise means that detection thresholds are masked thresholds.
- Masking is frequency specific.
- Tonotopic organization starts with the cochlea.
Buick Lucerne, 1 meter, idle

15 dB
beeps
60 dBA
Buick Lucerne, 1 meter, 2000-1500 RPM

72 dBA
DC9, next to the intake

30 dB

hp
Sound Localization

- Azimuth
  -90 degrees to 90 degrees
- Elevation – including front back
  0 to 360 degrees
- Distance
Azimuth

ITD = Interaural time difference
ILD = Interaural level difference

$x_L$ $x_R$
ITD

\[ \theta = 41 \text{ deg} \Rightarrow 500 \text{ us} \]

\[ \Rightarrow ITD < 1200 \text{ Hz} \]
ILD

>1300 Hz

Arago-Poisson, 1818

Head shadow

Bright spot

N Left and Right

Right ear

Left ear
Source angle (degrees)

 ILD

Source number

ILD (dB)
Listener Responses
Amplitude modulation

- $f_c = 1500$ Hz, $f_m = 100$ Hz
- 100% modulation
AM
$f_m = 100$ Hz
100%

Listener N

Listener Responses

Source angle (degrees)

Response number

Source number

N AM 100%
Direct sound retains coherence
Max cross-correlation = 1.
Room reflections destroy coherence
Coherence $\gamma$

Bandwidth (BW) $(\tau - \tau_o)$

ITD $\tau_o$
Interaural coherence required to distinguish an ITD of +200µs from an ITD of −200µs for high-pass noise. Listeners W, X, and Z.

⇒ Need high coherence to use envelope cues.
PRECEDEENCE EFFECT

- Competition between Sound 1 & Sound 2
- Sound 1: Localization cues, ITD, ILD, etc.
- Sound 2: Localization cues, ITD, ILD, etc.

Sound 1 takes precedence.
Sound 1 takes precedence.

- Sound 1: Localization cues, ITD, ILD, etc.
- Sound 2: Localization cues, ITD, ILD, etc.

Fused image near location of Sound 1.
Elevation and front/back

- Anatomical filtering – cues at high frequencies > 8 kHz or > 2 kHz.
- Problem for elderly listeners.
- Front/back
  - Turn the head. Otherwise lost.
Distance

- Intensity of known sounds
- Air absorption attenuates high frequencies
- Direct-to-reverberant ratio
- Low-frequency ILD (< 1 meter).
Buick Lucerne, at grille, idle

66 dBA
Conclusions

- To be detectable, signals should have frequencies different from masking. Practically, this means mid to high.
- To be localizable (AZ), signals should have low frequencies. Both ITD and ILD work for the widest range of azimuths. Signals should be impulsive to elicit precedence.
- To disambiguate front/back signals should have high frequencies too.
The End

Thanks to the NIDCD
Listener does not use ear signals independently.
1500 Hz is too fast for ongoing ITD or IPD.