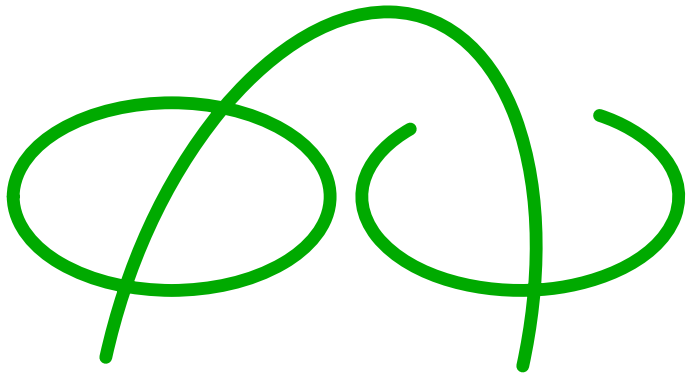
Human Detection and Localization of Sounds in Complex Environments

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.

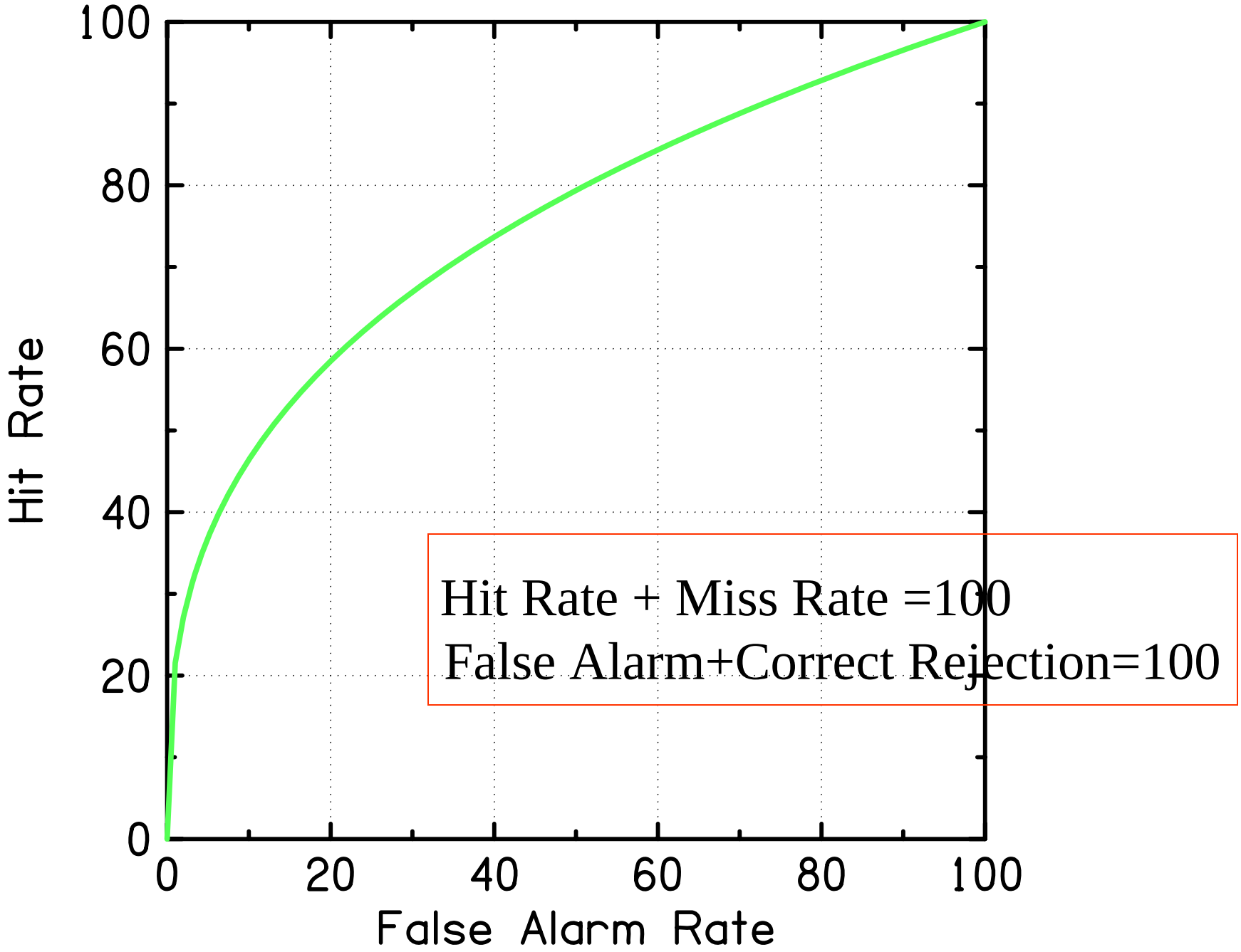
But

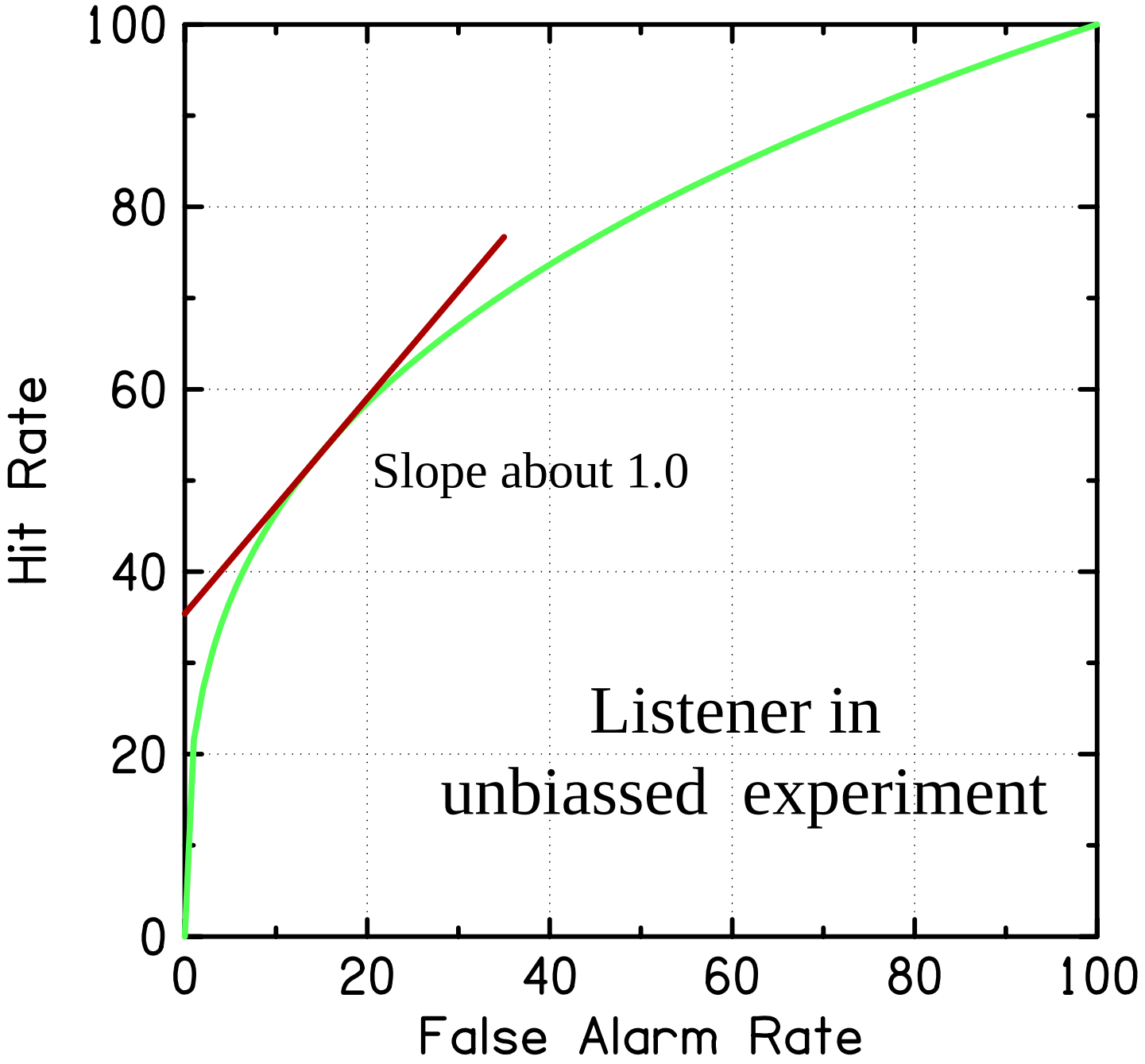
- Relevant to the blind pedestrian on the street corner?

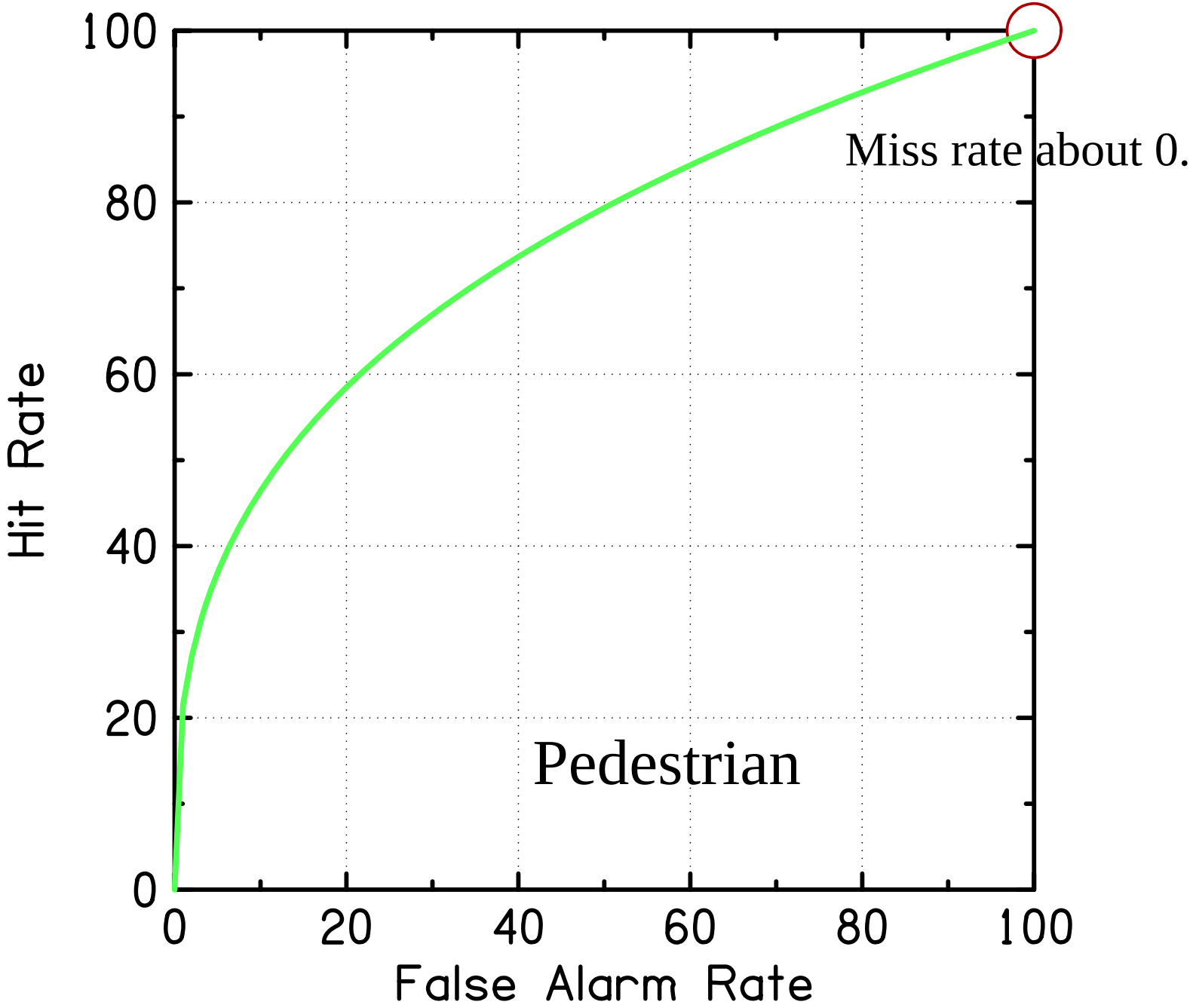
Single interval/ No interval

- Maybe there's a signal…. Maybe not.

	YES	NO
Signal present:	HIT	MISS
No signal:	FALSE ALARM	CORRECT REJECTION


100
80
60
40
20
0
Hit Rate
0
20
40
60
80
100
False Alarm Rate
Hit Rate + Miss Rate =100
False Alarm+Correct Rejection=100


100
80
60
40
20
0
Hit Rate
0
20
40
60
80
100
False Alarm Rate
Slope about 1.0
Listener in
unbiassed experiment


100
80
60
40
20
0
Hit Rate
0
20
40
60
80
100
False Alarm Rate
Miss rate about 0.
Pedestrian

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:

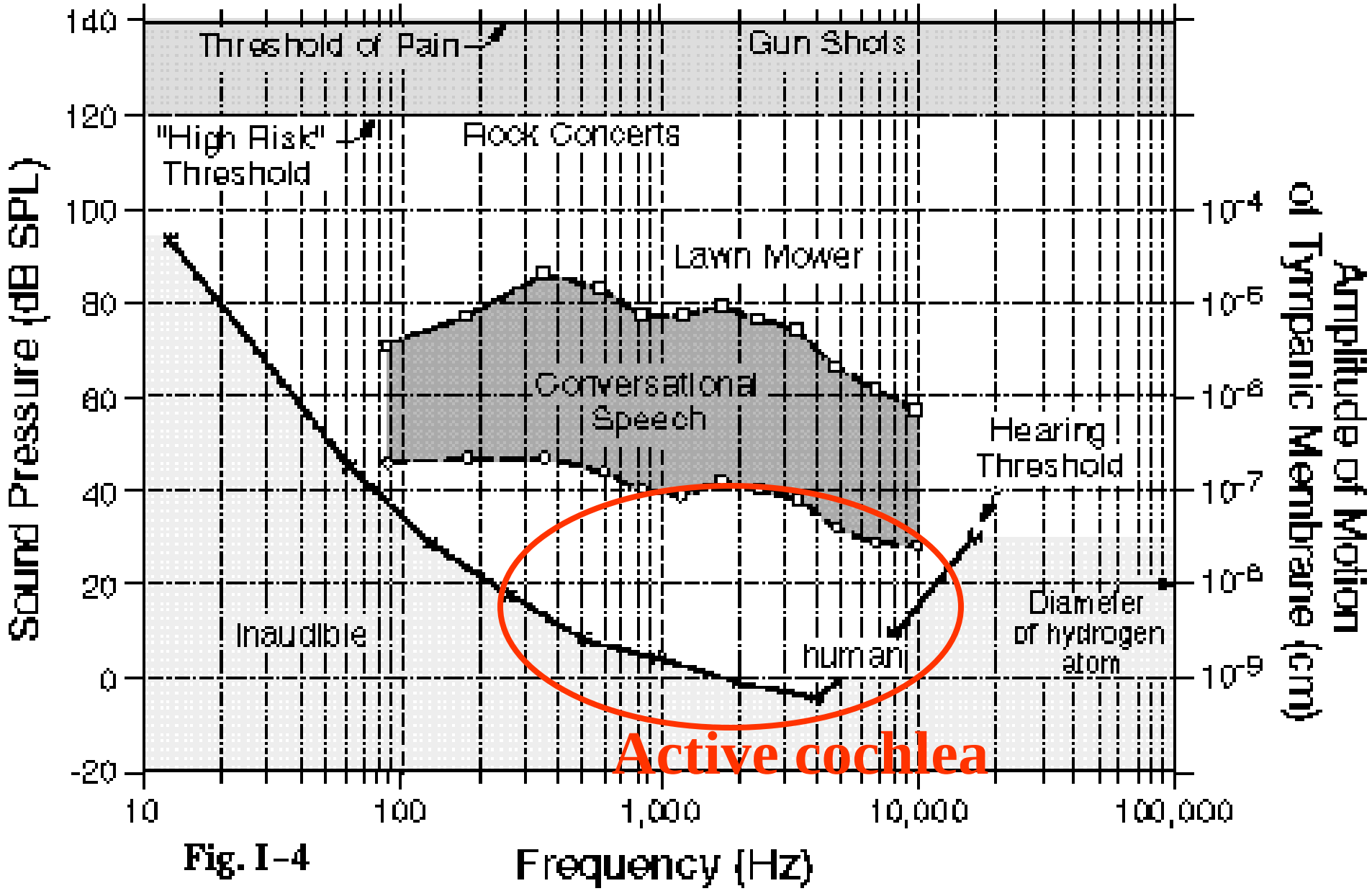
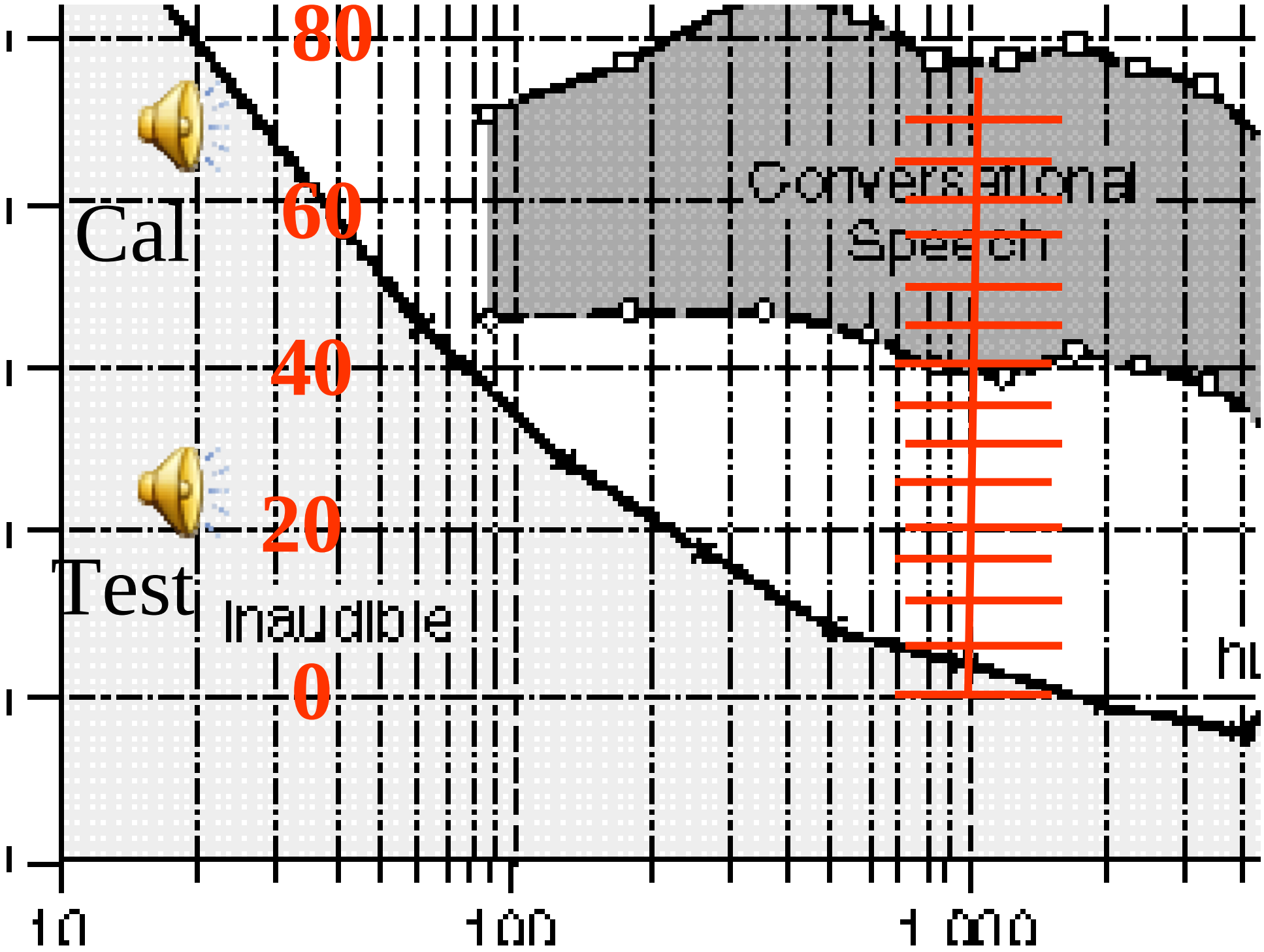
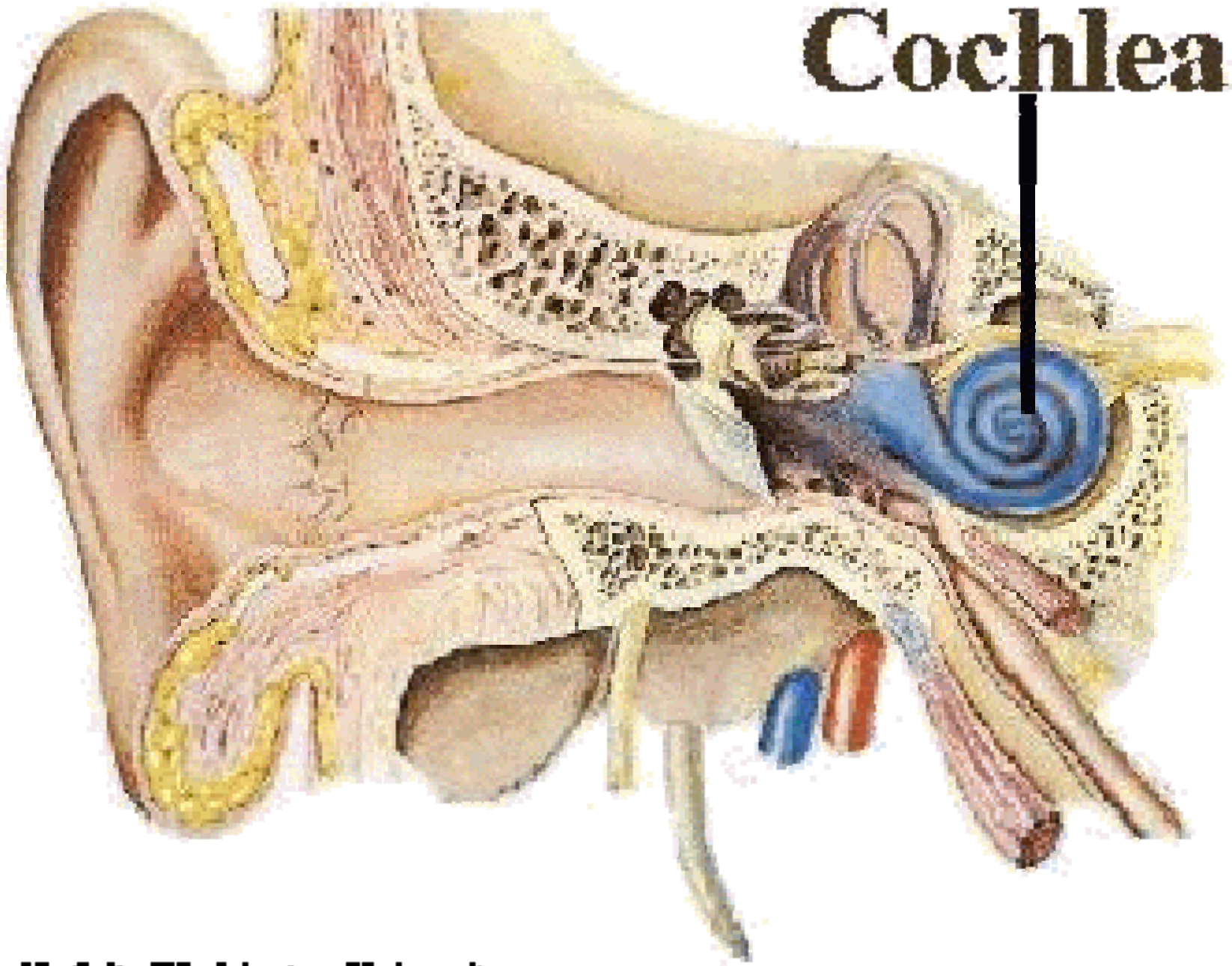

Threshold of Pain
Gun Shots
"High Risk" Threshold
Rock Concerts
Lawn Mower
Conversational Speech
Hearing Threshold
Inaudible
human
Diameter of hydrogen atom
Active cochlea
Sound Pressure (dB SPL)
Amplitude of Motion of Tympanic Membrane (cm)
Frequency (Hz)
140
120
100
80
60
40
20
0
-20
10
100
1,000
10,000
100,000
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}

Fig. I-4

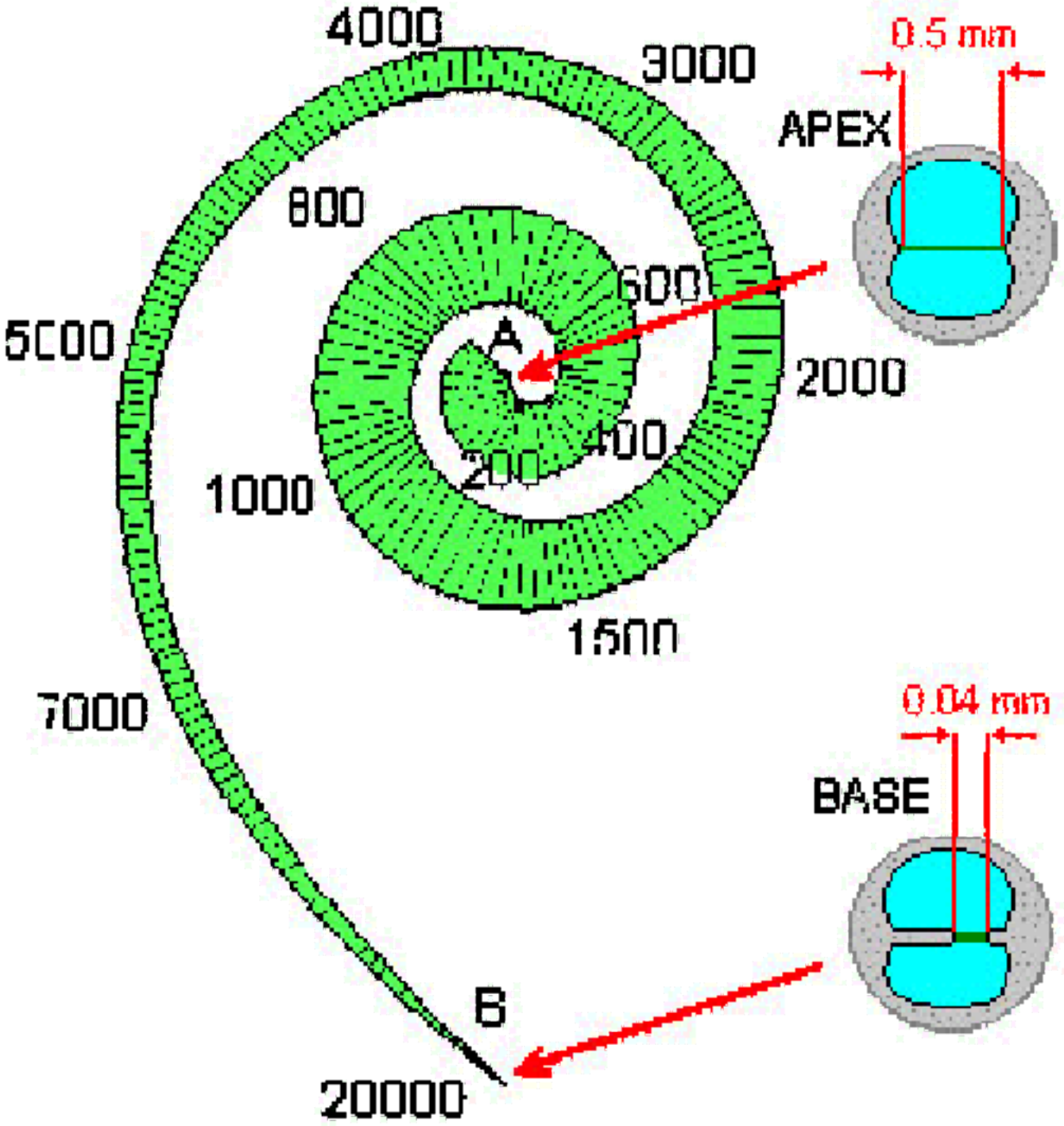

80
60
40
20
0
Cal
Test
Inaudible
Conversational
Speech
10
100
1 000

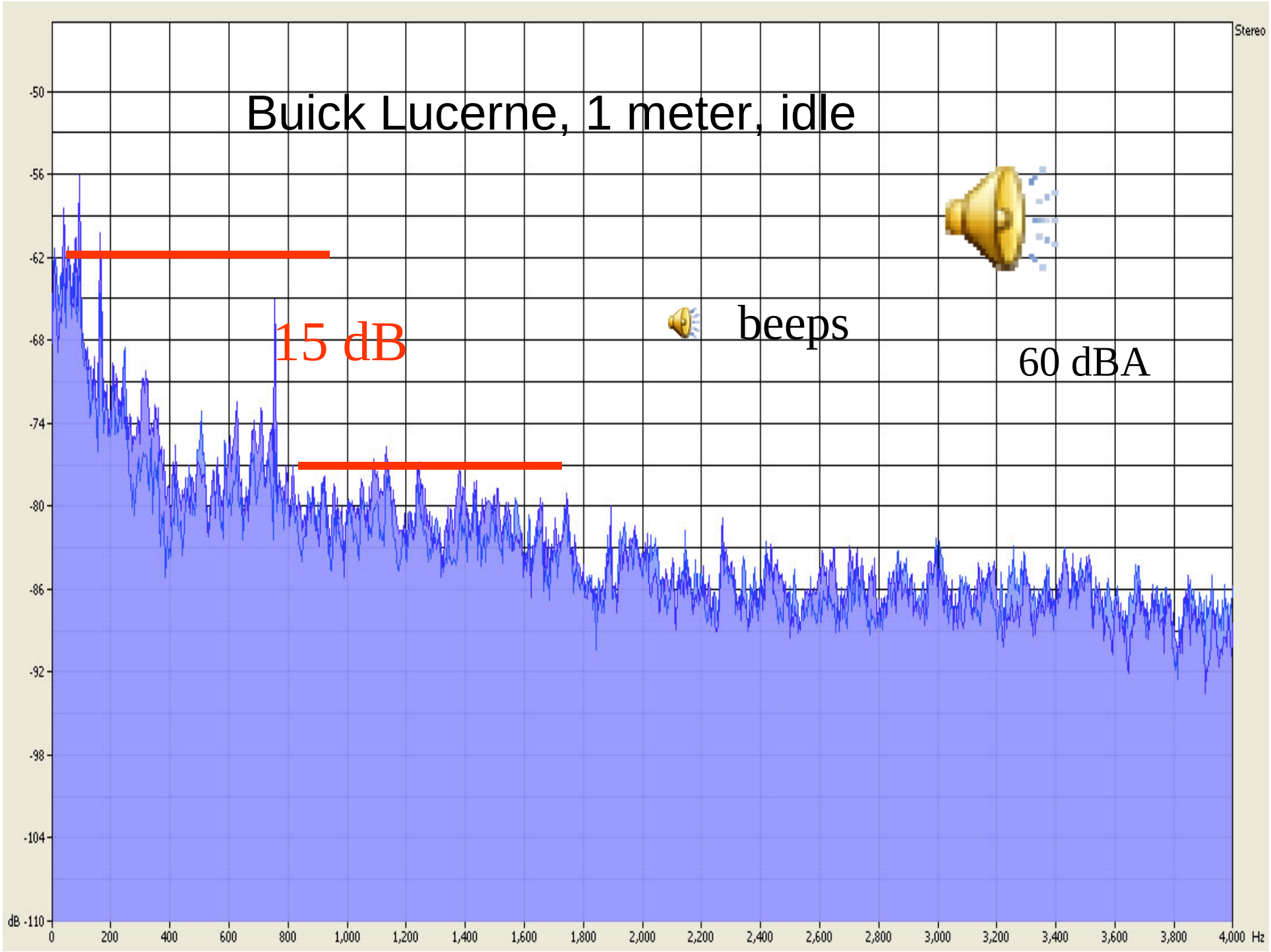
Detection

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

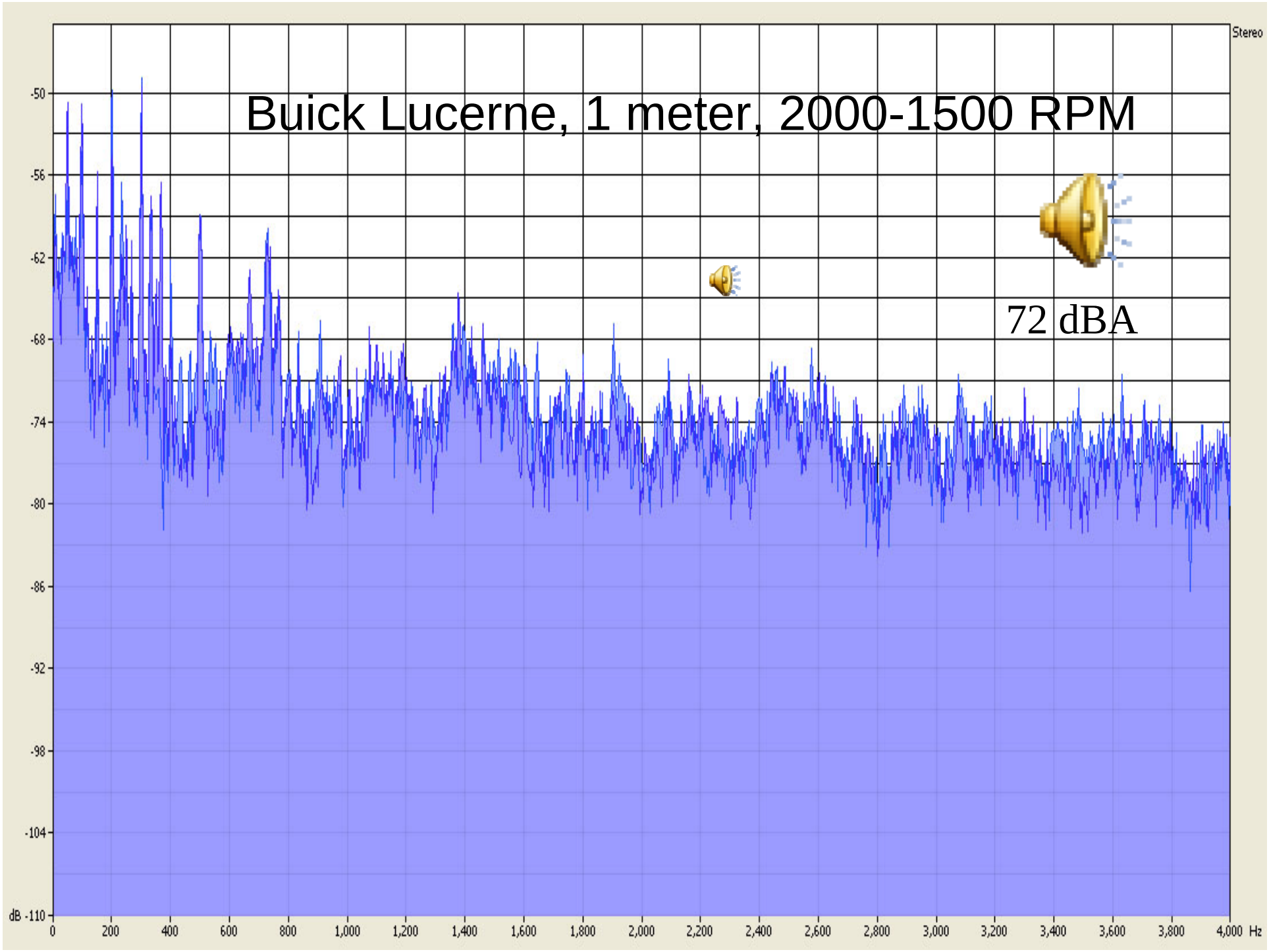


Alec N. Salt, Washington University

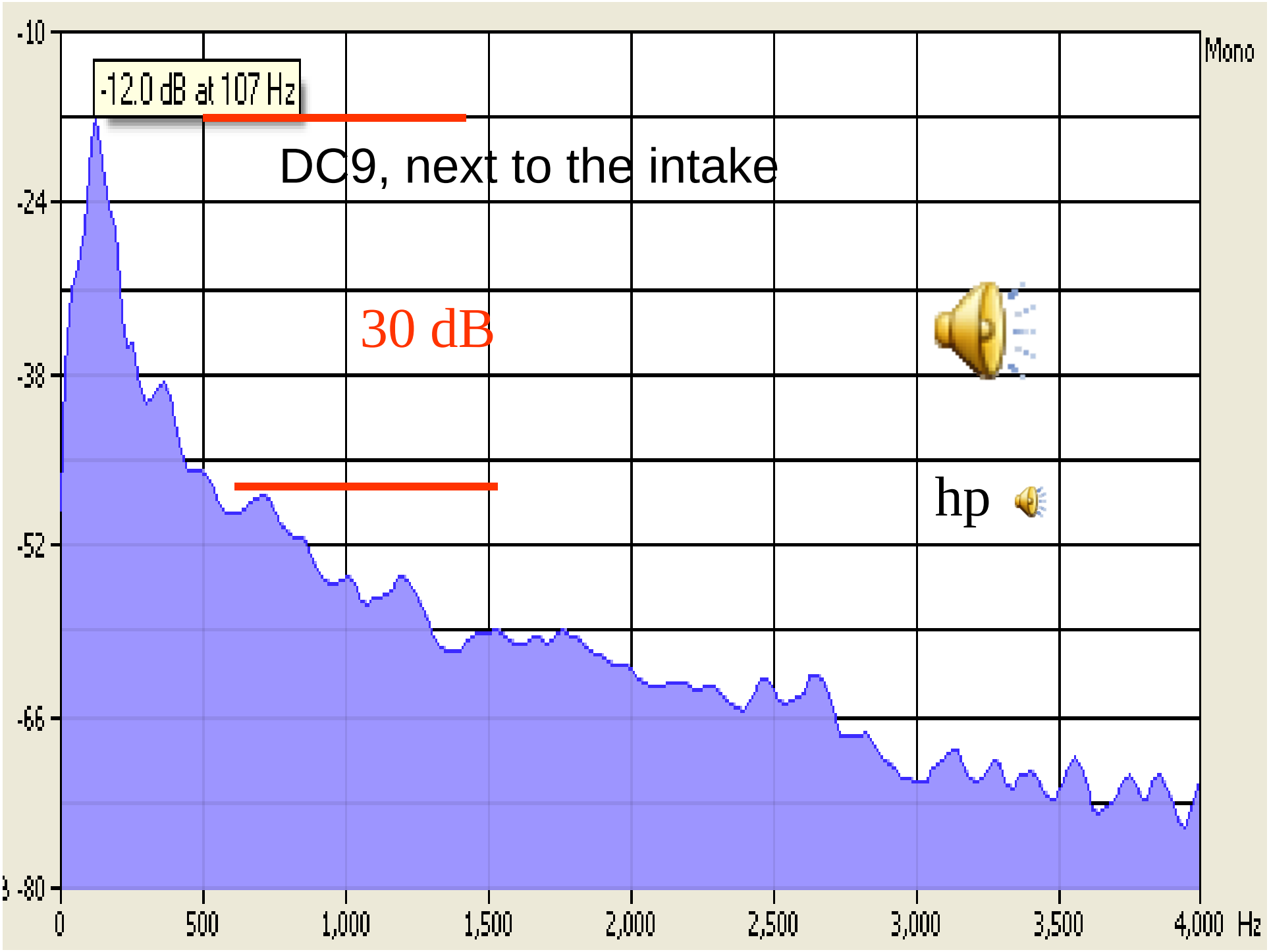

4000
3000
800
600
5000
2000
A
400
200
1000
1500
7000
B
20000
0.5 mm
APEX
0.04 mm
BASE


Buick Lucerne, 1 meter, idle
15 dB
beeps
60 dBA
Stereo
-50
-56
-62
-68
-74
-80
-86
-92
-98
-104
dB -110
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
2,200
2,400
2,600
2,800
3,000
3,200
3,400
3,600
3,800
4,000
Hz


100 dBA


Buick Lucerne, 1 meter, 2000-1500 RPM
72 dBA
Stereo
-50
-56
-62
-68
-74
-80
-86
-92
-98
-104
dB -110
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
2,200
2,400
2,600
2,800
3,000
3,200
3,400
3,600
3,800
4,000 Hz


DC-9
9

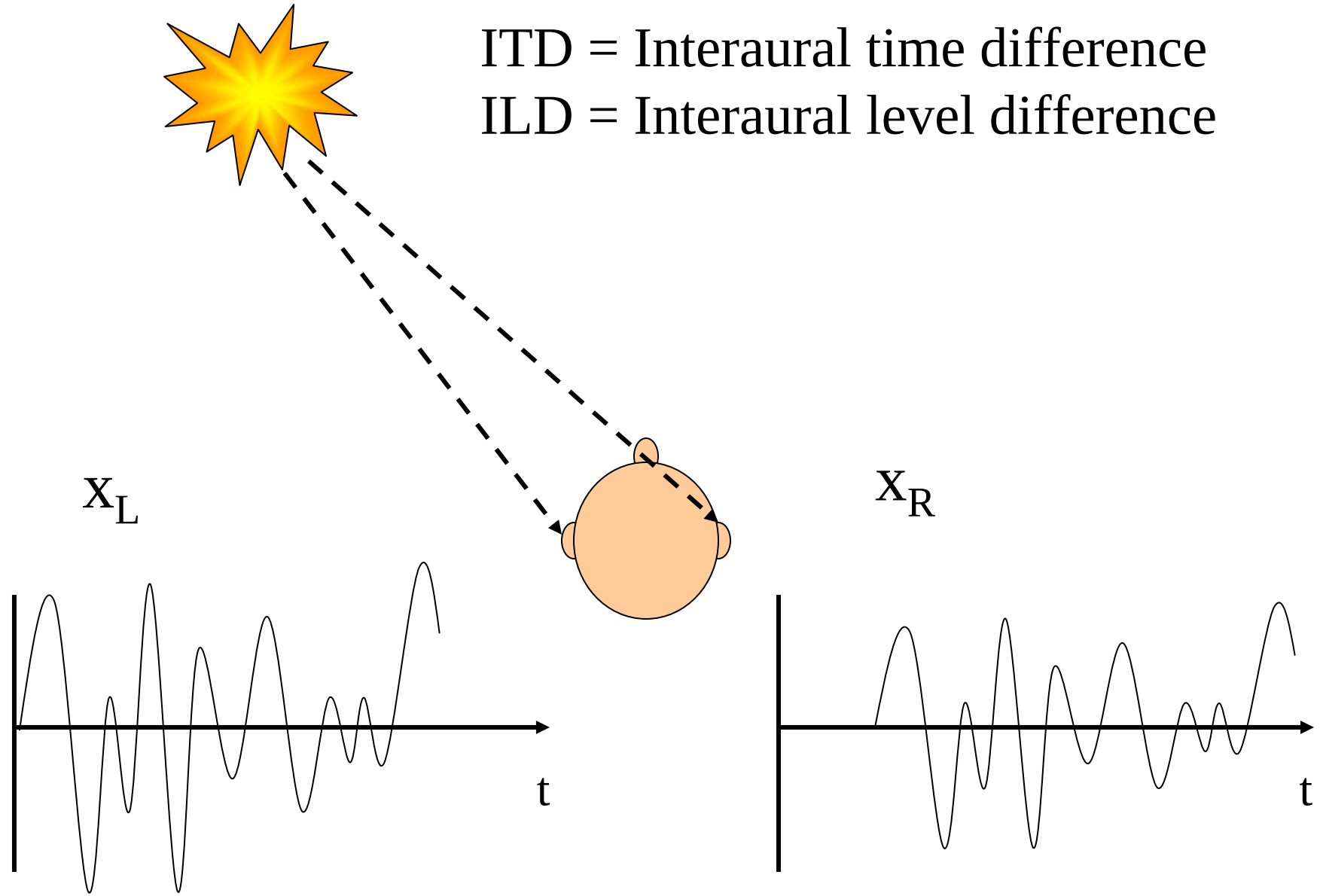

-10
-24
-38
-52
-66
-80
Mono
-12.0 dB at 107 Hz
DC9, next to the intake
30 dB
hp
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000 Hz

Sound Localization

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

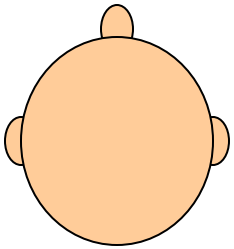
Azimuth

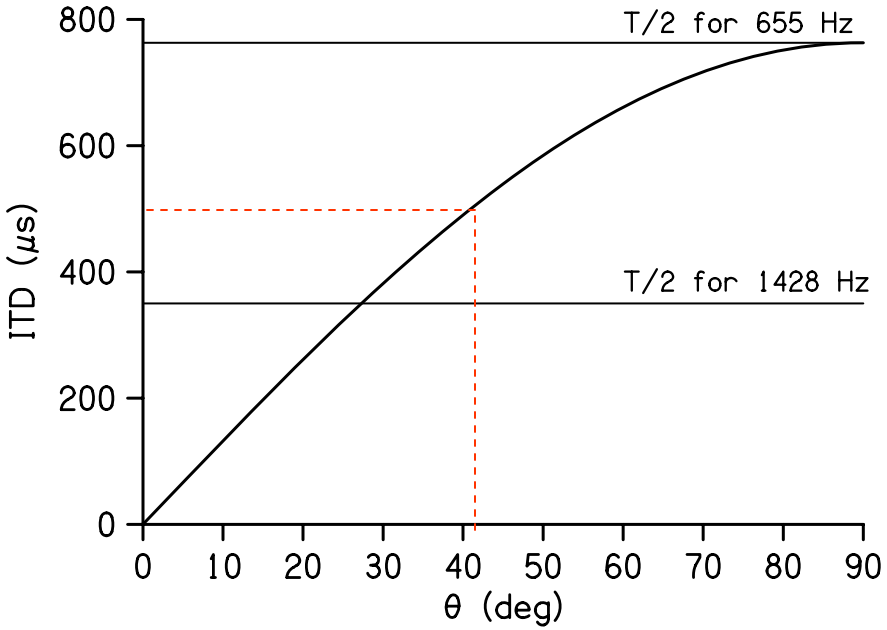
ITD = Interaural time difference
ILD = Interaural level difference

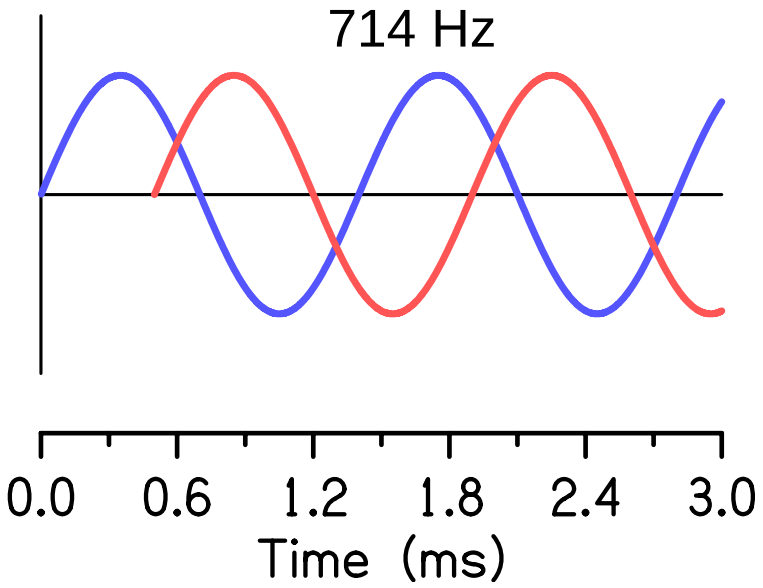


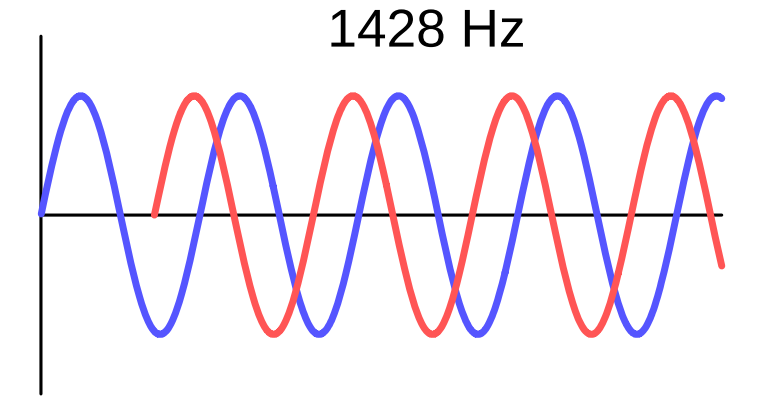
ITD

θ=41 deg => 500 us





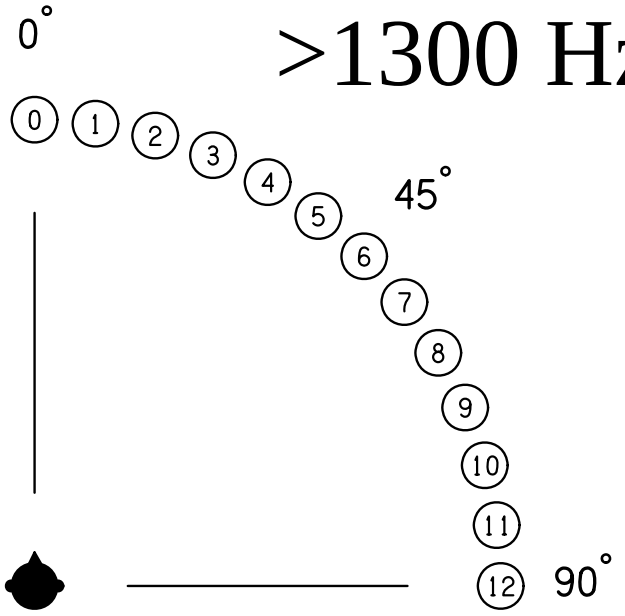


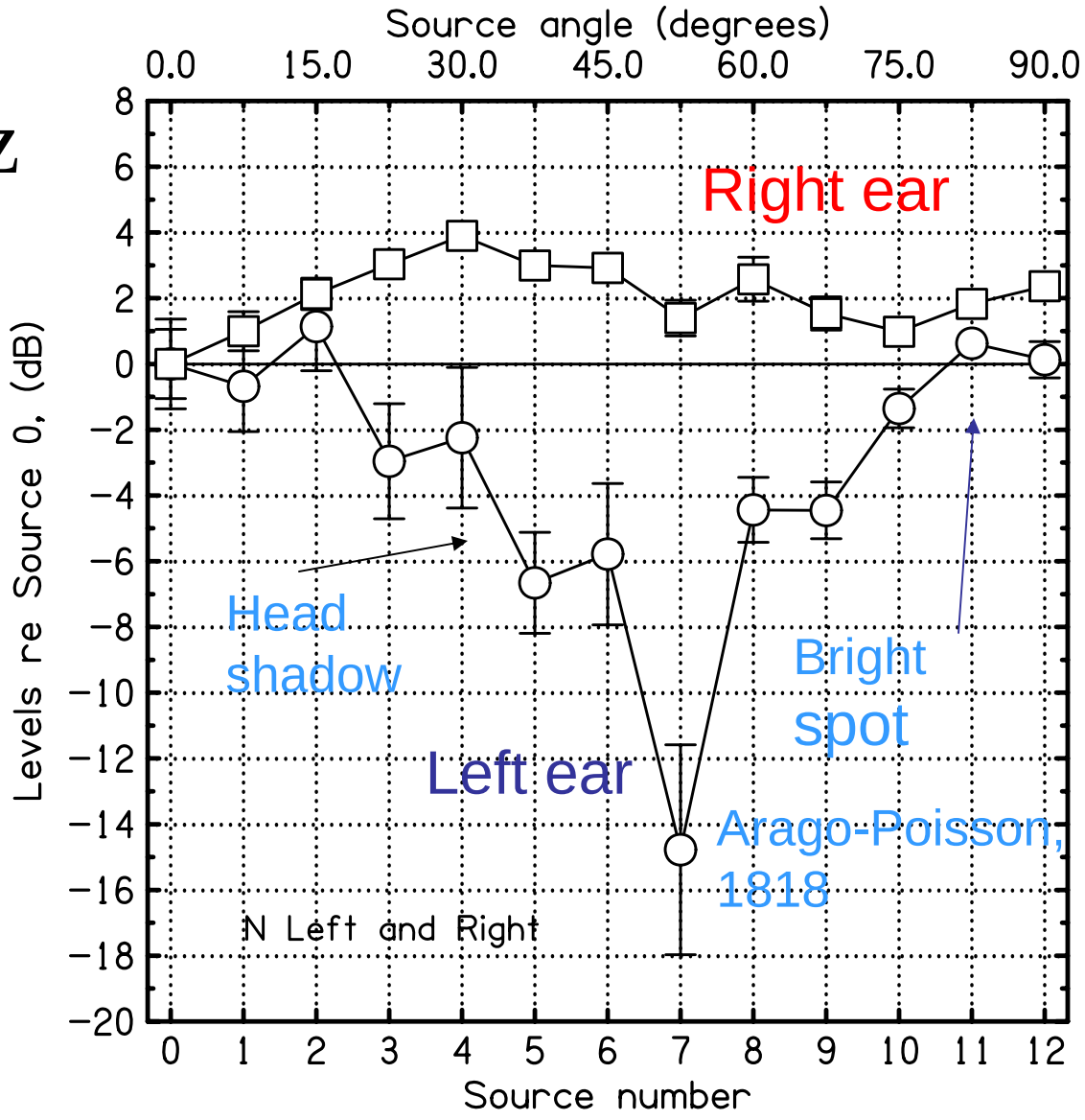


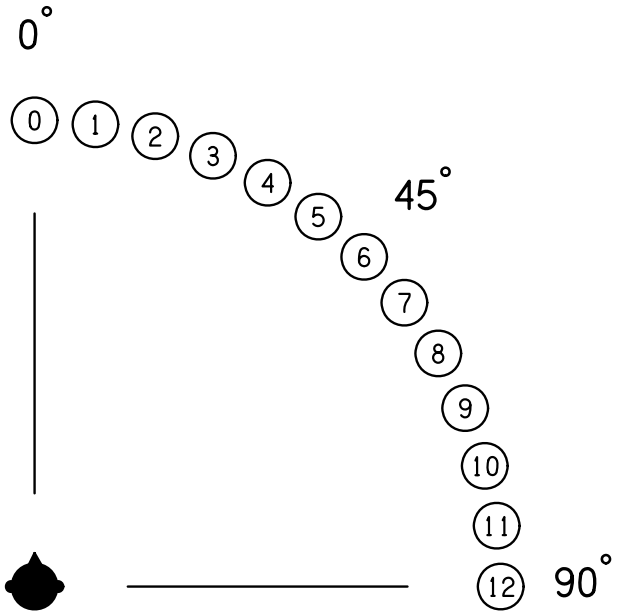
=> ITD < 1200 Hz

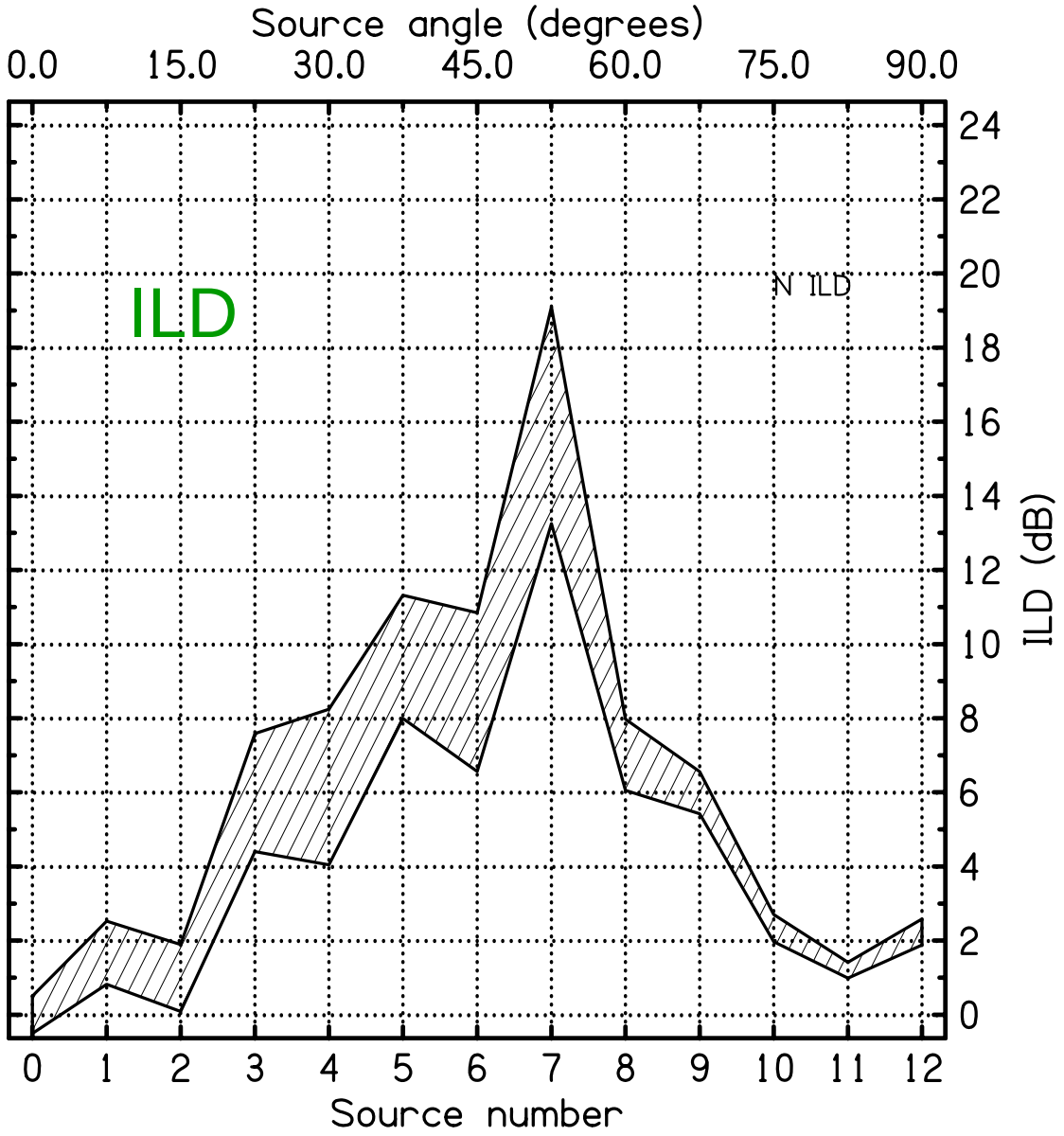
ILD

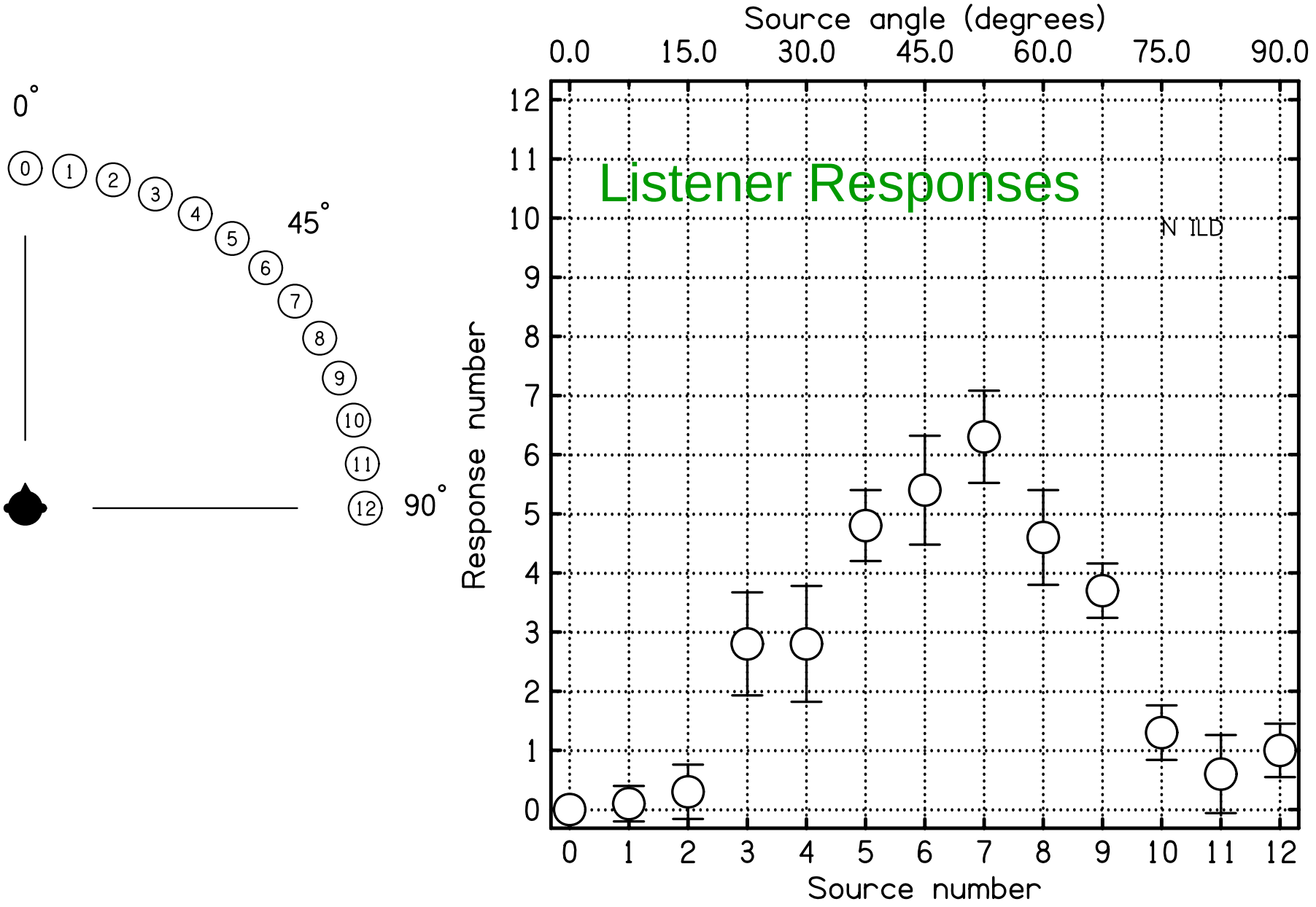
>1300 Hz


0°
0
1
2
3
4
5
45°
6
7
8
9
10
11
12
90°


Source angle (degrees)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
Levels re Source 0, (dB)
8
6
4
2
0
−2
−4
−6
−8
−10
−12
−14
−16
−18
−20
0
1
2
3
4
5
6
7
8
9
10
11
12
Source number
Right ear
Head shadow
Bright spot
Left ear
Arago-Poisson, 1818
N Left and Right

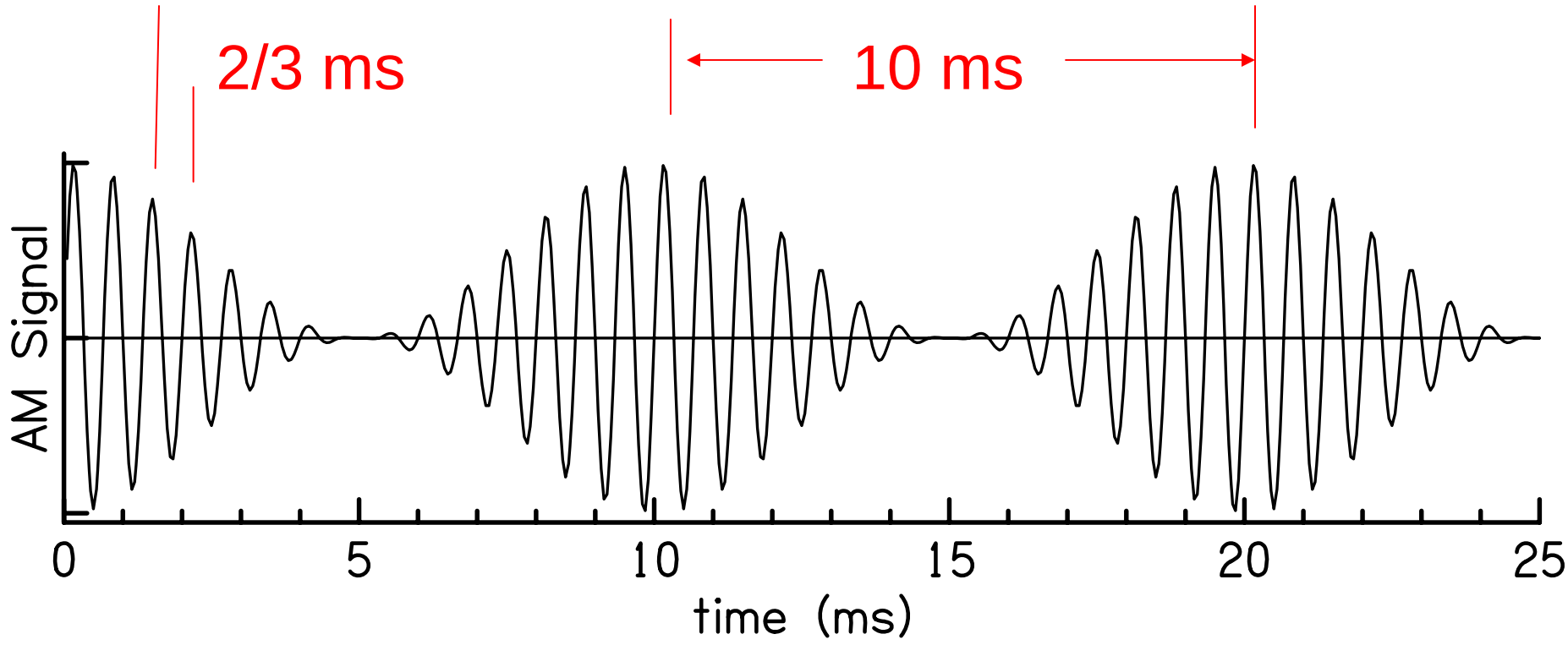

0°
0
1
2
3
4
45°
5
6
7
8
9
10
11
12
90°


Source angle (degrees)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
ILD
N ILD
24
22
20
18
16
14
12
10
8
6
4
2
0
ILD (dB)
0
1
2
3
4
5
6
7
8
9
10
11
12
Source number


0°
0
1
2
3
4
45°
5
6
7
8
9
10
11
12
90°
Source angle (degrees)
0.0
15.0
30.0
45.0
60.0
75.0
90.0
Listener Responses
N ILD
Response number
Source number
0
1
2
3
4
5
6
7
8
9
10
11
12

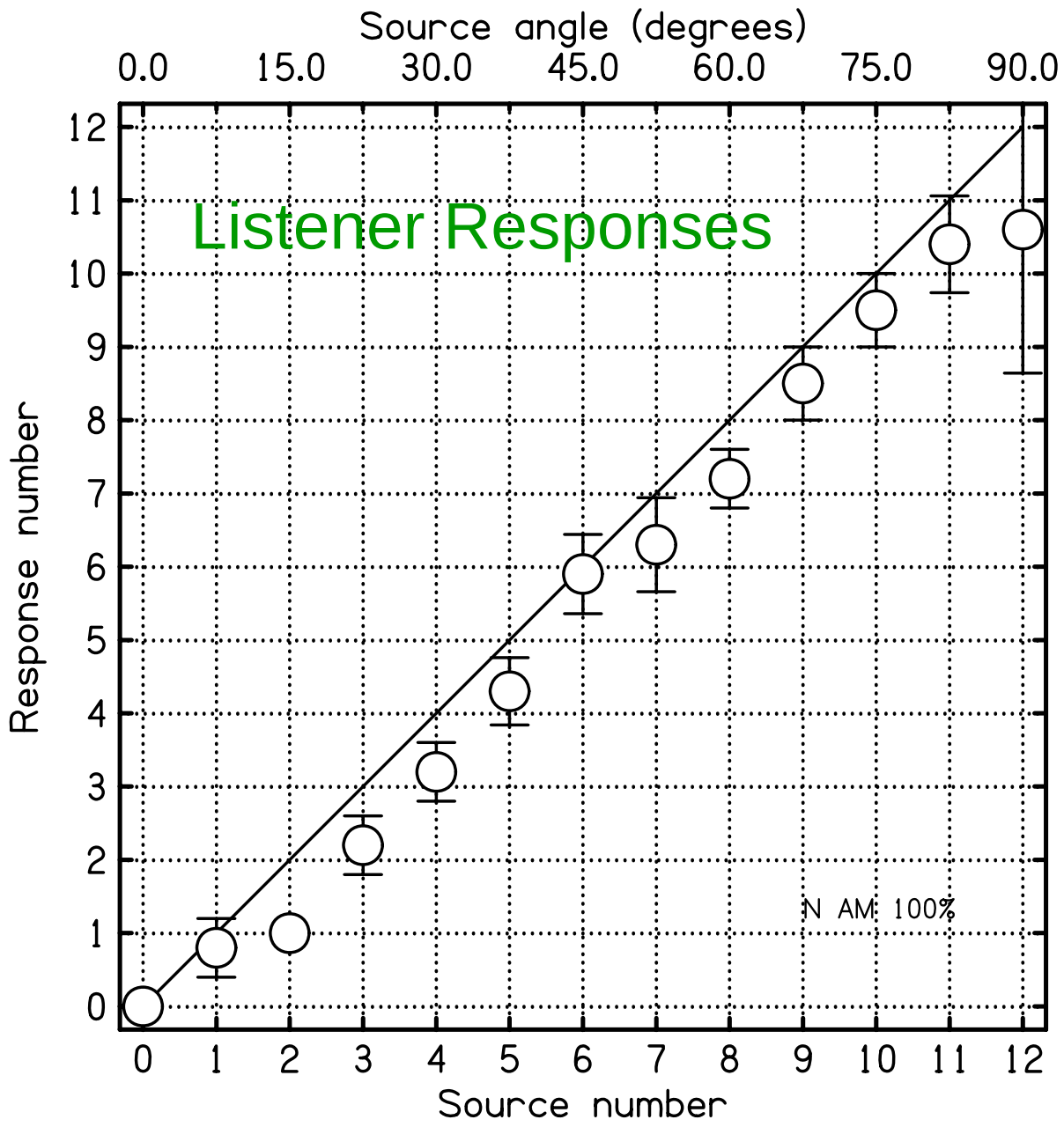
Amplitude modulation

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

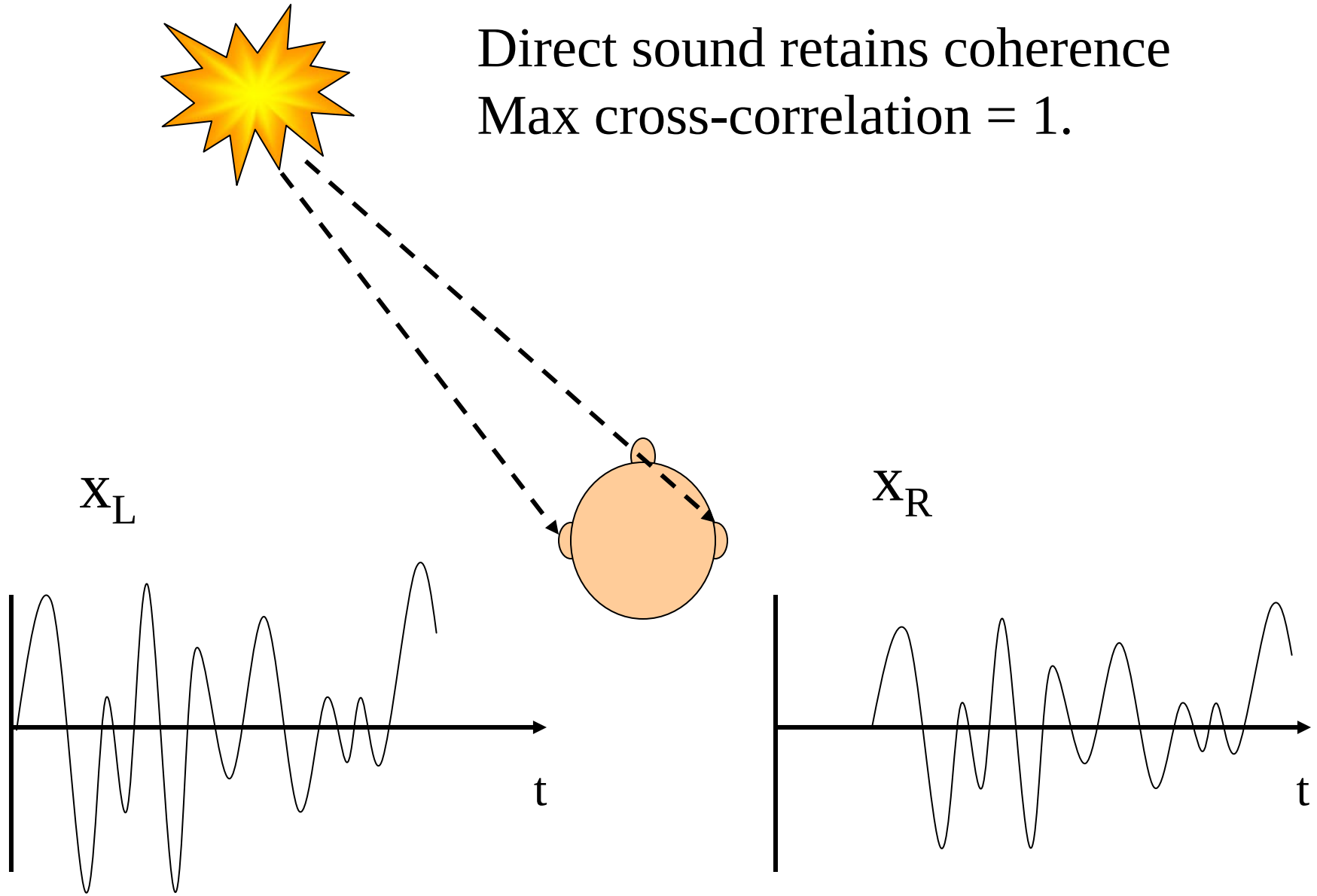


AM
f_m=100 Hz
100%

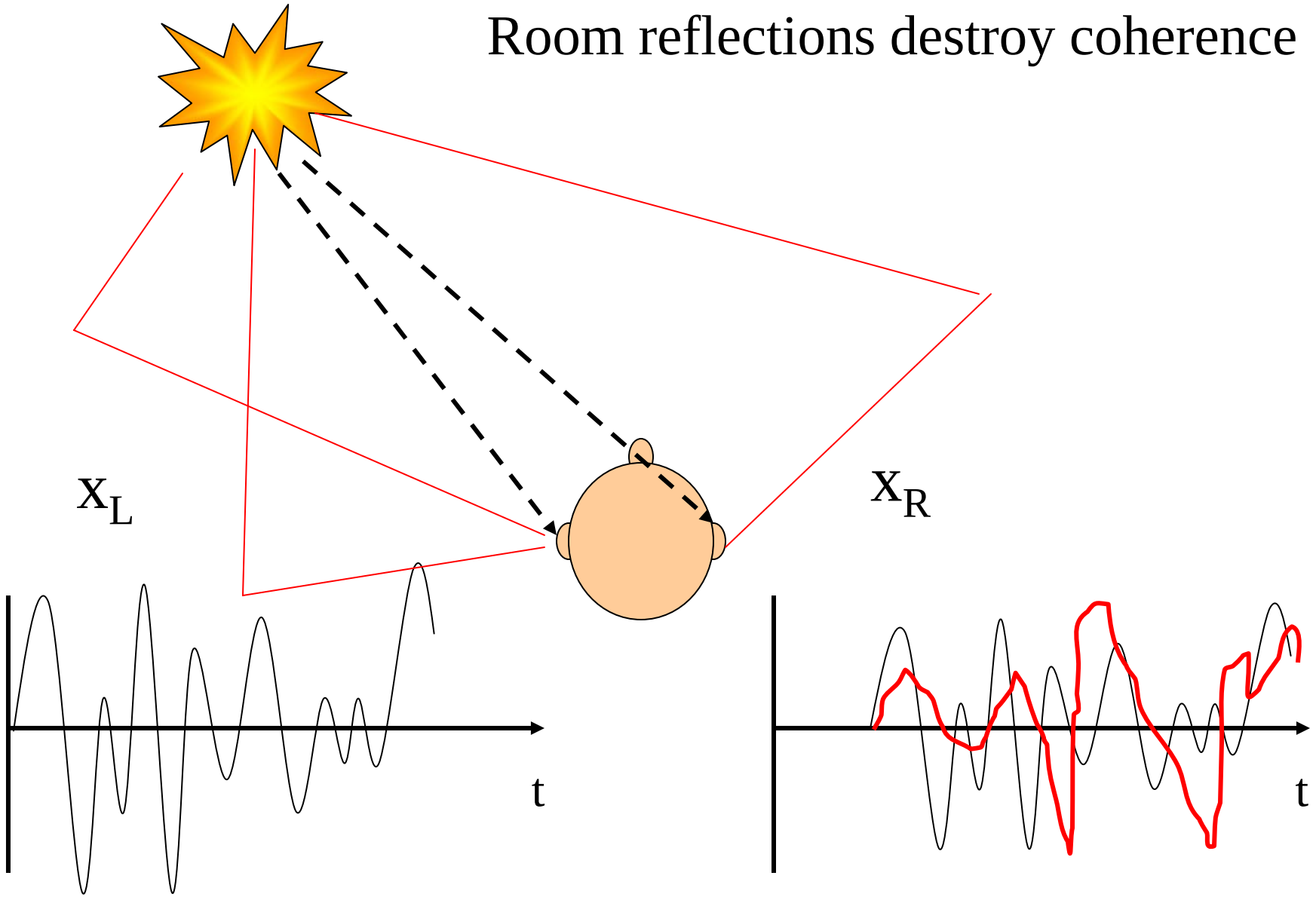
Listener N

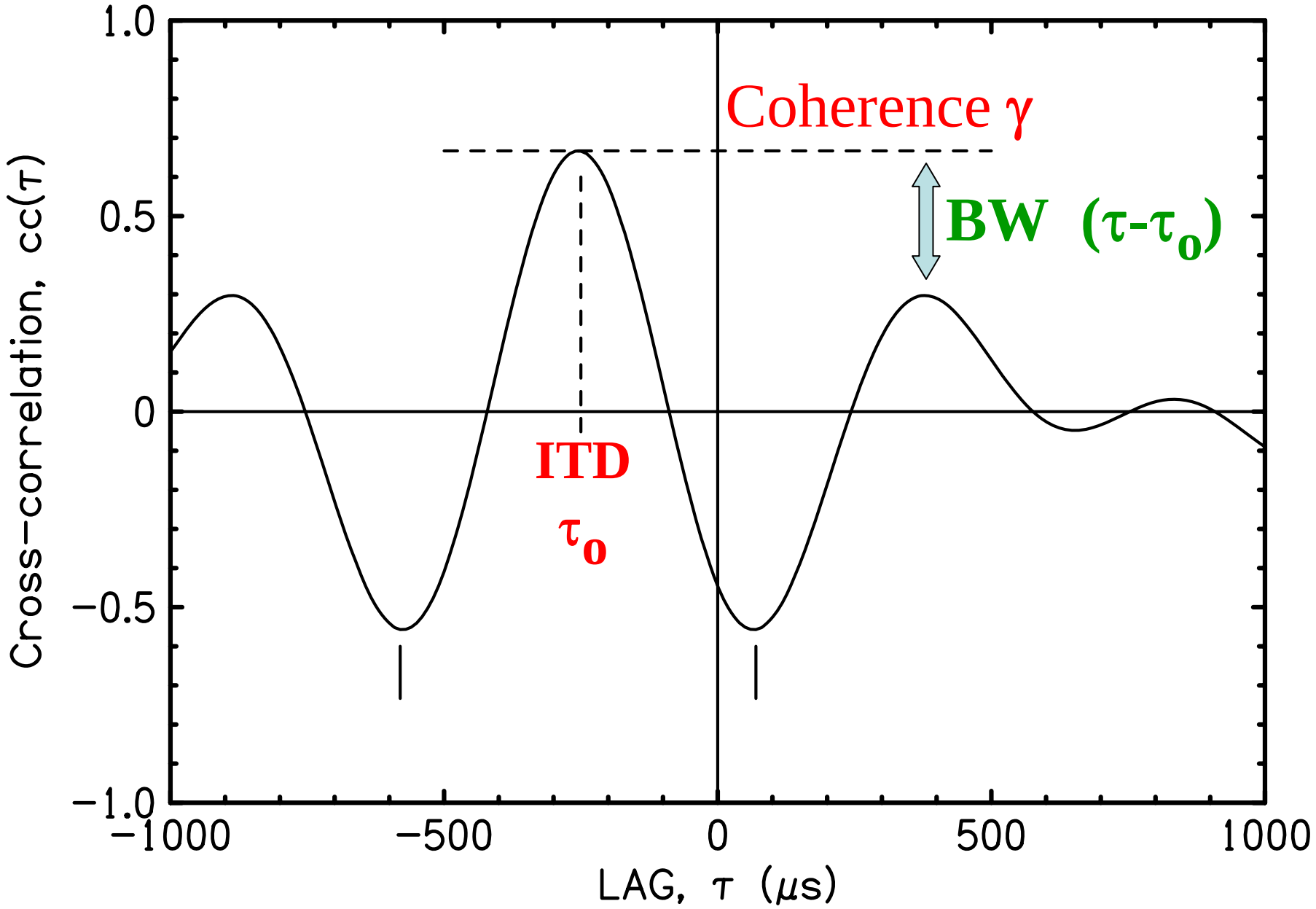


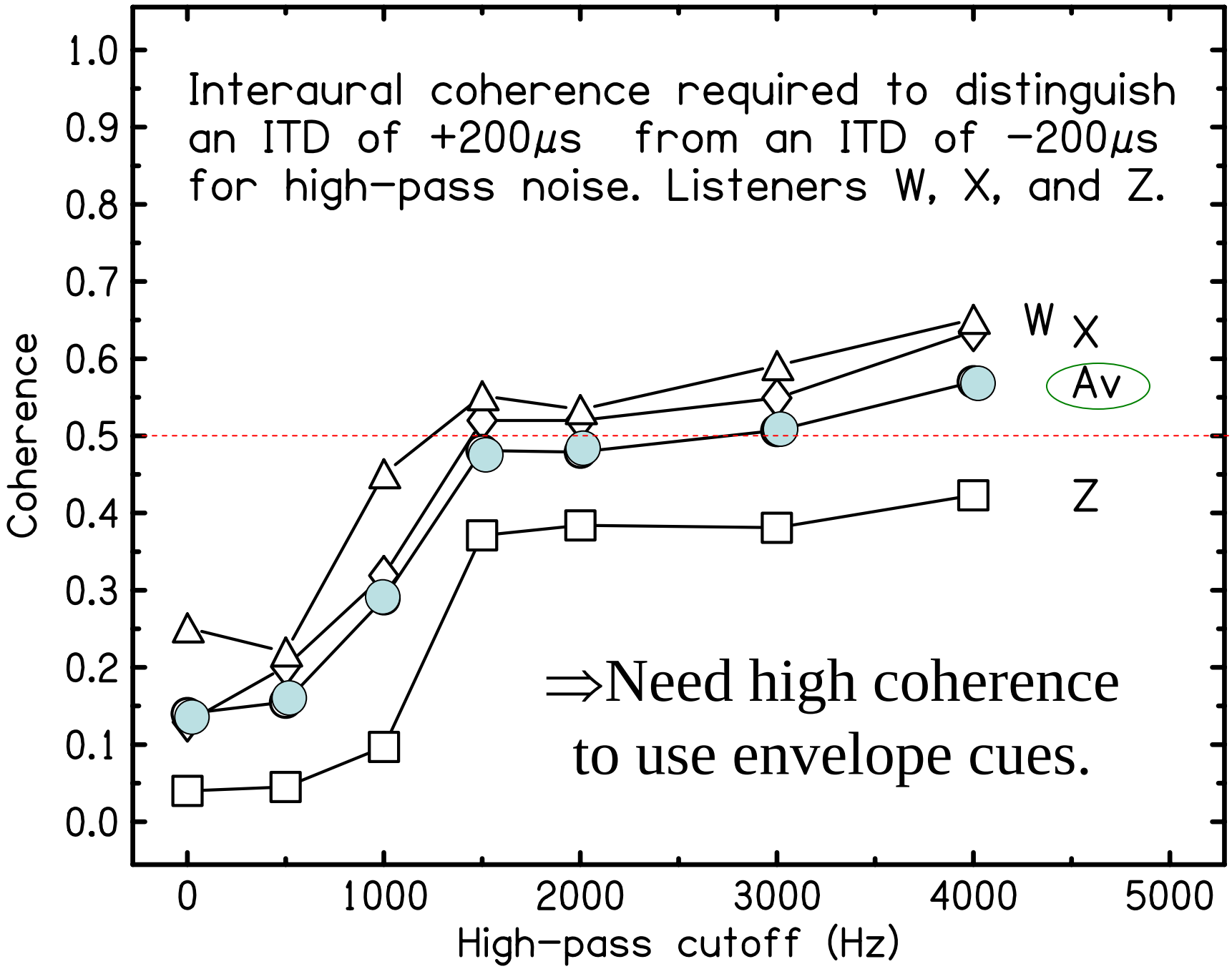
Direct sound retains coherence
Max cross-correlation = 1.



Room reflections destroy coherence




Coherence γ
BW (τ-τo)
ITD
τo
Cross-correlation, cc(τ)
LAG, τ (μs)
1.0
0.5
0
−0.5
−1.0
−1000
−500
0
500
1000


Interaural coherence required to distinguish
an ITD of +200μs from an ITD of −200μs
for high-pass noise. Listeners W, X, and Z.
W
X
Av
Z
⇒Need high coherence
to use envelope cues.
Coherence
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1000
2000
3000
4000
5000
High-pass cutoff (Hz)

PRECEDENCE EFFECT

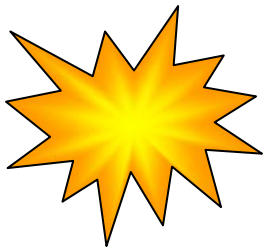
- Competition between Sound 1 & Sound 2
- Sound 1: Localization cues, ITD, ILD, etc.
- Sound 2: Localization cues, ITD, ILD, etc.

Sound 1 takes precedence.

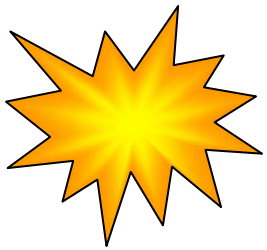
Sound 1 takes precedence.

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

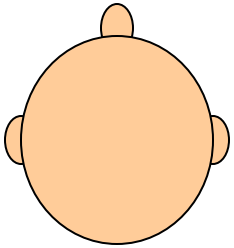
Fused image near location of Sound 1.



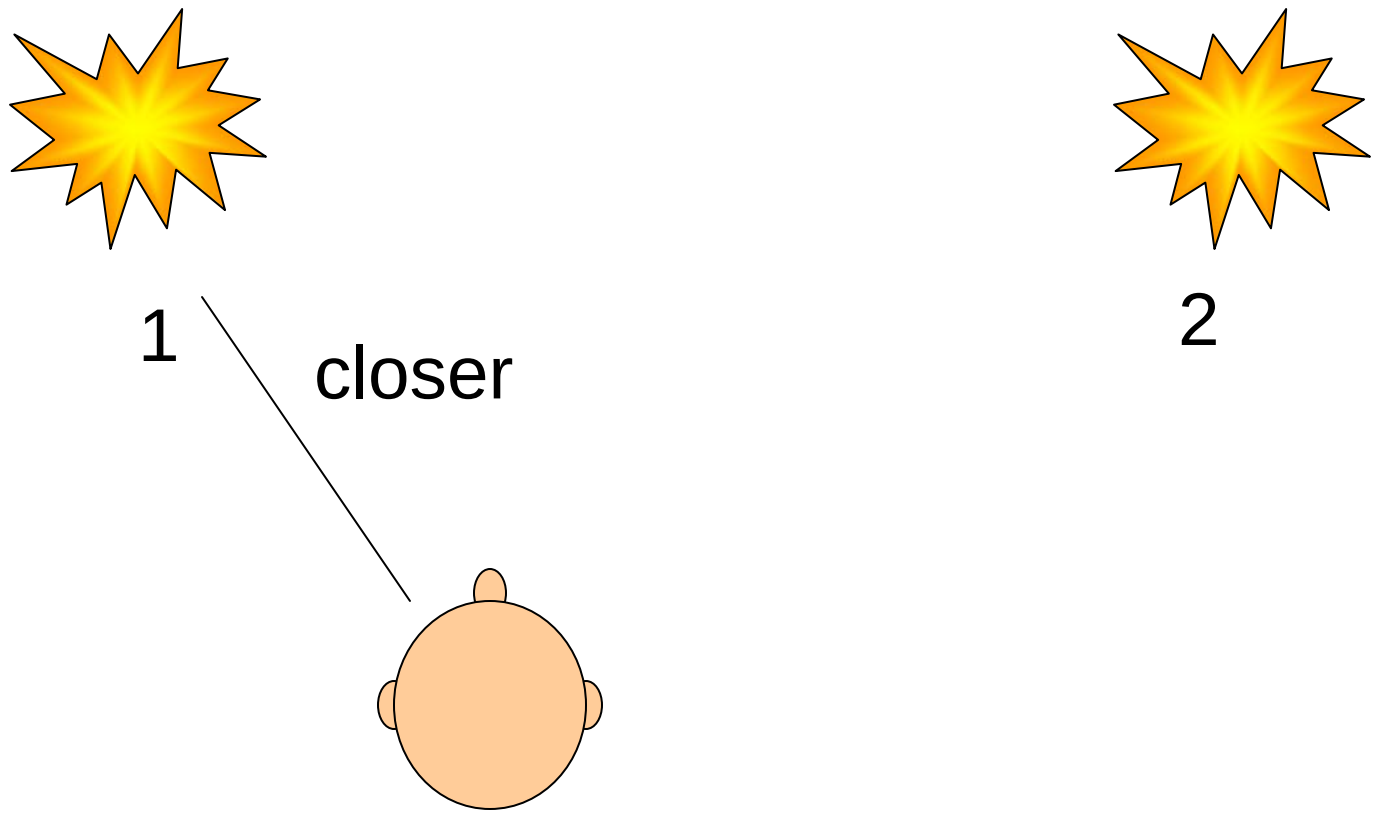
1



2



Yvor Winters, 1900-1968

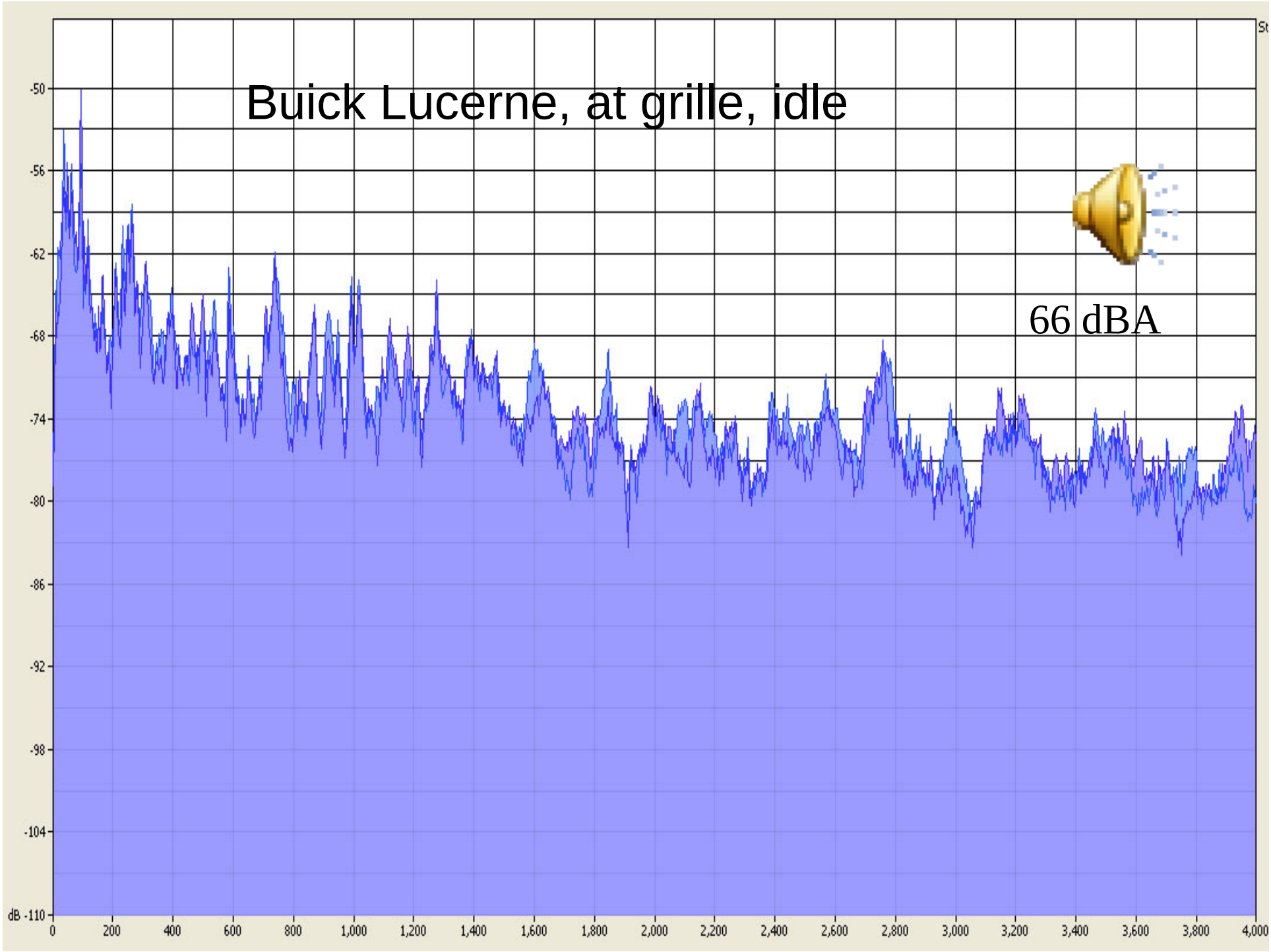

1
closer
2

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).

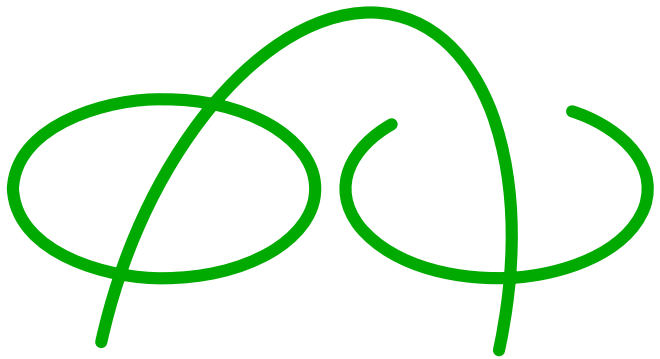

Buick Lucerne, at grille, idle
66 dBA
-50
-56
-62
-68
-74
-80
-86
-92
-98
-104
dB -110
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
2,200
2,400
2,600
2,800
3,000
3,200
3,400
3,600
3,800
4,000

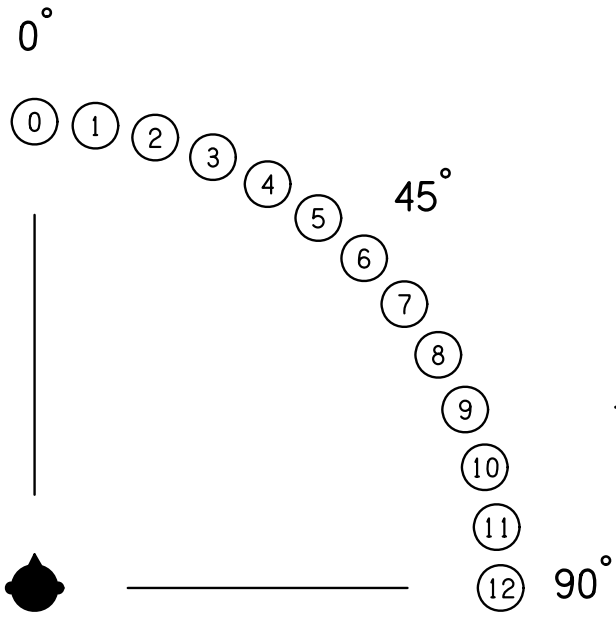
Conclusions

- 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.

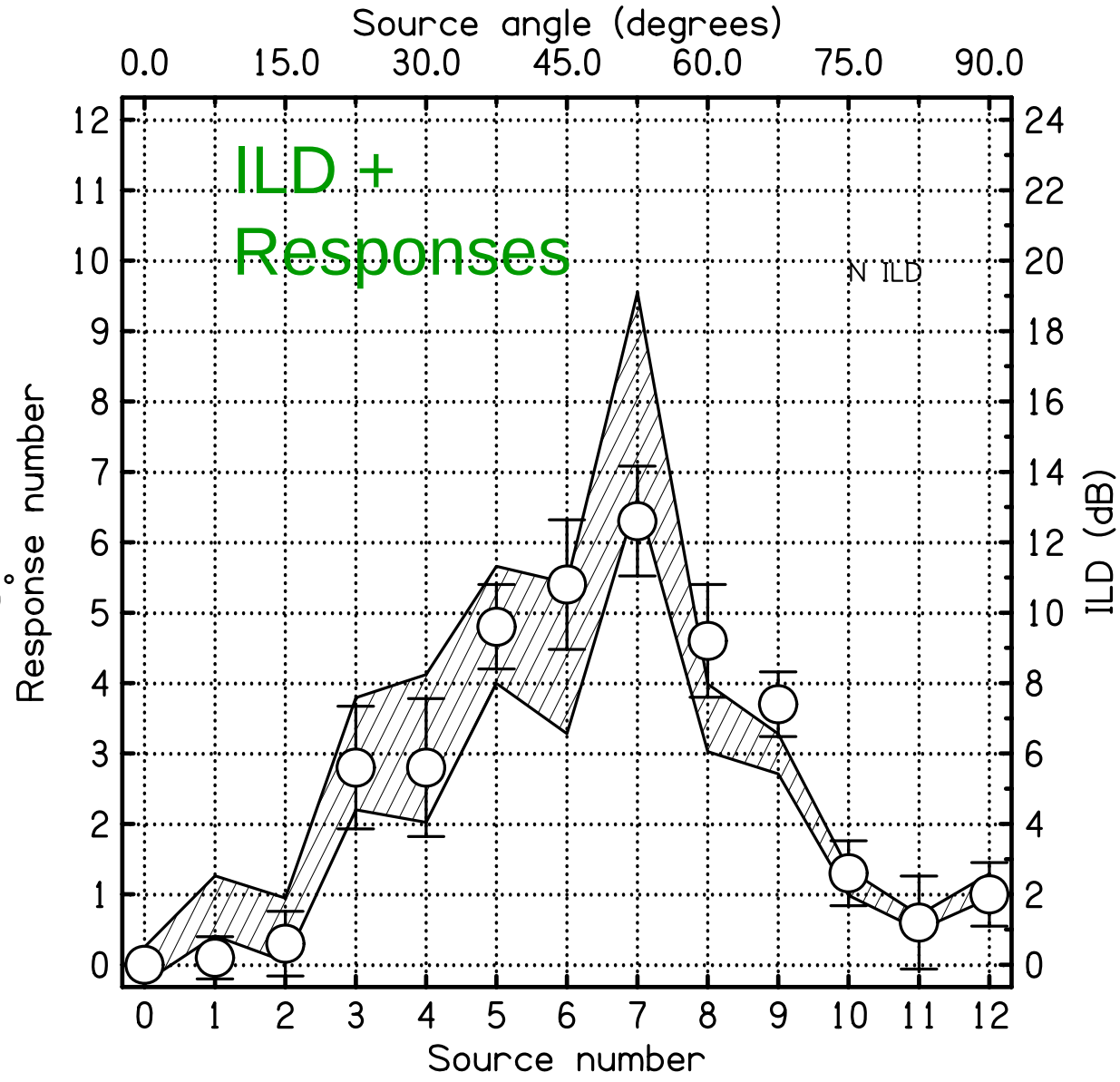
The End

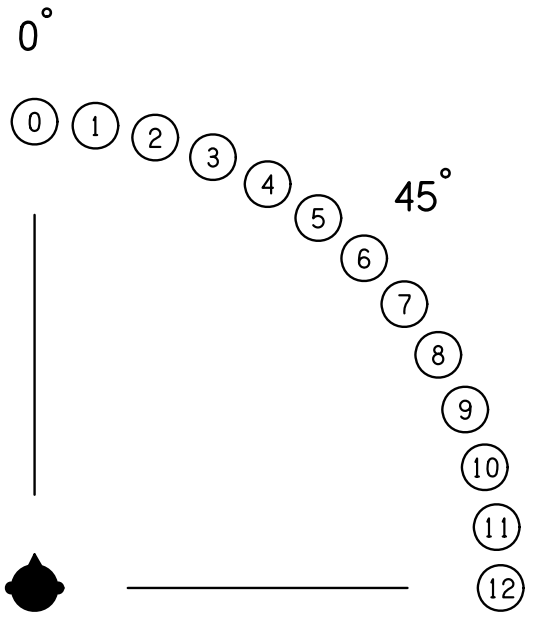
Thanks to
the NIDCD





Listener does not use ear signals independently.





1500 Hz is too fast for ongoing ITD or IPD.