AUDL 4007 & GS12
Auditory Perception

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.
- 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 …
    - We could have discussed the …
  - 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

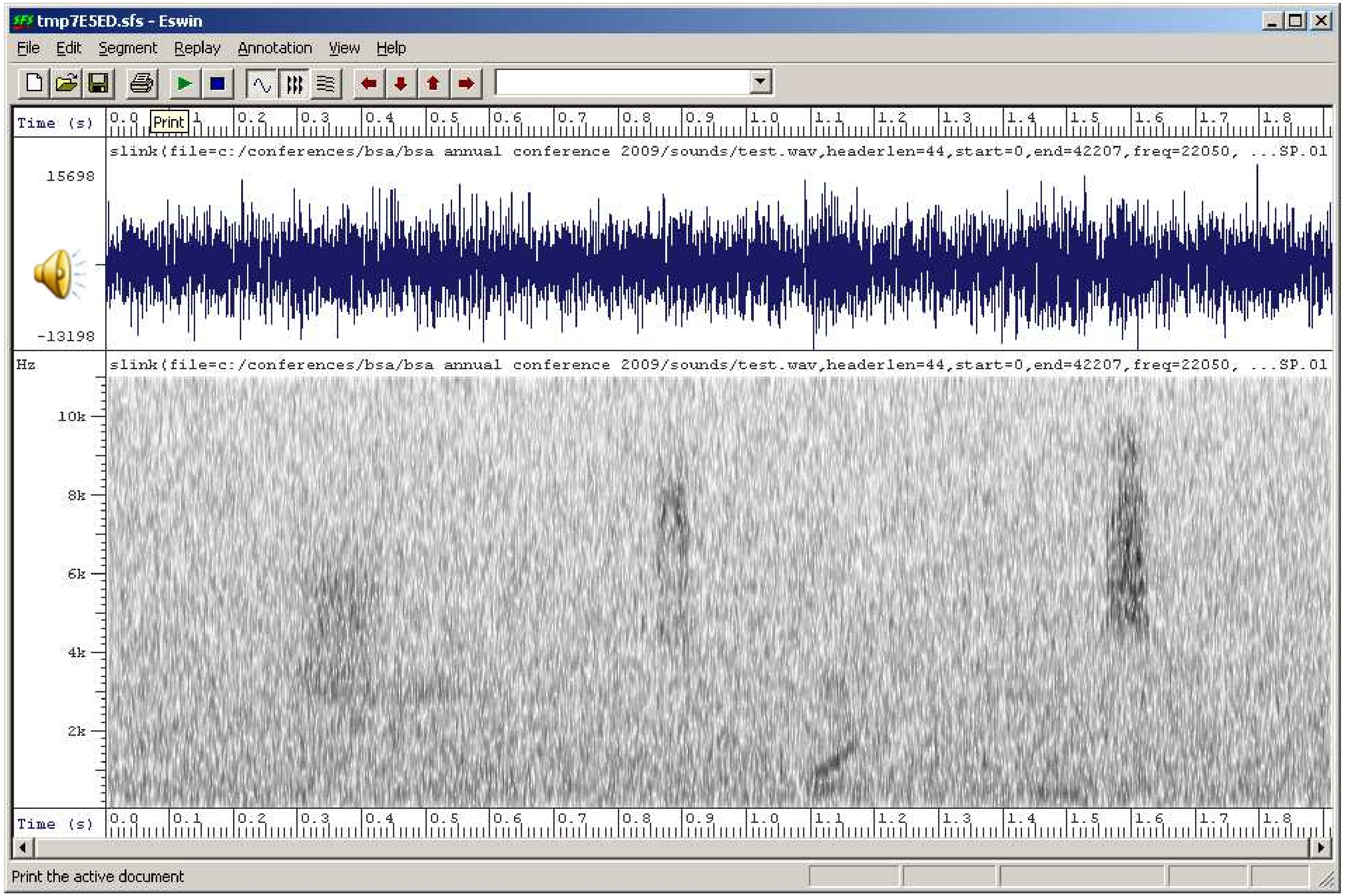
- Talker characteristics
  - Talkers vary considerably in intrinsic intelligibility
  - Talkers can vary their own speech depending upon demands of the situation (hyper/hypo distinction of Lindblom, 1990)
    - manipulations in vowel space, prosody, rate
  - Match between talker and listener accents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

Some determinants of performance: V

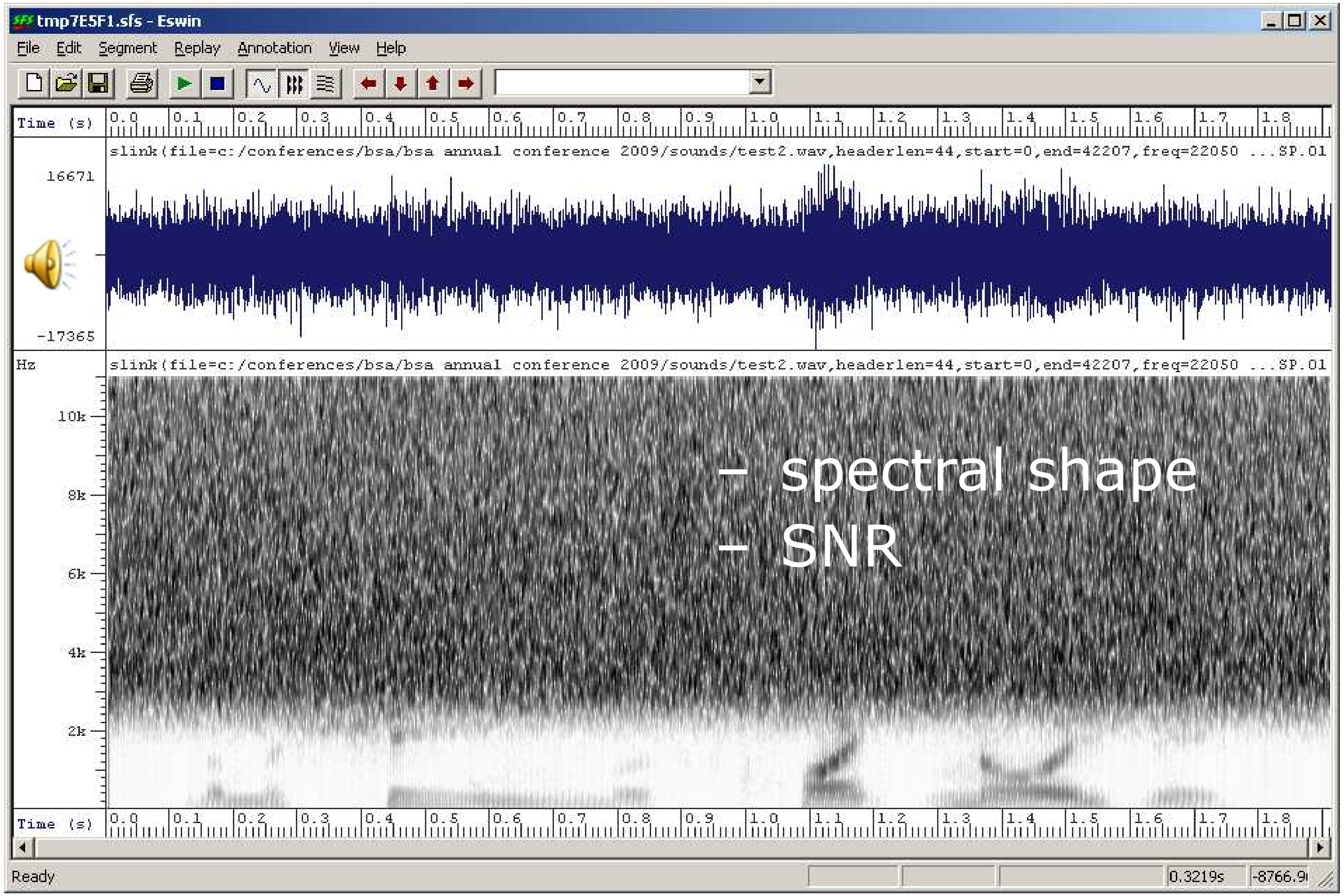
- Listener characteristics
  - Linguistic development
    - L1 vs L2
    - vocabulary knowledge
    - ability to use context
  - Hearing sensitivity and any hearing prosthesis used

Focus on factors more centrally related to audiology

The simplest case:
A steady-state background noise



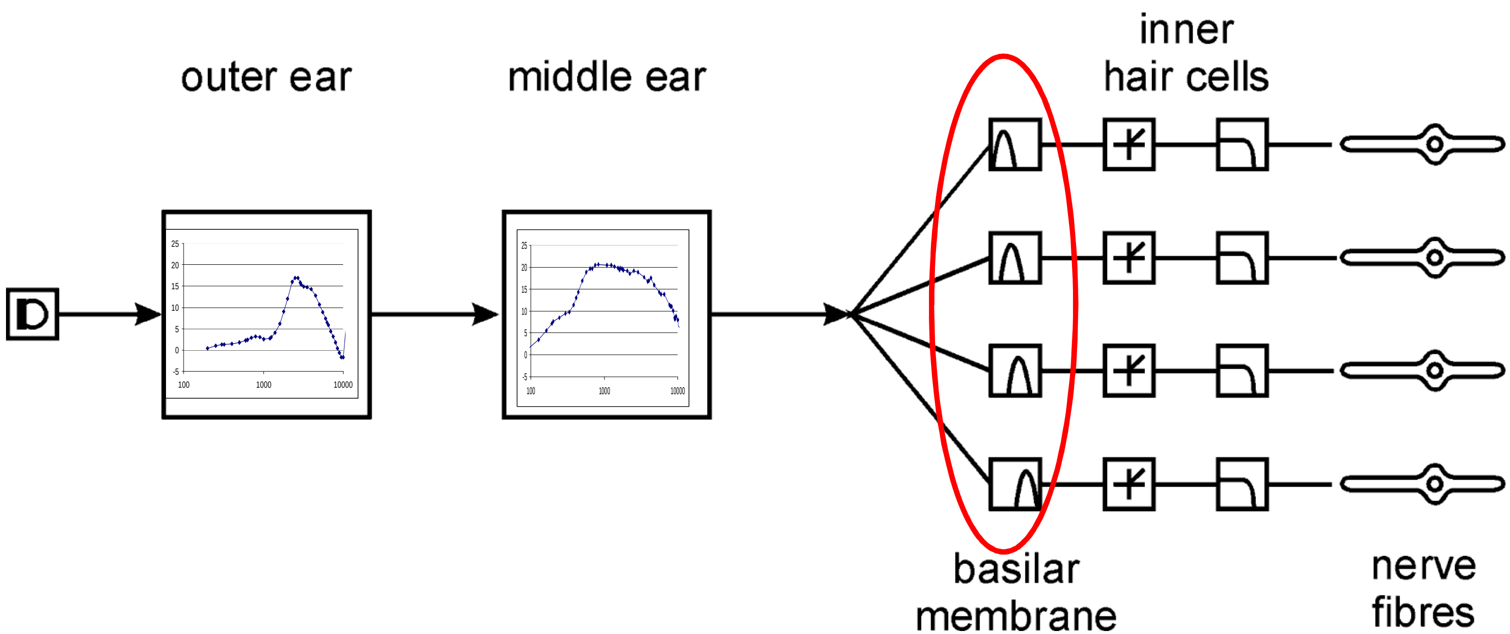
Much is understood about what makes one steady noise more or less interfering than another

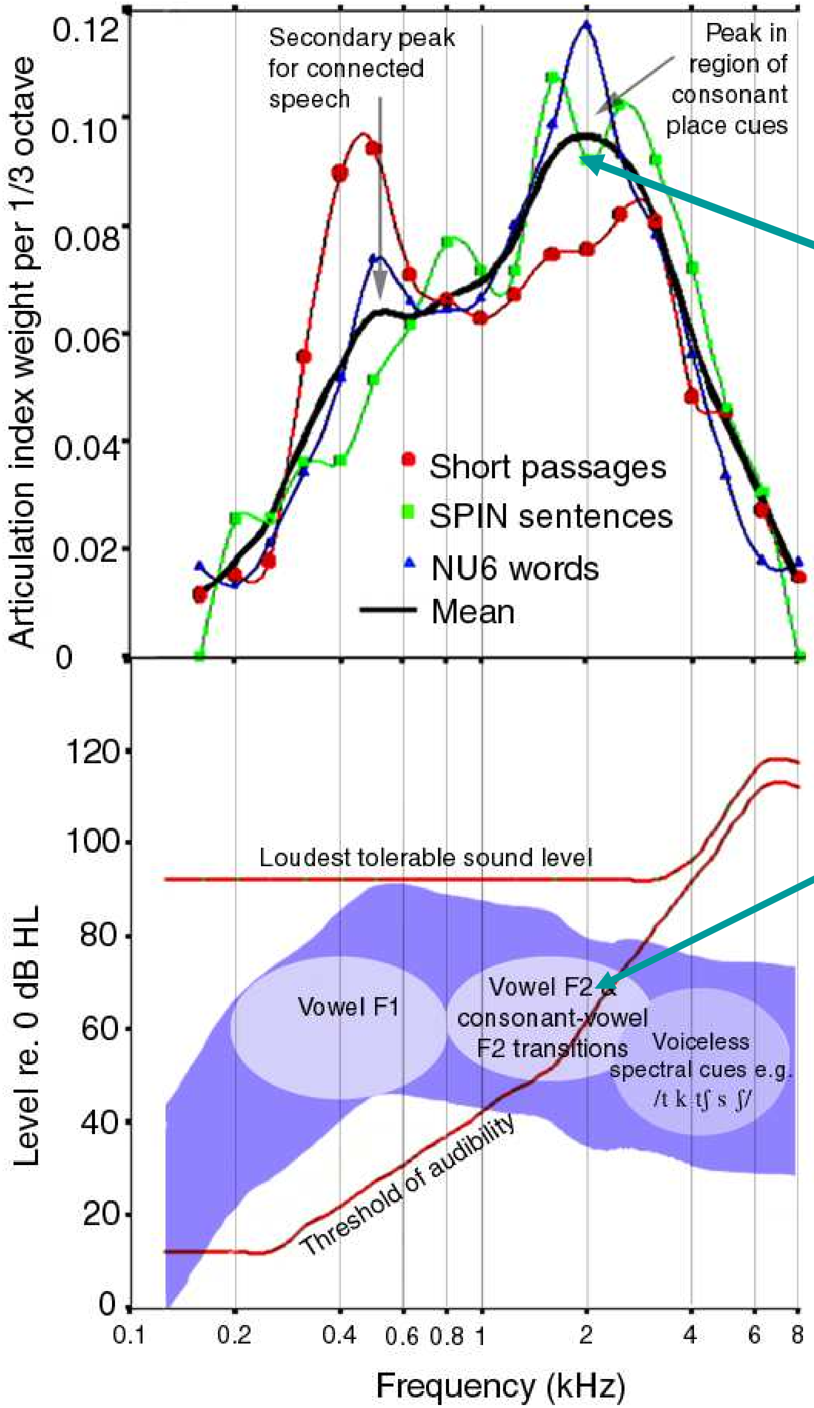


‘Energetic’ masking

- 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)

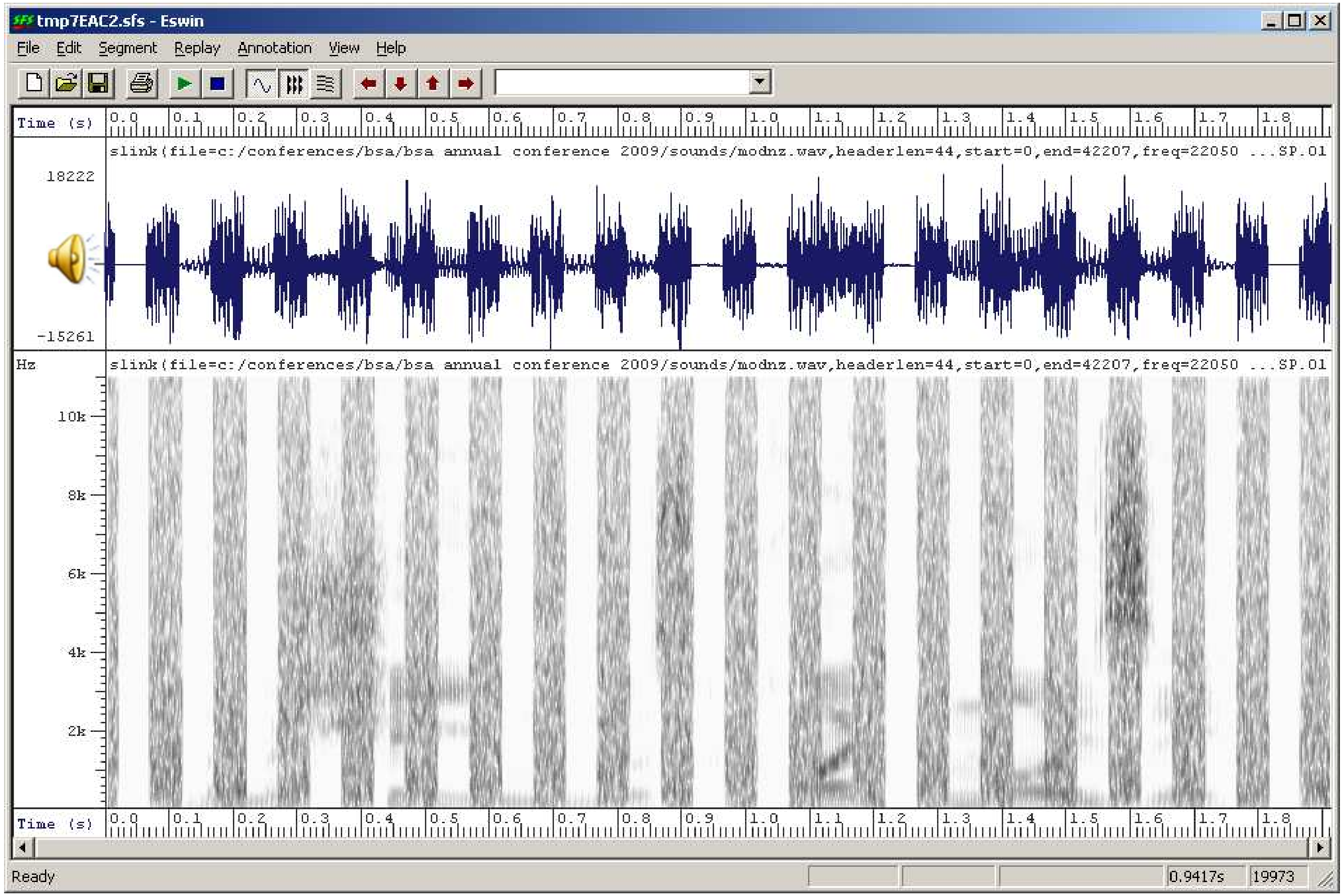
$$A = \sum_{i=1}^{n} I_i \quad W_i,$$

–W (2000 Hz) – here W is approx 0.6

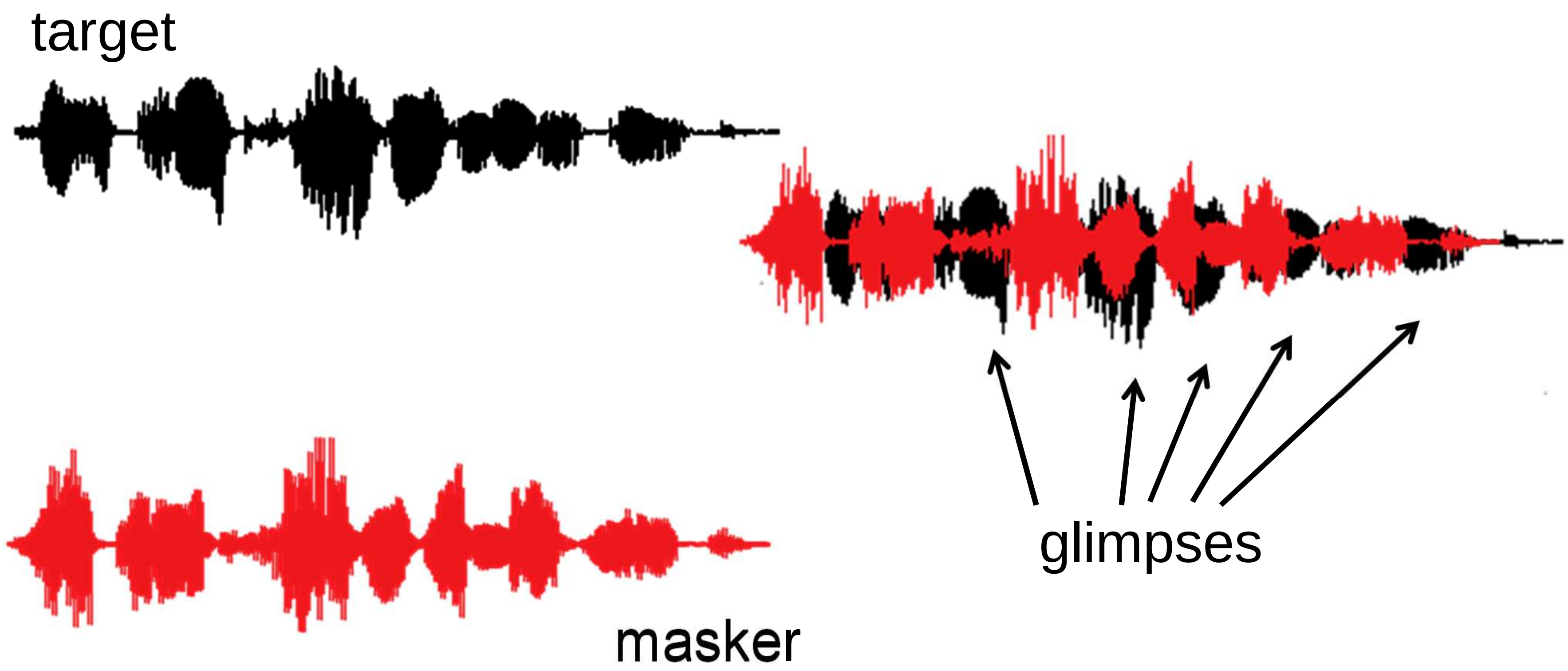
–***A*** is the Articulation Index (predicted intelligibility).

–***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 …



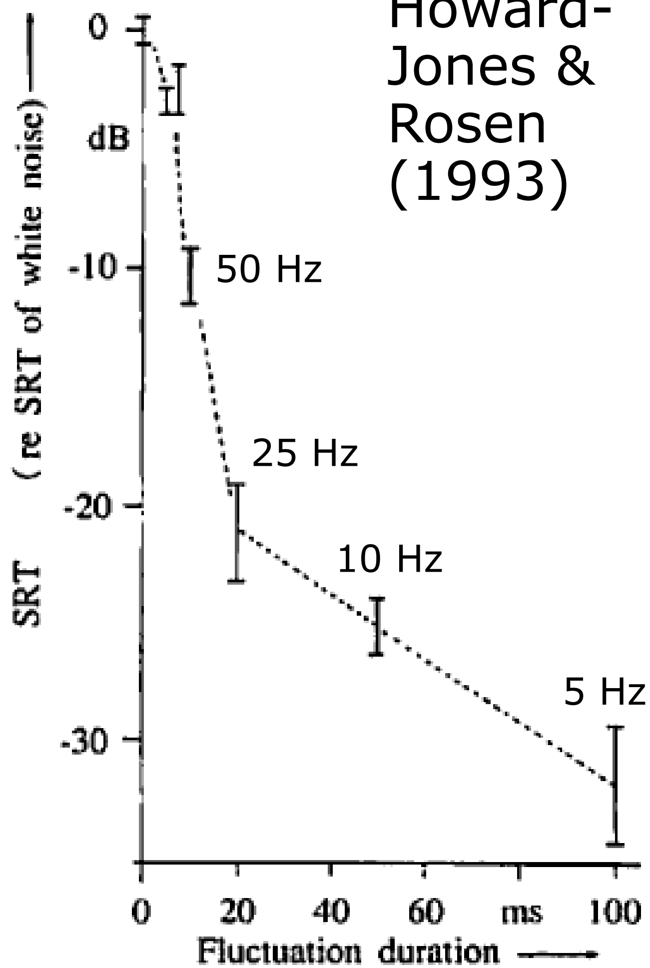
Fluctuating maskers afford 'glimpses' of the target signal



‘dip listening’ or ‘glimpsing’

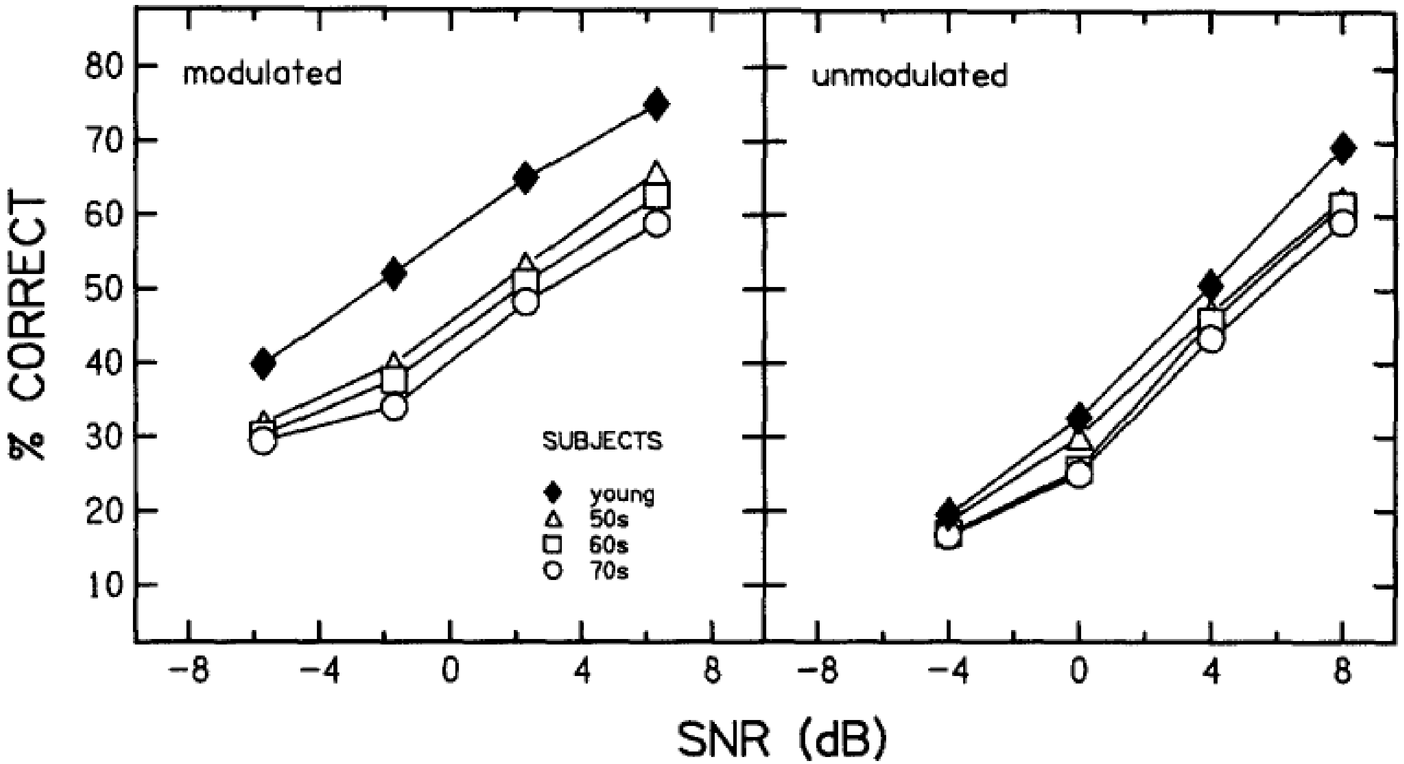
People with normal hearing can listen in the ‘dips’ of an amplitude modulated masker

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)

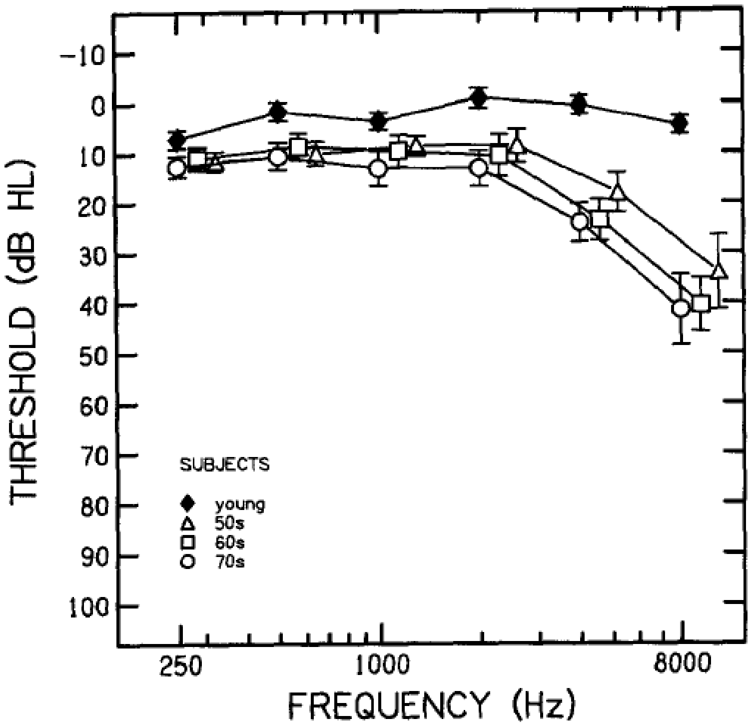


FIGURE 1. Mean pure-tone audiometric thresholds (in dB HL) for each subject group. The three older groups are represented by open symbols and the young group by closed symbols. Error bars indicate ±1 standard error. Data for the older groups have been shifted horizontally.

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

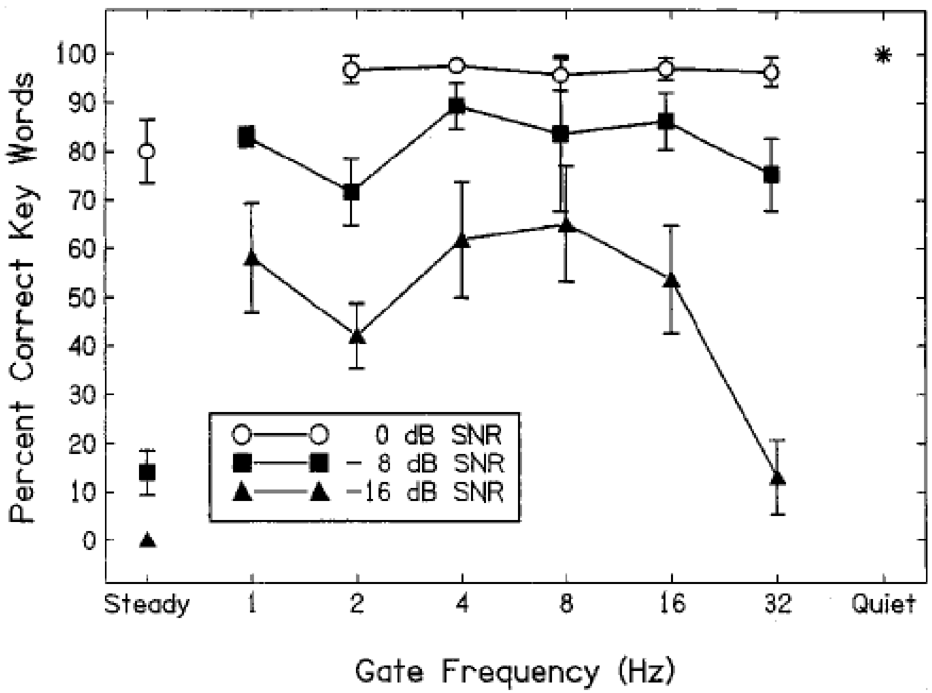
Why is ‘dip’ listening limited in hearing-impaired listeners?

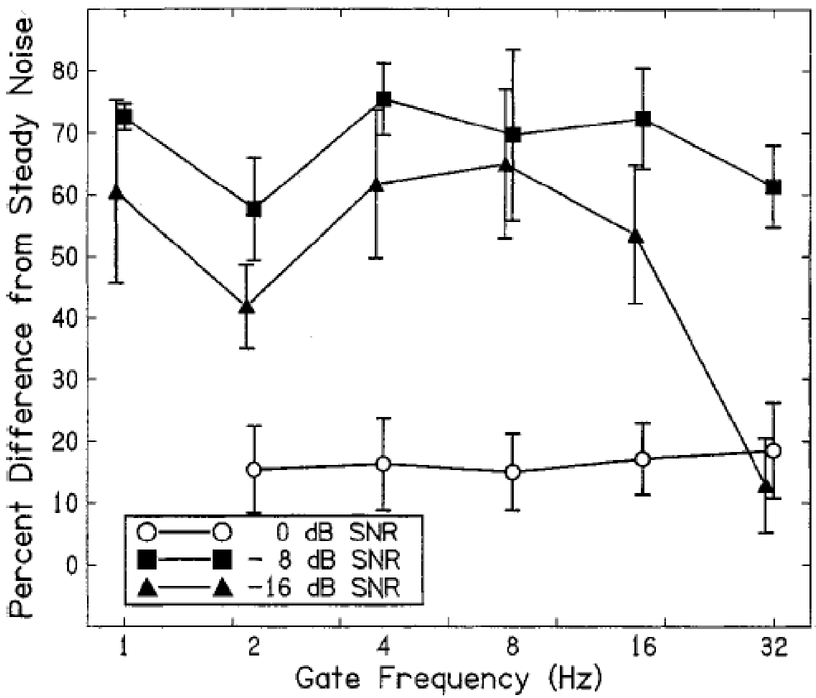
- Audibility can be an influence
- Some of the lack of masking release may be due to SNRs being higher for HI listeners.

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

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

normal listeners

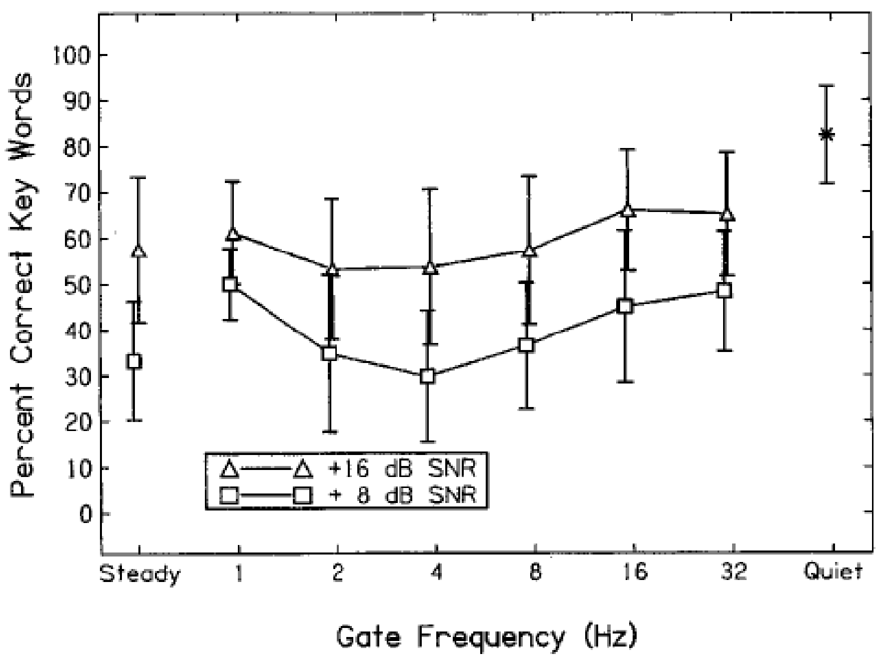


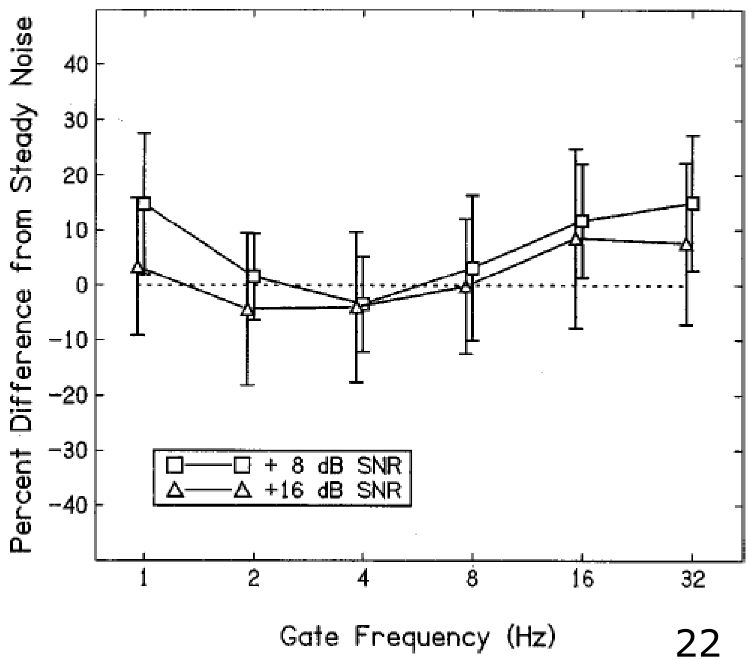


CI users

Note much higher SNRs

(+8 and +16 vs -8 and -16 dB)



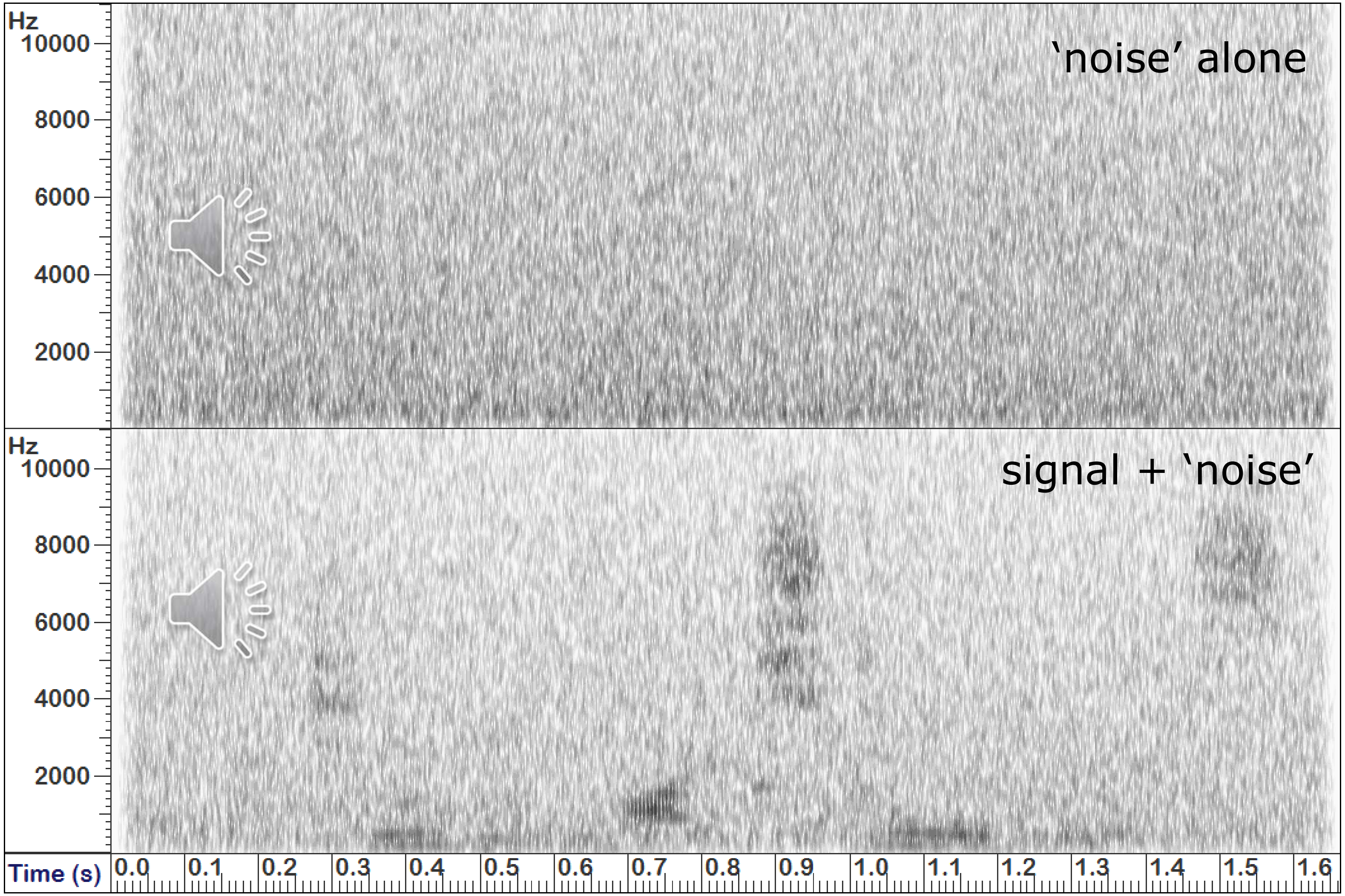


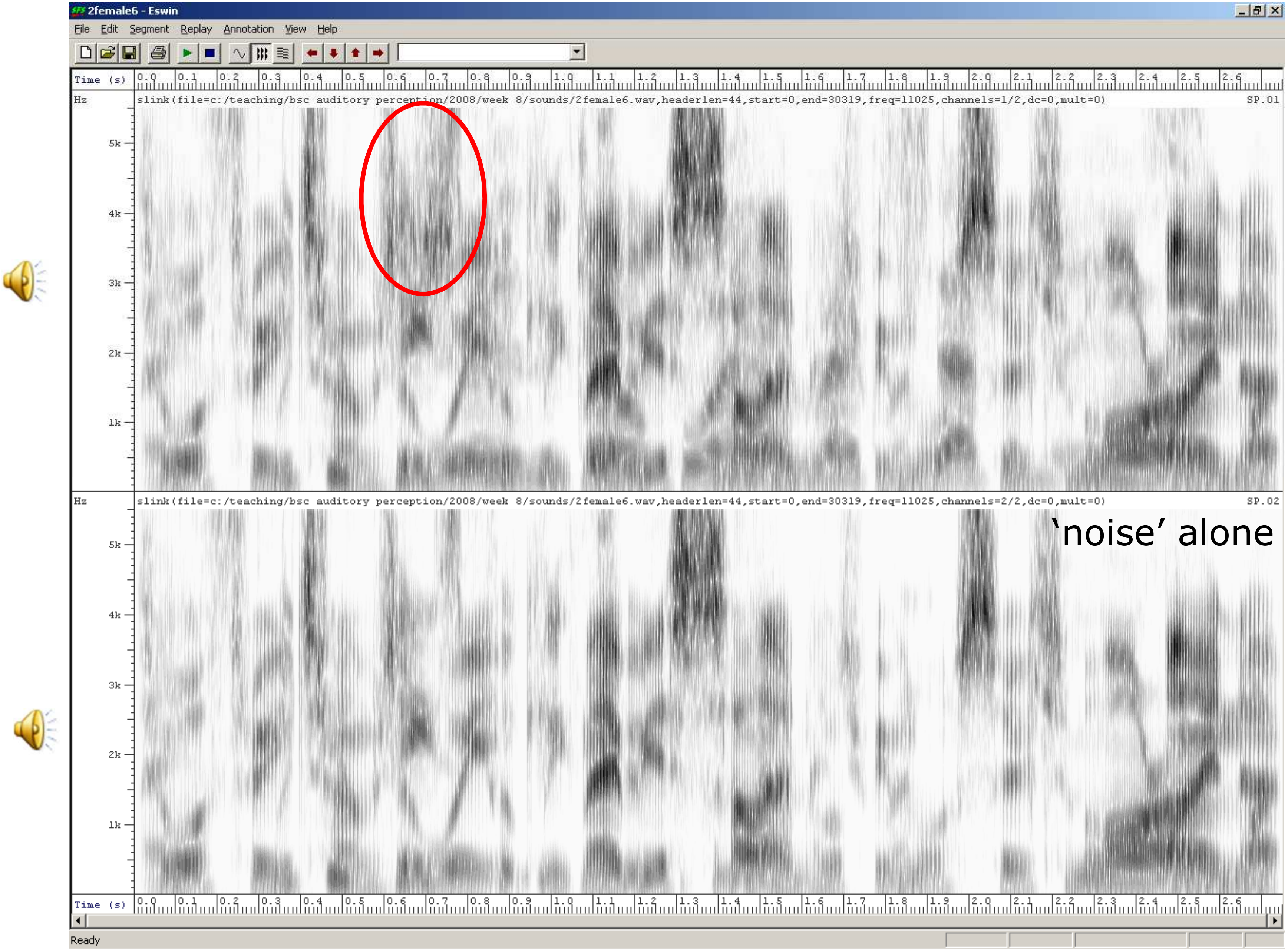
But maskers can be periodic too, most importantly, when speech is in the background.

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.

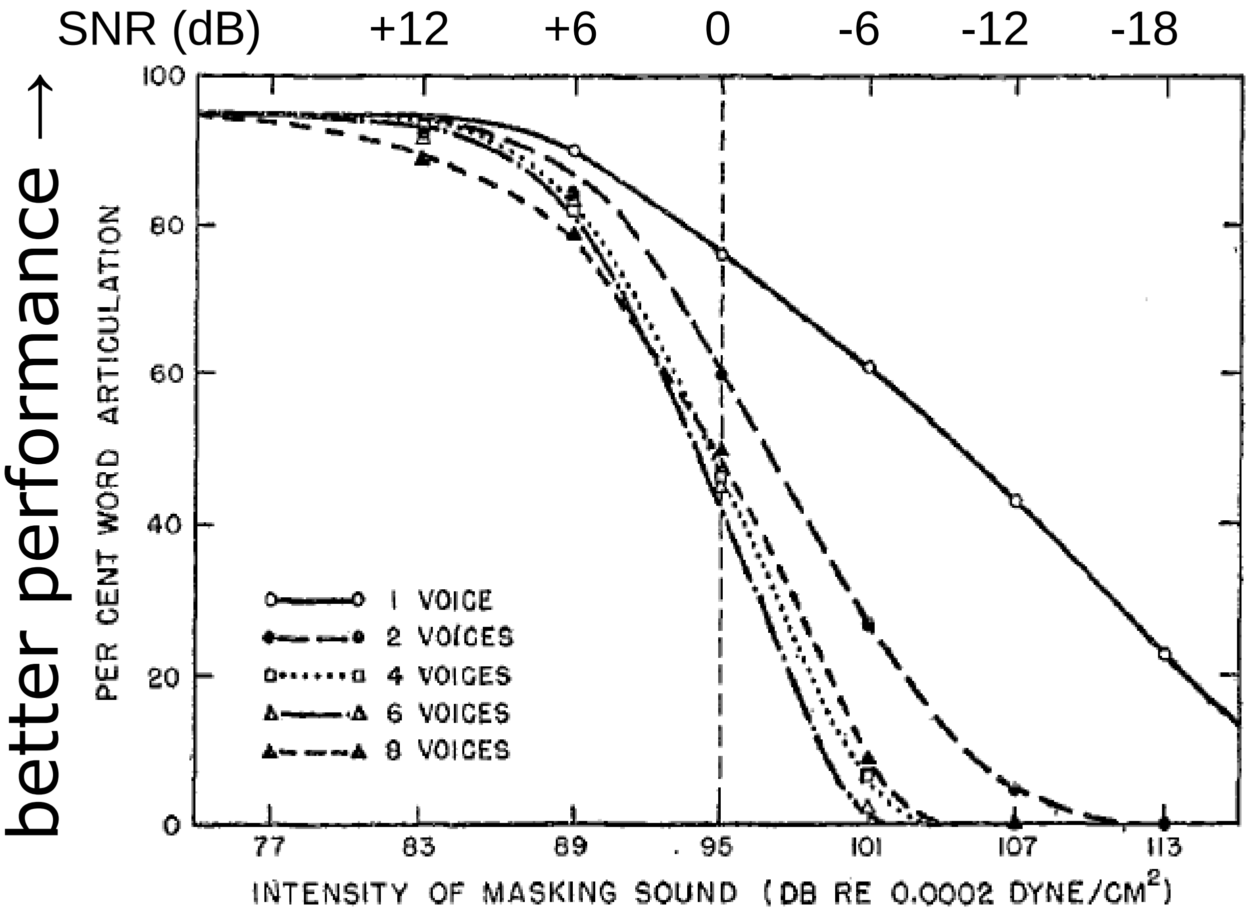
There are lots of different kinds of 'noises'





–'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

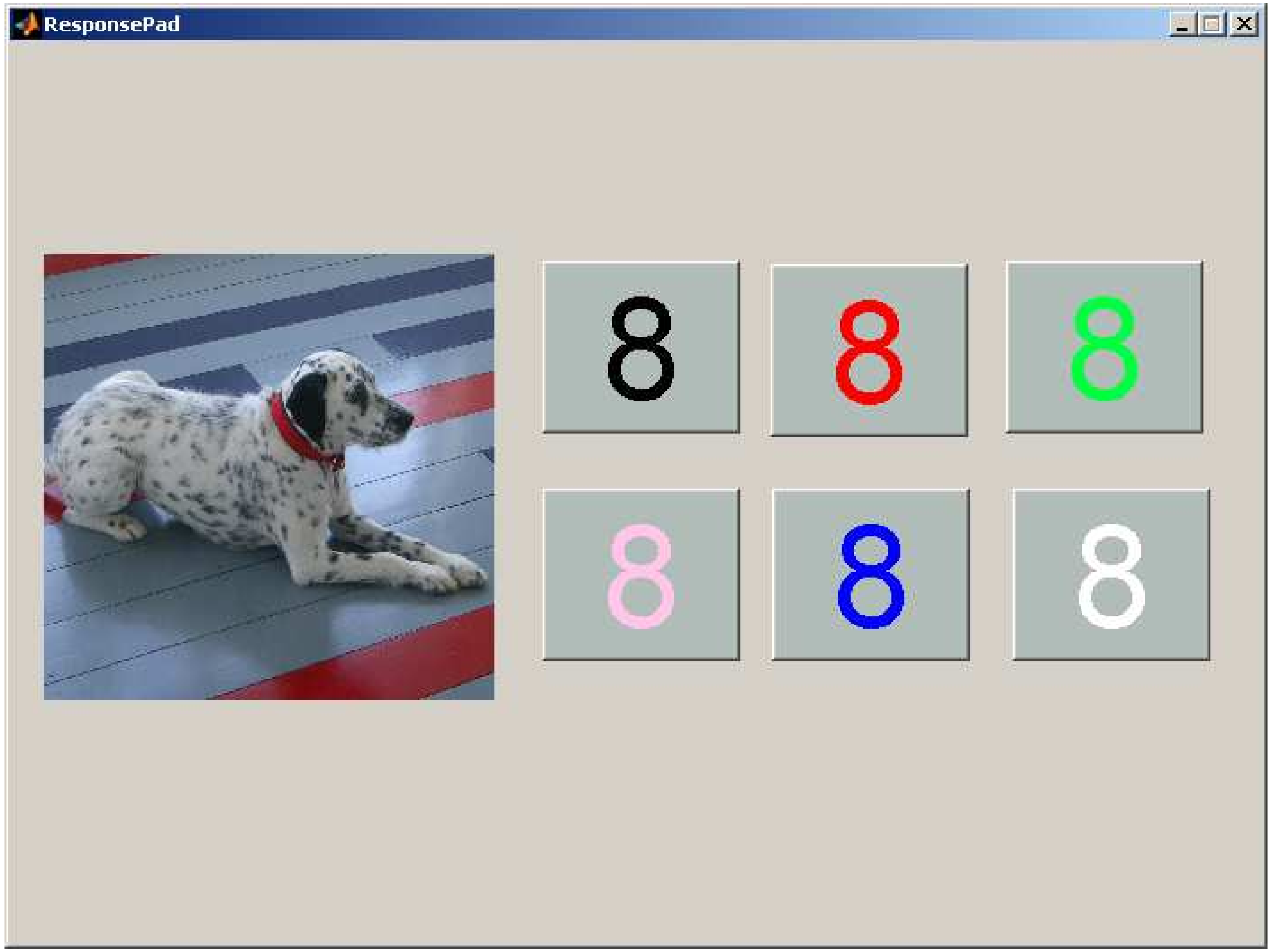
1 woman, 1 man 2 men

- 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'

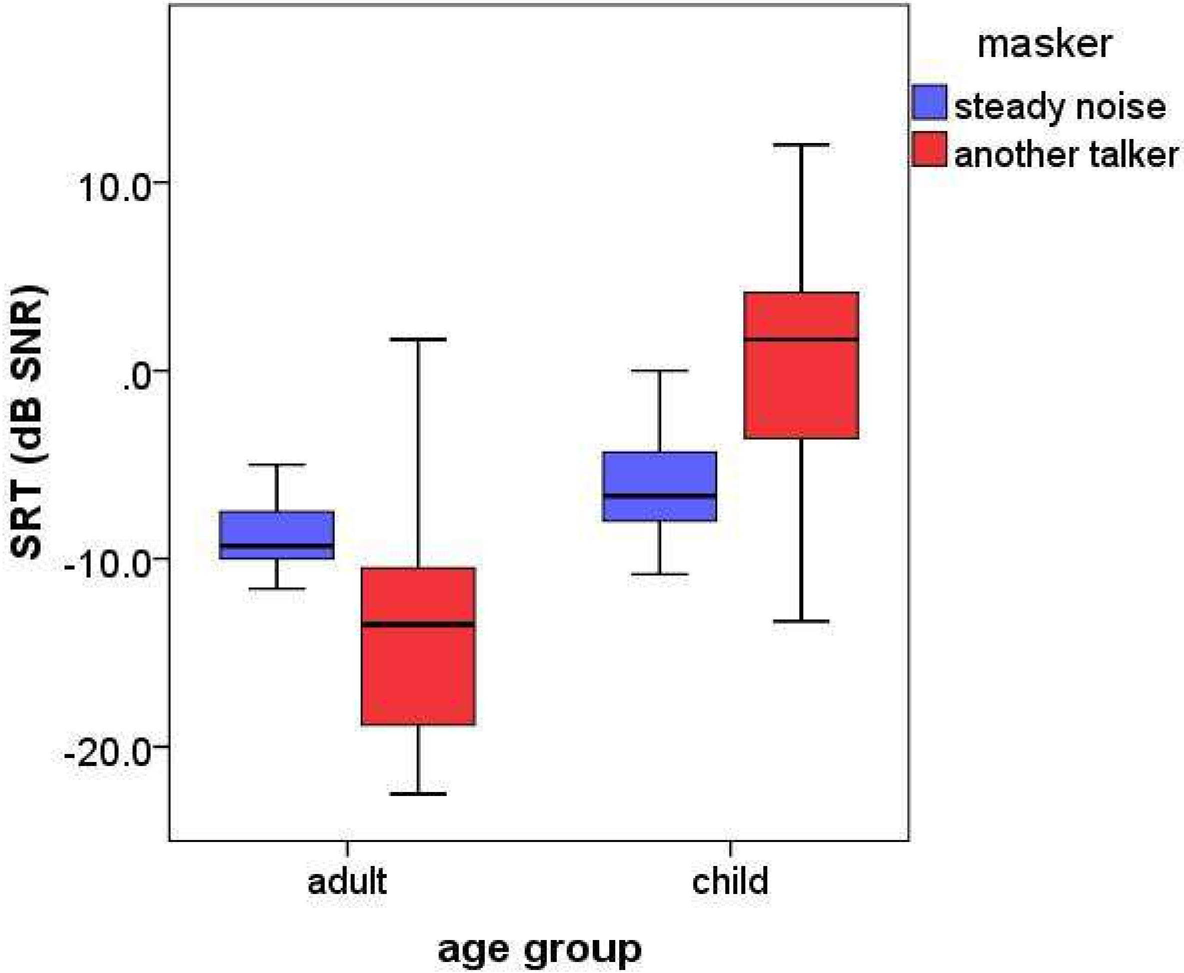


in quiet

in steady noise

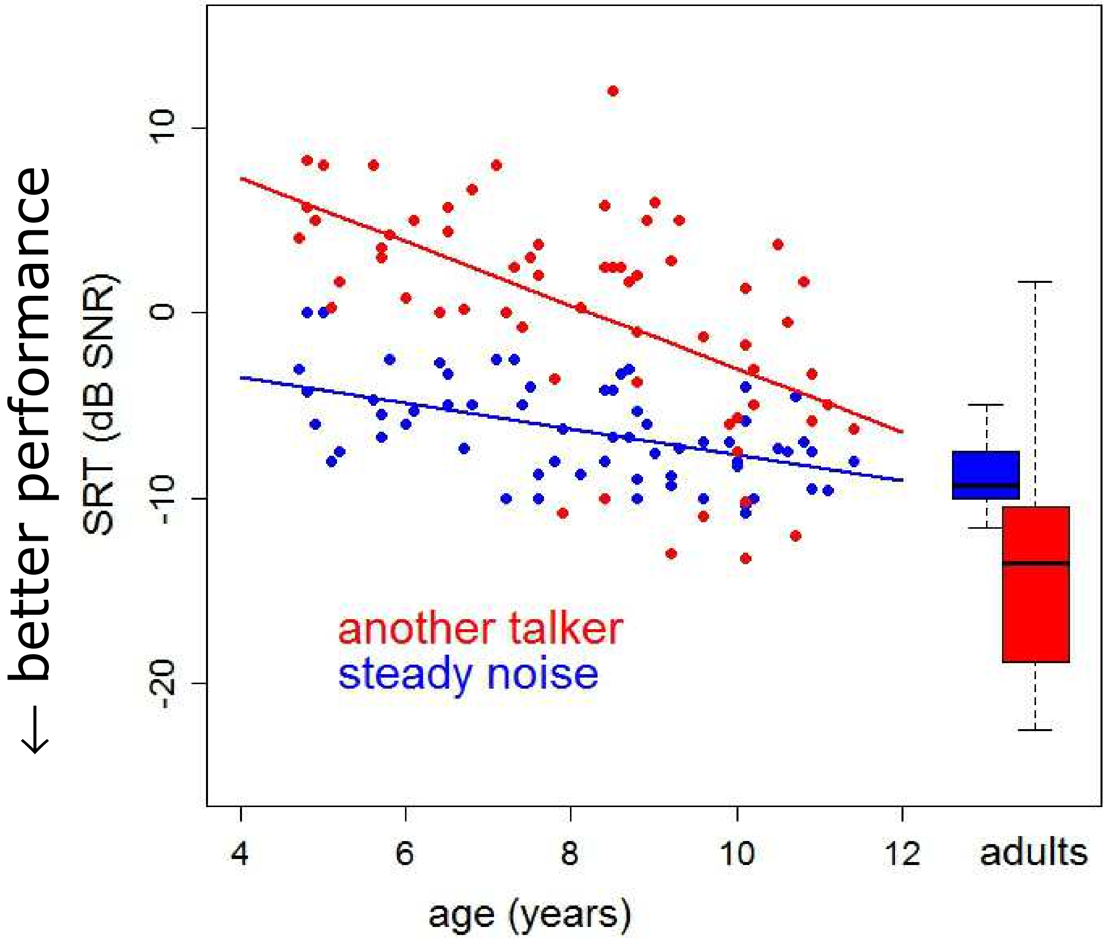
against another talker

Children find it hard to ignore another talker



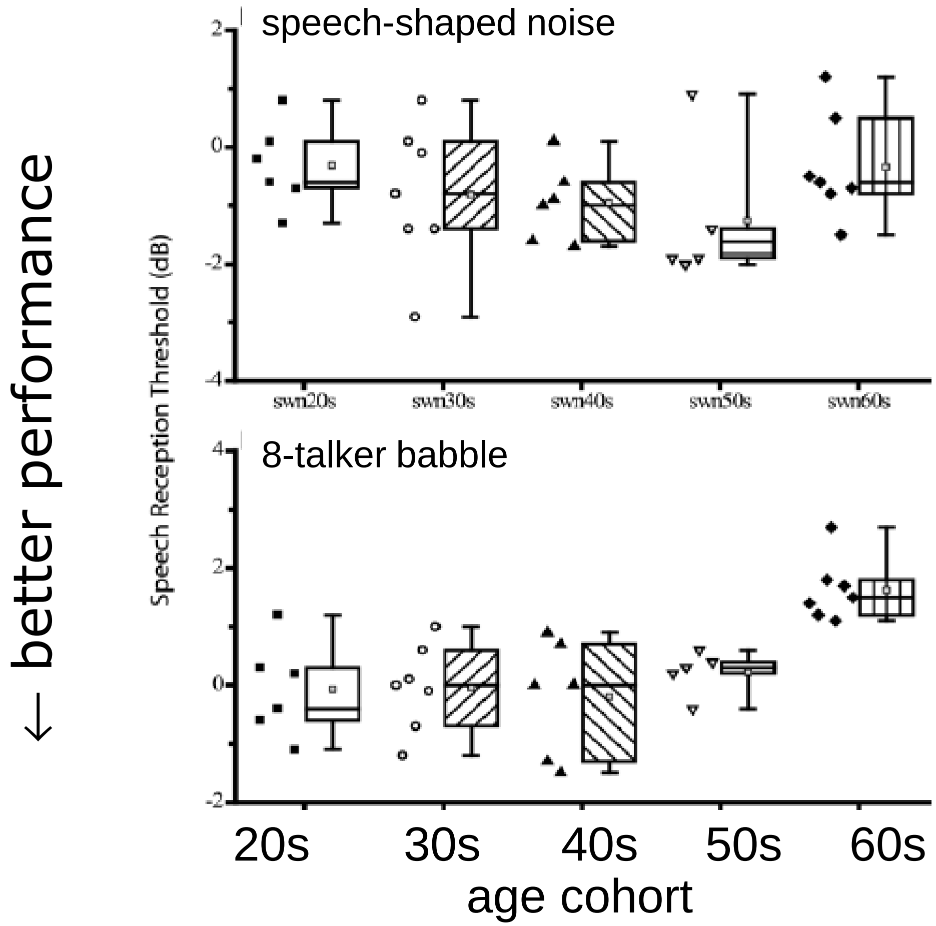
↓ better performance

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)

Specific Language Impairment (SLI)

- A *specific* deficit
  - occurs in the context of other cognitive abilities that are more or less normal.
- SLI — Late-developing and severely impaired language abilities alongside more-or-less normal cognitive abilities.

Email message from AZ when 15-years old

Subject: RE: A New Warriors

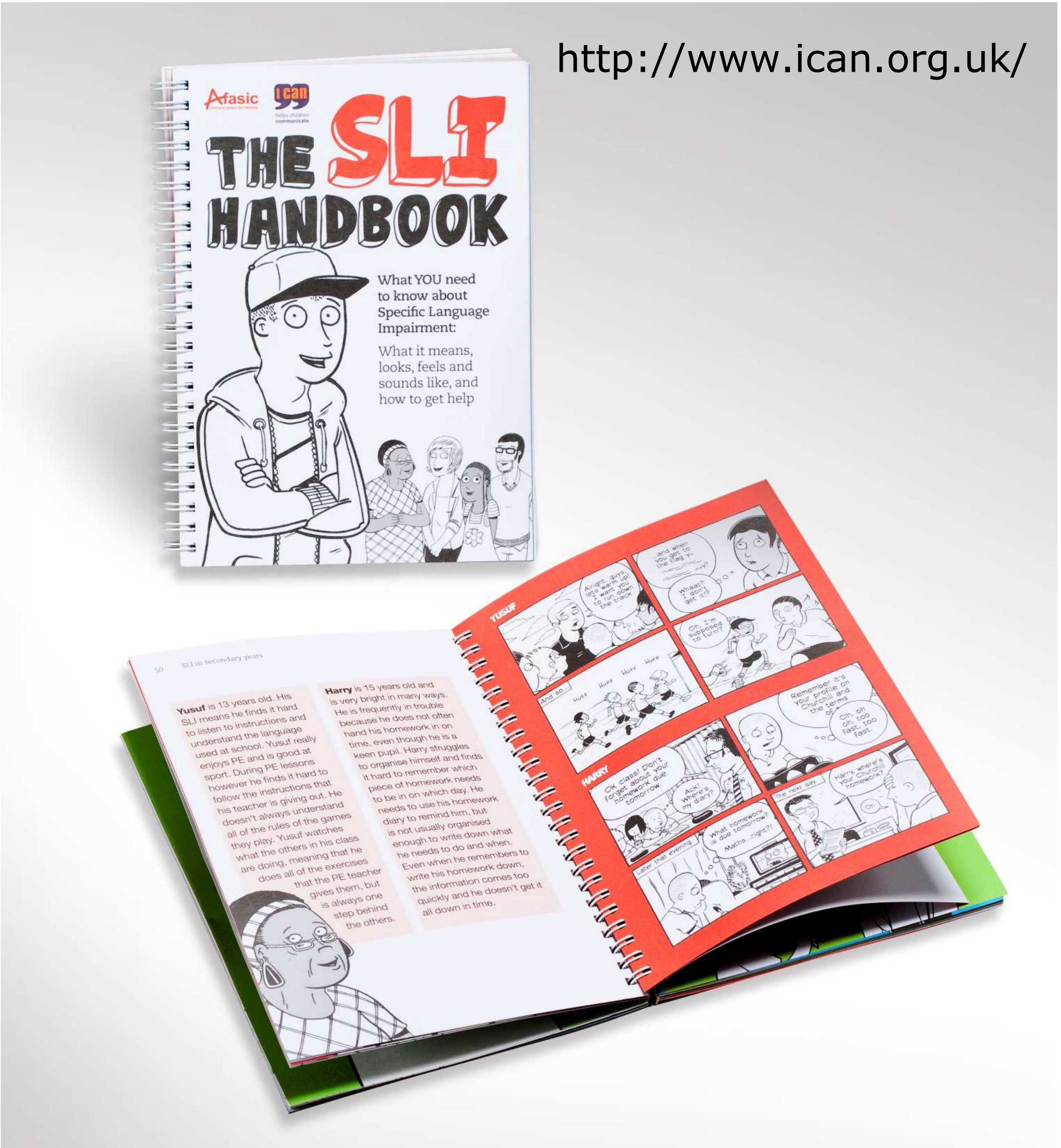
Date: Fri, 8 Aug 1997 11:07:40 -0400

Can you make a new warriors? (e.g I have planed a new space marine for the Blood Angels and I have called him The Blood Hurter. He got a jump pack on him and a rip arm from the tyranids and a orks head on his belt.)

AZ: a child with SLI

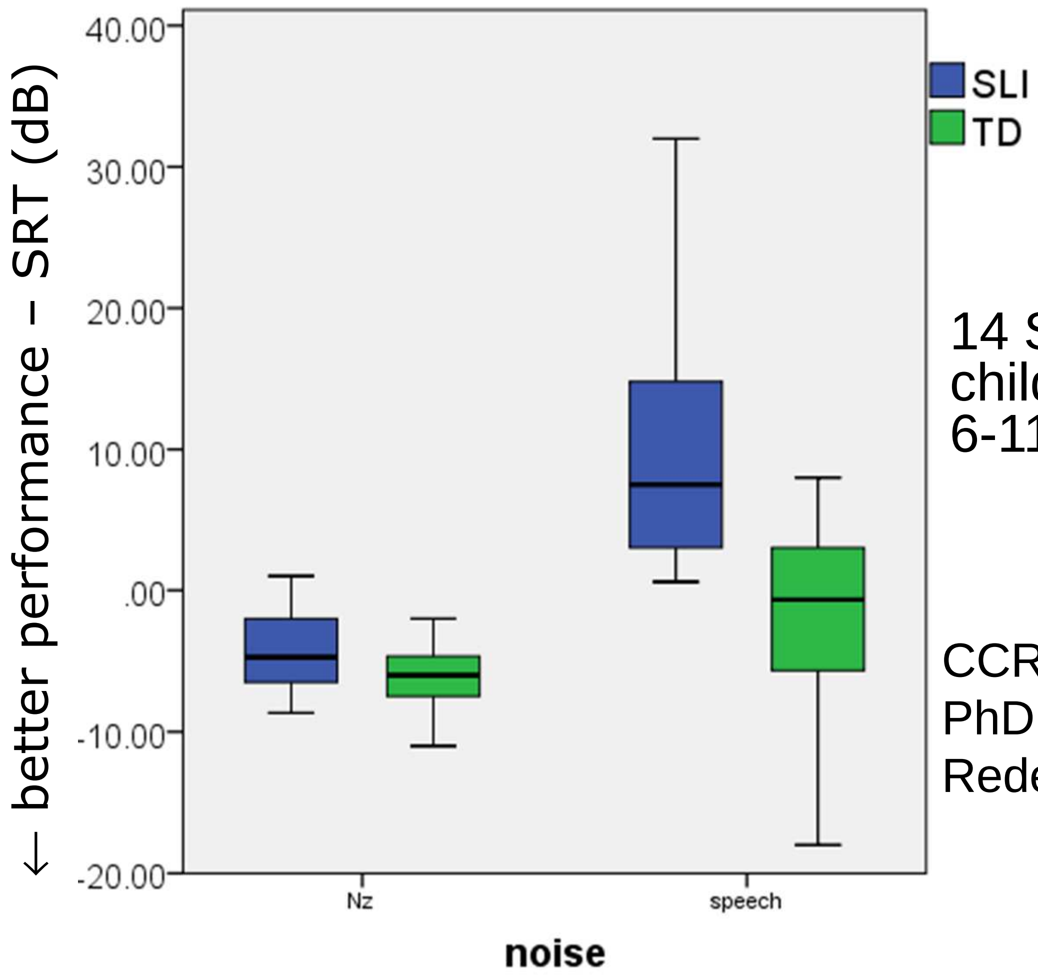
- Late to develop language, using 3 words at age 5: *Mummy, Daddy, Gangan*
- Errors in plural forms: *Two mens. Two foots.*
- Errors in verb tense: *My dad go to work. This is what they ated. My mum make the breakfast.*
- Errors in the use of embedded phrases: *Which cat Mrs White stroked? What did Mrs Brown dropped.*
- Nonverbal IQ ≈ 120-130

http://www.ican.org.uk/



A book for parents

Children with SLI find it hard ignore another talker



14 SLI & 17 TD children aged 6-11 years

CCRM sentences
PhD work of Csaba Redey-Nagy

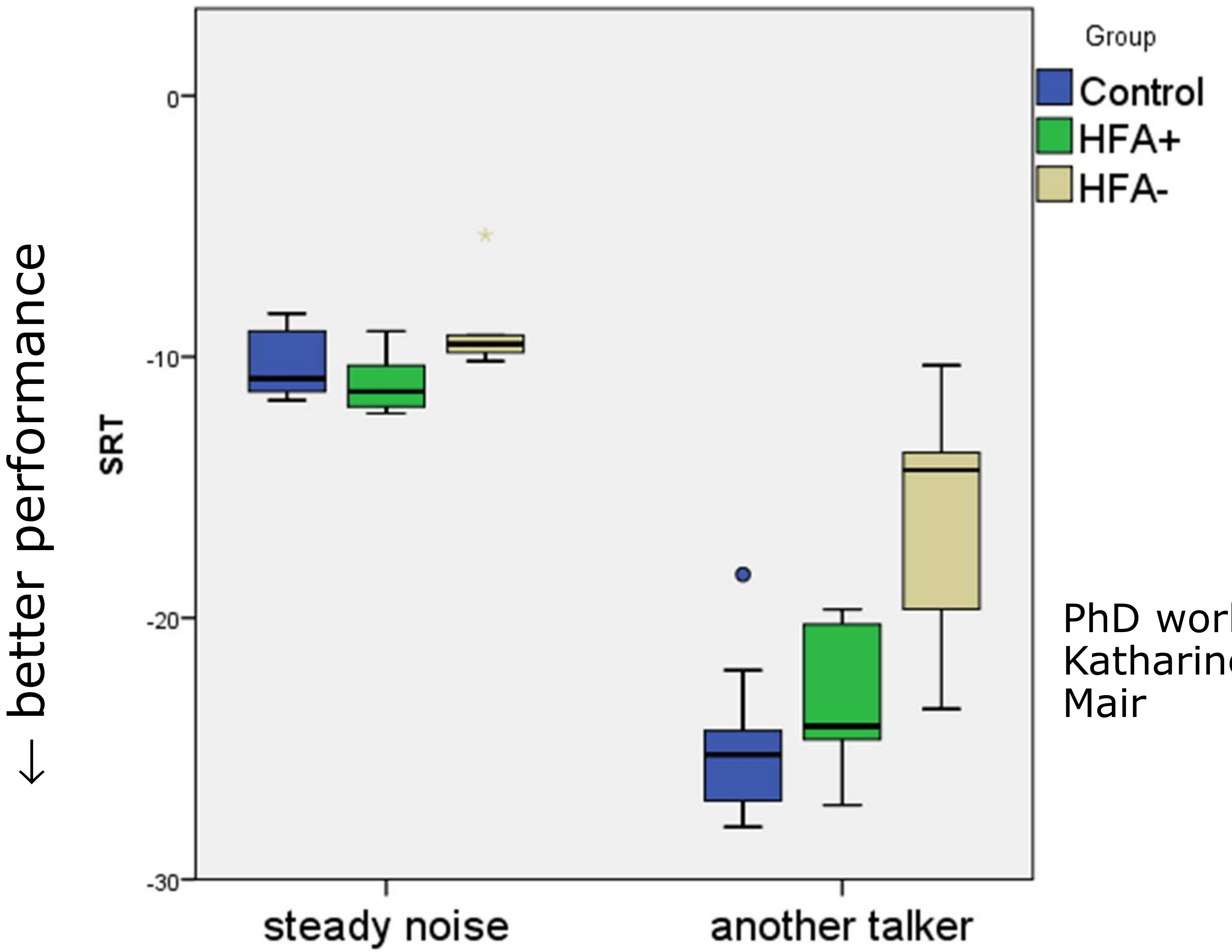
Autism Spectrum Disorders (ASD)

- Primary problems in aspects of social interactions
- Often accompanied by ‘sensory’ symptoms, e.g. hyperacusis

A personal report from someone with ASD

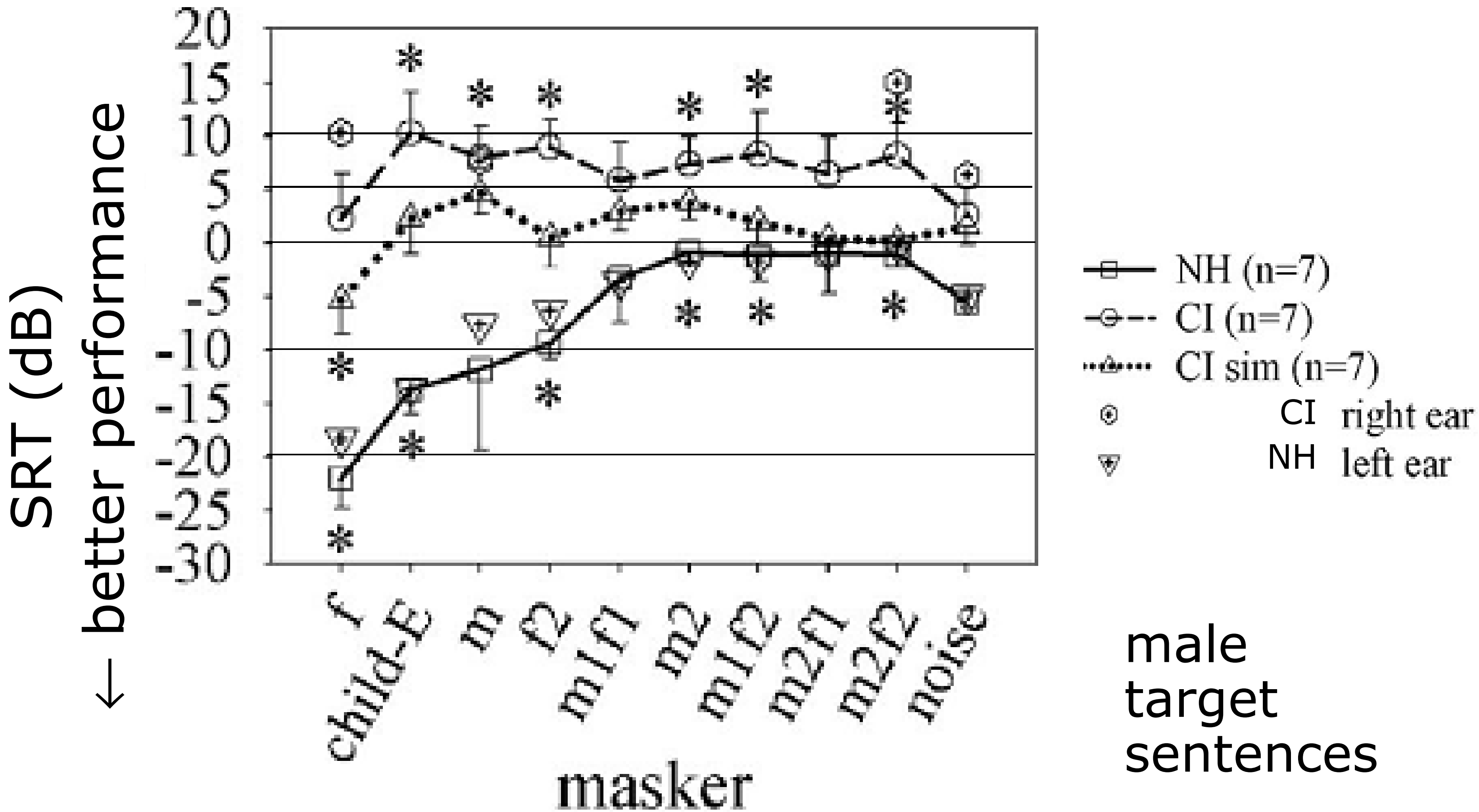
Being with a group of friends, and sitting in the middle, one conversation going on to my left and one to my right, no excessive background noise … when two are going at once, I can't follow either, so I usually just sit in silence.

Increased IM in some people with High Functioning Autism (HFA)



PhD work of Katharine Mair

CI users show little variation in SRT for different maskers



male
target
sentences

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'
- 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)

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

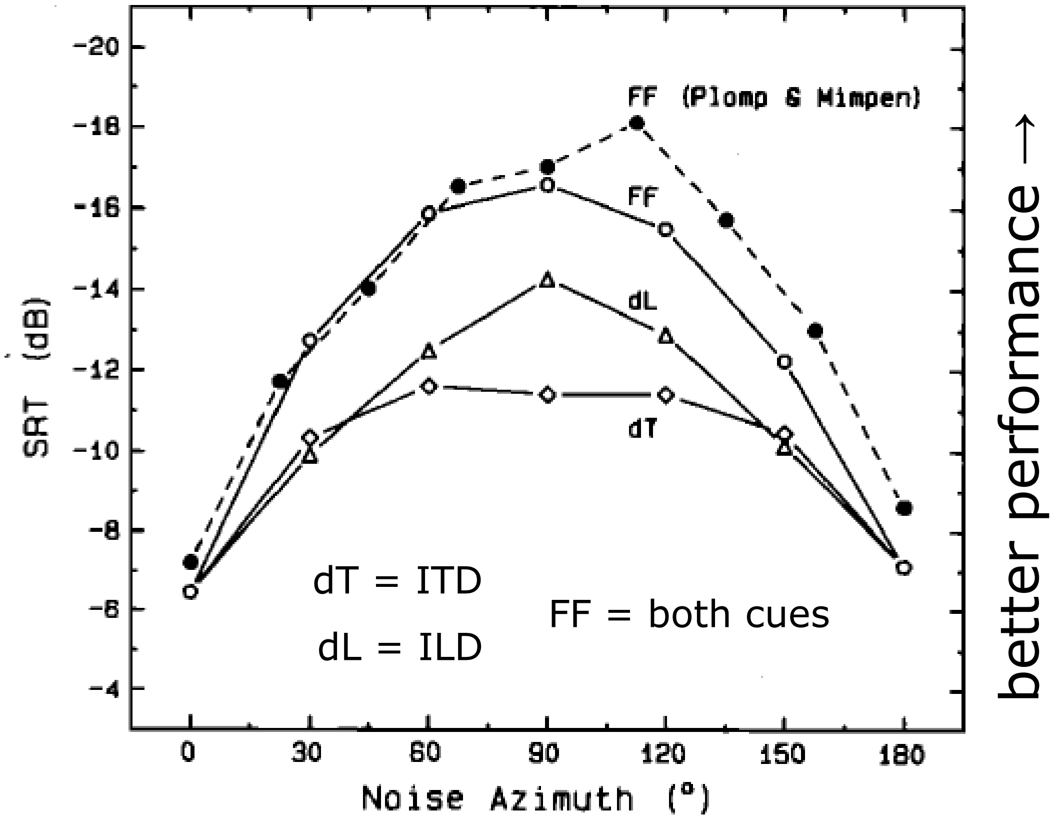


FIG. 5. Mean speech reception thresholds obtained in experiment I for three different noise types : FF (free field), dL (headshadow only), and dT (ITD only). The closed data points represent results of Plomp and Mimpen (1981) obtained in a free field.

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.