# AUDL GS08/GAV1
# Signals, systems, acoustics and the ear

## Psychoacoustics of hearing impairment
## &
## Perceiving speech in noise

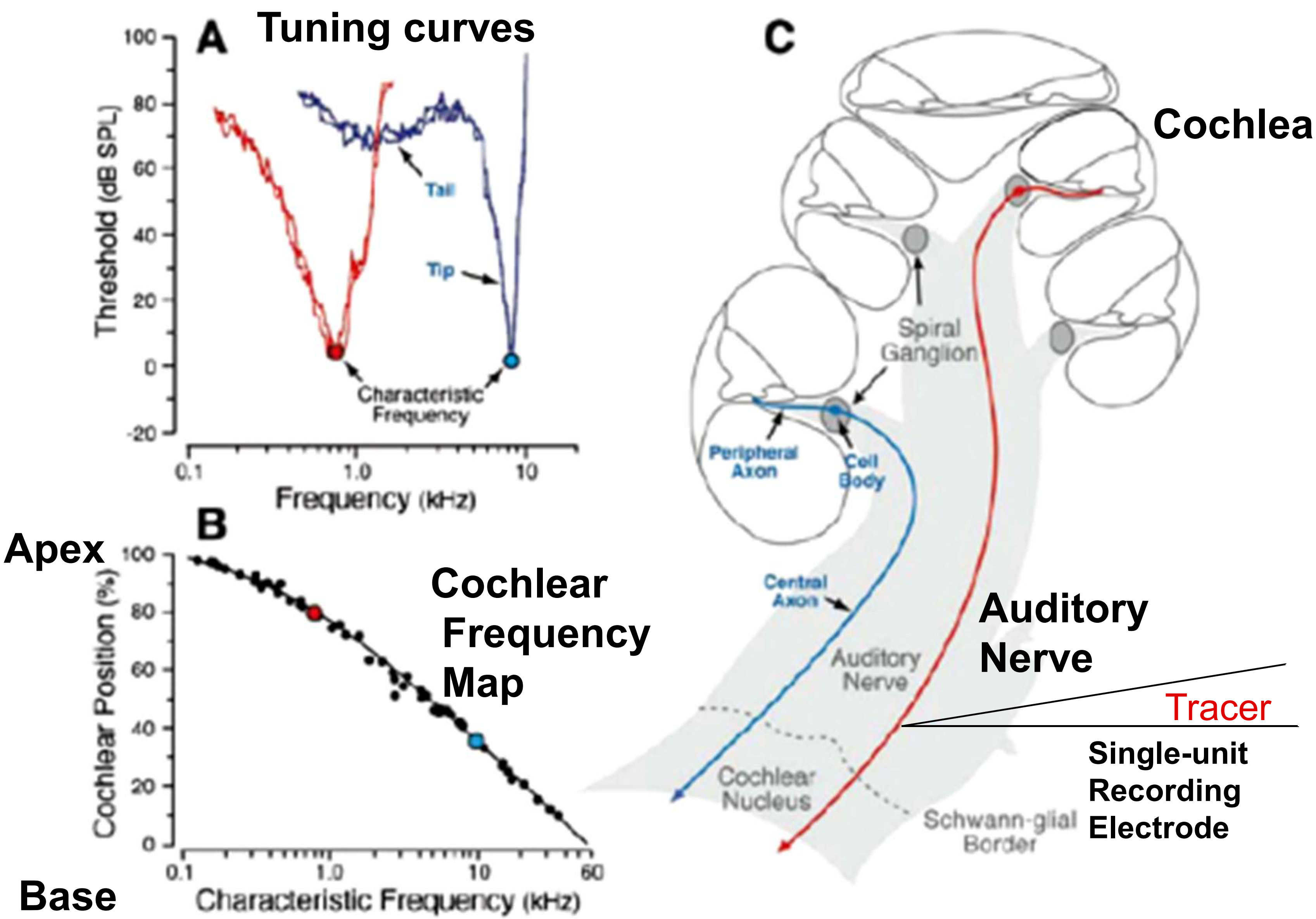
# Three main types of hearing impairment

- Conductive
  - Sound is not properly transmitted from the outer to the inner ear
- Sensorineural
  - Damage to the inner ear
- Retrocochlear
  - Damage to the auditory nerve and beyond

# What do we know about physiological reflections of sensori-neural hearing loss?

focus on hair cell damage

# Auditory Nerve Structure and Function




slide courtesy of Chris Brown, Mass Eye & Ear

Liberman (1982)

# Outer Hair Cells are relatively vulnerable to damage, leading to ...

- Decreases in basilar membrane movement and hence increased thresholds to sound
  - *hearing loss*
- A loss of cochlear compression (a *linearised* input/output function)
  - *reduced dynamic range*
  - *loudness recruitment*
- Loss of frequency tuning (analogous to widened filters in an auditory filter bank).
  - *degraded frequency selectivity*

# Input/ Output functions on the basilar membrane near CF in an impaired ear

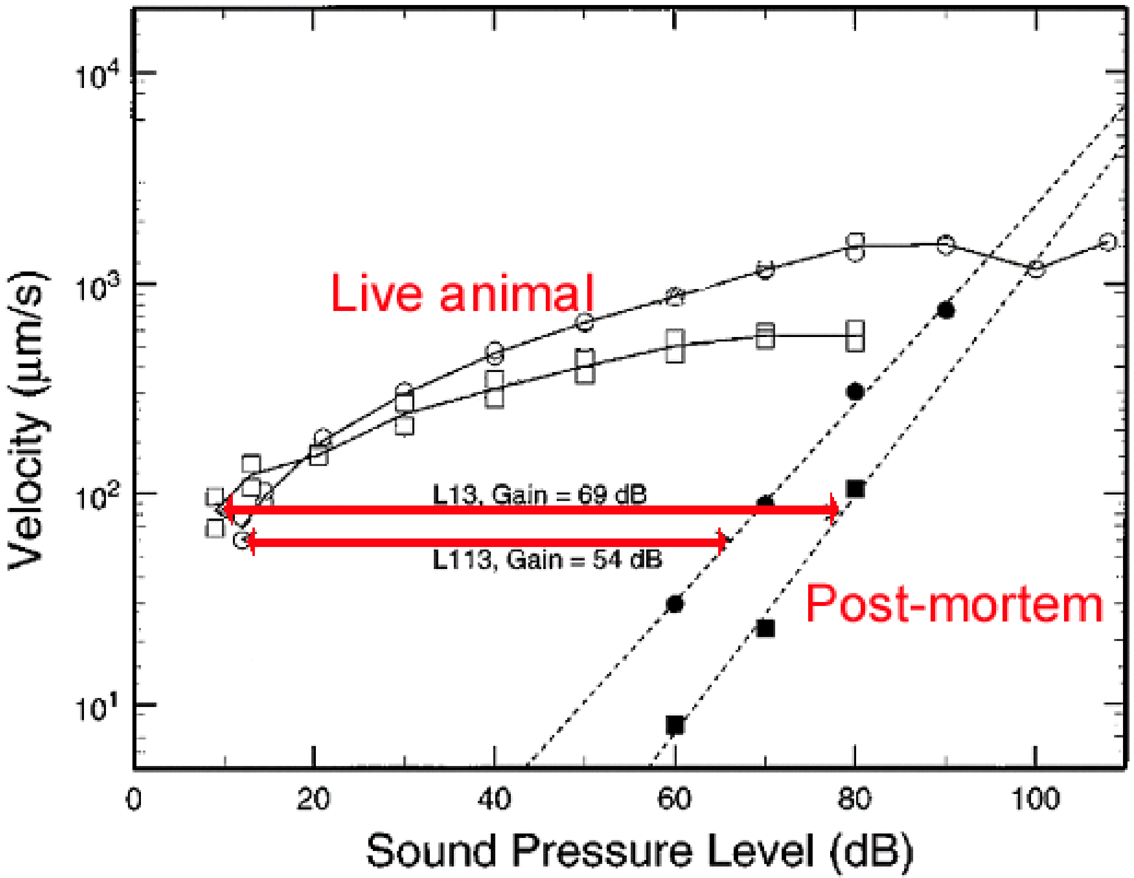
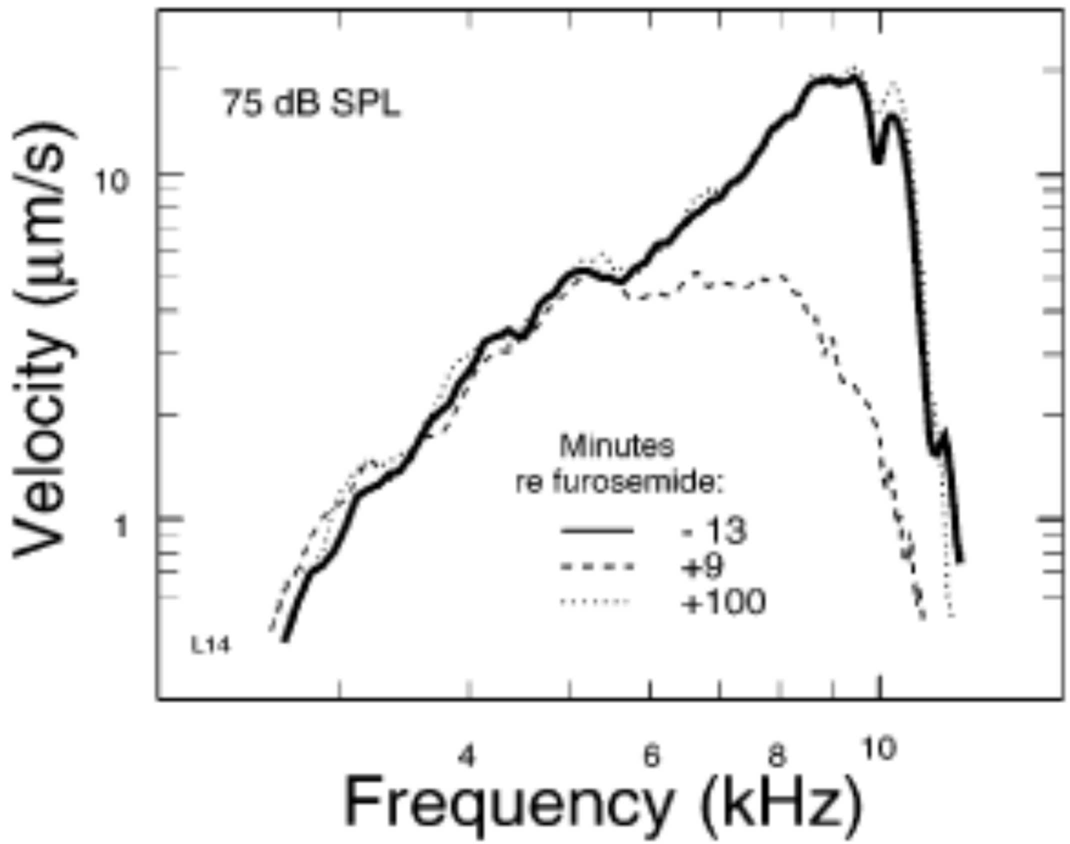



FIG. 16. Stability and vulnerability of responses to CF and near-CF tones. The open symbols depict the peak velocities of responses to CF tones (L13: squares; L113: circles) recorded in the sensitive cochleae of two live chinchillas. The filled symbols represent the CF responses recorded immediately after (within minutes of) death. Responses to CF tones in both cochleae were measured both early in the experiment and 160–240 min later.

# Frequency response of a single place on the BM in an impaired ear (furosemide)




Ruggero and Rich (1991)

# Inner Hair Cell (IHC) damage ...

- Leads to a more sparse representation of *all* auditory information passed on to higher auditory centres.
- There are possibly even regions of the cochlea without any IHCs — so-called *dead regions*.
- Hence, there may be a degradation of a wide variety of auditory abilities *(e.g*. temporal resolution).

# Relation of Hair Cell loss to audiogram

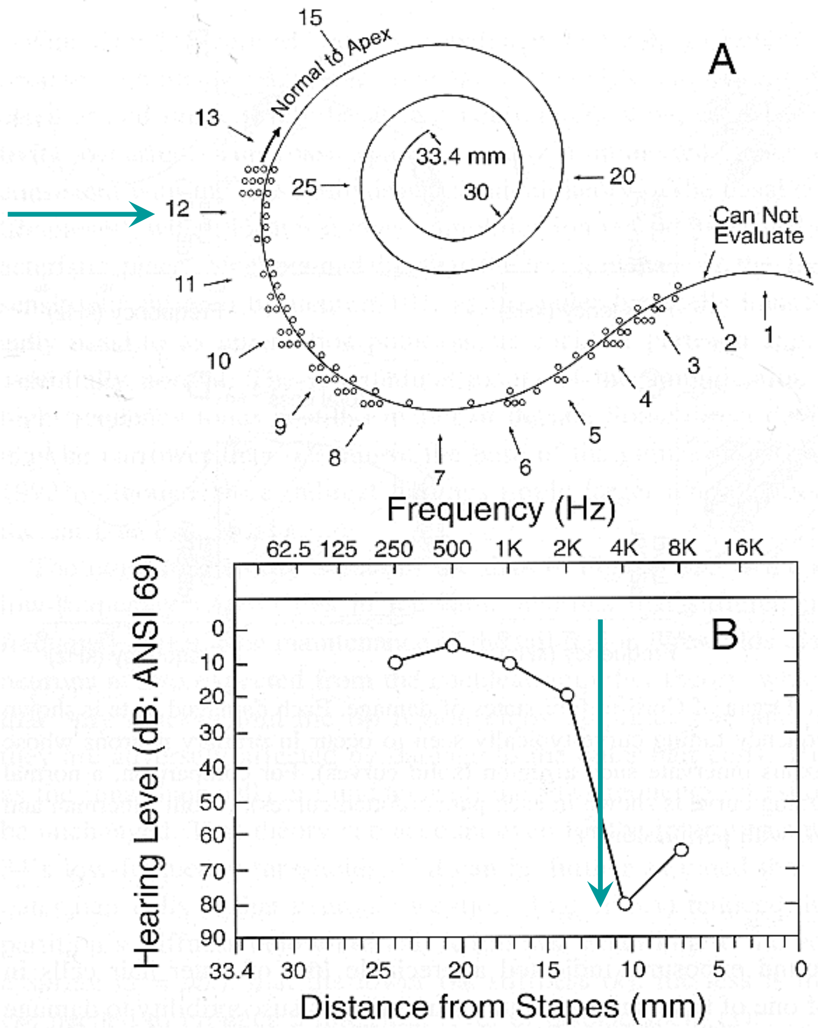
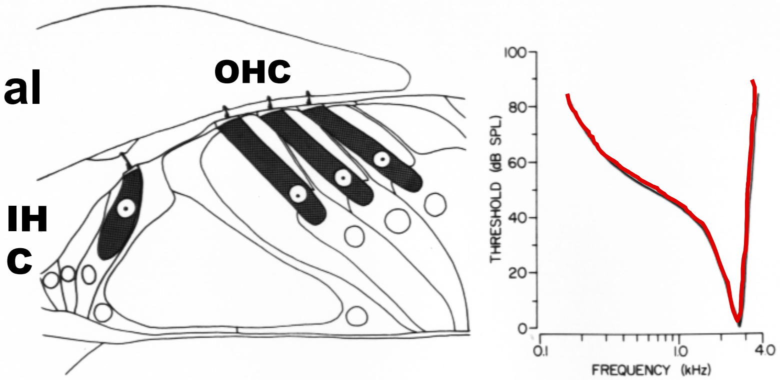



**Figure 16.5.** Comparison of cochlear pathology with the audiogram of a human patient. *A*. Patient's cytocochleogram, showing in pictorial form the hair cells (circles) remaining in each of the four rows of hair cells, regardless of their condition, plotted as a function of distance from the stapes. Note the extensive hair cell loss in the most basal 12 mm. *B*. Patient's audiogram, showing a profound hearing loss above 2 kHz (top scale of abscissa). The apical border of the extensive hair cell loss corresponds well with the 3 kHz place on the characteristic-frequency/location map for primary auditory neurons in humans (bottom scale of abscissa). (From Schuknect, 1993, with permission.*)

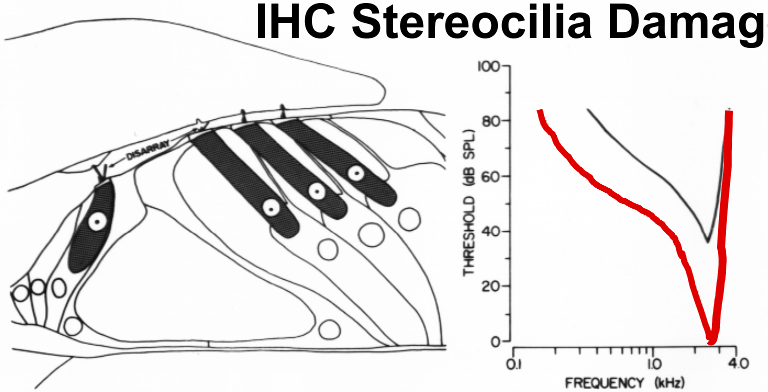
# Auditory Nerve Fiber Responses From Damaged Cochleae

**Normal**

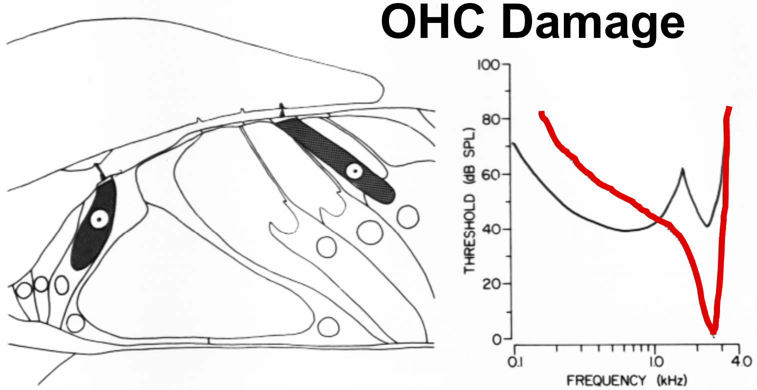



slide courtesy of Chris Brown, Mass Eye & Ear

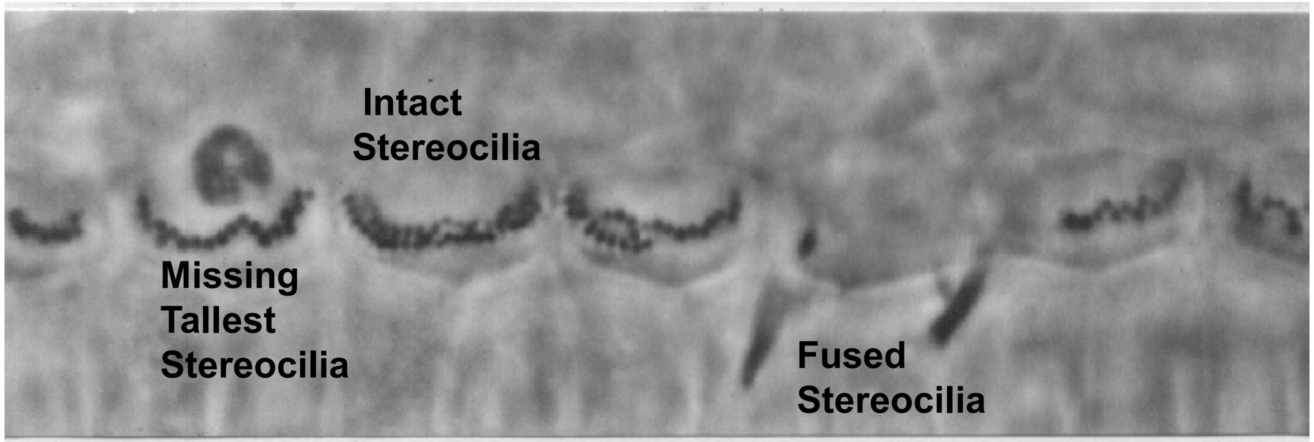
**IHC Stereocilia Damage**




**OHC Damage**

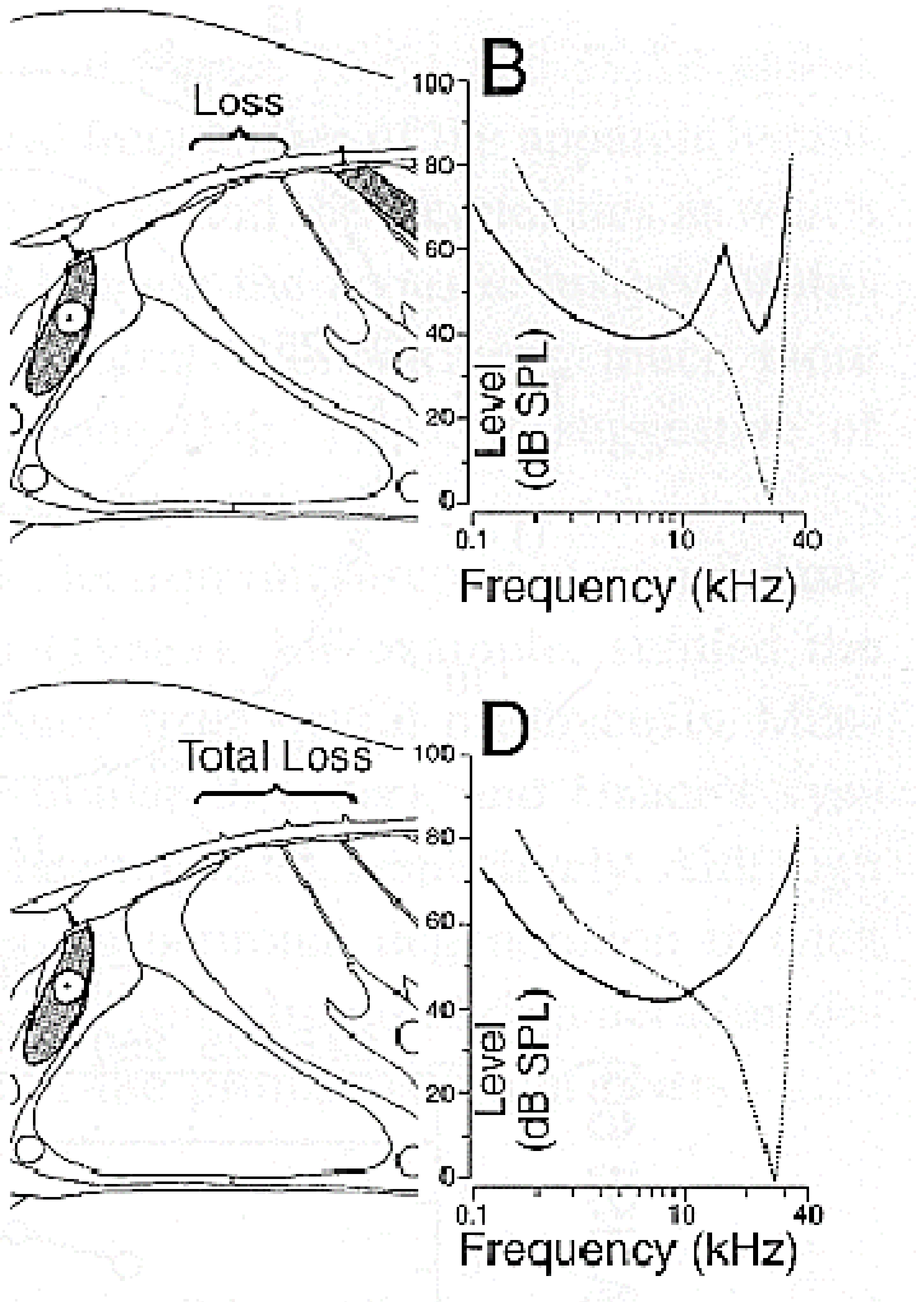



Stereocilia on IHCs




Liberman and Kiang (1978)
Liberman and Beil (1979)
Liberman and Dodds (1984)

Effects of OHC damage and total loss on tuning in the auditory nerve

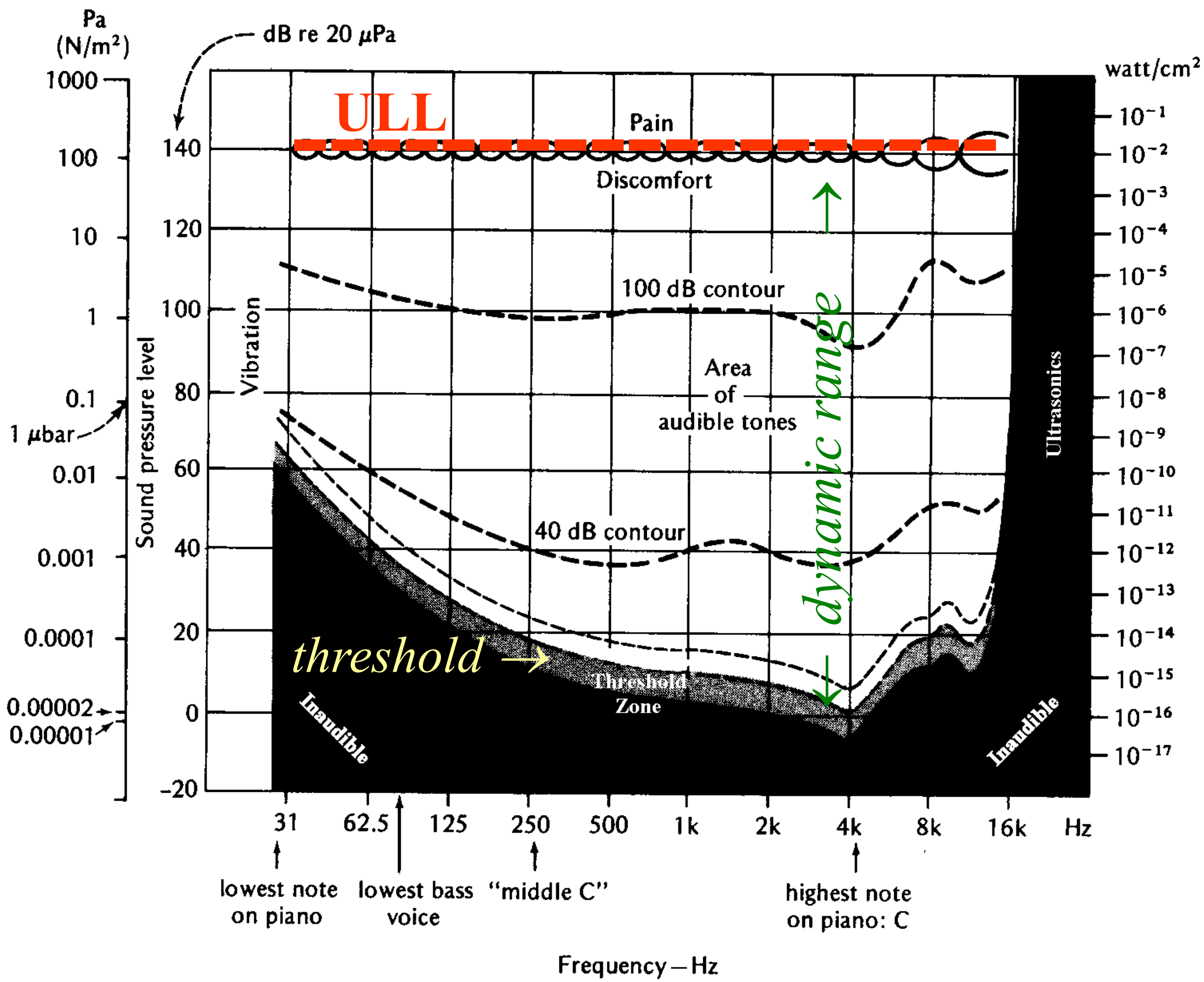


# Psychoacoustic consequences of sensori-neural (cochlear) hearing loss

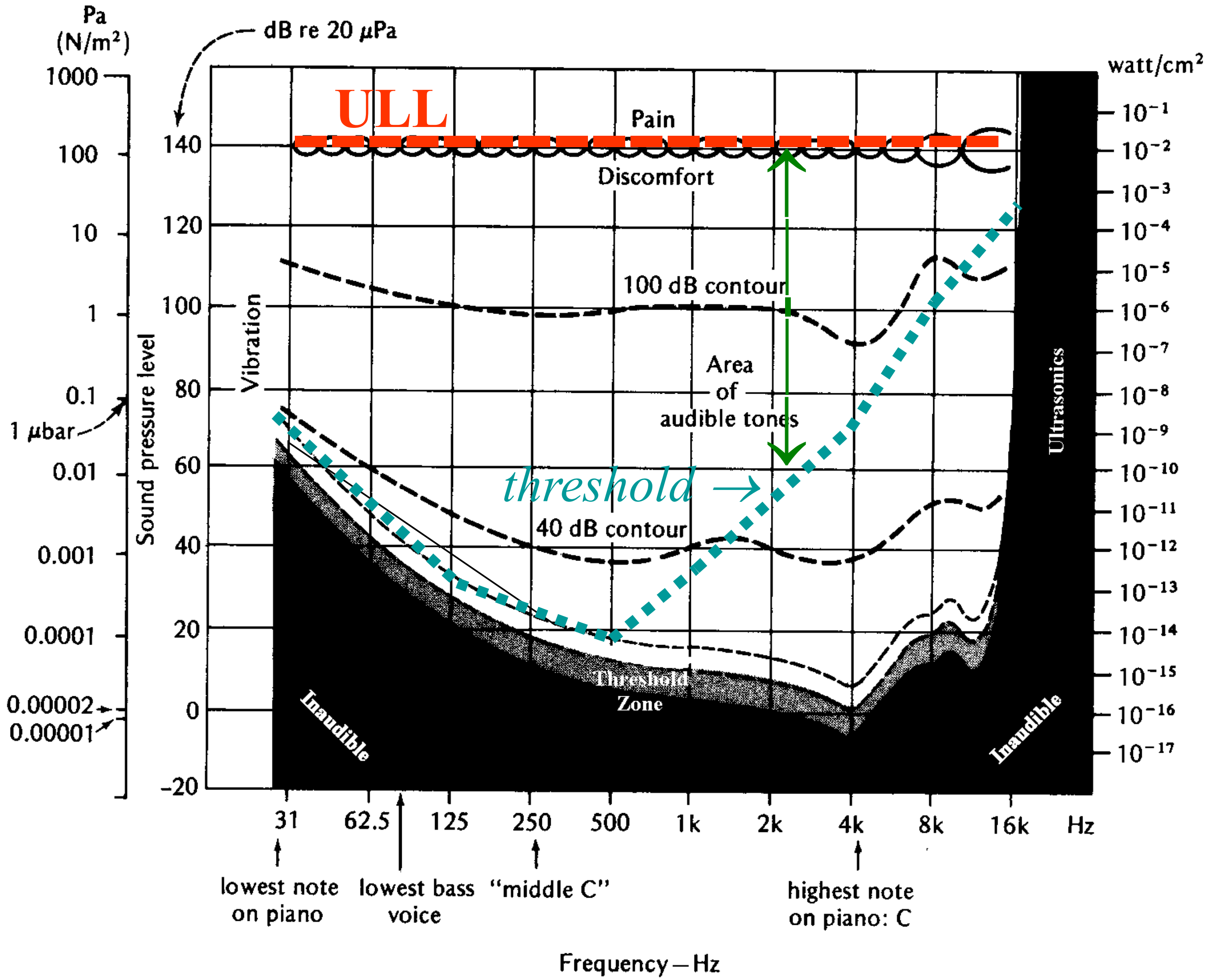
- Raised thresholds
- Reduction of dynamic range and abnormal loudness growth
- Impaired frequency selectivity

What is the impact on speech perception?

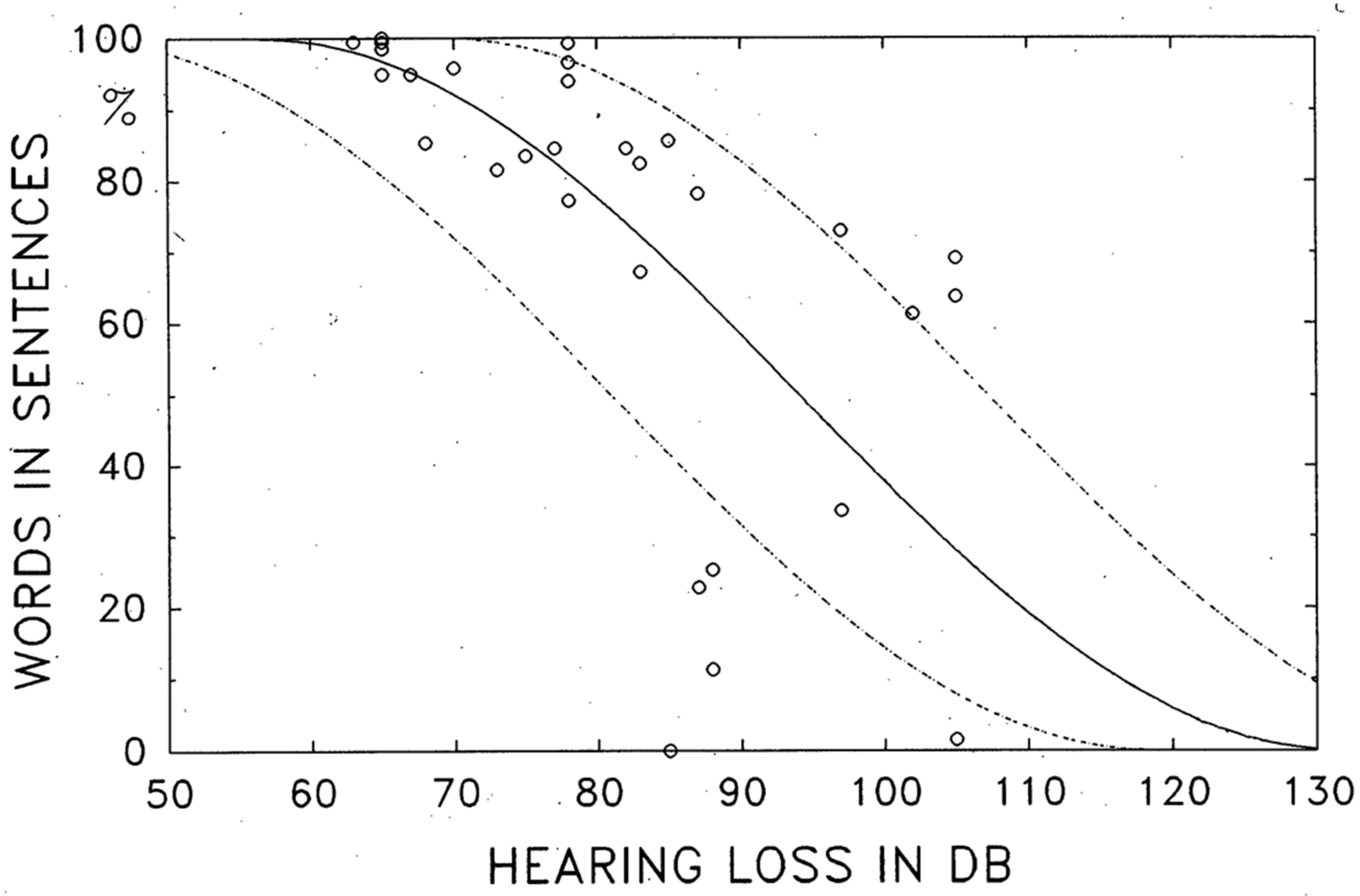
# A normal *auditory area*



# An *auditory area* in sensori-neural loss



# Hearing Loss & Speech Perception

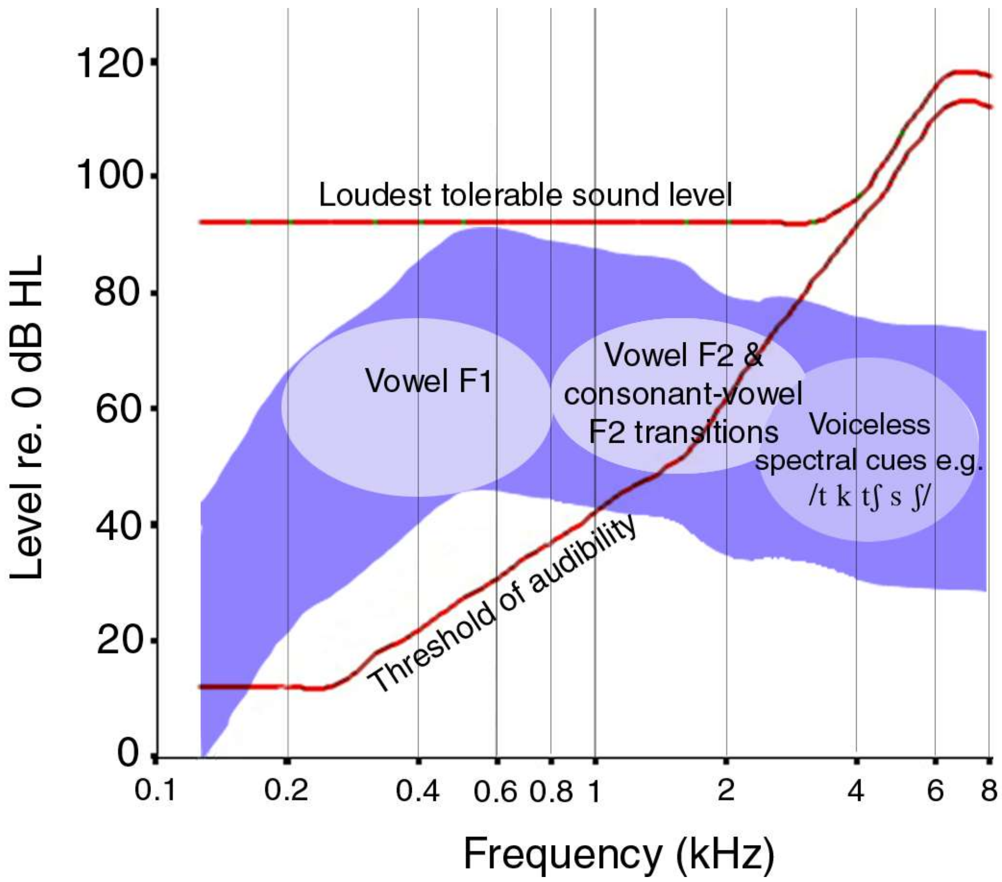



Words recognised from simple sentences in **quiet** by **aided** hearing impaired adults as a function of average hearing loss at 0.5, 1 and 2 kHz. (After Boothroyd, 1990)

# The Role of Audibility

- Much of the impact of hearing loss is thought of in terms of ***audibility***
- How much of the information in speech is audible?
  - Over frequency
  - Over intensity
- Consider the audible area of frequency and intensity in relation to the range of frequencies and intensities in speech

# Speech energy and audibility



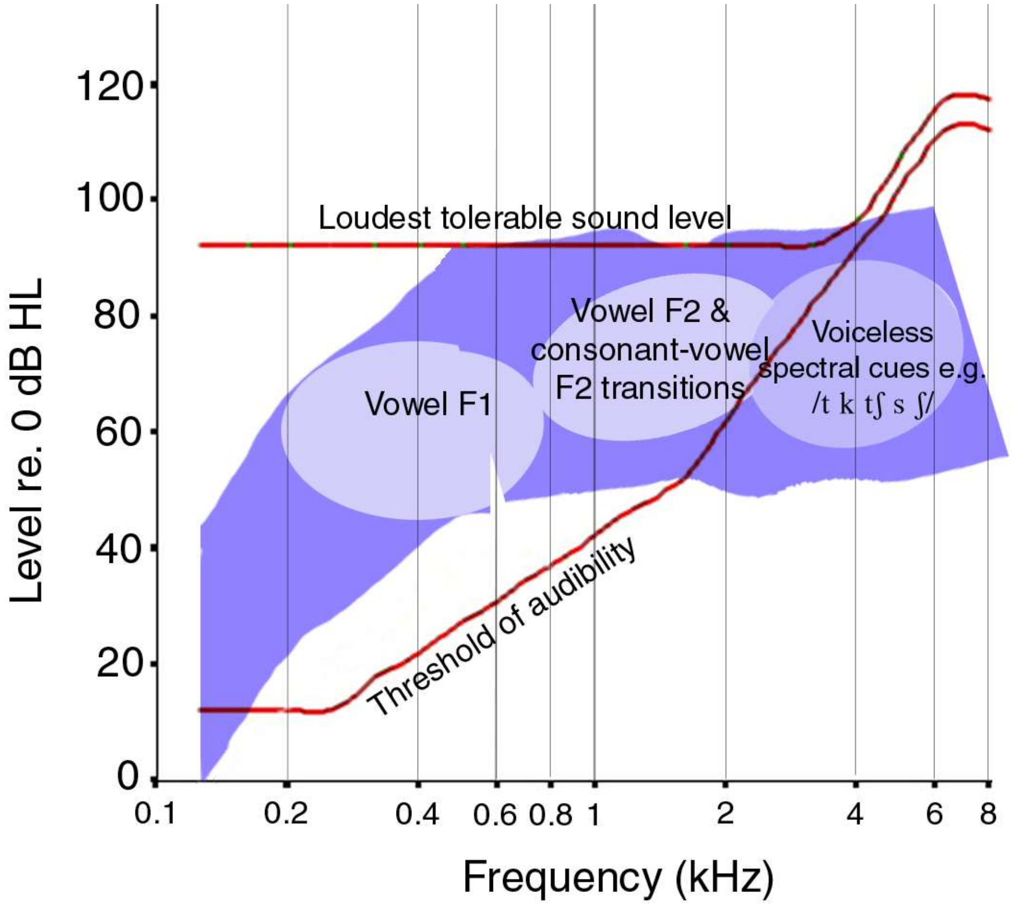

**blue**: the energy range of speech according to frequency relative to the normal threshold of hearing.

**red**: the range of audible levels over frequency for a typical moderate sloping hearing loss.

Intelligibility can be predicted from the portion of the speech range that is audible.

Hearing aids can be set to increase audibility by overall amplification and by shaping of frequency response

# Speech energy and audibility

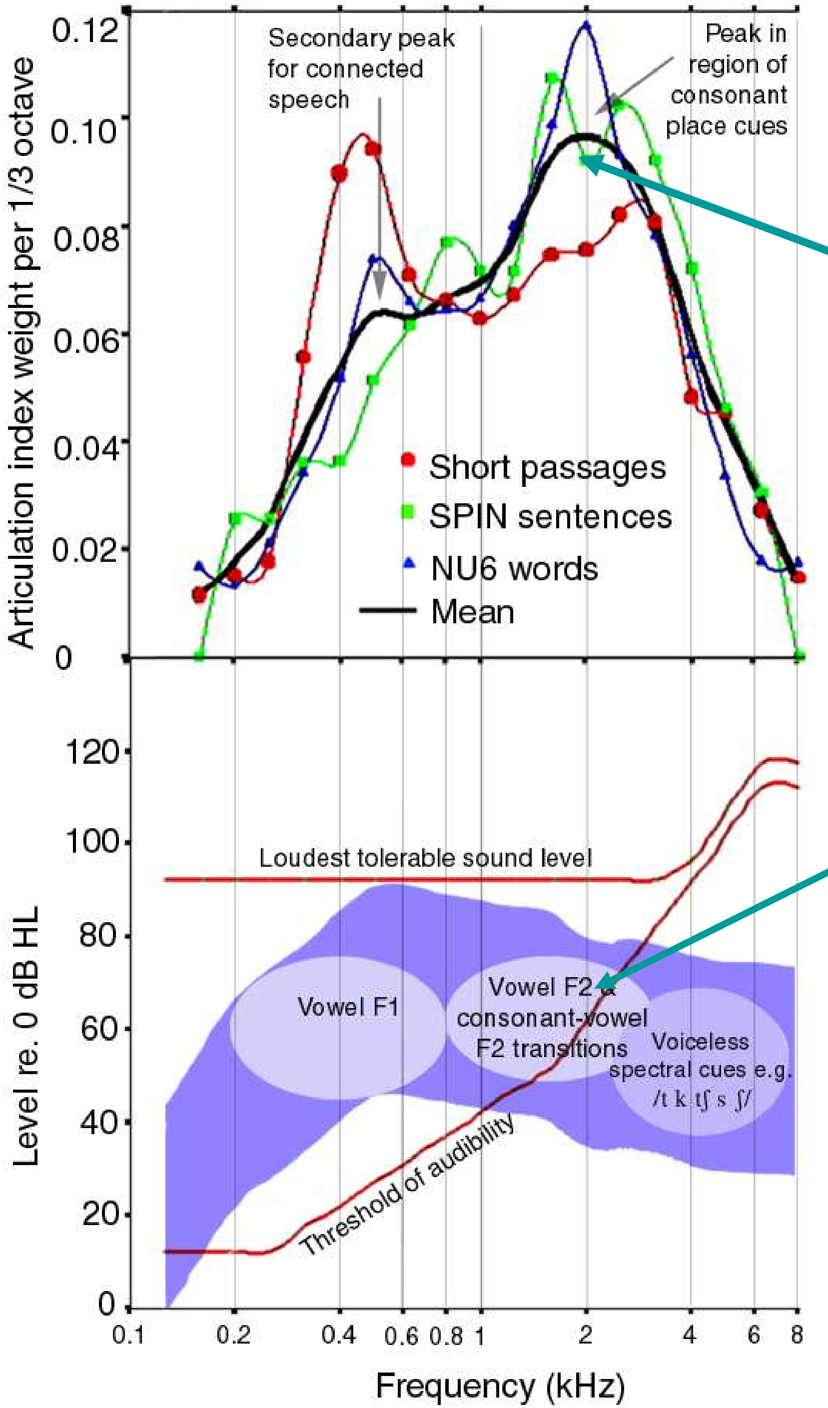



**blue**: the energy range of speech ***after amplification.***

# Articulation Index (AI) or Speech Intelligibility Index (SII)

- An attempt to quantify the role of audibility in speech perception
- Related to standard rules for setting HA frequency response
- Intelligibility is assumed to relate to a simple sum of the contributions from different frequency bands
- Some frequency bands are more important than others

# Frequency importance weightings: AI




I (2000 Hz)

$$A = \sum_{i=1}^{n} I_i \quad W_i,$$

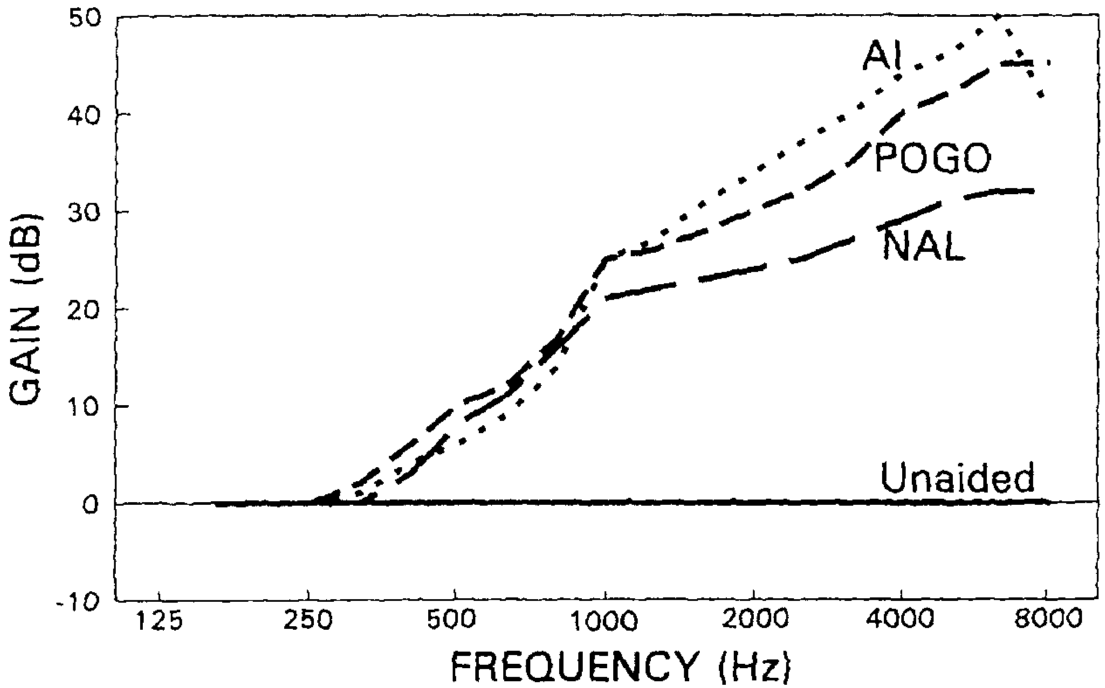
W (2000 Hz) – here W is approx 0.6

***A*** is the Articulation Index (predicted intelligibility).

***A*** is determined by adding up ***W x I*** over frequency bands, where ***I*** is the band importance weight and ***W*** is the proportion of a 30 dB dynamic range of speech in that band that is audible.

AI can predict the frequency response of a hearing aid for a given audiogram that should maximise intelligibility.

Similar to standard HA fitting rules, although these generally recommend less gain than AI where losses are more severe.



# AI predictions

AI predictions reasonable for mild and moderate hearing losses. But the effects of audibility in severe and profound losses are not enough to explain limits to speech recognition.

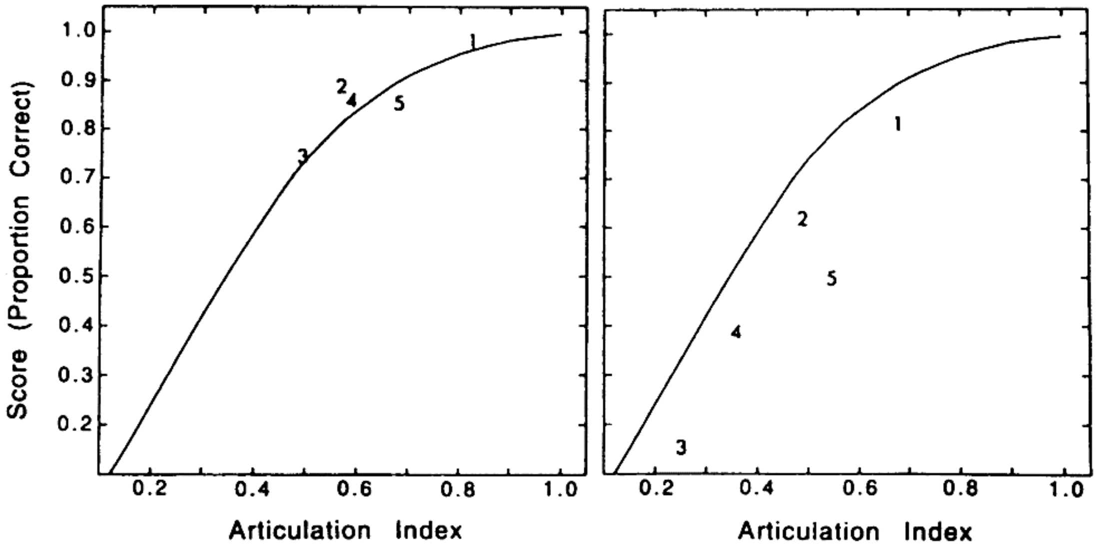
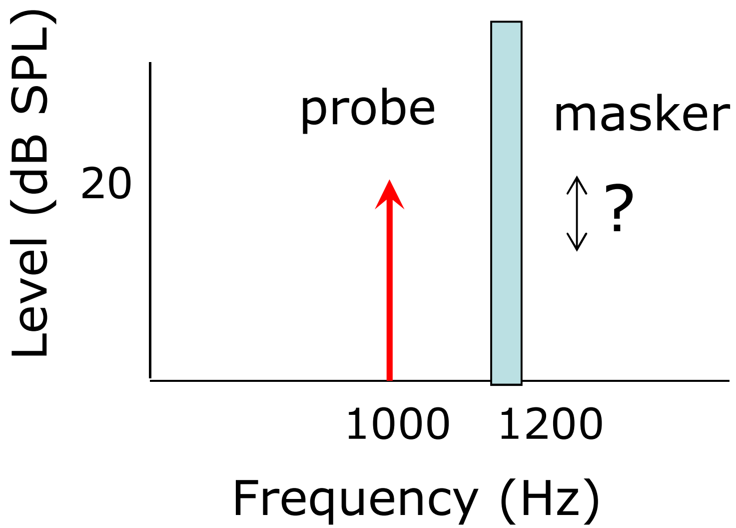


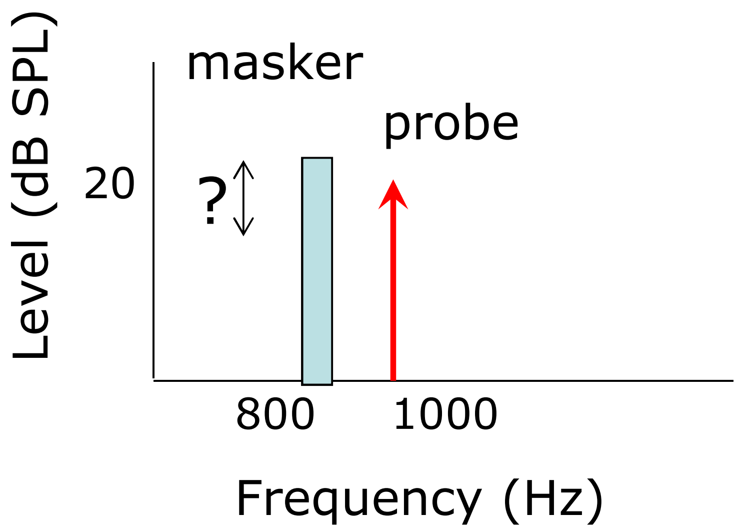

Figure 8.1: Results of Pavlovic (1984) comparing speech recognition scores of hearing-impaired subjects with predictions based on the AI. Each number represents the mean score across subjects for a specific condition of filtering/background noise. For subjects with mild losses, the predictions are accurate (left panel); for subjects with more severe losses, the obtained scores fall below the predicted values (right panel)

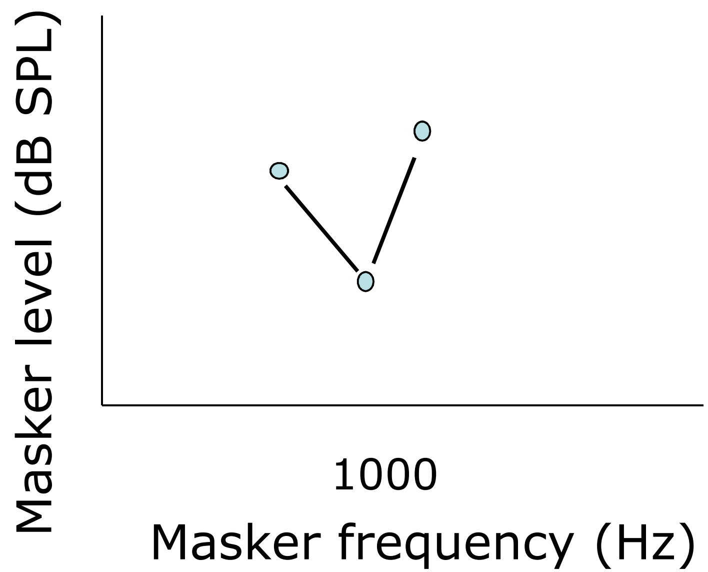
# ‘Dead’ regions: An extreme case of increased threshold

- Regions in the inner ear with absent or non-functioning inner hair cells (IHCs)
- No BM vibrations in such regions are directly sensed
- But spread of BM vibration means that tones can be detected ‘off-place’
  - by auditory nerve fibres typically sensitive to a different frequency region
- Most clearly seen when measuring PTCs
  - directly interpretable

# Psychophysical tuning curves (PTCs)

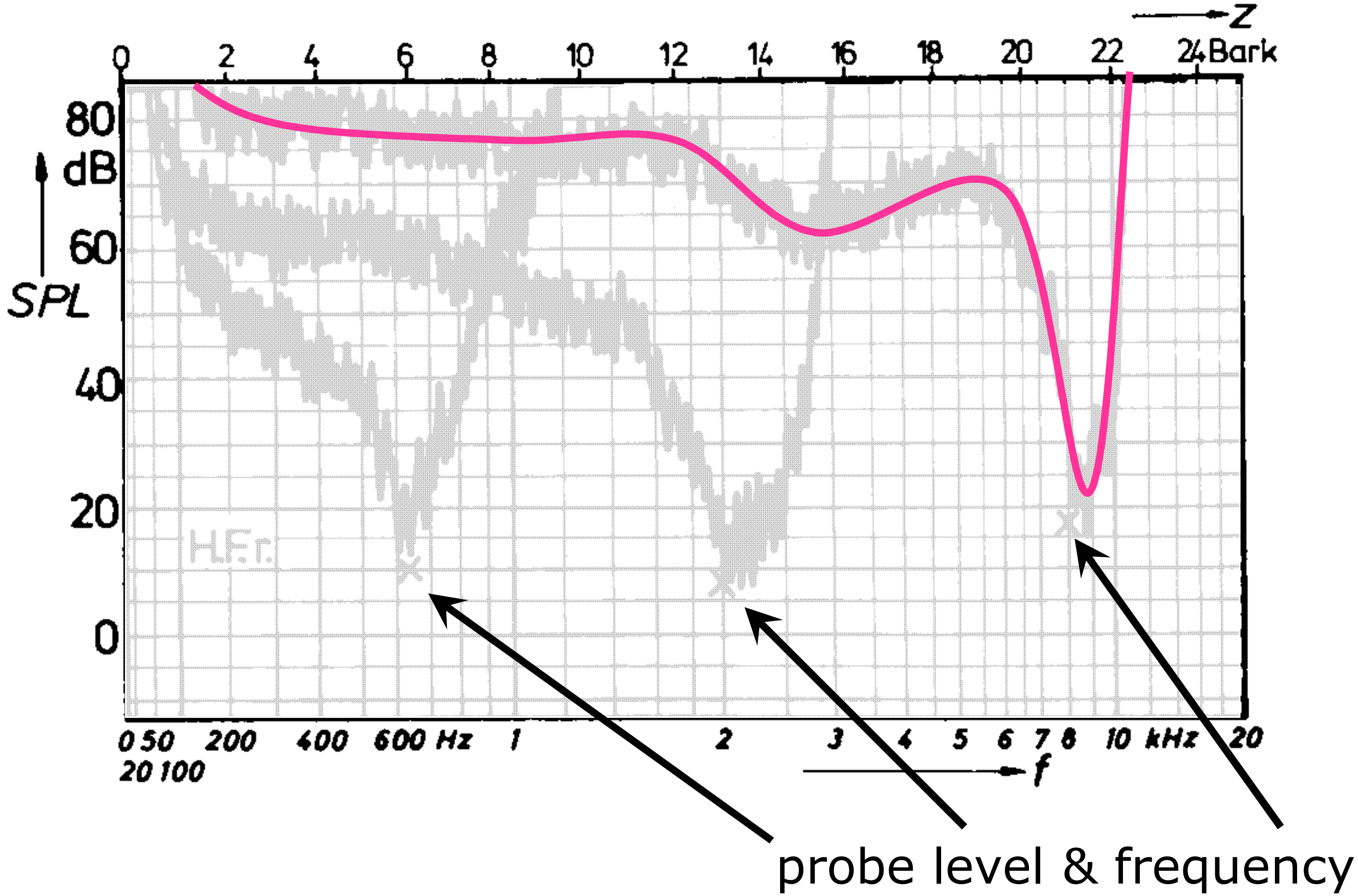


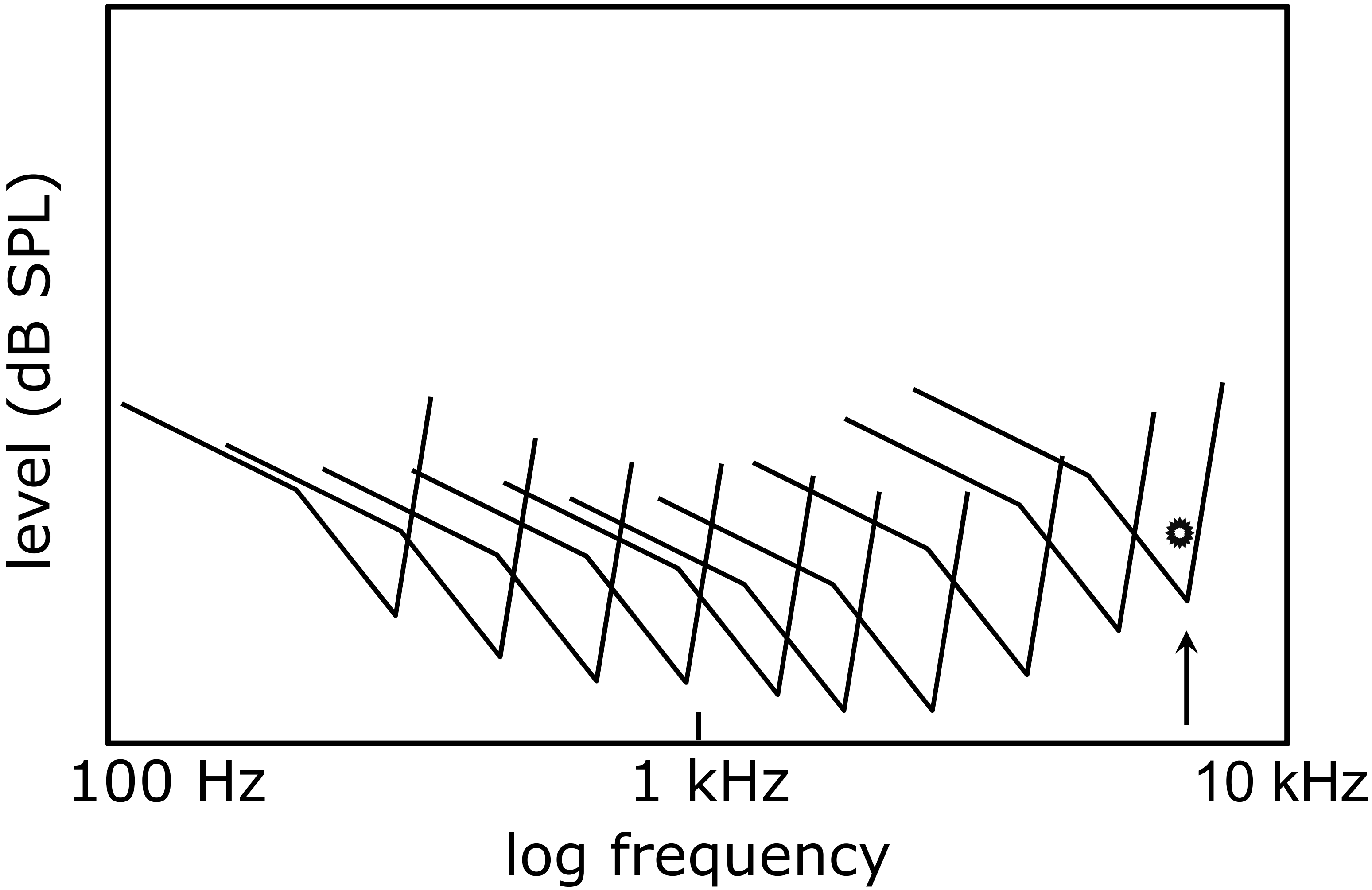



# Psychophysical tuning curves (PTCs)

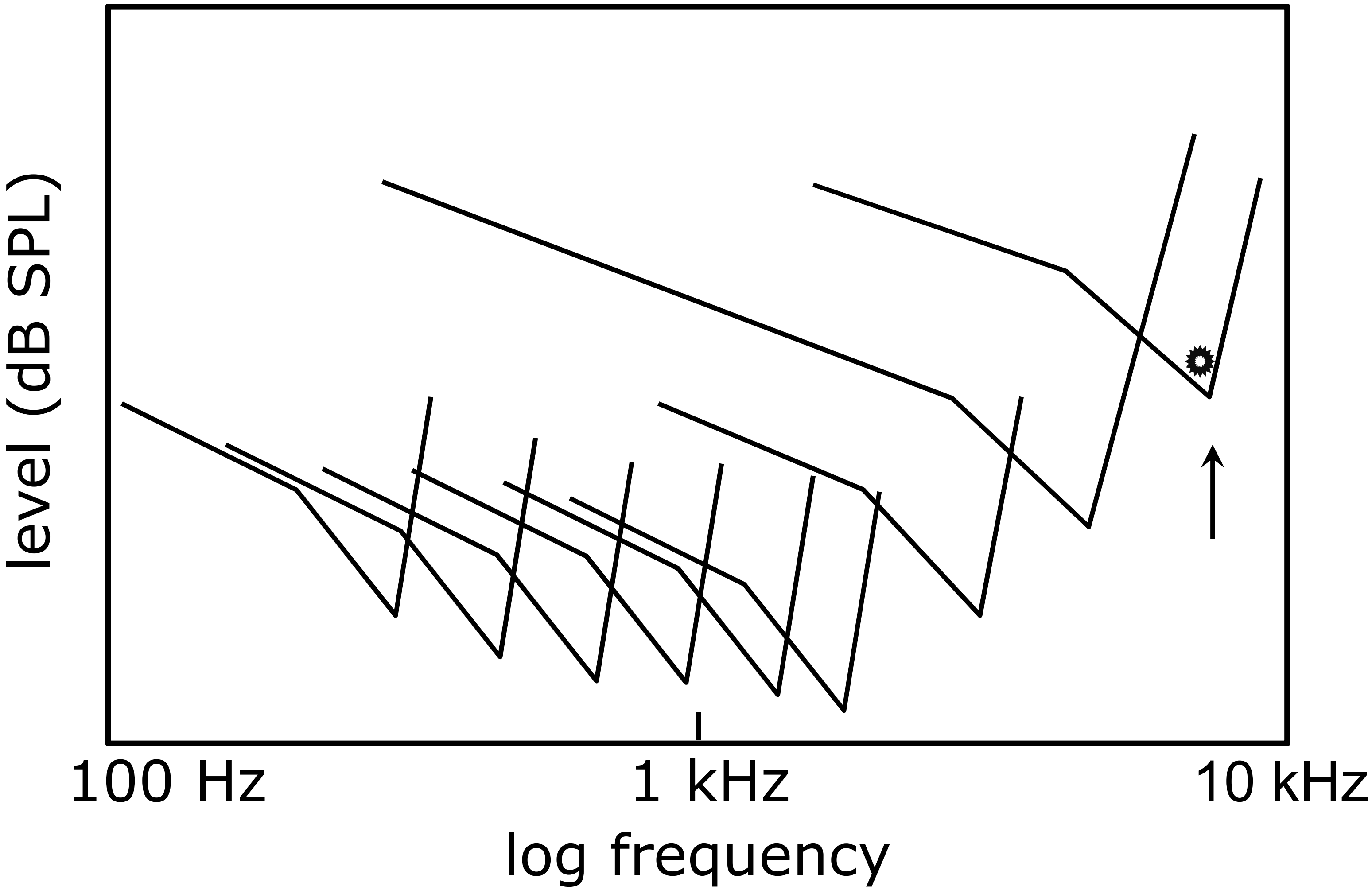
*Determine the minimum level of a narrow-band masker at a wide variety of frequencies that will just mask a fixed **low-level** sinusoidal probe.*



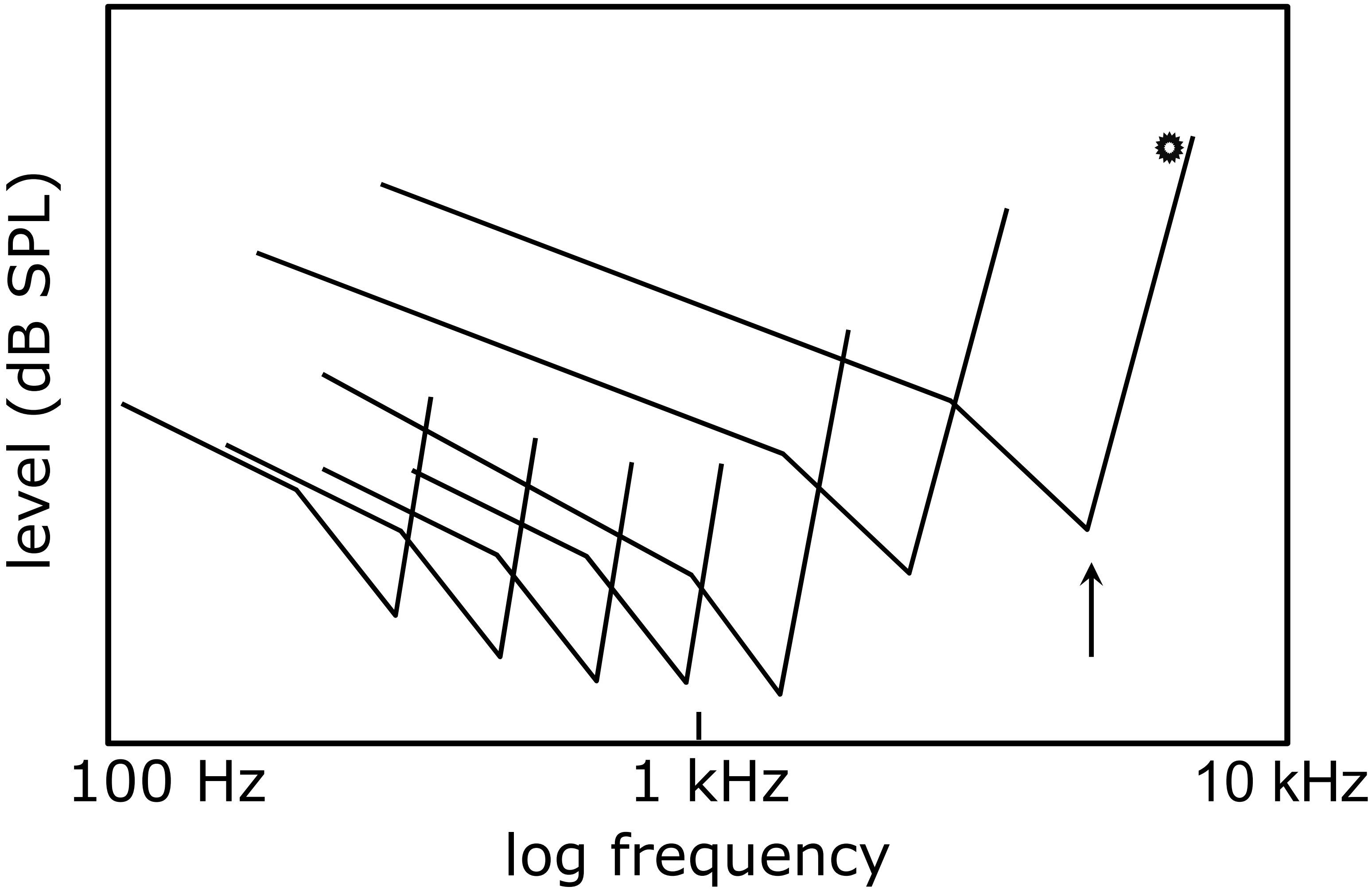
# Physiological TCs for a range of auditory nerve fibres: Normal hearing



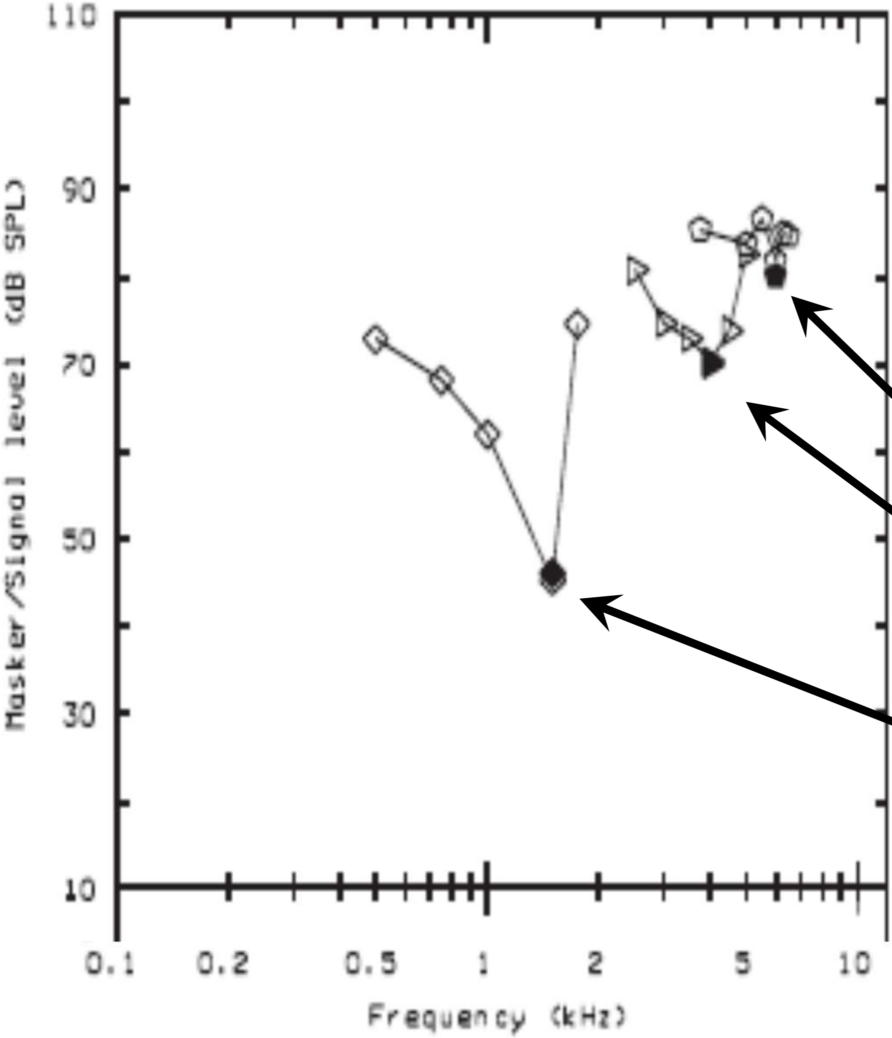
# Hearing loss *without* a dead region

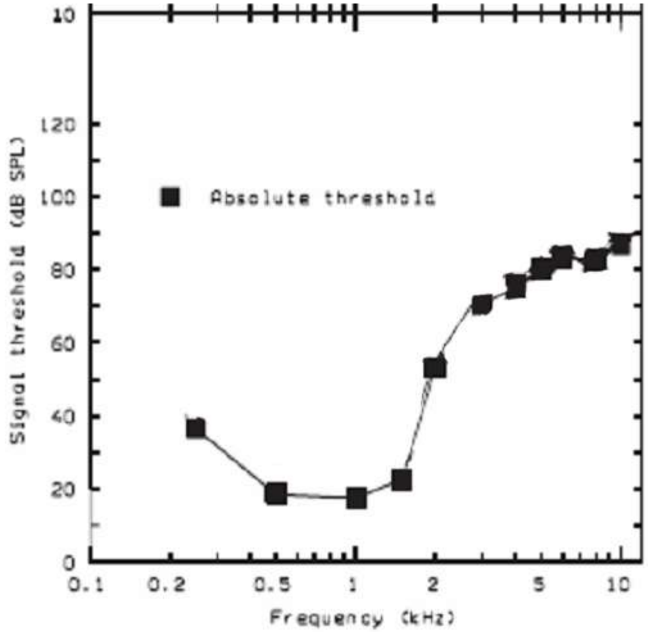


# Hearing loss *with* a dead region

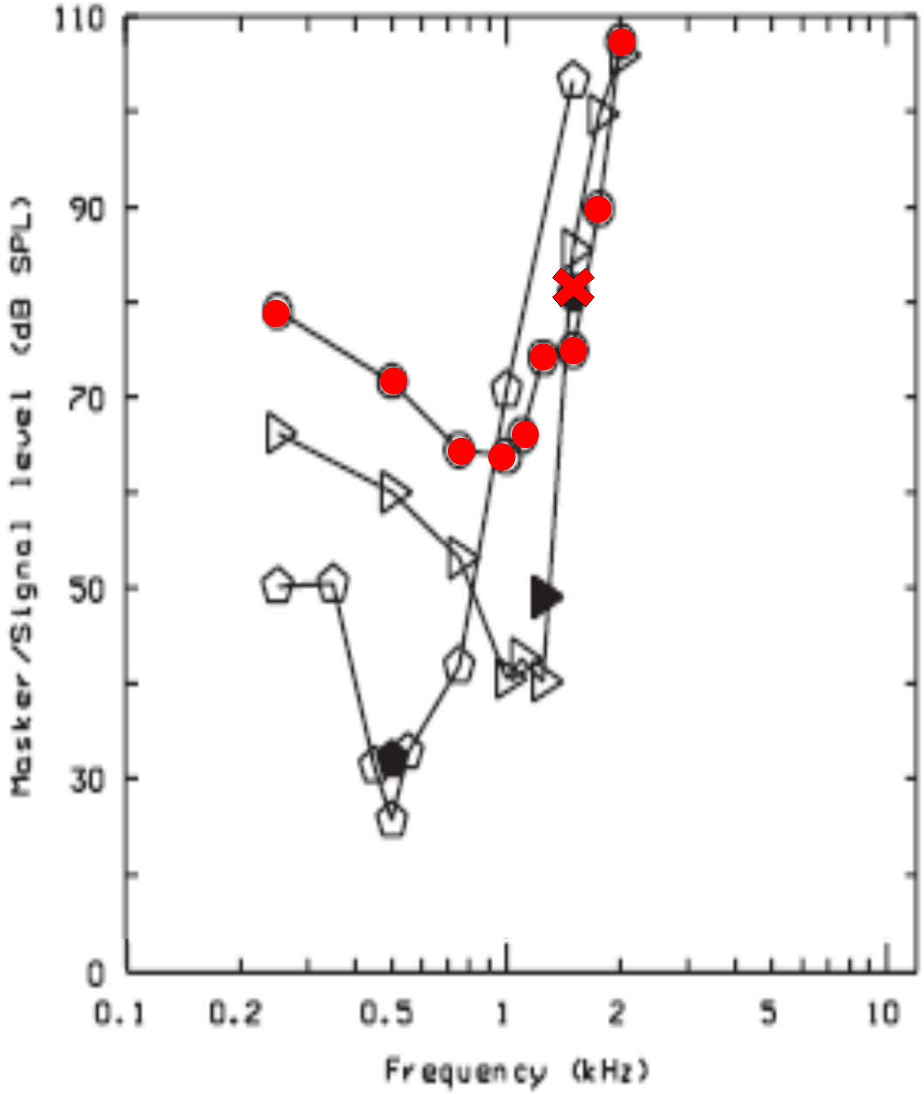


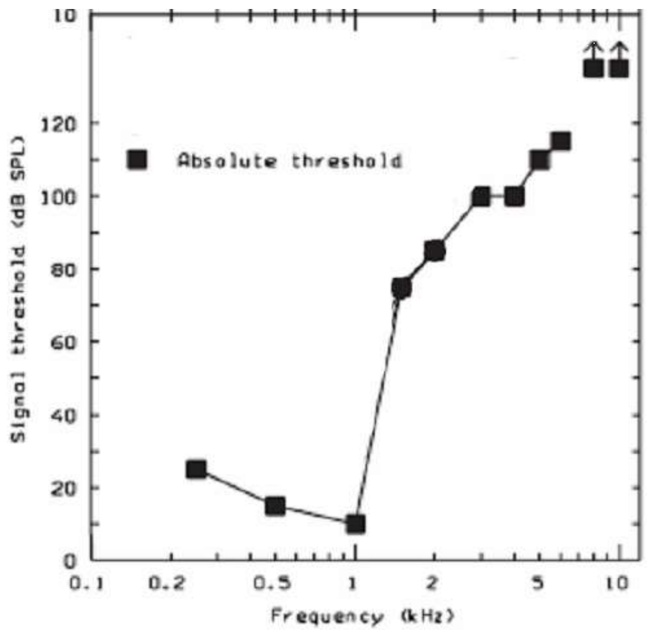
# SNHL without dead region: PTCs



# SNHL with dead region: PTCs



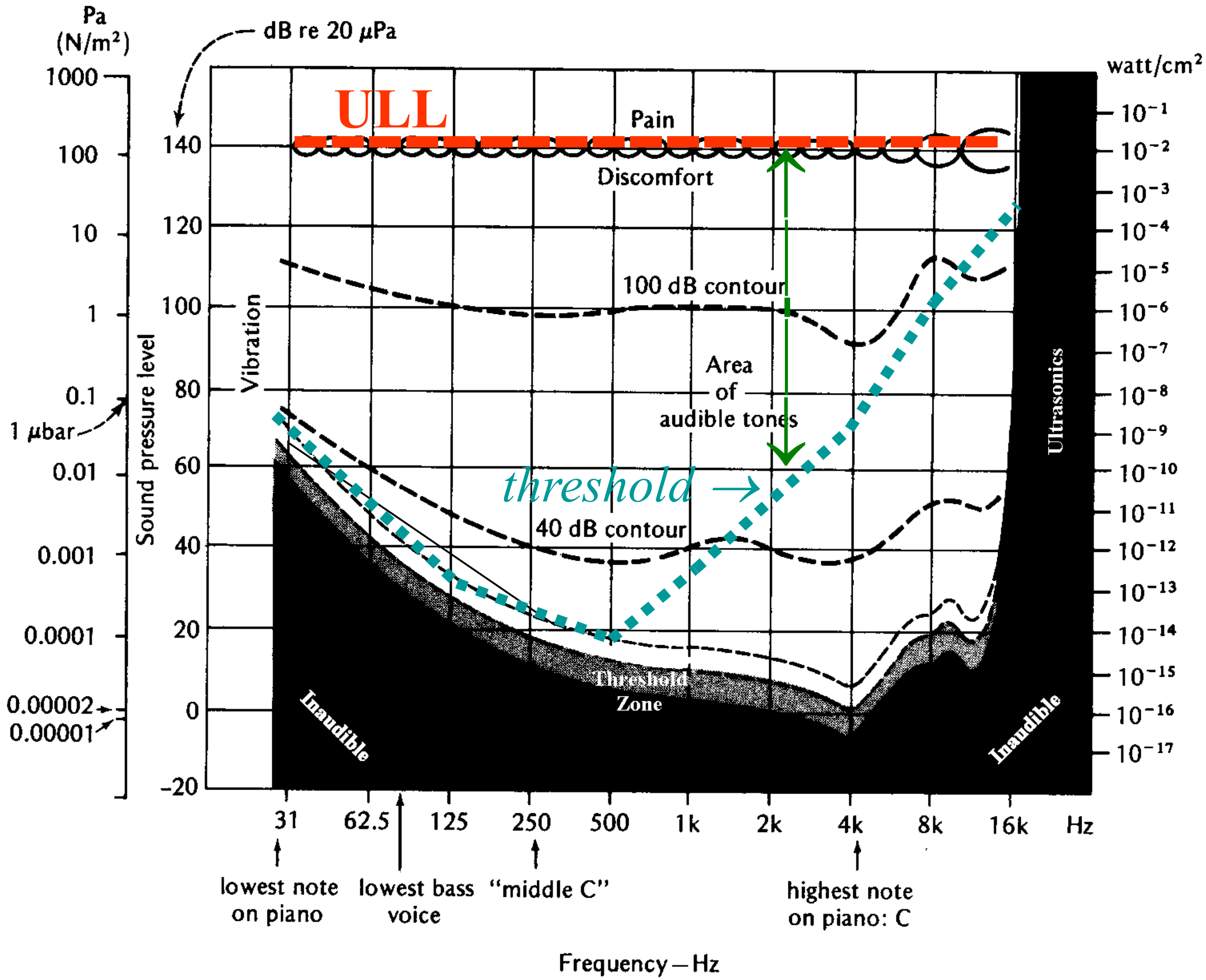
# Diagnosing dead regions

- PTCs perhaps clinically impractical
- TEN test (threshold equalizing noise)

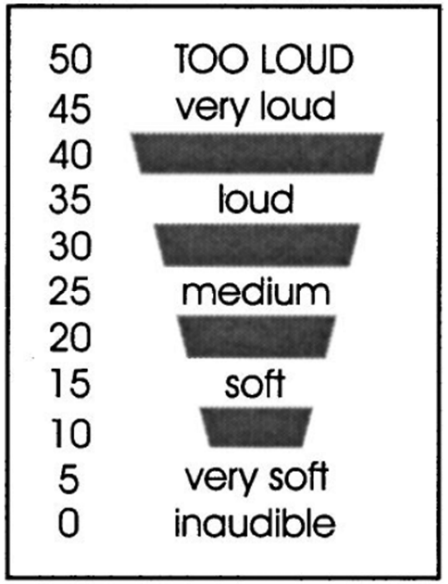
# Audibility accounts don't explain everything

- Good predictions of speech intelligibility from audibility hold only for mild to moderate hearing losses
- Complete restoration of audibility with more severe losses cannot restore intelligibility
- And these predictions only hold for speech in quiet

# Reduced dynamic range in sensori-neural hearing loss



# Categorical scaling of loudness
# ACALOS (adaptive categorical loudness scaling)




ACALOS category scale. Subjects do not see the numbers.
Brand and Hohmann (2002) JASA 112, 1597-1604

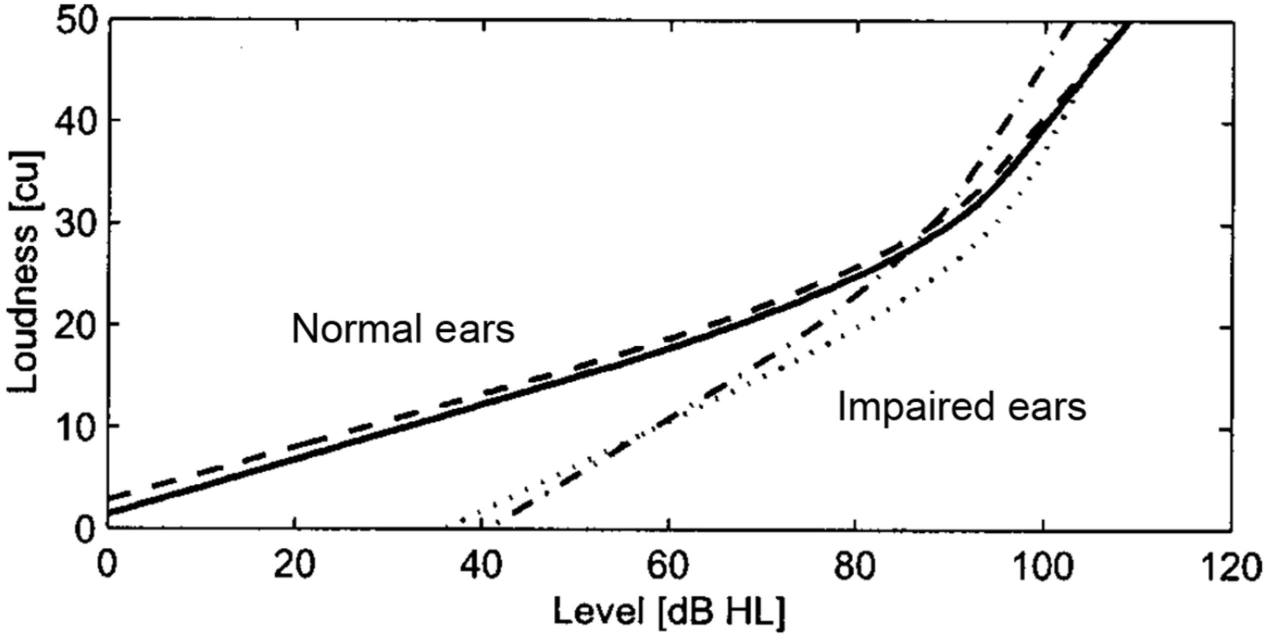



FIG. 5. Loudness functions with the median parameters displayed in Table I. Normal-hearing subjects with adaptive procedure (solid), normal-hearing subjects with constant stimuli procedure (dashed), subjects with hearing impairment with adaptive procedure (dotted), subjects with hearing impairment with constant stimuli procedure (dash-dotted).

Recruitment requires compression as well as amplification to maximize audibility

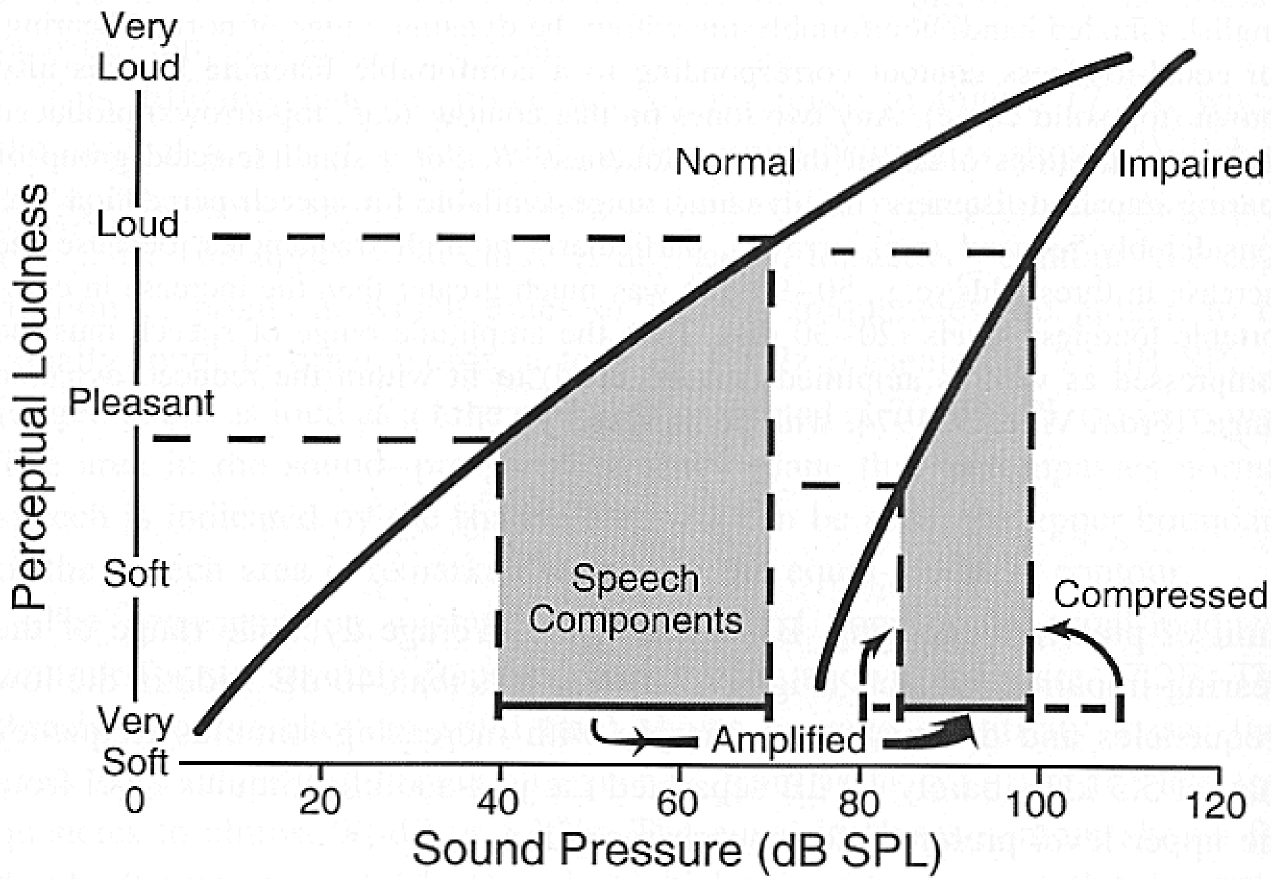
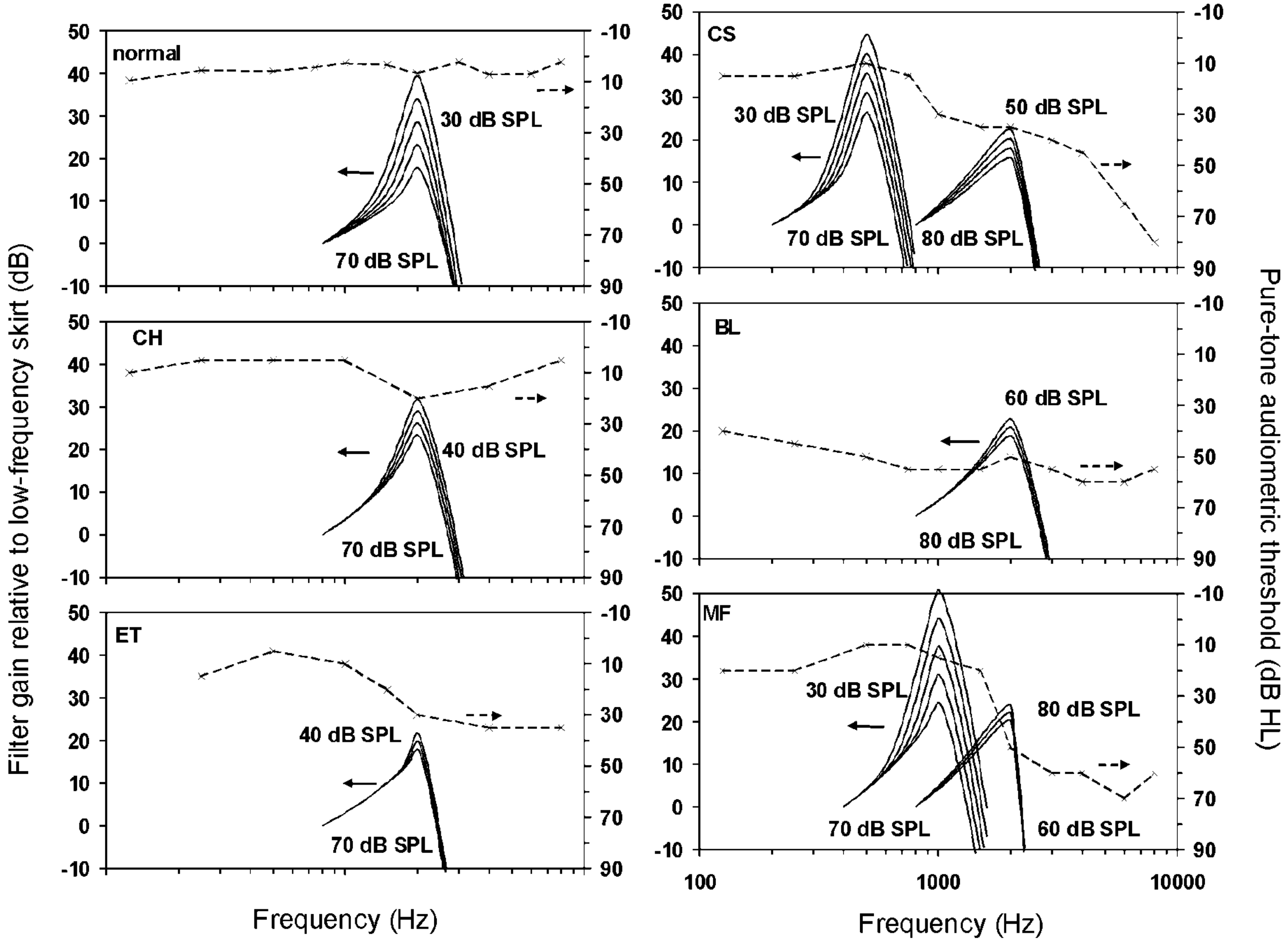
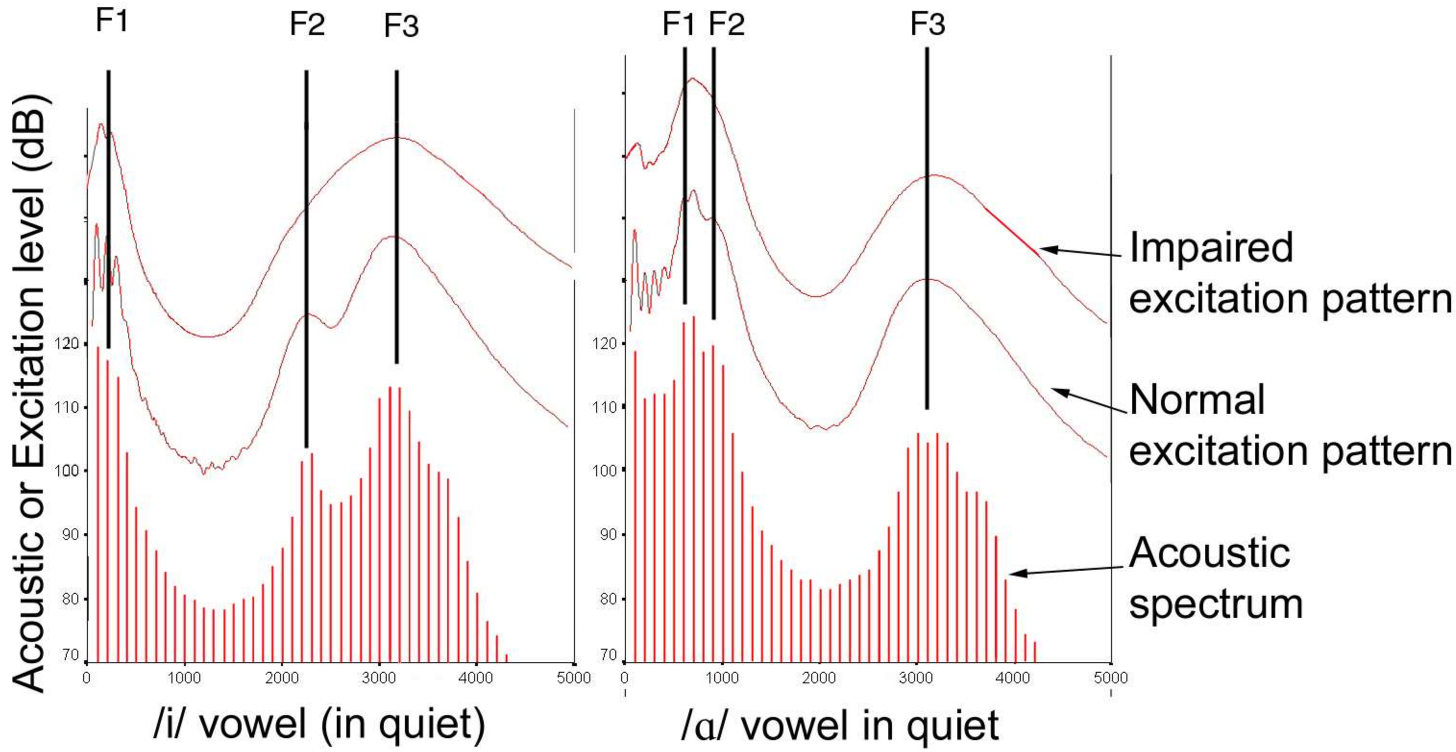



**Figure 17.3.** Idealized relations between sound pressure and perceptual loudness for subjects with normal hearing (left curve) and those with severely impaired hearing (right curve) for a representative band of frequencies (e.g., around 2 kHz). To produce the same levels of subjective loudness as those experienced by normally hearing listeners, speech for the hearing impaired must be both amplified *and* compressed. (Adapted from Pluvinage, 1994.)

# Changes in frequency selectivity reflect loss of nonlinearity
# Rosen & Baker (2002)

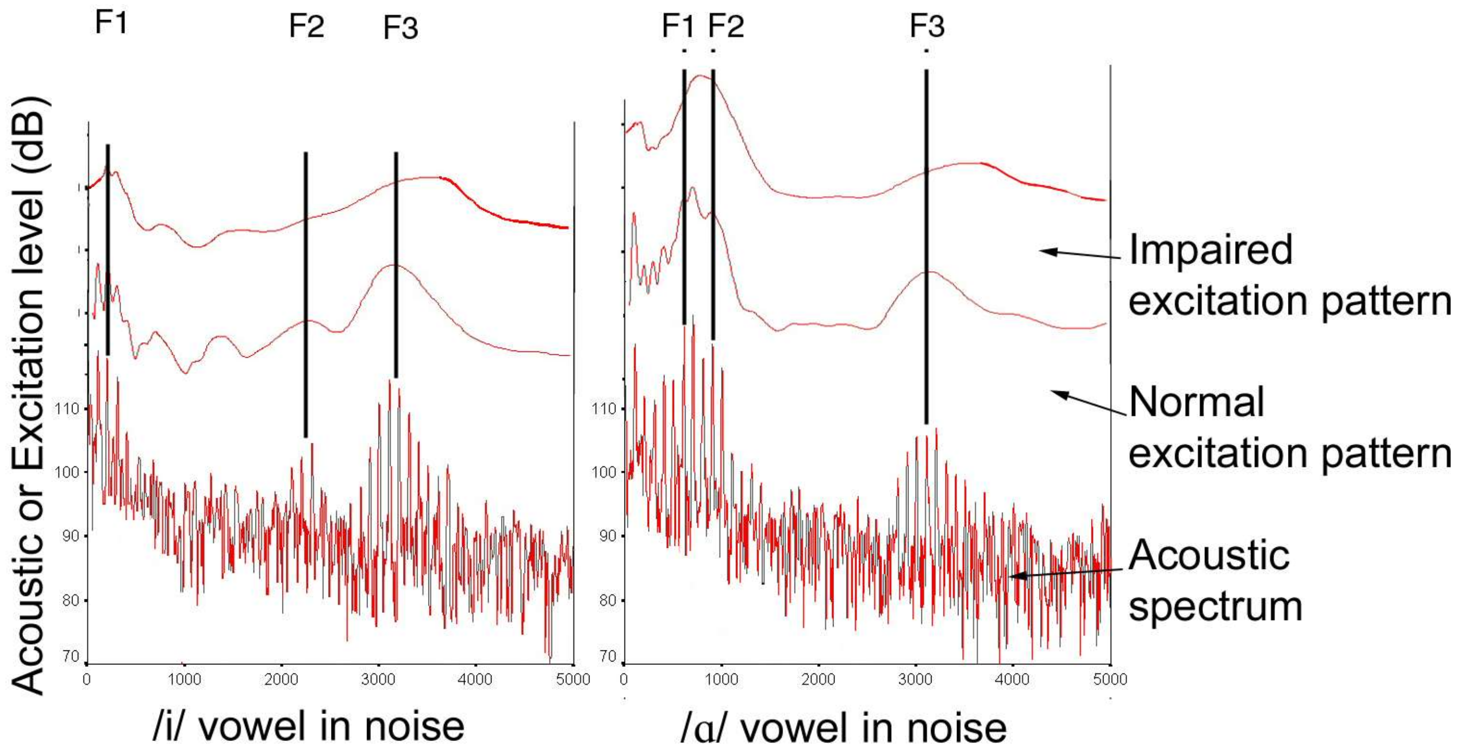


# Normal compared to impaired excitation patterns - quiet




Impaired excitation pattern - retains much of formant structure in quiet

# Normal compared to impaired excitation patterns - noise




SNR = +6 dB

Normal excitation pattern retains much of formant structure in noise

Impaired excitation pattern has reduced formant structure in noise

# What can current hearing aids do for ...

- Hearing loss
  - *?*
- Reduced dynamic range & loudness recruitment
  - *?*
- Degraded frequency selectivity
  - *?*
- Dead regions
  - *?*

# Hearing speech in noise

# Essential terminology

- ***signal*** or ***target***
  - what you are trying to listen to
  - typically speech or music or …

- ***'noise'*** or ***masker***
  - what you are trying to ignore
  - can be noise like from a hoover, but also other speech

- ***signal to noise ratio***
  - The amount of energy in the signal divided by the amount of energy in the noise
  - expressed in dB

# Why is listening to speech in noisy backgrounds interesting?

- Most speech is not heard in quiet.
  - Classrooms can be really noisy.
- People vary a lot in how well they can understand speech in the presence of other sounds.
- Difficulties in understanding speech in noise are a very common complaint in the clinic
- Lots of developmental disorders seem to have an impact on this ability
  - Language impairment
  - Autism spectrum disorders
  - Auditory processing disorder (APD)?
- Hearing impairment makes perceiving speech in noise difficult.
  - Cochlear implant users have great difficulties
- Being a non-native speaker makes it harder
- Effects of age
  - Ageing itself (≥60 y.o.) may lead to poorer speech perception in certain kinds of noise.
  - Younger children (≤12 y.o.) appear to be more affected by certain kinds of noise

# Some determinants of performance: I

- The nature of the target speech material
  - context
    - e.g., the so-called SPIN test, Kalikow *et al*., 1977
      - Throw out all this useless **junk** …
      - We could have discussed the **junk** …
  - number of alternative utterances
    - listening for digits when given a telephone number vs. an individual’s name
    - ‘easy’ (*mouth*) vs ‘hard’ (*mace*) words (see Bradlow & Pisoni, 1999)
      - tied to frequency of usage and size of lexical ‘neighbourhoods’

# Some determinants of performance: II

- The nature of the background noises
  - level (SNR)
  - spectral characteristics
  - genuine ‘noise’: periodic or aperiodic?
  - and/or other talkers
    - how many there are
    - speaking your own language or a language you don’t know
  - How ‘attention-grabbing’ the background noises are

# Some determinants of performance: III

- The configuration of the environment
  - Open air or in a room?
  - How ‘dry’ is a room?
    - effects of reverberation
  - spatial separation between target and noise
- or, the transmission system (e.g. mobile telephone)
  - distortion, reverberation, noise

# Some determinants of performance: IV

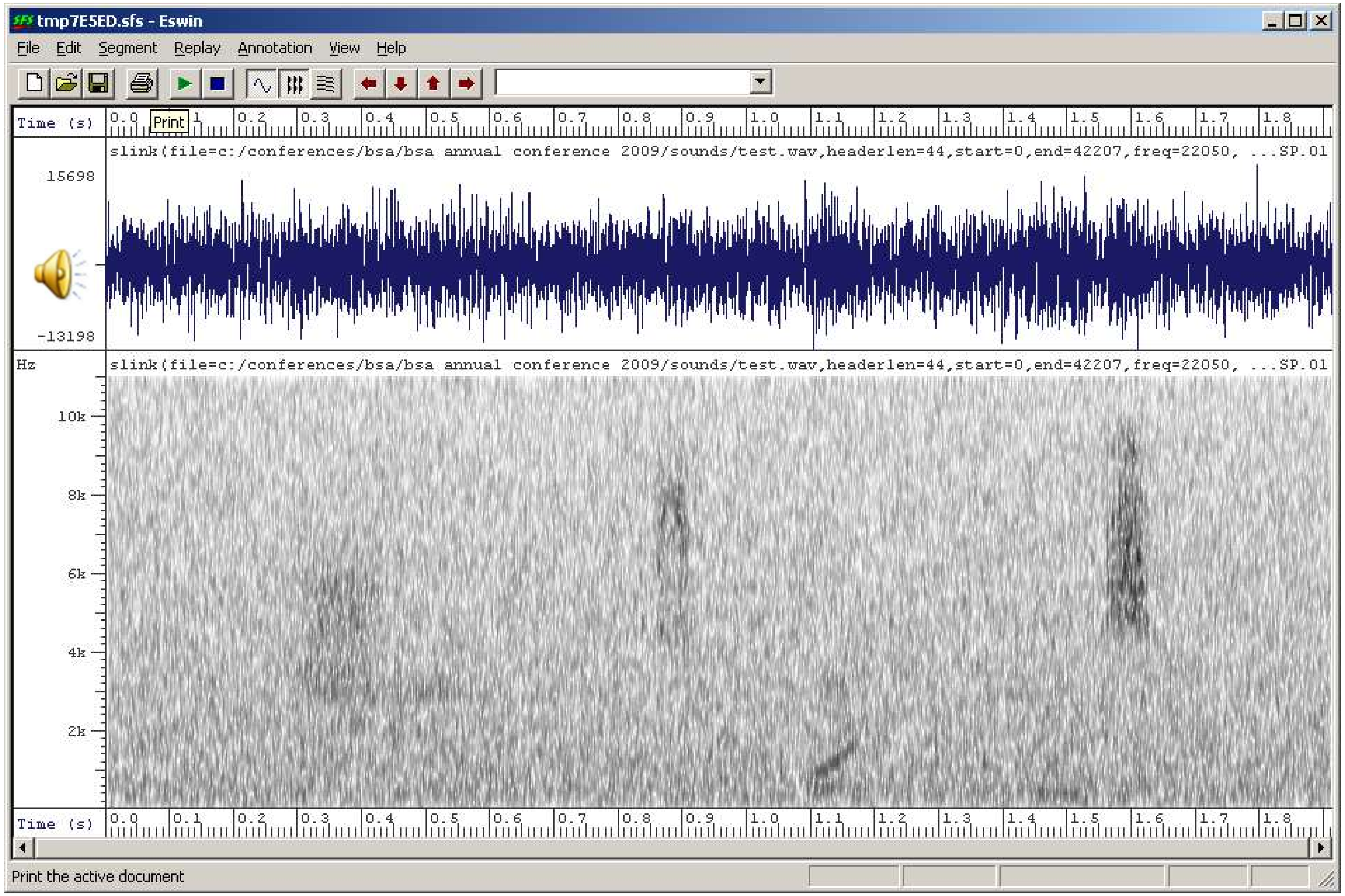
- Talker characteristics
  - Talkers vary considerably in intrinsic intelligibility
  - Talkers can vary their own speech depending upon demands of the situation (hyper/hypo distinction of Lindblom, 1990)
    - manipulations in vowel space, prosody, rate
  - Match between talker and listener accents
  - Individual familiarity

# Some determinants of performance: V

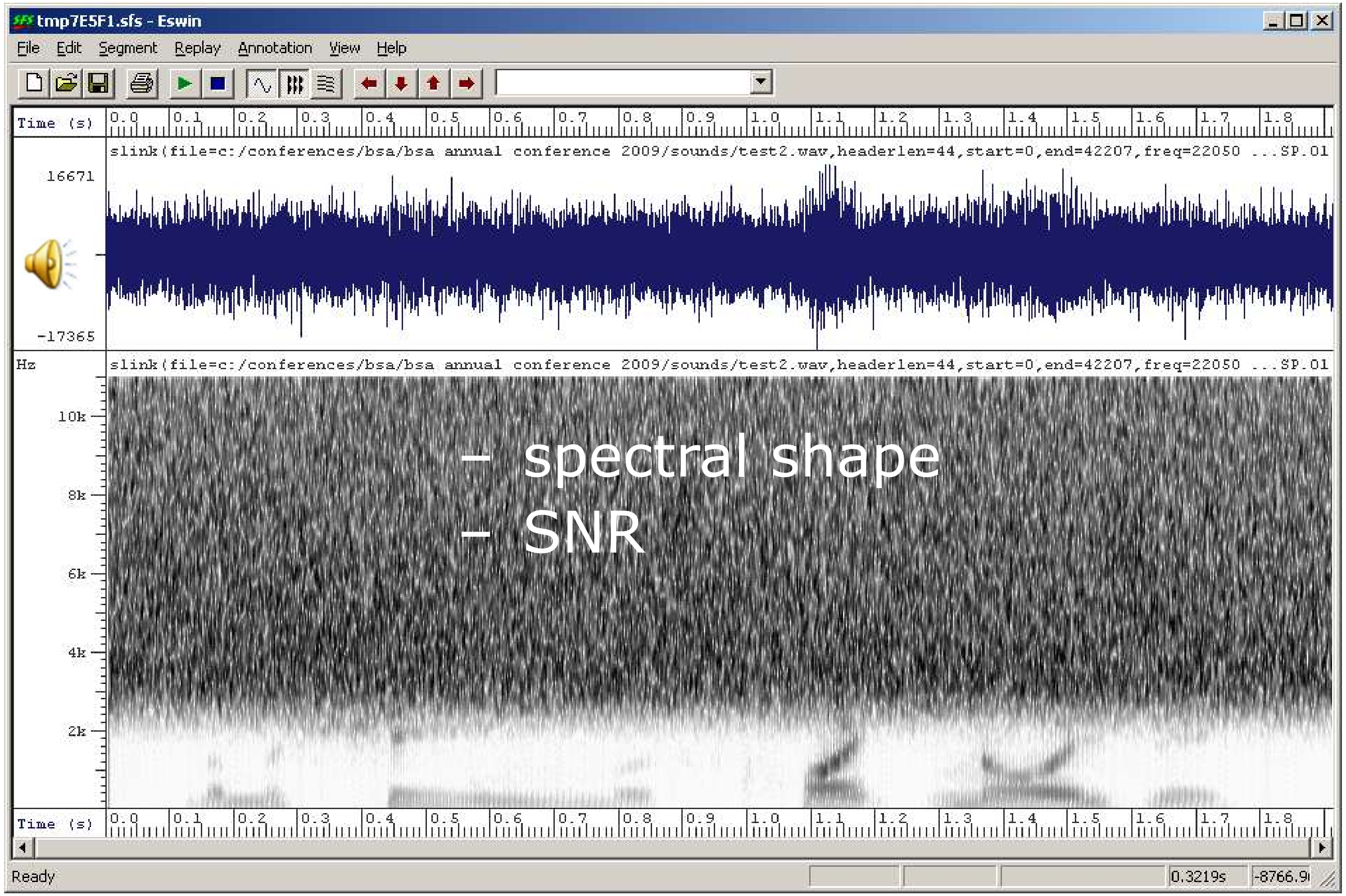
- Listener characteristics
  - Linguistic development
    - L1 vs L2
    - vocabulary knowledge
    - ability to use context
  - Hearing sensitivity and any hearing prosthesis used
  - Cognitive abilities
    - working memory
    - linguistic closure skills: piecing together a sensible message from incomplete information

Focus on factors more centrally related to audiology

# The simplest case:
# A steady-state background noise



# Much is understood about what makes one steady noise more or less interfering than another

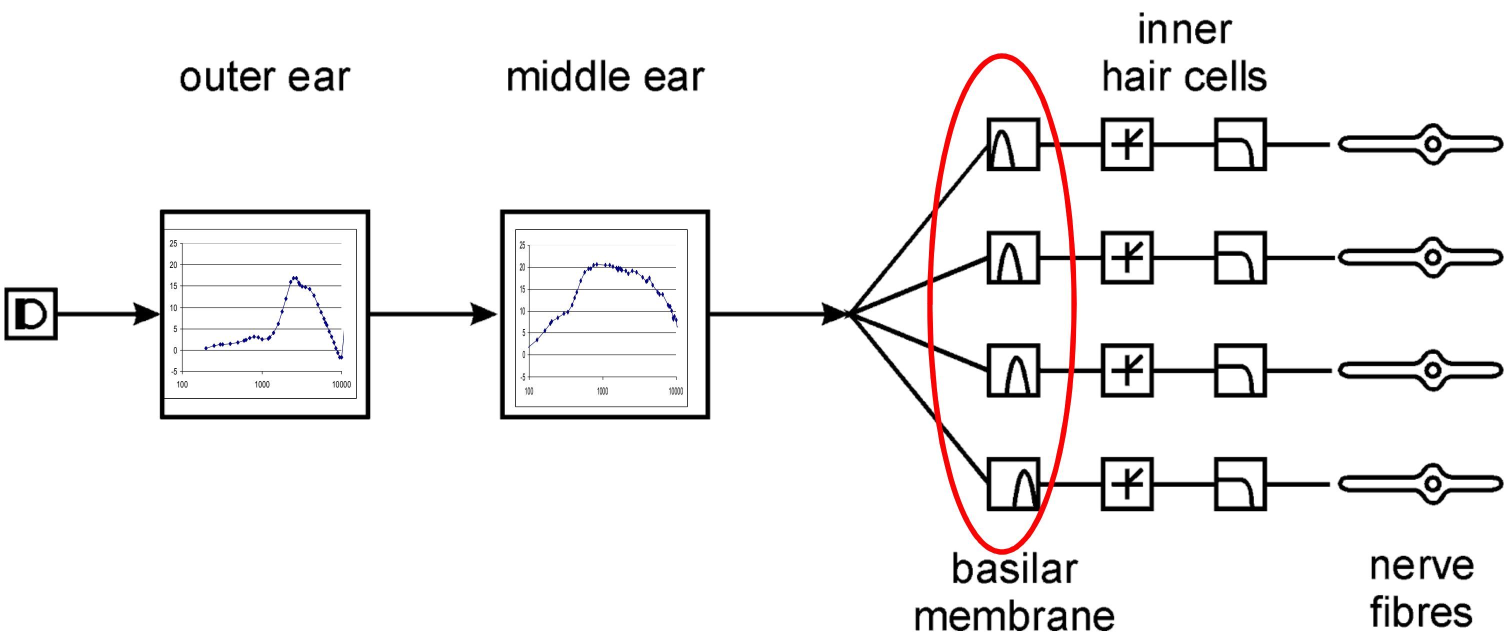



– spectral shape
– SNR

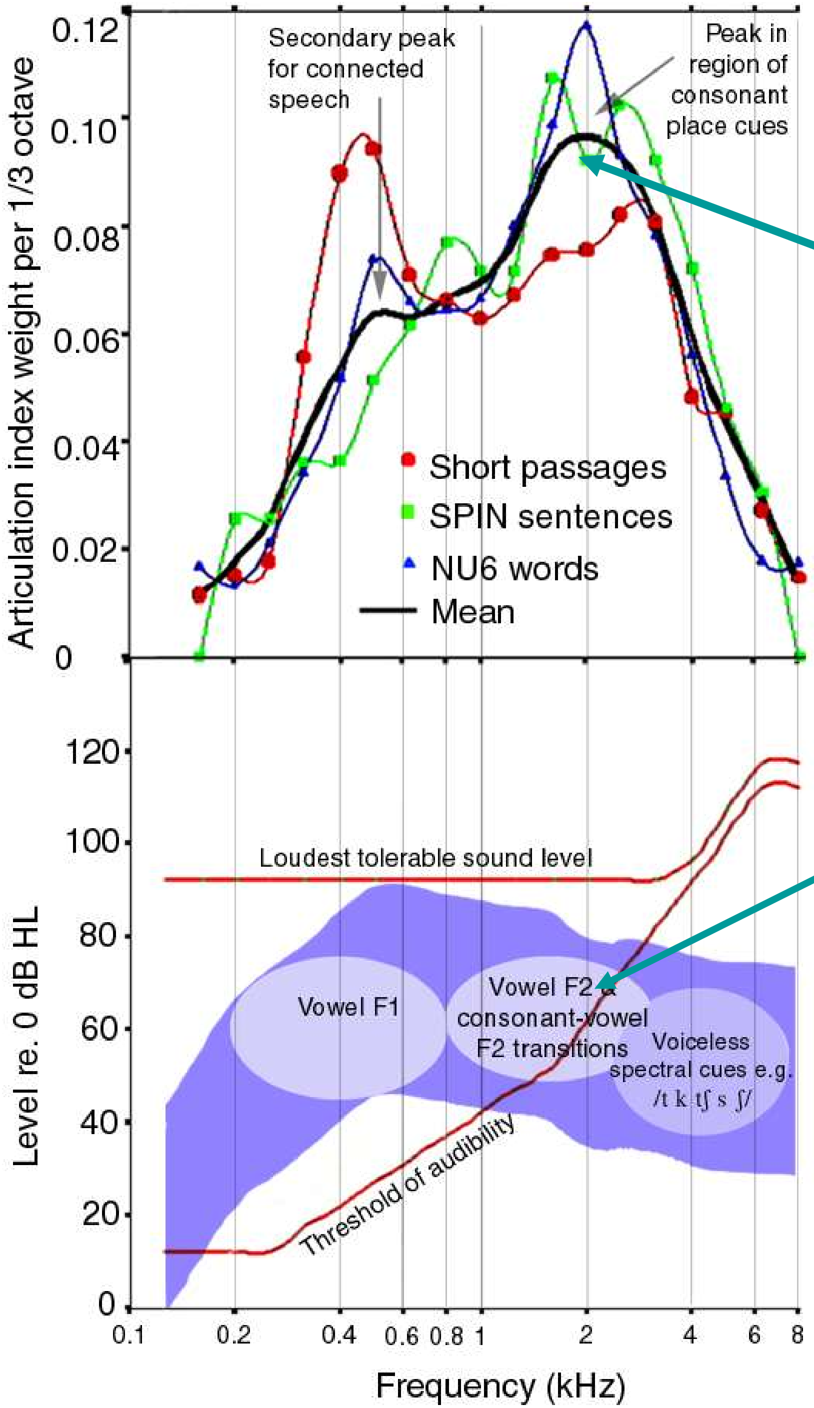
# ‘Energetic’ masking

- Noises interfere with speech to the extent that have energy in the same frequency regions
- Can be quantified in the ‘articulation index’
- Reflects direct interaction of masker and speech in the cochlea, which acts as a frequency analyser
- Hearing impaired listeners are more affected by steady noises …
  - because they typically have impaired frequency selectivity (wider auditory filters).

# Better frequency selectivity keeps noise in its place



# Frequency importance weightings: AI




–I (2000 Hz)

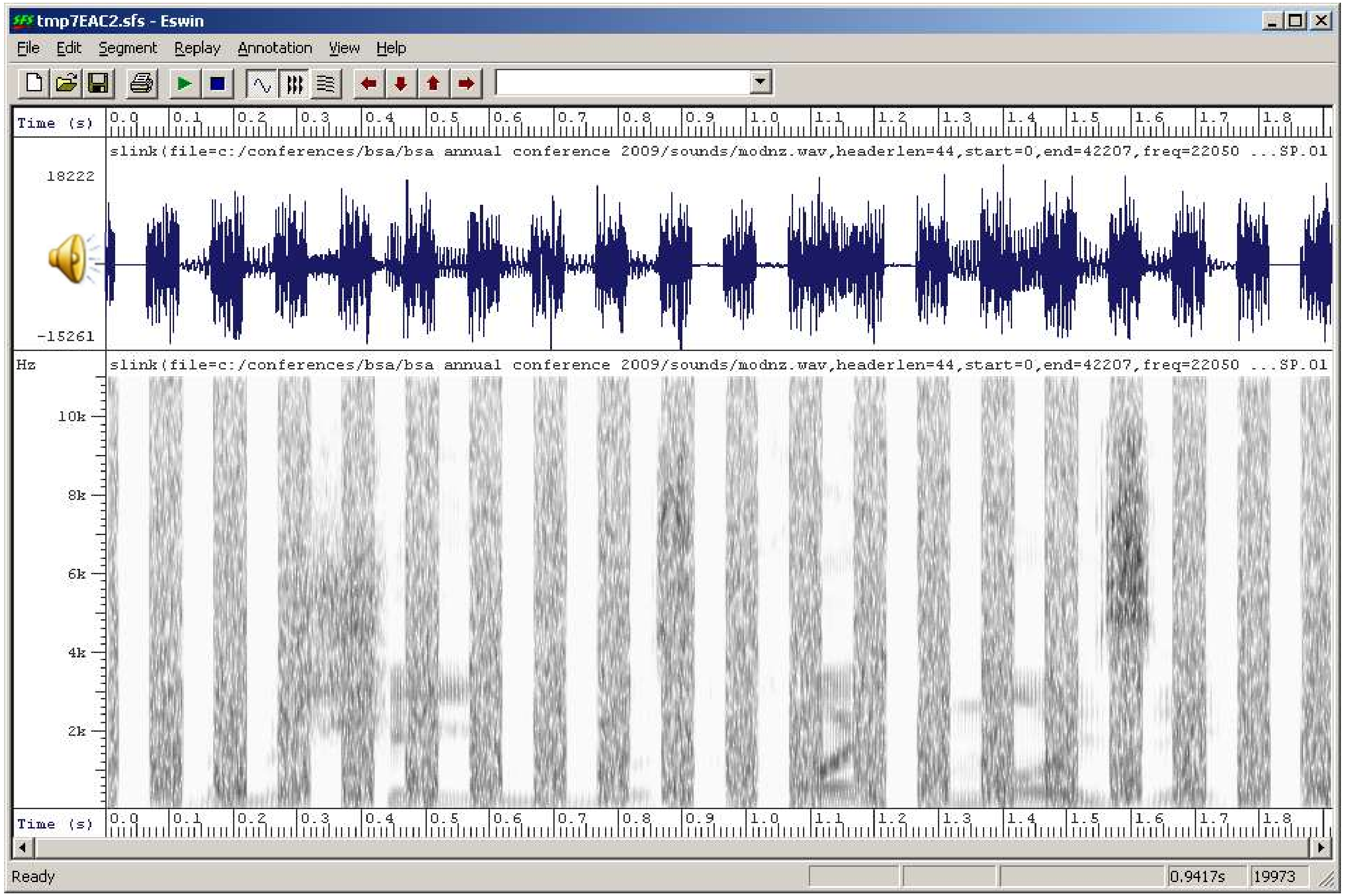
$$A = \sum_{i=1}^{n} I_i \quad W_i,$$

–W (2000 Hz) – here W is approx 0.6

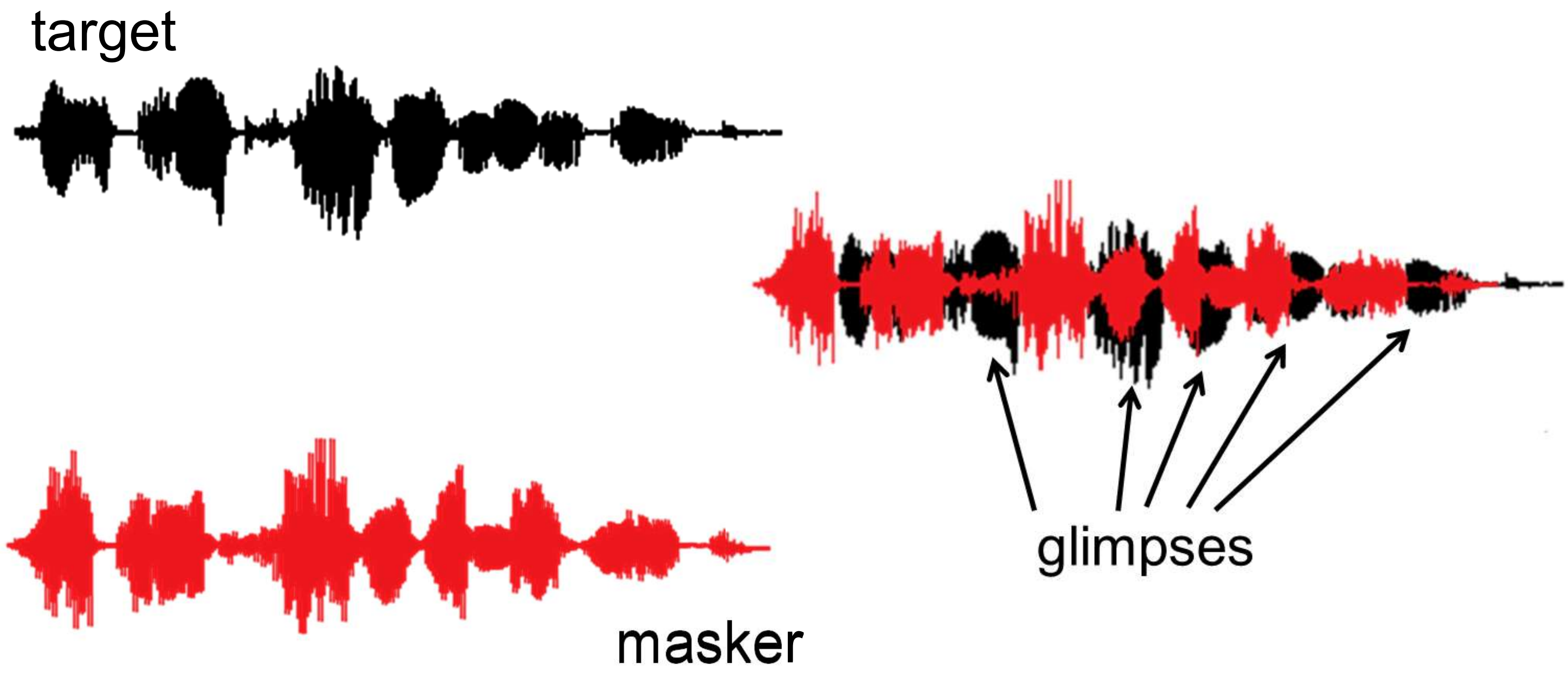
–***A*** is the Articulation Index (predicted intelligibility).

–***A*** is determined by adding up ***W x I*** over frequency bands, where ***I*** is the band importance weight and ***W*** is the proportion of a 30 dB dynamic range of speech in that band that is audible.

# But noises are typically not steady …



# Fluctuating maskers afford 'glimpses' of the target signal

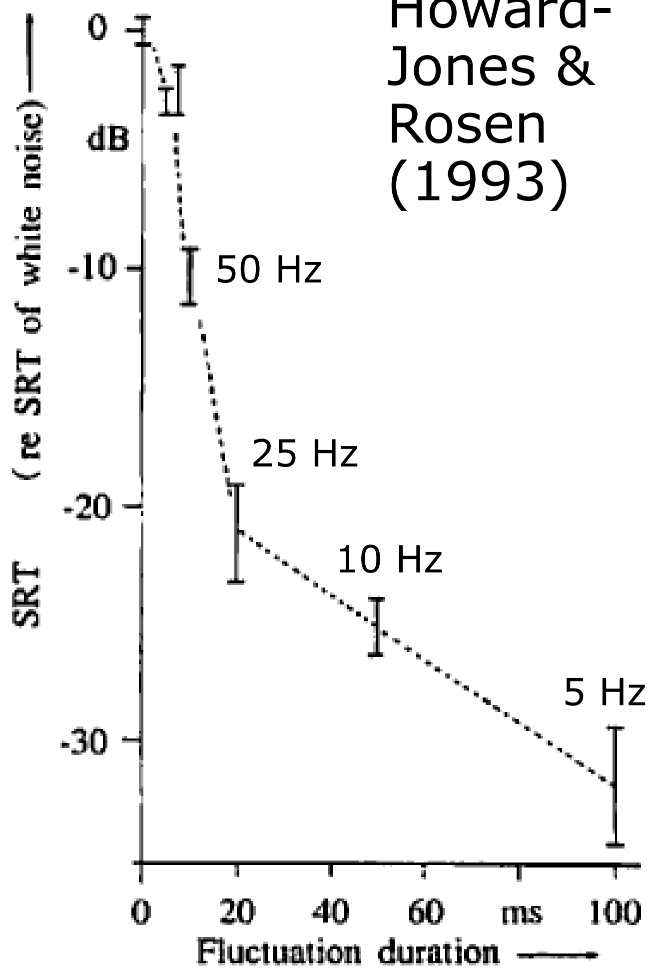


# ‘dip listening’ or ‘glimpsing’

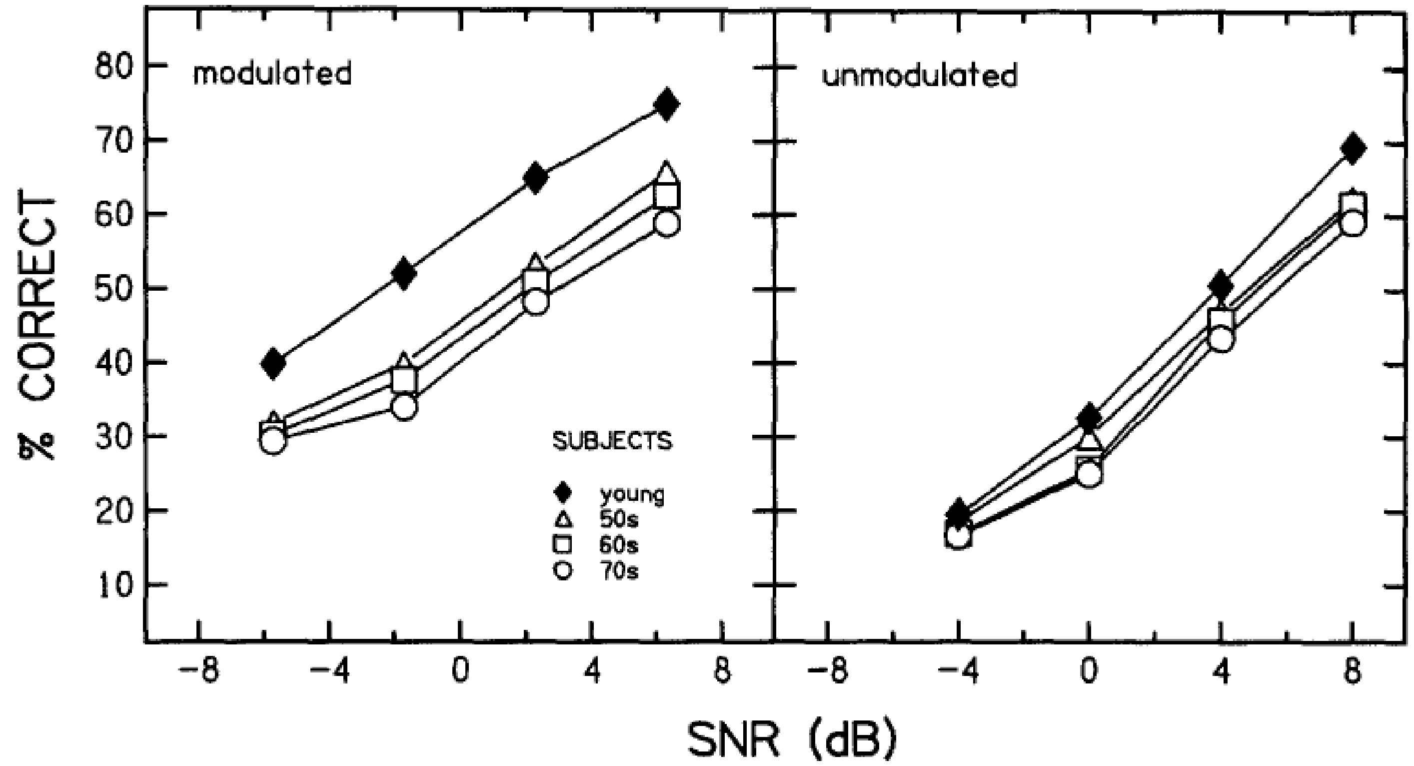
People with normal hearing can listen in the ‘dips’ of an amplitude modulated masker

The speech reception threshold for consonants in simple on/off fluctuations as a function of the duration of the fluctuation.

Howard-Jones & Rosen (1993)



# Hearing impaired listeners have limited ‘glimpsing’ capabilities




Performance in the SPIN task as a function of SNR for modulated and unmodulated noises (not an effect of ageing) Takahashi & Bacon (1992)

# Takahashi & Bacon (1992)

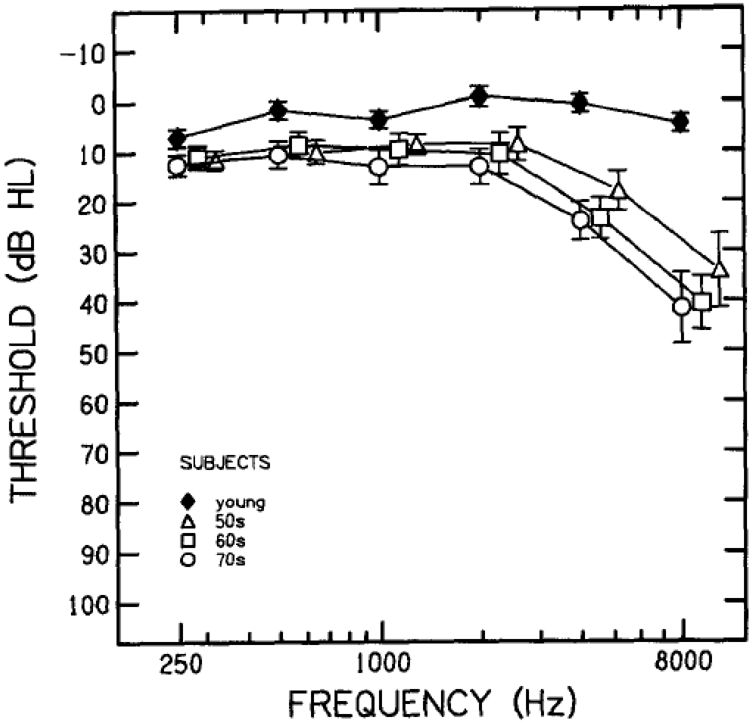



**FIGURE 1. Mean pure-tone audiometric thresholds (in dB HL) for each subject group. The three older groups are represented by open symbols and the young group by closed symbols. Error bars indicate ±1 standard error. Data for the older groups have been shifted horizontally.**

- SPIN low probability sentences
- SAM noise at 8 Hz, 100% modulation

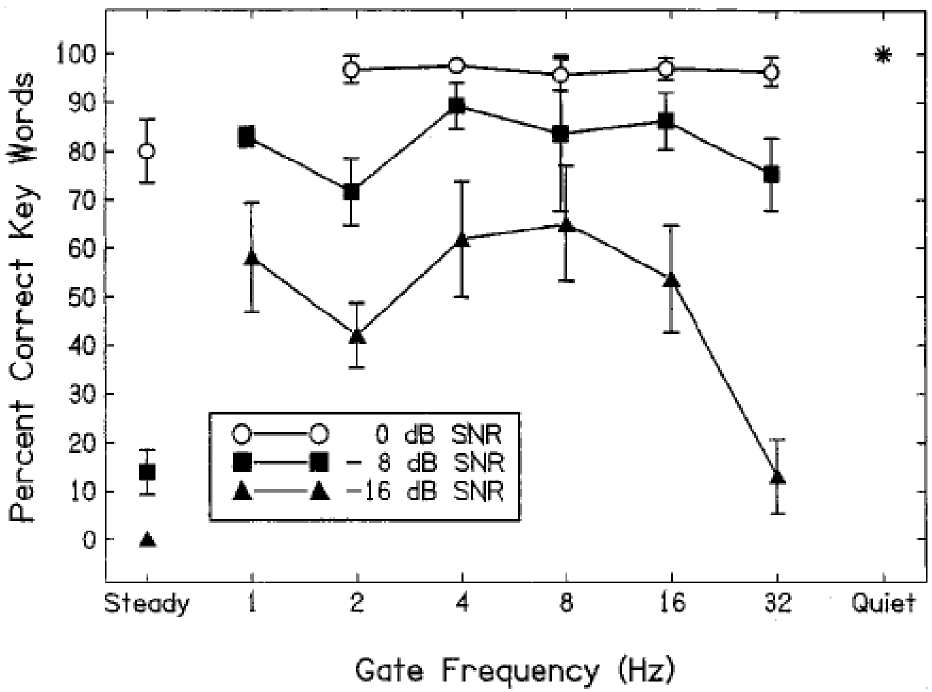
# Why is ‘dip’ listening limited in hearing-impaired listeners?

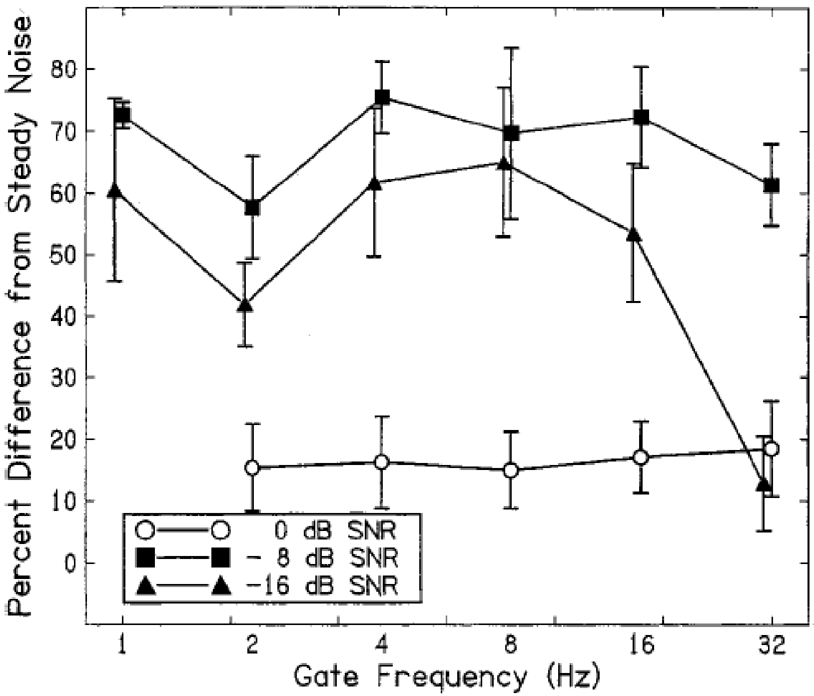
- Audibility can be an influence
- Some of the lack of masking release may be due to SNRs being higher for HI listeners.

# little glimpsing for CI users
# Nelson *et al.* (2003)

## speech-spectrum-shaped masking noise square-wave modulated added to IEEE sentences

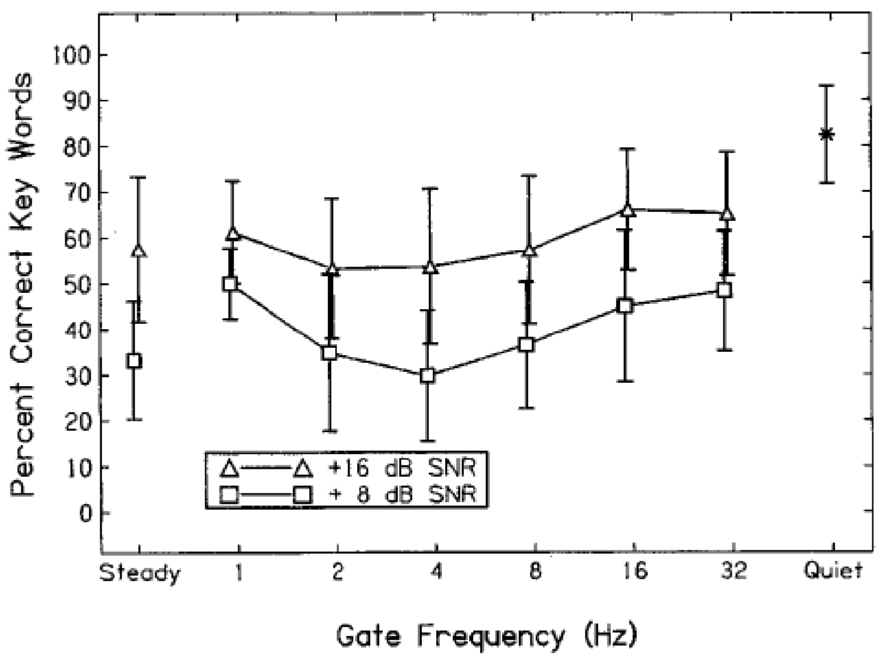
normal listeners

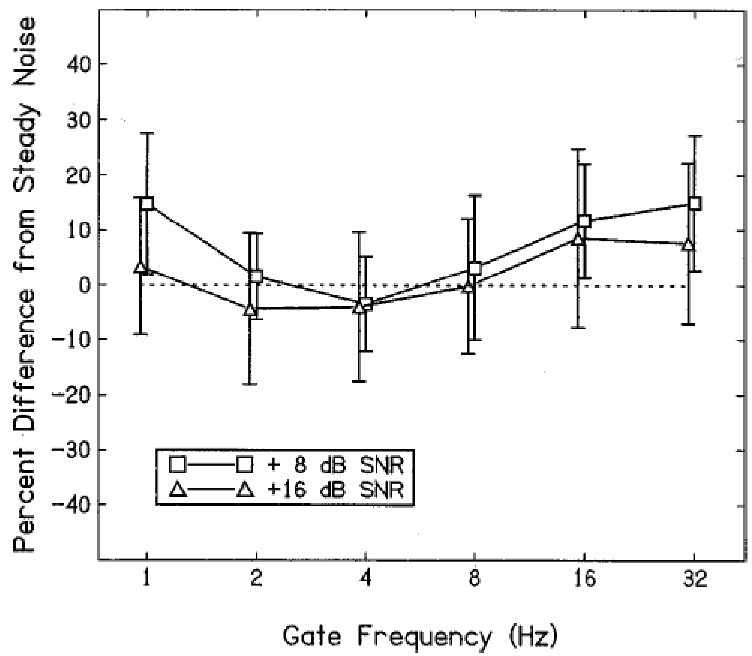


# CI users

Note much higher SNRs
(+8 and +16 vs -8 and -16 dB)



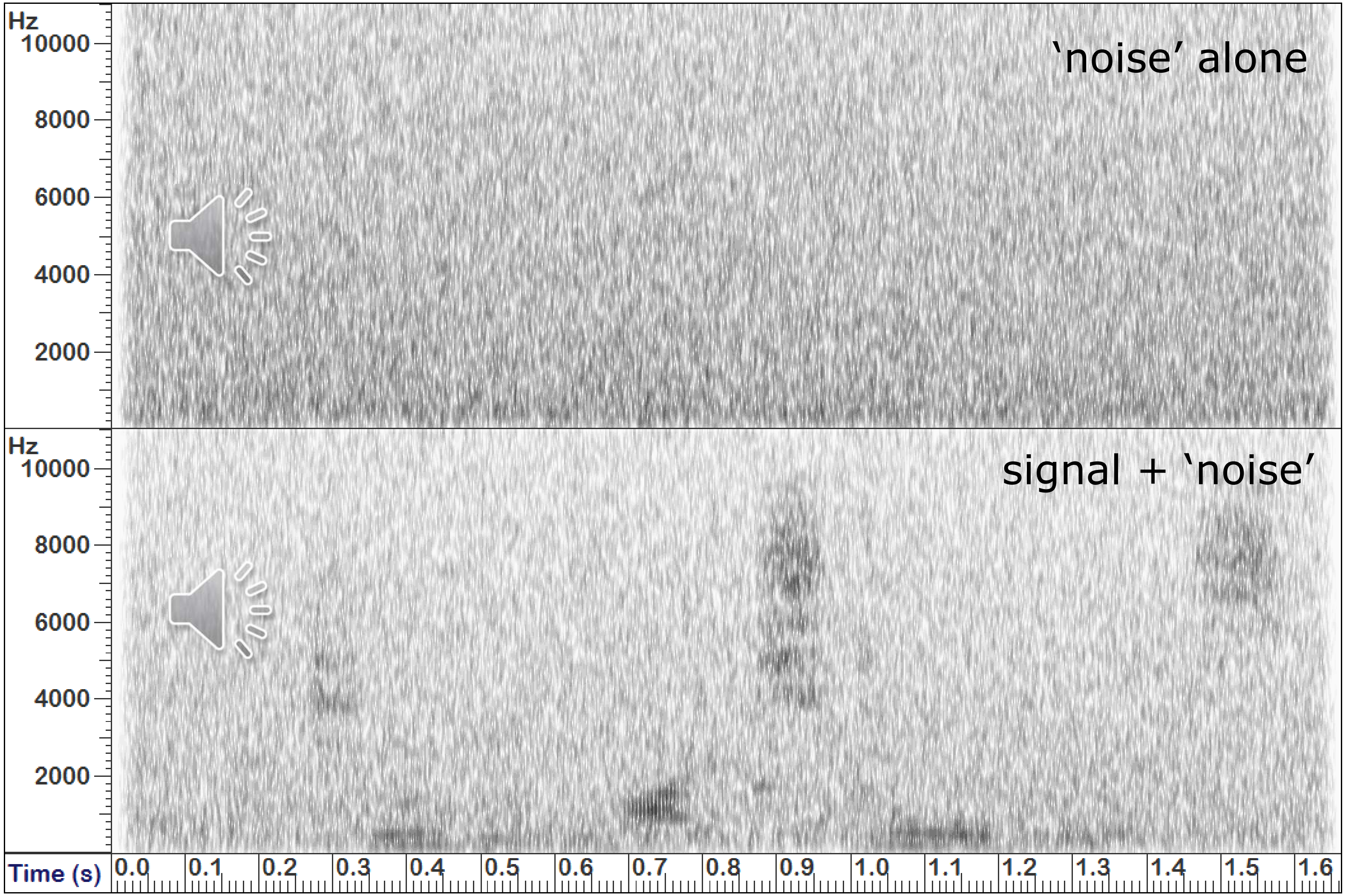



But maskers can be periodic too, most importantly, when speech is in the background.

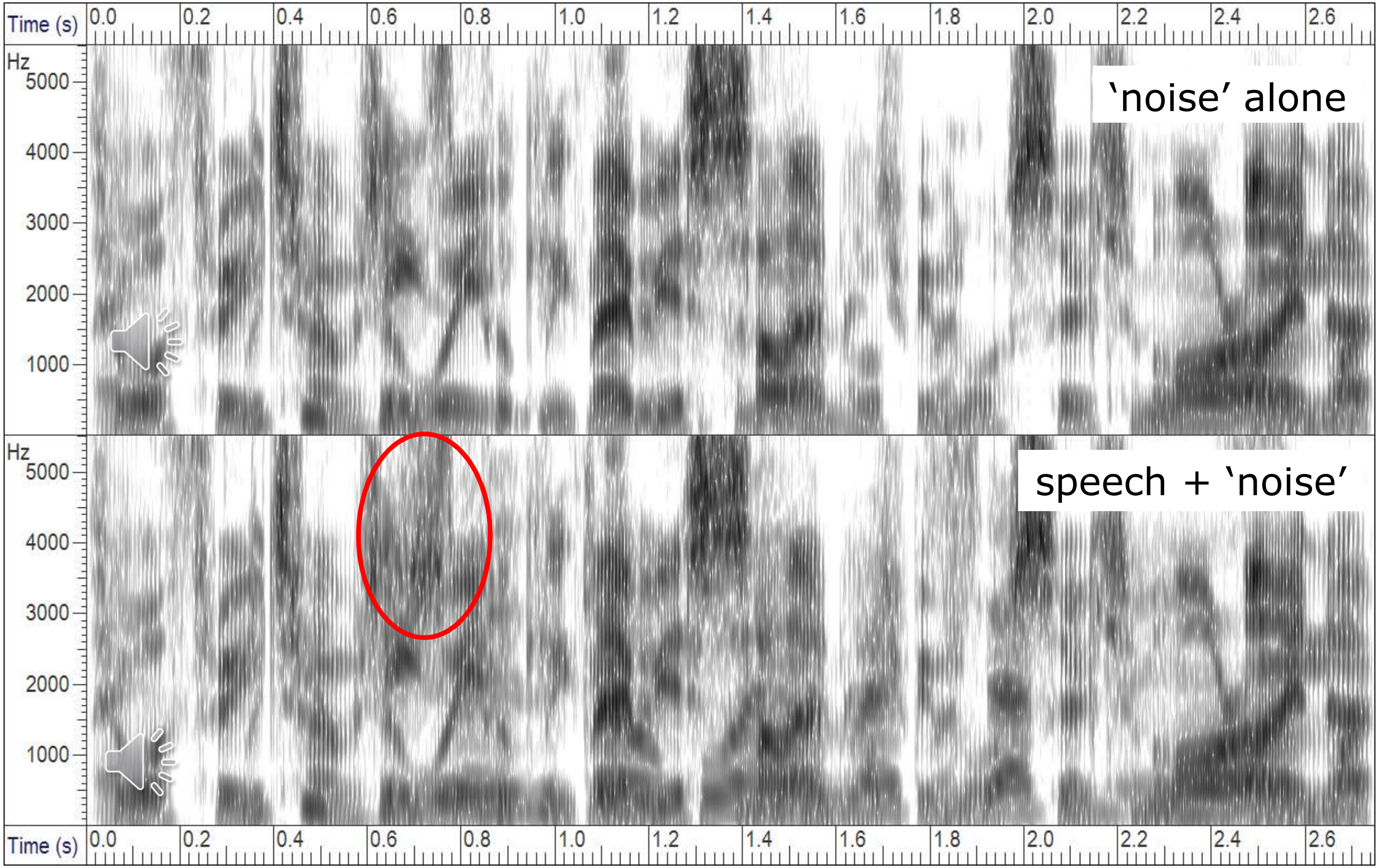
# Miller (1947)
# ‘The masking of speech’

*It has been said that the best place to hide a leaf is in the forest, and presumably the best place to hide a voice is among other voices.*

# There are lots of different kinds of ‘noises’

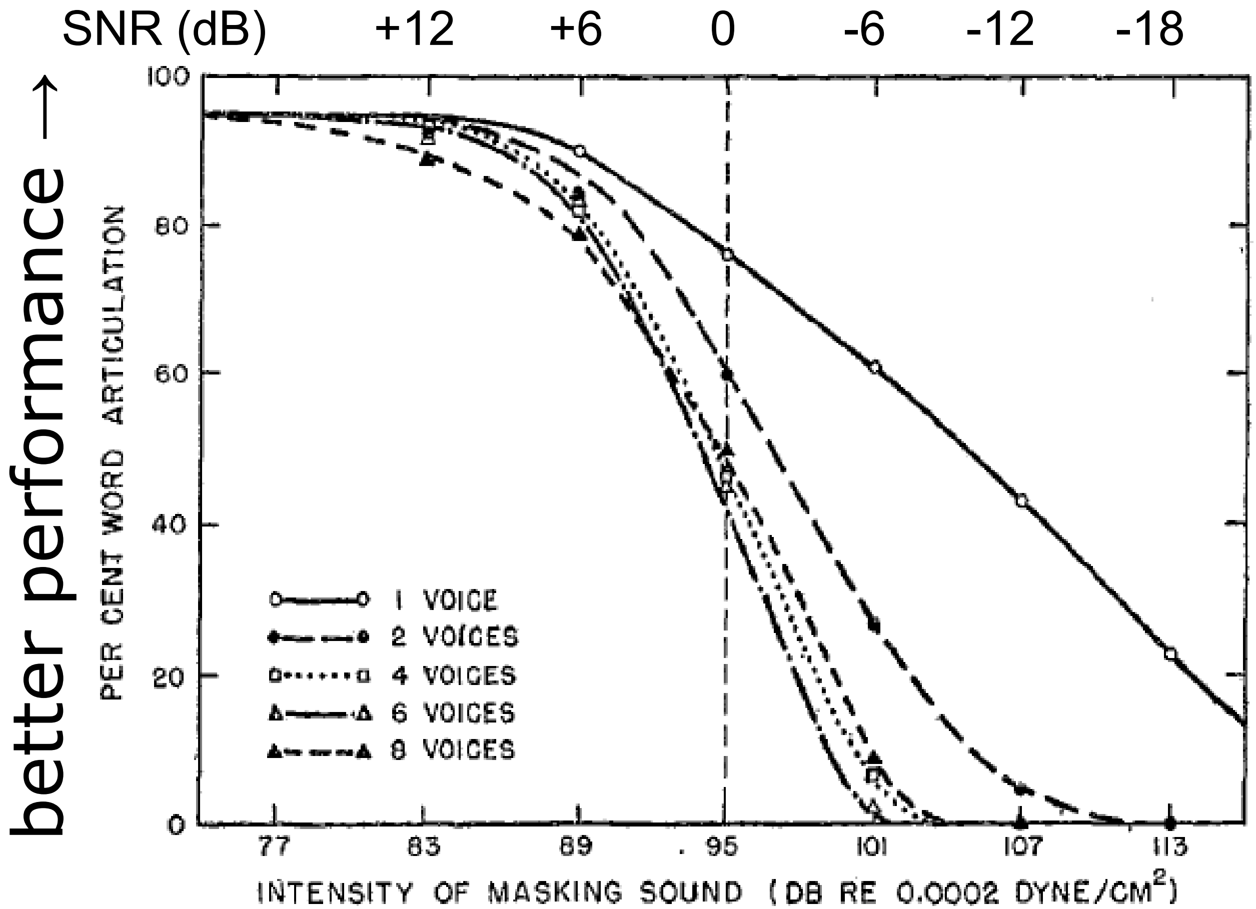


# Another kind of ‘noise’




‘show’ starts at t≈0.65 ms

# Miller (1947)
# Increasing the number of talkers in the masker




*'It is relatively easy for a listener to distinguish between two voices, but as the number of rival voices is increased the desired speech is lost in the general jabber*.'

- target words from multiple males
- babble: equal numbers of m/f
  (1 VOICE is male)

# Why is it easy to ignore one other talker and not more?

- More opportunities to glimpse with one talker
- Differences in pitch contour for two talkers makes it easier to ignore one and attend to the other

# A useful distinction

- Energetic masking
  - maskers interfere with speech to the extent that have energy in the same time/frequency regions
  - primarily reflecting direct interaction of masker and speech in the cochlea
  - relevance of glimpsing/dip listening
    - Temporal and/or spectral ‘dips’ in the masker allow ‘glimpses’ of target speech
- Informational masking
  - everything else!

# Caveat: Another kind of masking

- What we have called 'energetic masking' may in fact be two different things
  - Genuinely energetic masking (as described before)
  - *Modulation* masking (MM)
- MM is the disruptive effect that modulations in the masker have on the modulations in the target
  - So it's not the *energy* in the masker that is so important
  - Similar to EM, in happening at the periphery (needing to be in the same time/frequency)
- For the details
  - Stone, M. A., Fullgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. J Acoust Soc Am, 132, 317-326.

# Informational masking

- Something to do with target/masker similarity?
  - signal and masker ‘are both audible but the listener is unable to disentangle the elements of the target speech from a similar-sounding distracter’ (Brungart, 2005)

# Informational masking: a finer distinction (Shin-Cunningham, 2008)

- Problems in ‘object formation’
  - Related to auditory scene analysis
  - similarities in auditory properties make segregation difficult
    - voice pitch, timbre, rate

1 woman, 1 man

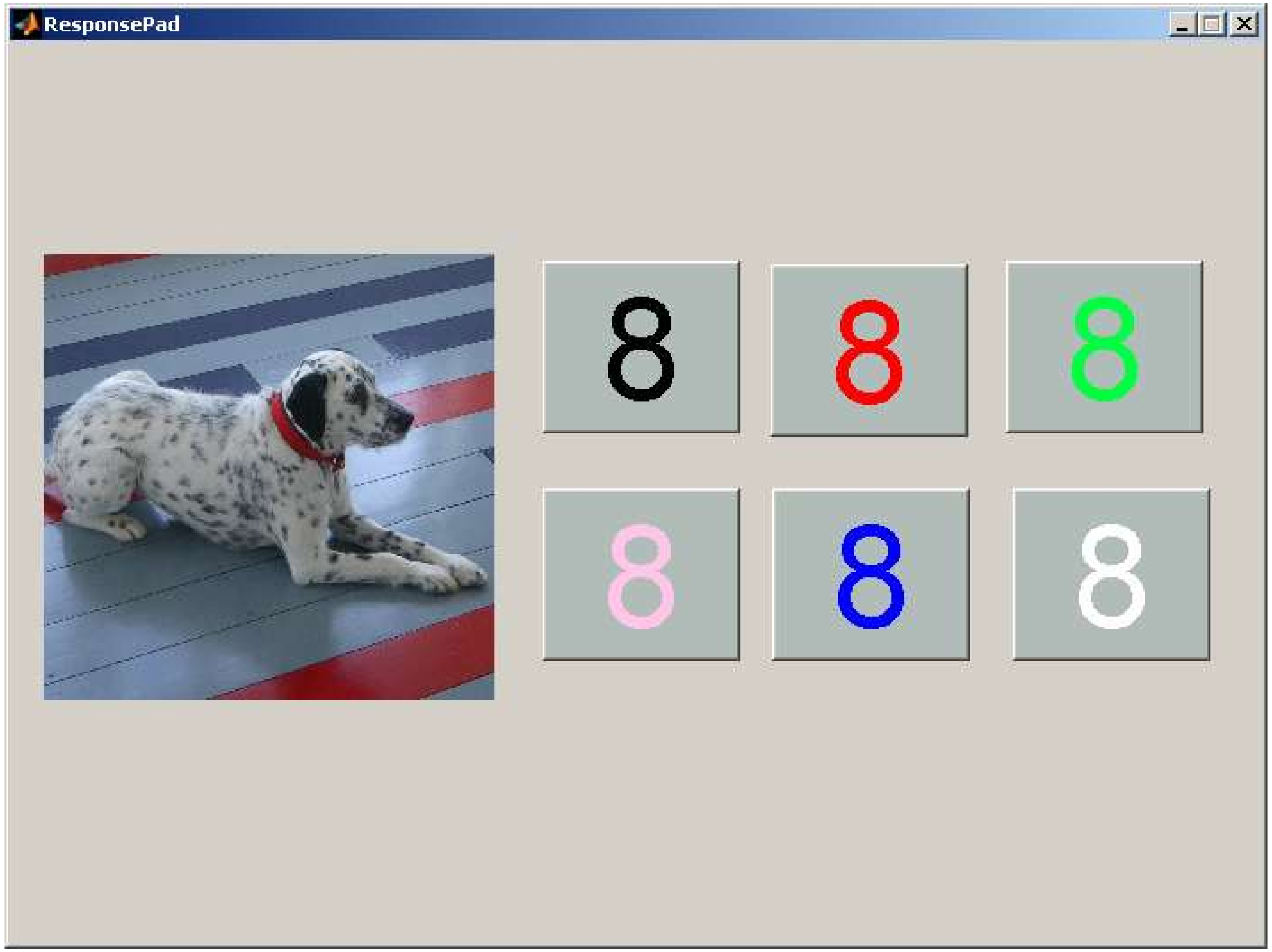
2 men

- Problems in ‘object selection’
  - Related to attention and distraction
  - the masker may distract attention from the target
    - e.g., more interference from a known as opposed to a foreign language

# EM & IM appear to operate at different parts in the auditory pathway

- Energetic masking at the periphery, in the cochlea
  - Early developing abilities
  - Increased EM from hearing impairment
- Informational masking at higher centres
  - Late developing abilities?
  - Increased IM in younger and older listeners?
  - But aspects of IM can be made difficult by peripheral factors
    - *e.g.*, CI users difficulties with auditory scene analysis

# Listening to speech in 'noise'

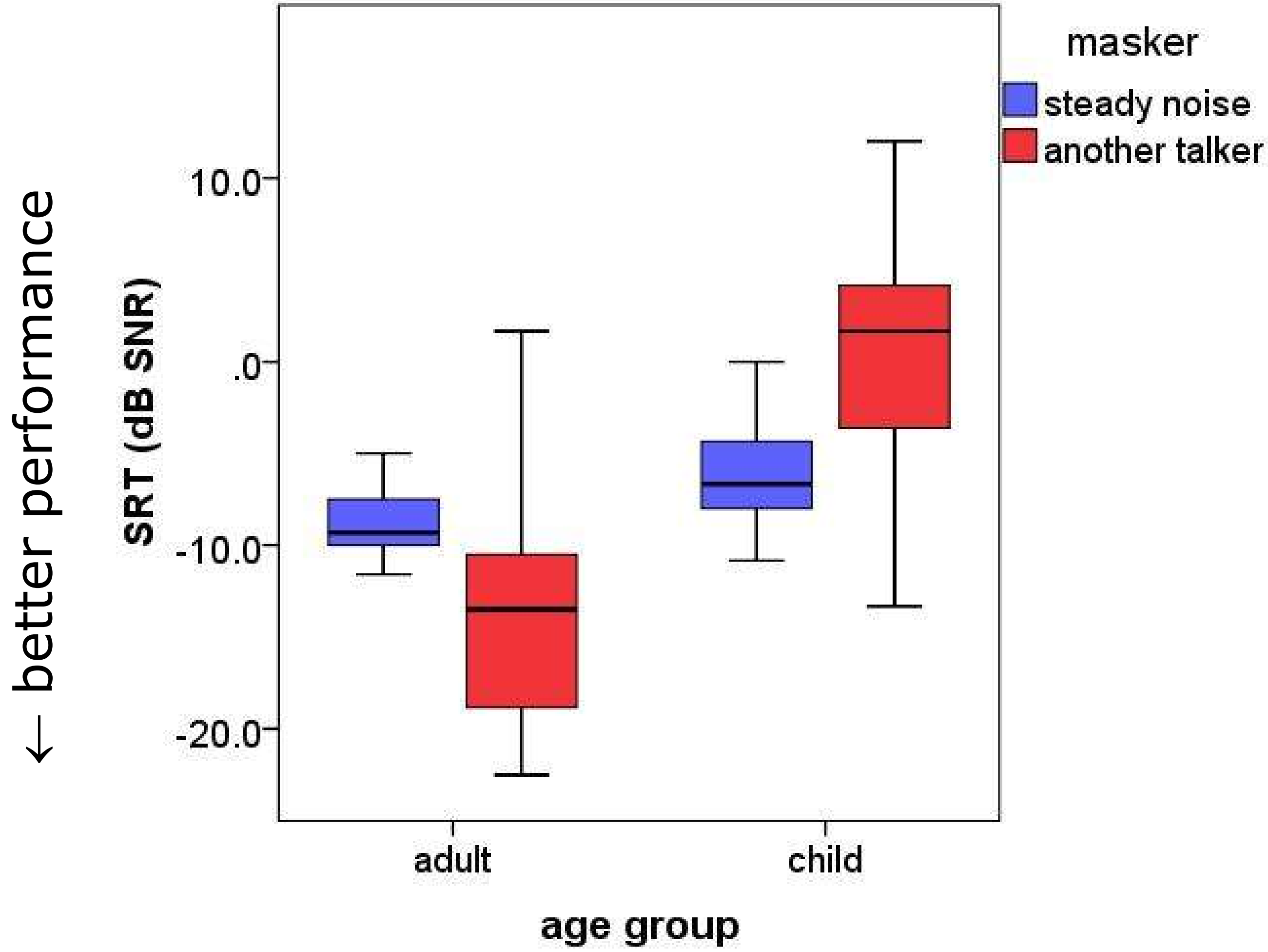



in quiet

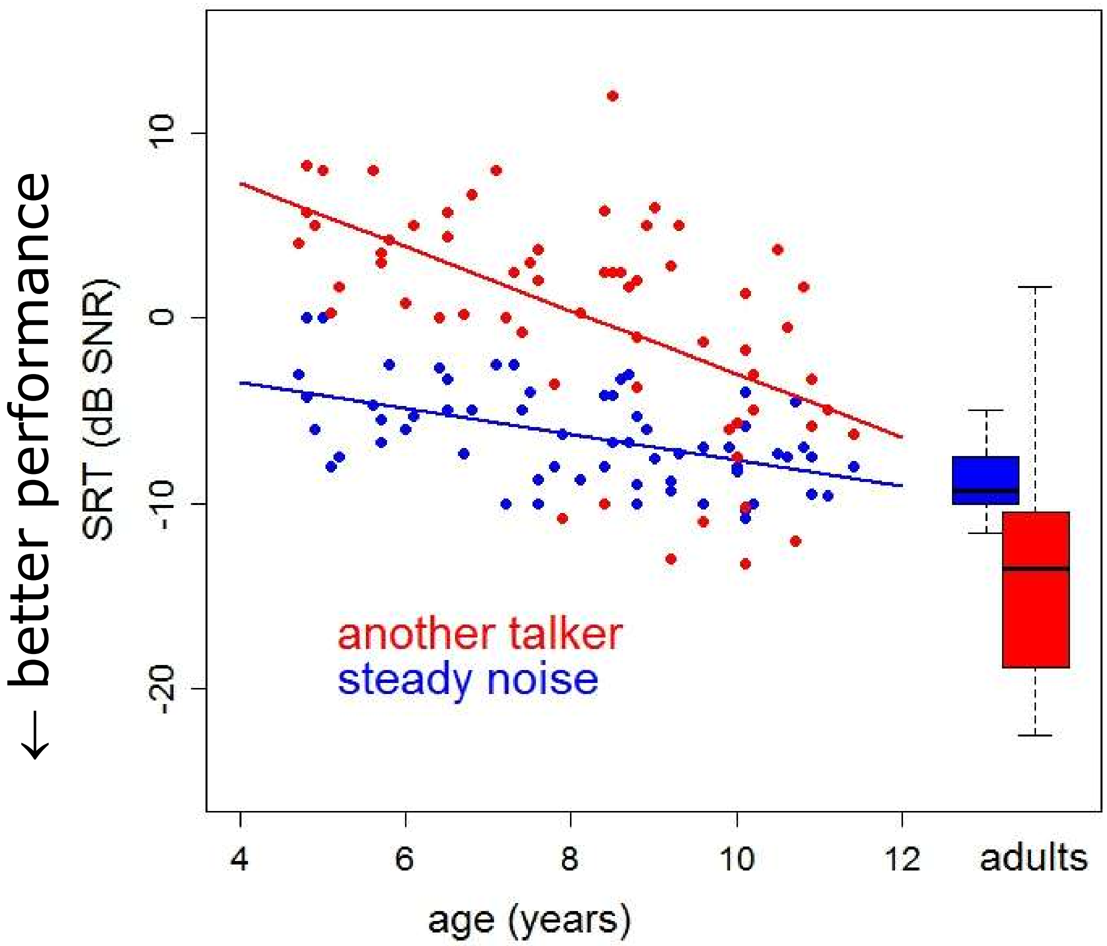
in steady noise

against another talker

# Children find it hard to ignore another talker

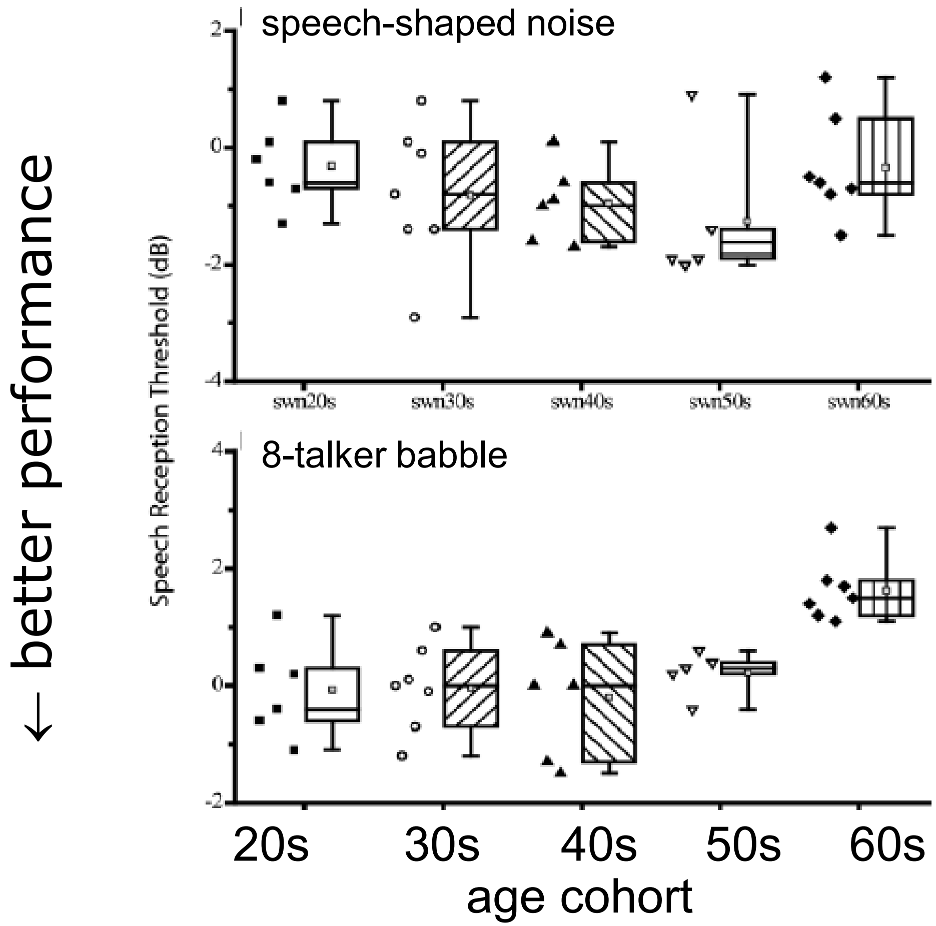


# Slow development of abilities that minimise IM



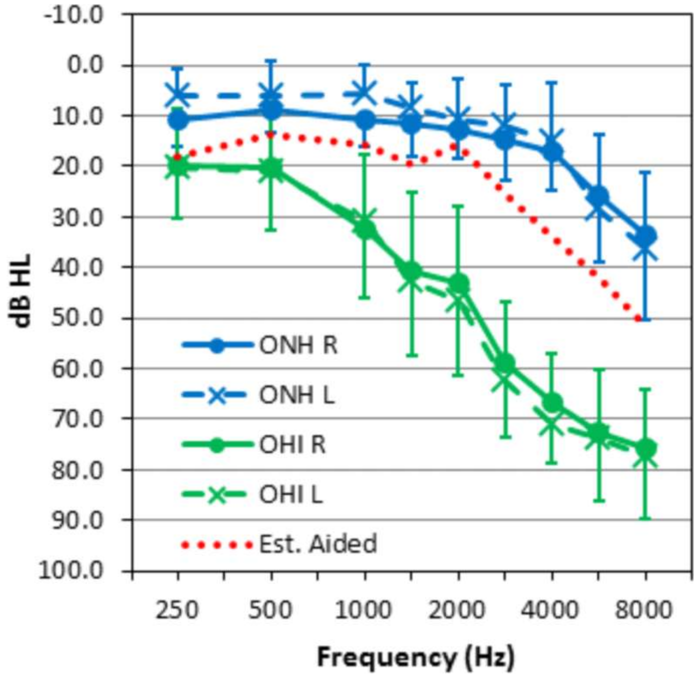

With contributions from Jude Barwell & Zoe Lyall

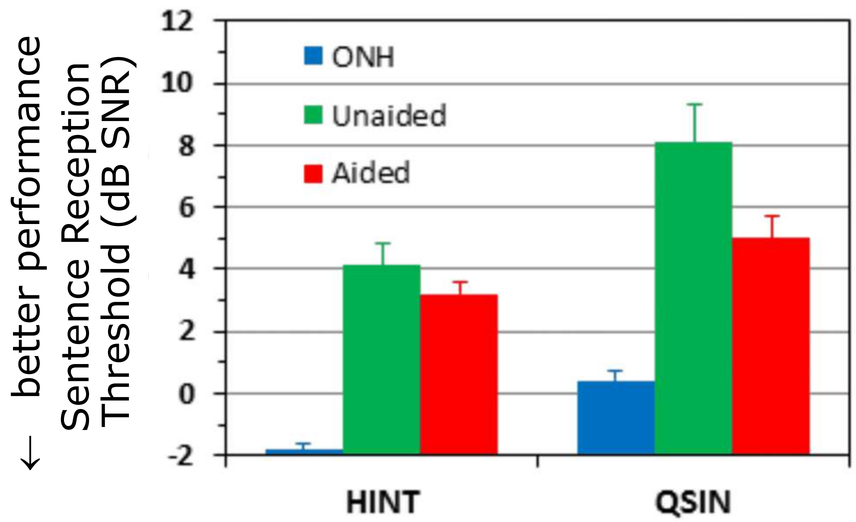
# Increased difficulty in older listeners for some noises




Rajan & Cainer (2008)

# Speech in Noise in HI

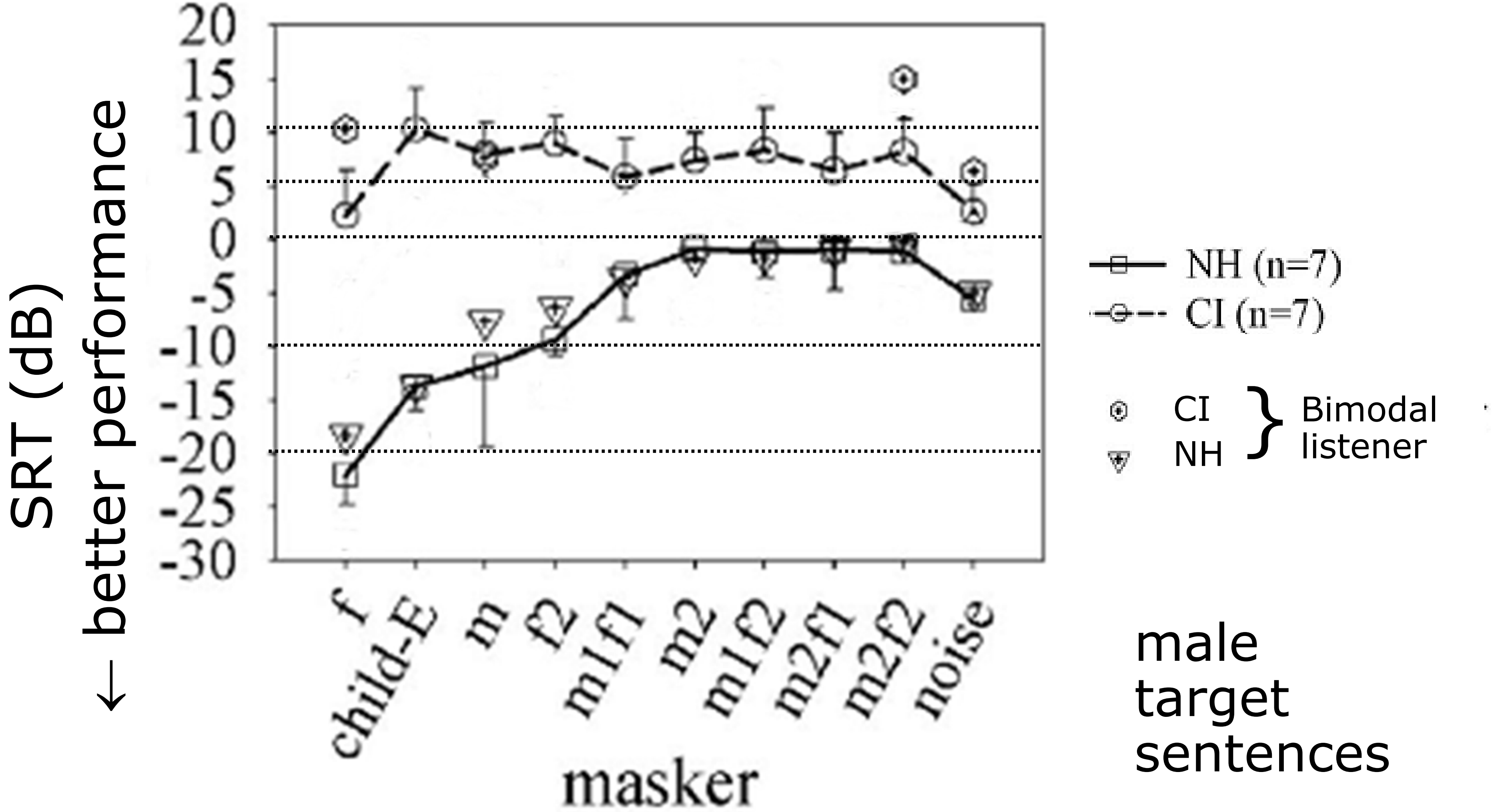



- 60-81 y.o.
- OHI: 24 males
- ONH: 14/16 females
- fixed target @ 70 dB SPL

Woods et al., 2015 *PLoS ONE*

# CI users show little variation in SRT for different maskers




Cullington & Zeng (2008)

# Spatial Release from Masking: when target and masker come from different directions

- Head-shadow effects often result in one ear having a better SNR than the other (the "better-ear" advantage).
  - not a result of genuine binaural interaction
- Additionally, binaural mechanisms can produce improvements in speech comprehension as well as detection of tones (BMLD).
  - 'squelch' (aargghh!!)
- These operate optimally in different frequency regions
  - Why?
- Spatial separation reduces both EM and IM

# Bronkhorst & Plomp (1988)

- Measured HRTFs on an acoustic manikin to simulate spatial cues over headphones
- Allowed the separation of ITD from ILD cues so each could be presented in isolation
- Simple sentences in an adaptive procedure to measure SRT
- target speech always straight ahead; speech spectrum noise varied in position

# Bronkhorst & Plomp (1988)

- speech always at 0°
- ILD more important than ITD
  - why?
- But both really matter
- Implications for HI?
  - monaural fittings
  - mismatched hearing aids (*e.g.*, knee point of compression)

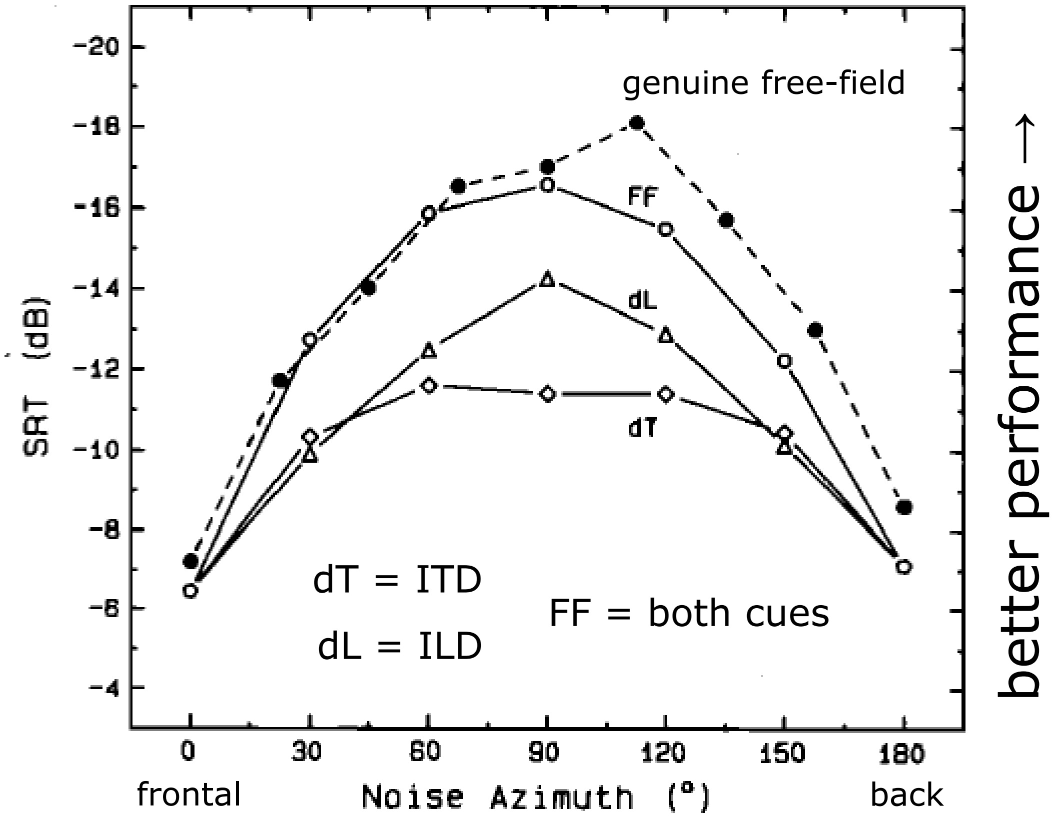



FIG. 5. Mean speech reception thresholds obtained in experiment I for three different noise types : FF (free field), dL (headshadow only), and dT (ITD only). The closed data points represent results of Plomp and Mimpen (1981) obtained in a free field.

# What you need to know

- Energetic vs. informational masking
- Object formation vs. object selection
- glimpsing/dip listening
  - What it is
  - That HI listeners find it harder
  - That CI listeners find it harder still

# References


- Bernstein, J. G. W. & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 125, 3358-3372.
- Bradlow, A. R. & Pisoni, D. B. (1999) 'Recognition of spoken words by native and non-native listeners:Talker-, listener-, and item-related factors' *J Acoust Soc Am*, 106(4).
- Bronkhorst & Plomp (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J Acoustical Society of America, 83*.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109,* 1101-1109.
- Cullington, H. E. & Zeng, F. G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *Journal of the Acoustical Society of America, 123,* 450-461.
- Howard-Jones, P. A. & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acustica, 78,* 258-272.
- Kalikow, Stevens, K. N., & Elliot (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.
- Lindblom, B. (1990) 'Explaining phonetic variation: A sketch of the H & H theory' in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 403–439.
- Miller, G. A. (1947). The Masking of Speech. *Psychological Bulletin, 44,* 105-129.
- Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 113,* 961-968.
- Rajan, R. & Cainer, K. E. (2008). Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. *Neuroscience, 154,* 784-795.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends In Cognitive Sciences, 12,* 182-186.
- Stone, M. A., Fullgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *J Acoust Soc Am*, 132, 317-326.
- Takahashi, G. A. & Bacon, S. P. (1992). Modulation Detection, Modulation Masking, and Speech Understanding in Noise in the Elderly. *J Speech & Hearing Res, 35,* 1410-1421.
- Woods, D. L., Arbogast, T., Doss, Z., Younus, M., Herron, T. J., & Yund, E. W. (2015). Aided and Unaided Speech Perception by Older Hearing Impaired Listeners. *Plos One*, 10.