

Binaural Hearing and Speech Understanding: Potential Measures for Hidden Hearing Loss

Jaime A. Undurraga

2020-07-16

Introduction

Cocktail-party problem (Cherry, 1953)

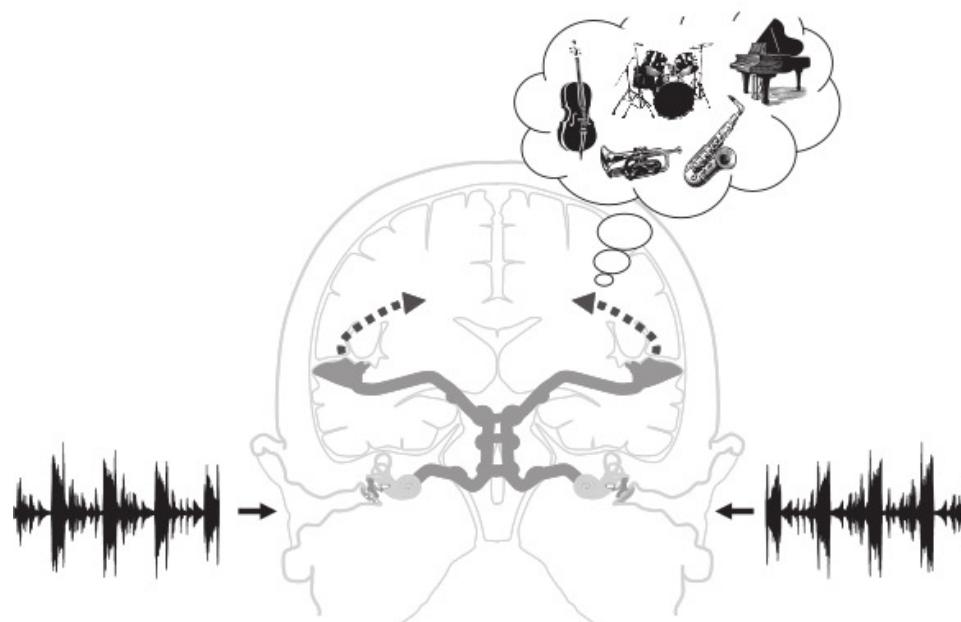
How do we recognize what one person is saying when others are speaking at the same time?

When everyone at a well-attended party talks at the same level, the speech of the attended talker at a distance of 0.7 m has a signal-to-noise ratio (SNR) of about 0 dB (the background is as intense as the target talker Plomp 1977). This level is sufficient to give adequate intelligibility for listeners with normal hearing (Miller, 1947).



Auditory scene analysis

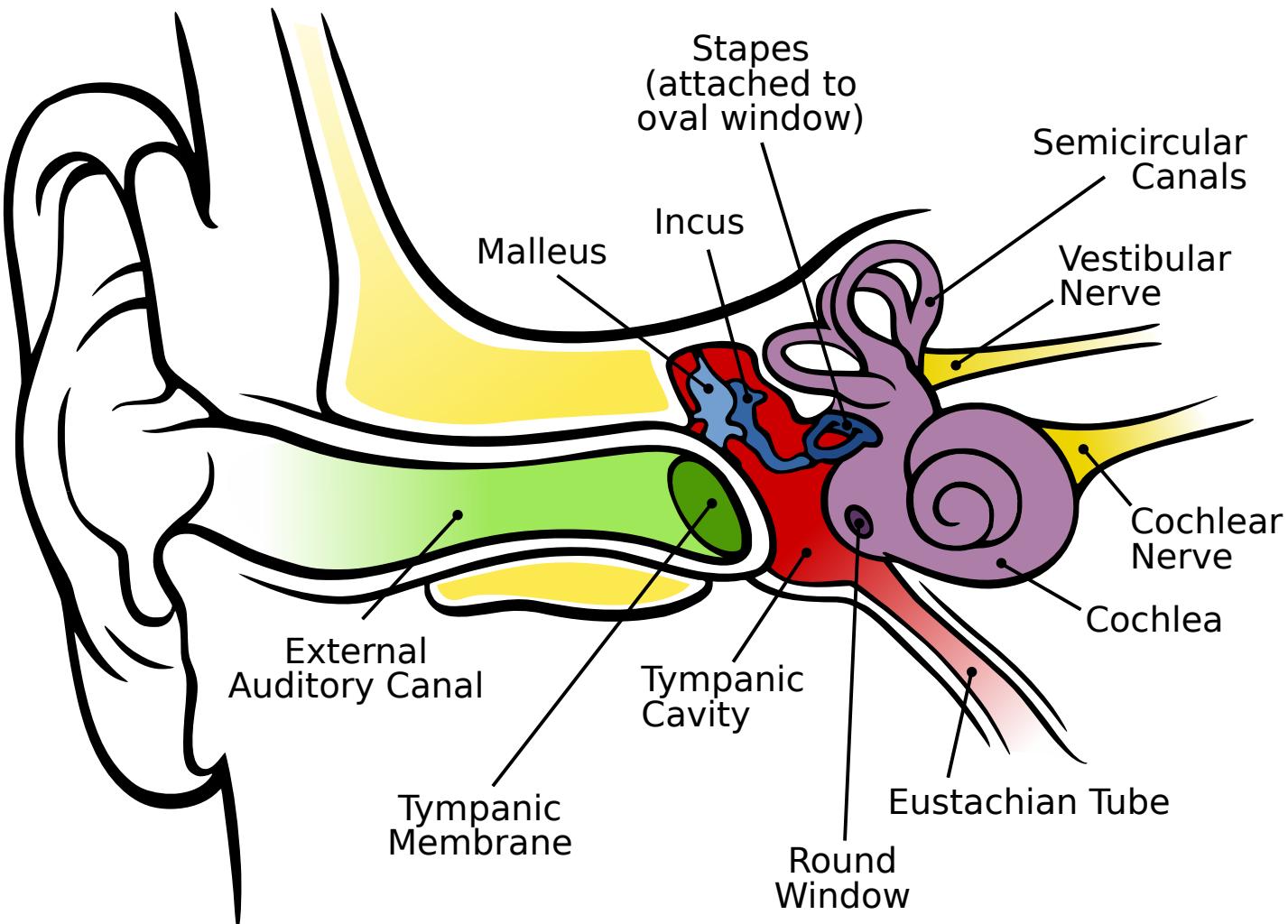
- A critical role of the auditory system is to parse the inputs from the left and right ears into auditory objects - auditory scene analysis. This is an essential role for survival, recognition, and communication.
- Binaural hearing provides cues that allow us to estimate the relative number and location of sources and objects in the environment.
- These cues also help us to estimate the dimensions and characteristics of rooms as well as to hear out speakers in the presence of interfering noise.

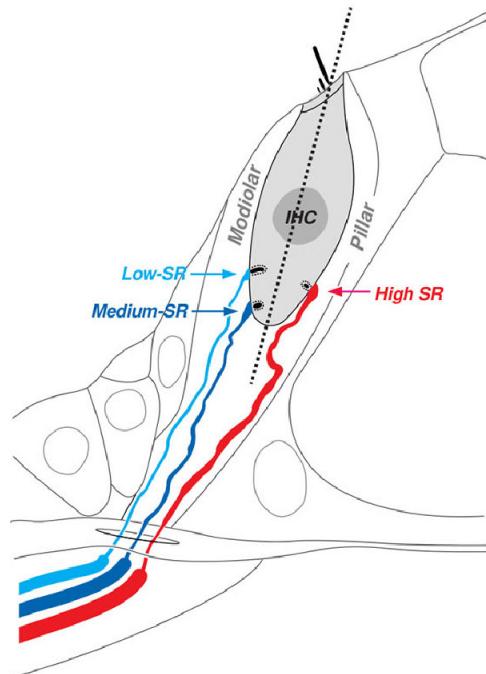
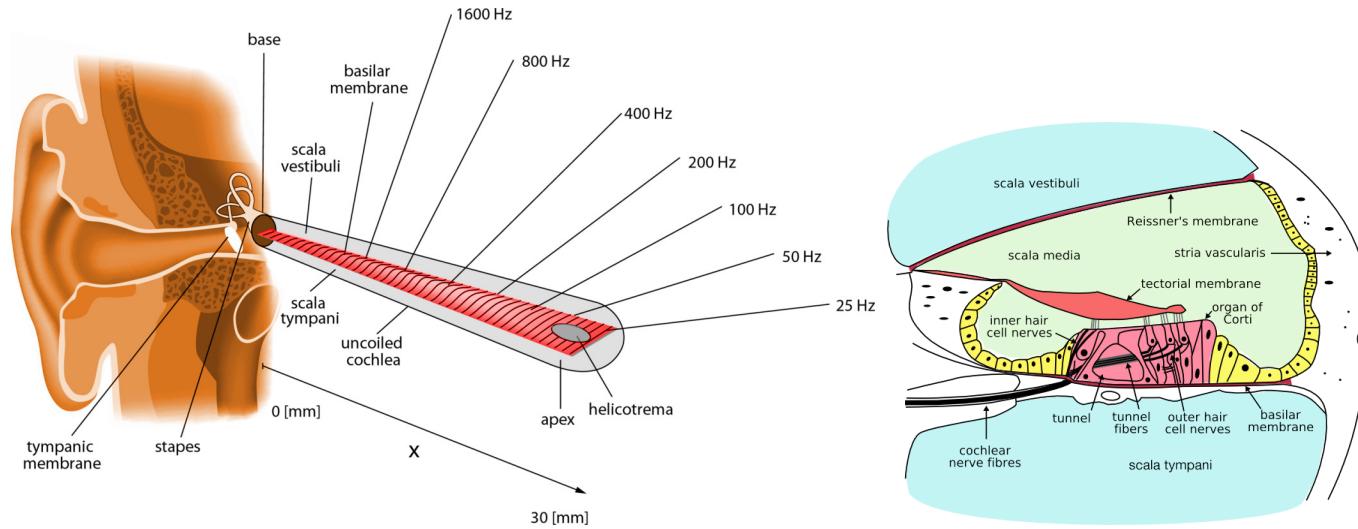


Grothe et al. (2010)

Encoding of sounds

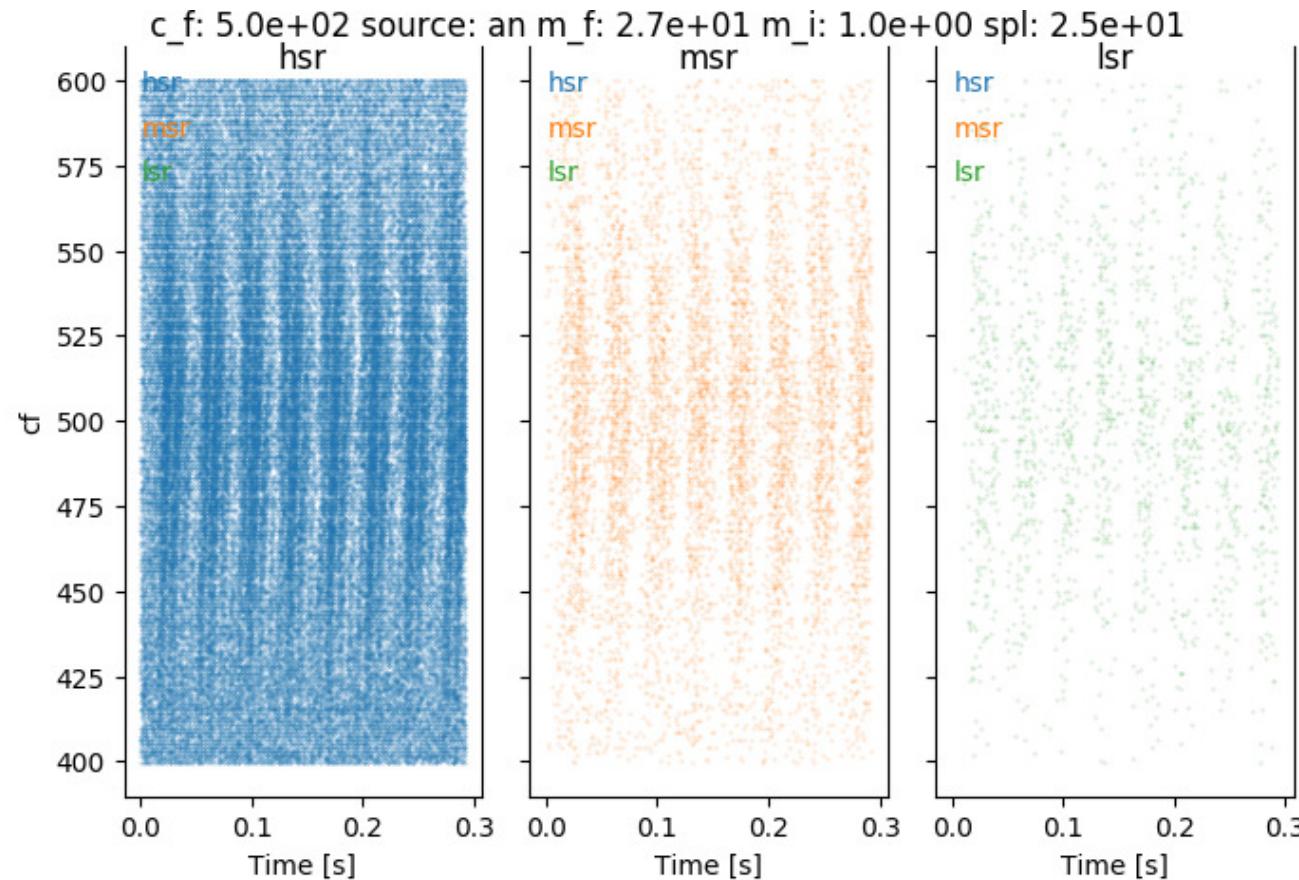
Monaural Pathway





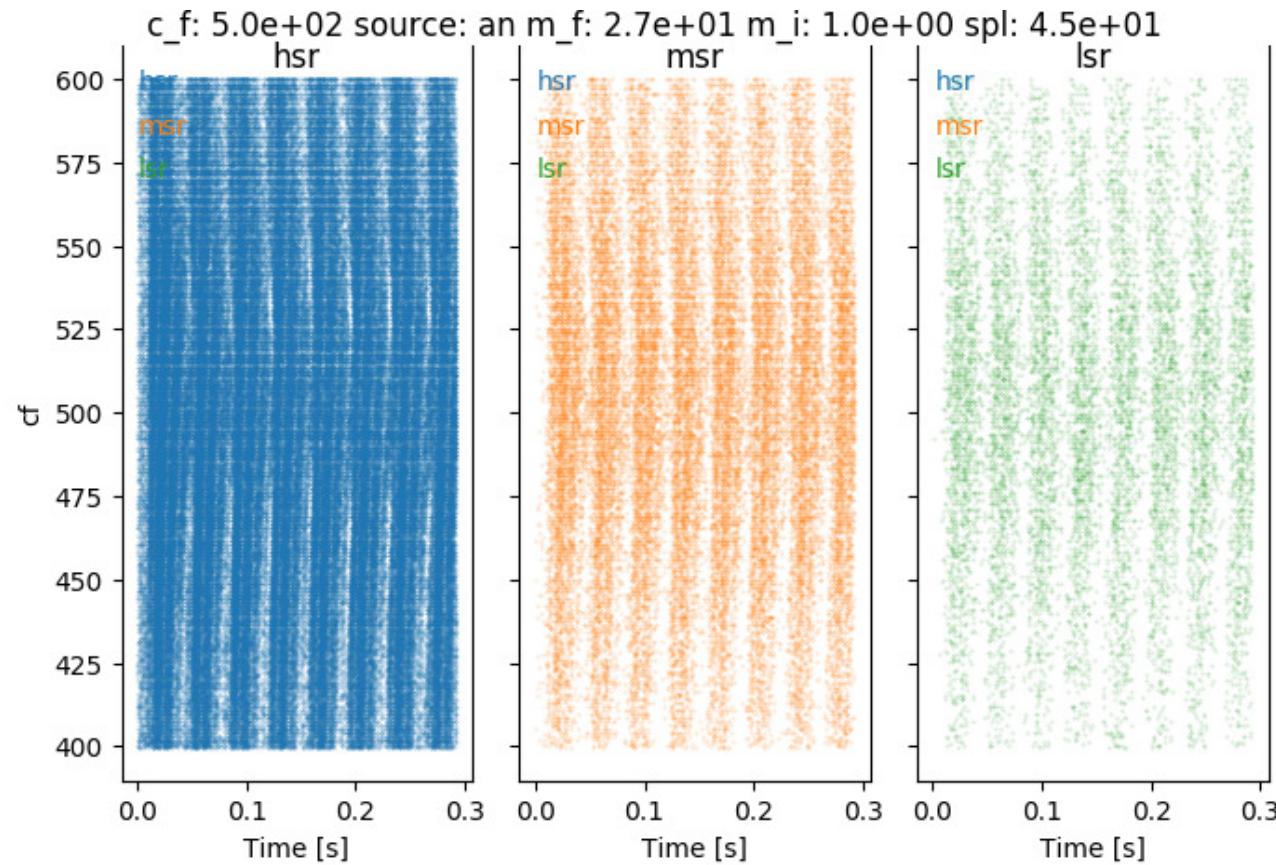
Coding at moderate intensity levels

- 500 Hz tone modulated at 40 Hz at 25 dB SPL



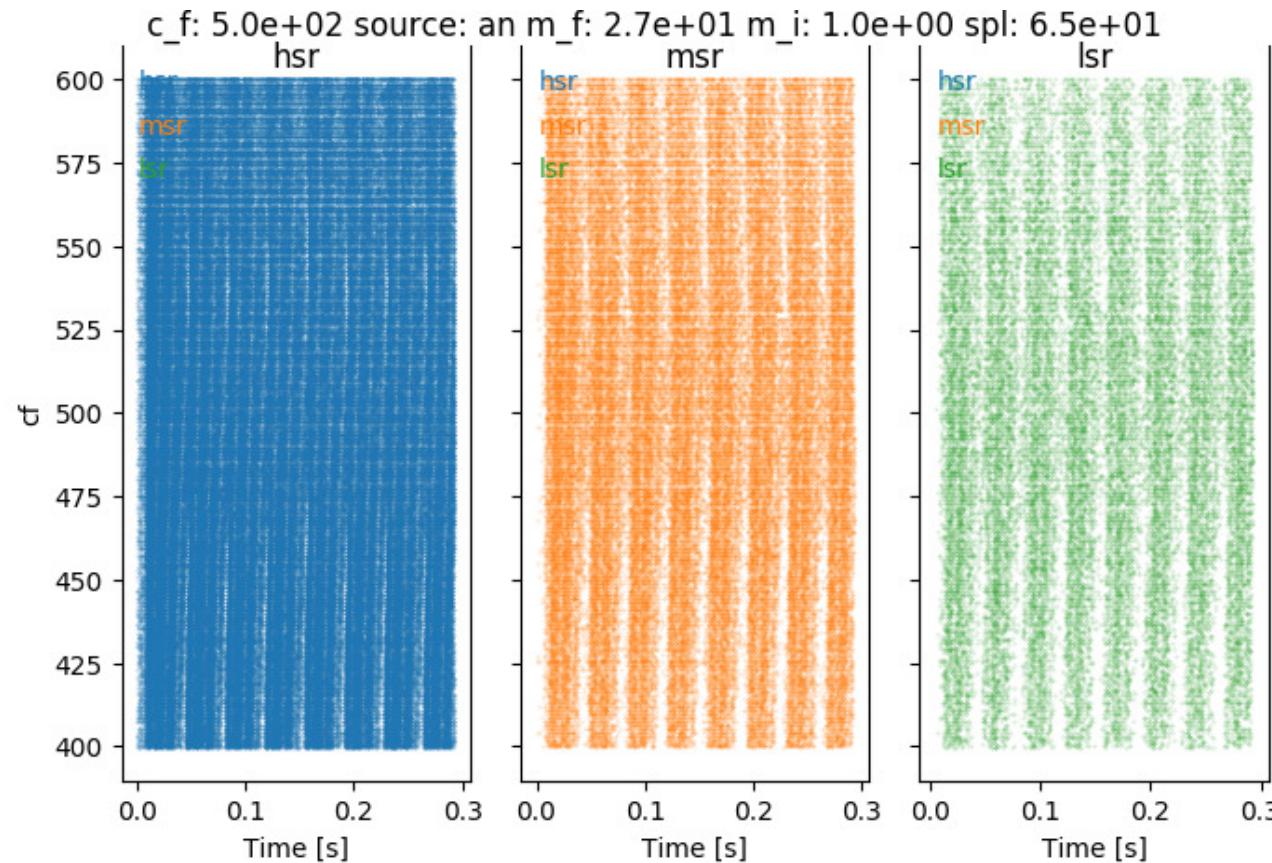
Coding at moderate-to-middle intensity levels

- 500 Hz tone modulated at 40 Hz at 45 dB SPL



Coding at middle intensity levels

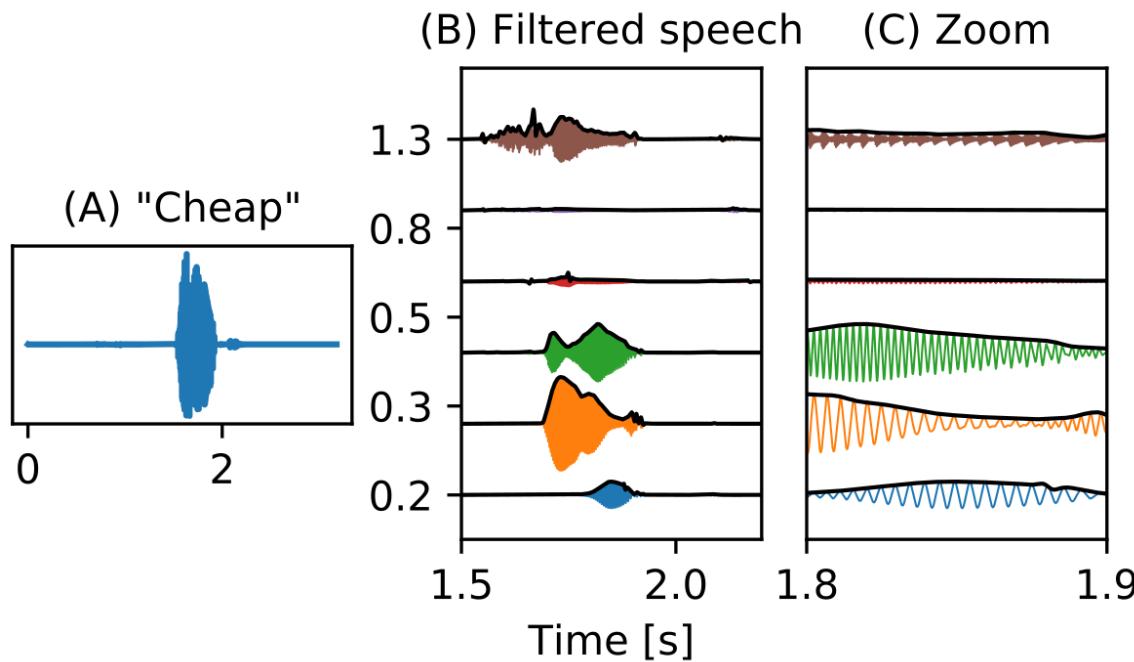
- 500 Hz tone modulated at 40 Hz at 65 dB SPL



- Cochlear synaptopathy reduces the ability to encode sound envelopes, specially at higher intensity levels.

Which features are conveyed by speech sounds?

- Temporal fine structure (TFS)
- Envelope information (ENV)
 - Play 6 Channel vocoder
 - Play Original

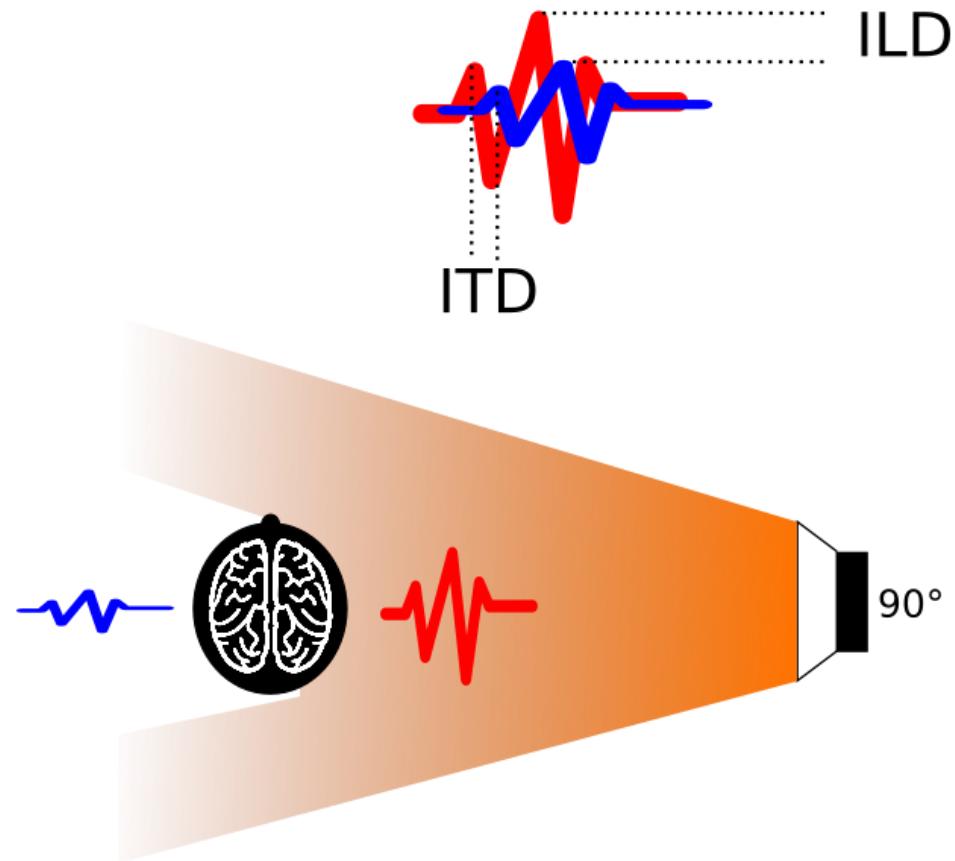


Binaural cues

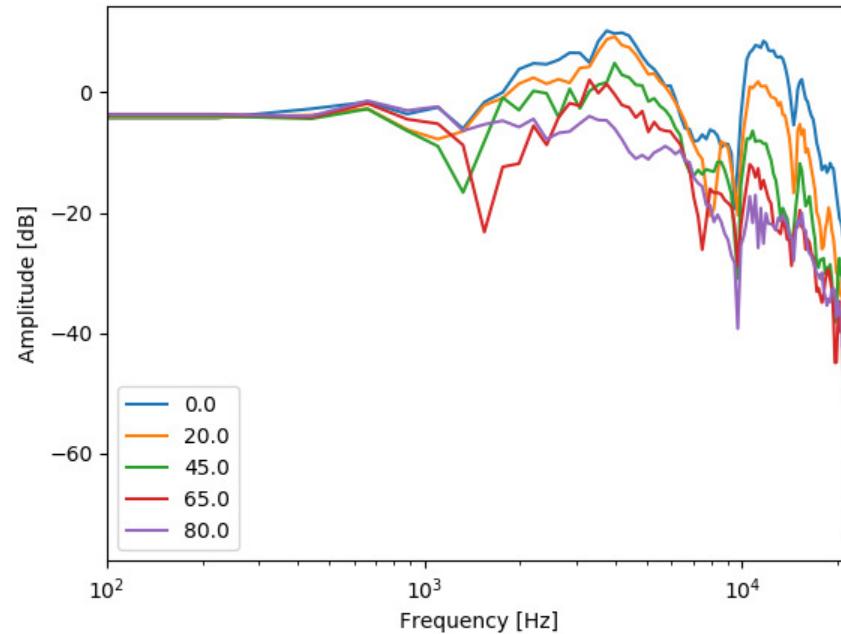
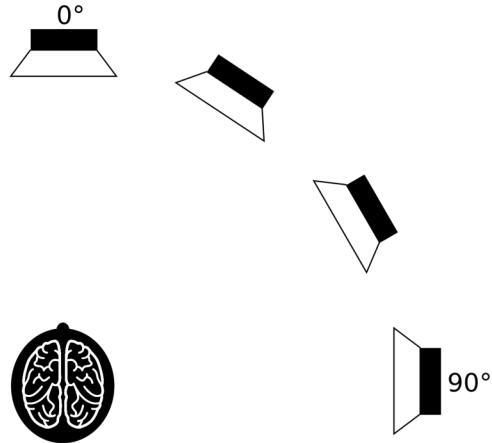
When listening to sounds, we rely on three mechanisms for both sound localization and auditory scene analysis

- Interaural level differences (ILDs)
- Interaural time differences (ITDs)
 - Interaural coherence

ITDs and ILDs

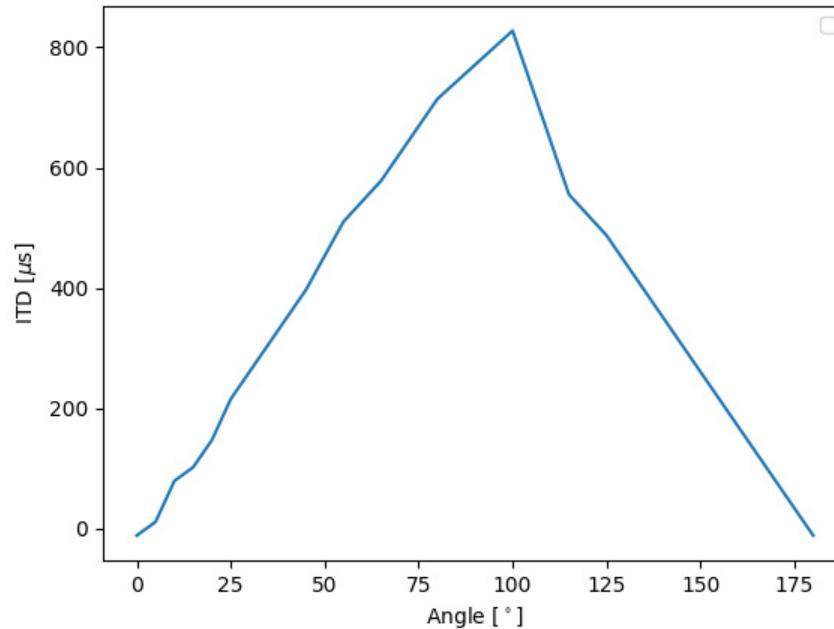
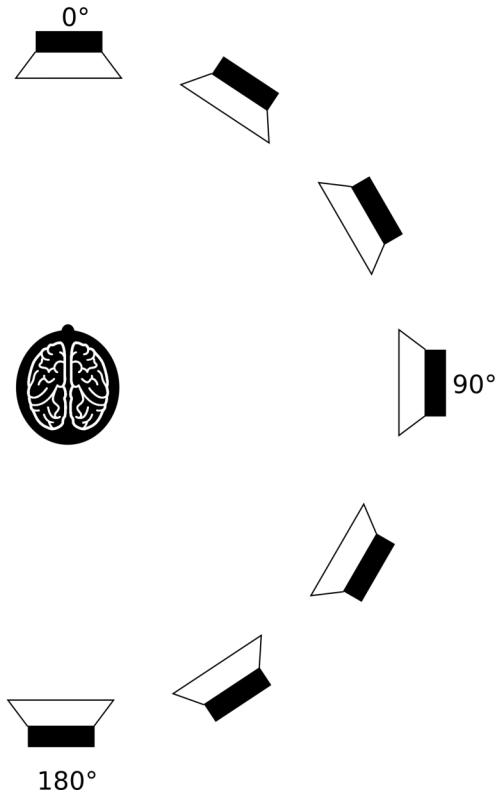


ILDs



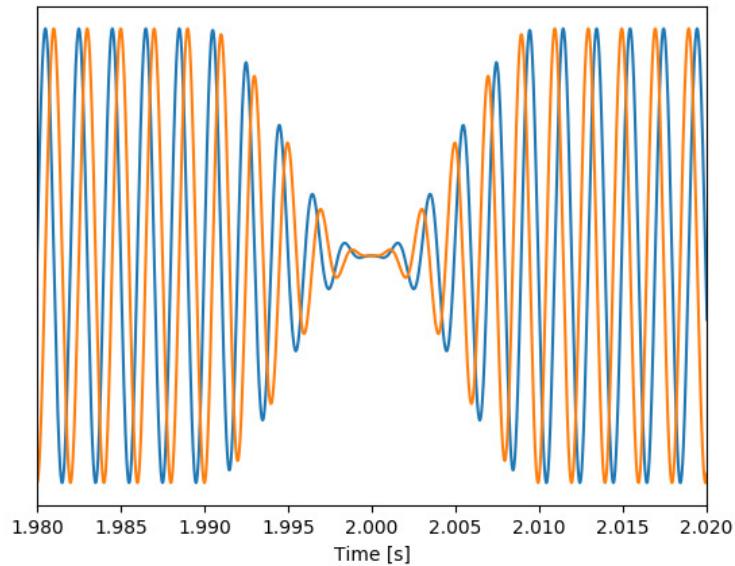
- ILDs only useful for frequencies higher than 1200 Hz
- ILDs can be as large as 20 dB (e.g. Bronkhorst & Plomp, 1988)
- ILDs resulting from the **head shadow** provide an advantage by sheltering the ear turned toward the target source from noise from the other side. However, listeners with unilateral hearing loss struggle when the target sound comes from the impaired side, specially in the presence of background noise.

Localization using ITDs

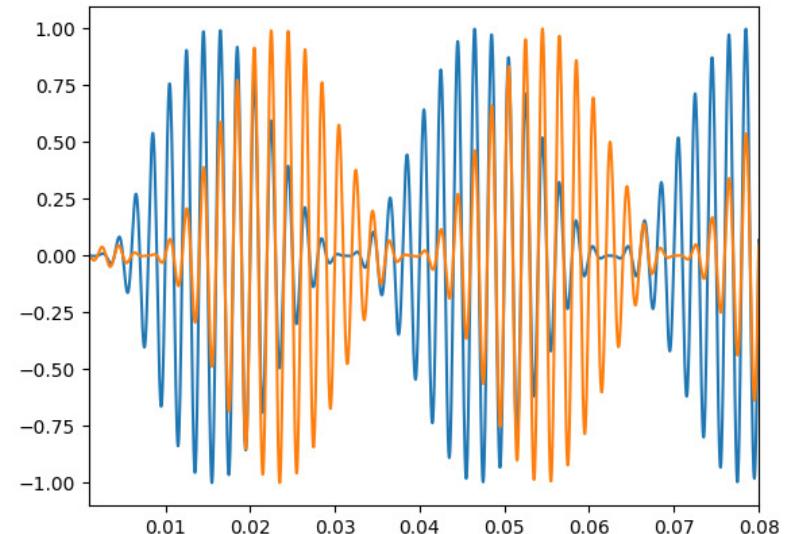


- ITDs within the physiological range experienced by human listeners are about $\pm 760 \mu\text{s}$; (e.g. Constan and Hartmann (2003) and Hartmann and Macaulay (2014))
 - ITDs (using fine structure) are useful for frequencies below 1500 Hz
- ITDs in the envelope of the signal are also used to determine the location of a source in both lower and higher frequencies.

Temporal fine structure ITDs

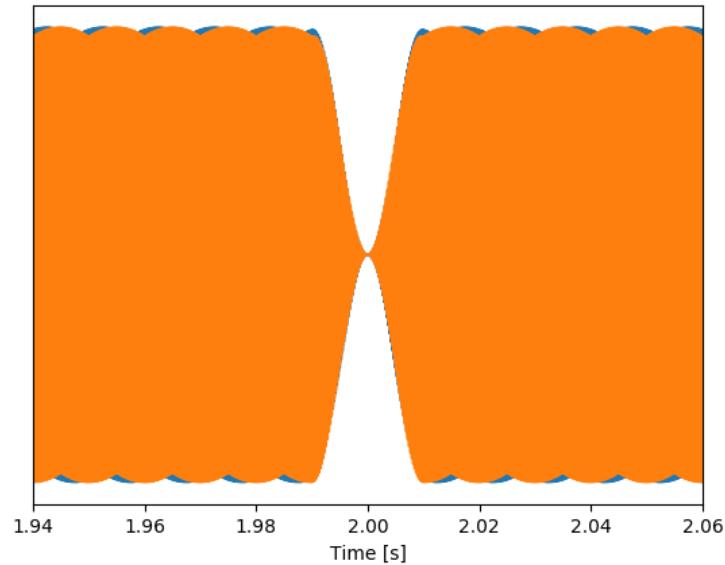


500 Hz tone TFS left-to-right

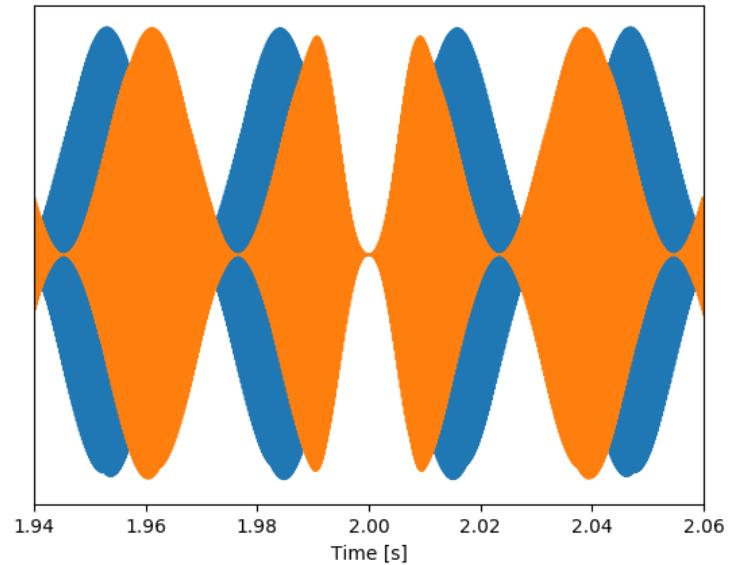


Modulated 500 Hz tone
with envelope itd (ipd) left-to-right

Envelope ITDs



4000 Hz tone TFS itd (ipd) left-to-right



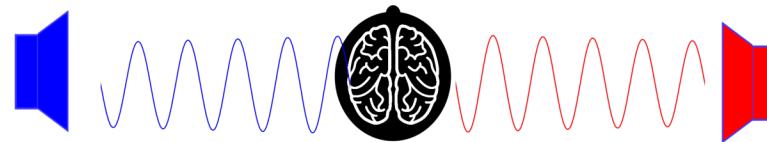
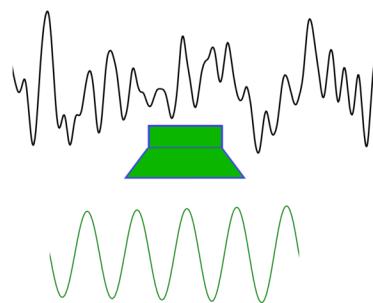
Modulated 4000 Hz tone
with envelope itd left-to-right

Binaural processing and speech understanding

Binaural Redundancy (aka Binarural summation)

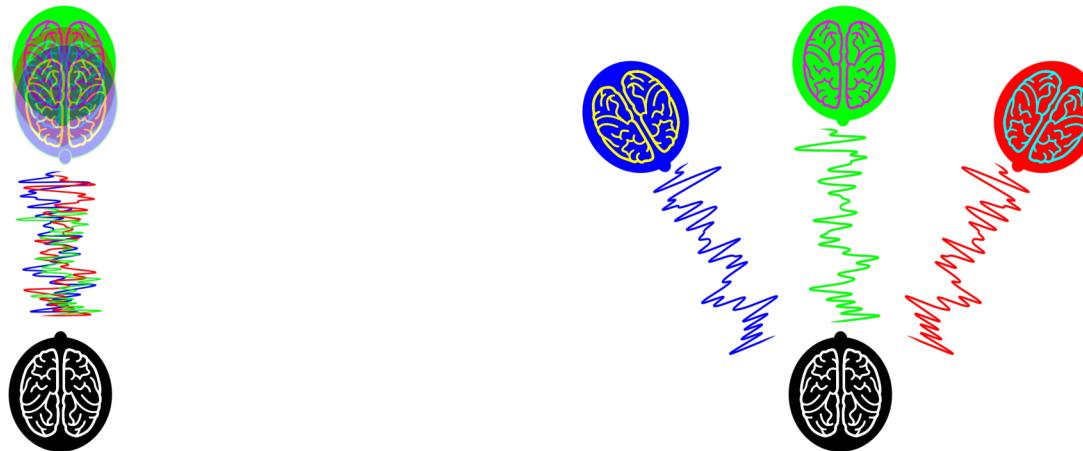
- Loudness doubles when the two ears are used instead of one ear for a sound coming from the front of the listener (a single ear would require an increase of about 10 dB; Fletcher and Munson, 1933)
- Just noticeable differences in intensity and frequency improve with signal redundancy
- Speech recognition in the presence of background noise improves (Marrone 2008, Neher 2009)
- Hearing impairment may lead to a slightly weaker binaural benefit in patients (Dillon, 2001)
- Binaural sounds can be louder than with a monaural presentation without causing discomfort (even true for CI-treated patient)

Binaural Release from Masking (or Binaural Squelch; or Hirsh effect)



- Binaural release from masking may improve detection threshold up to about 16 dB for frequencies around 250 Hz and around 3 dB at 1500 Hz

Spatial Release from Masking



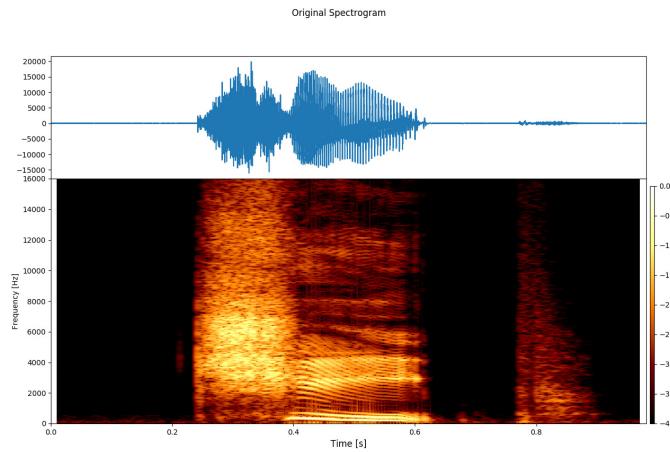
- Binaural release from masking may improve detection thresholds up to 12 dB for multiple speech interferers (Jones and Litovsky, 2011) and facilitates source segregation (Drennan, Gatehouse, and Lever, 2003).
- Segregation is always better for the combination of both ITDs and ILDs cues (Culling, Hawley, and Litovsky 2004).
- A separation of only 10° is already strong enough to allow segregation (Brungart and Simpson, 2007)
- ITD is a critical spatial cue for sound localization and speech perception in noise (Bronkhorst & Plomp, 1988; Wightman & Kistler, 1992) .

The role of spatial cues in Auditory scene analysis

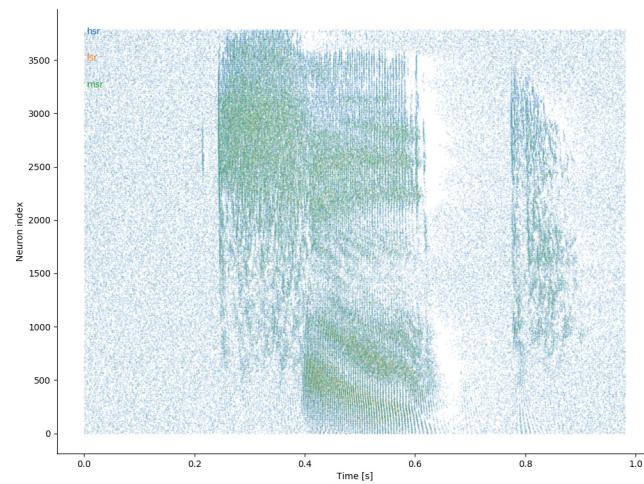
- When the speech sources are spatially separated, normal listeners can perceptually segregate and selectively attend to the source of interest if the sounds arrive from the same location - **spatial release from masking (SRM)** (Freyman et al., 1999; Brungart, 2001; Freyman et al., 2001; Hawley et al., 2004)
- This task is extremely challenging for listeners with sensorineural hearing loss (with or without hearing aids) or with cochlear implants (Loizou et al., 2009; Marrone et al., 2008).
- Moreover, listeners with “normal” hearing and elder listeners experience great difficulty when listening in the presence of background noise (Ruggles et al., 2011; Swaminathan et al., 2015; Gallun et al., 2013).
- Hearing impaired listeners with symmetric binaural hearing often demonstrate reduced SRM primarily due to increased thresholds in spatially separated conditions (Arbogast et al., 2005; Marrone et al., 2008b; Best et al., 2012)

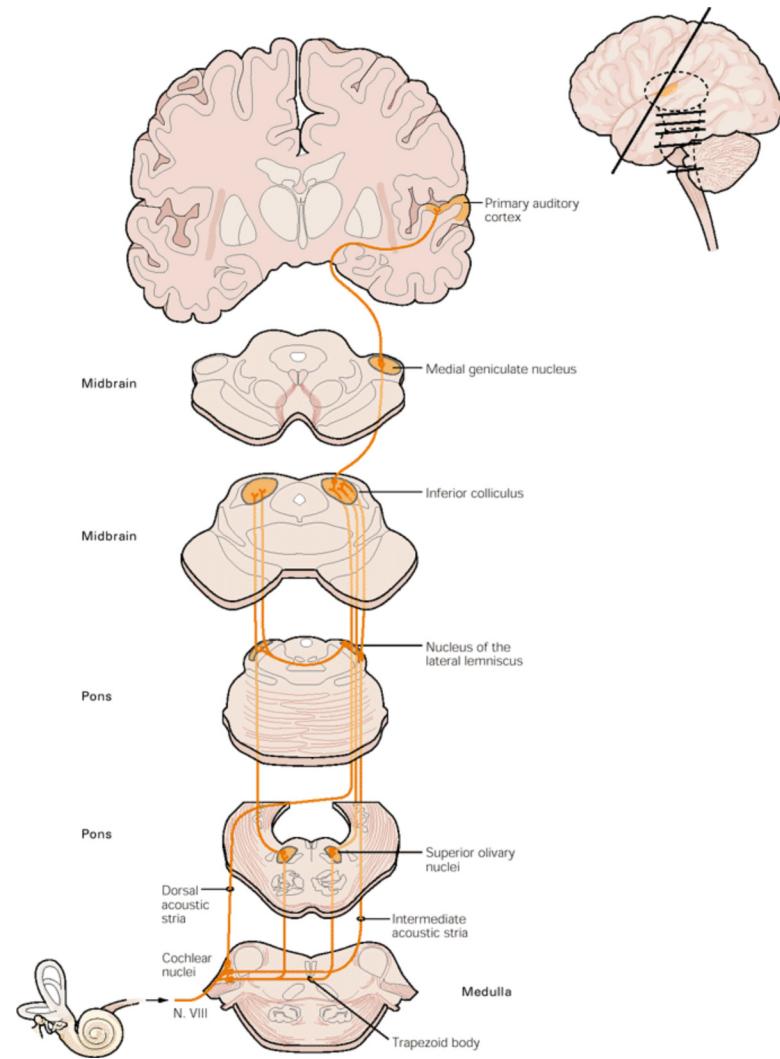
Auditory Pathway

Input to one ear

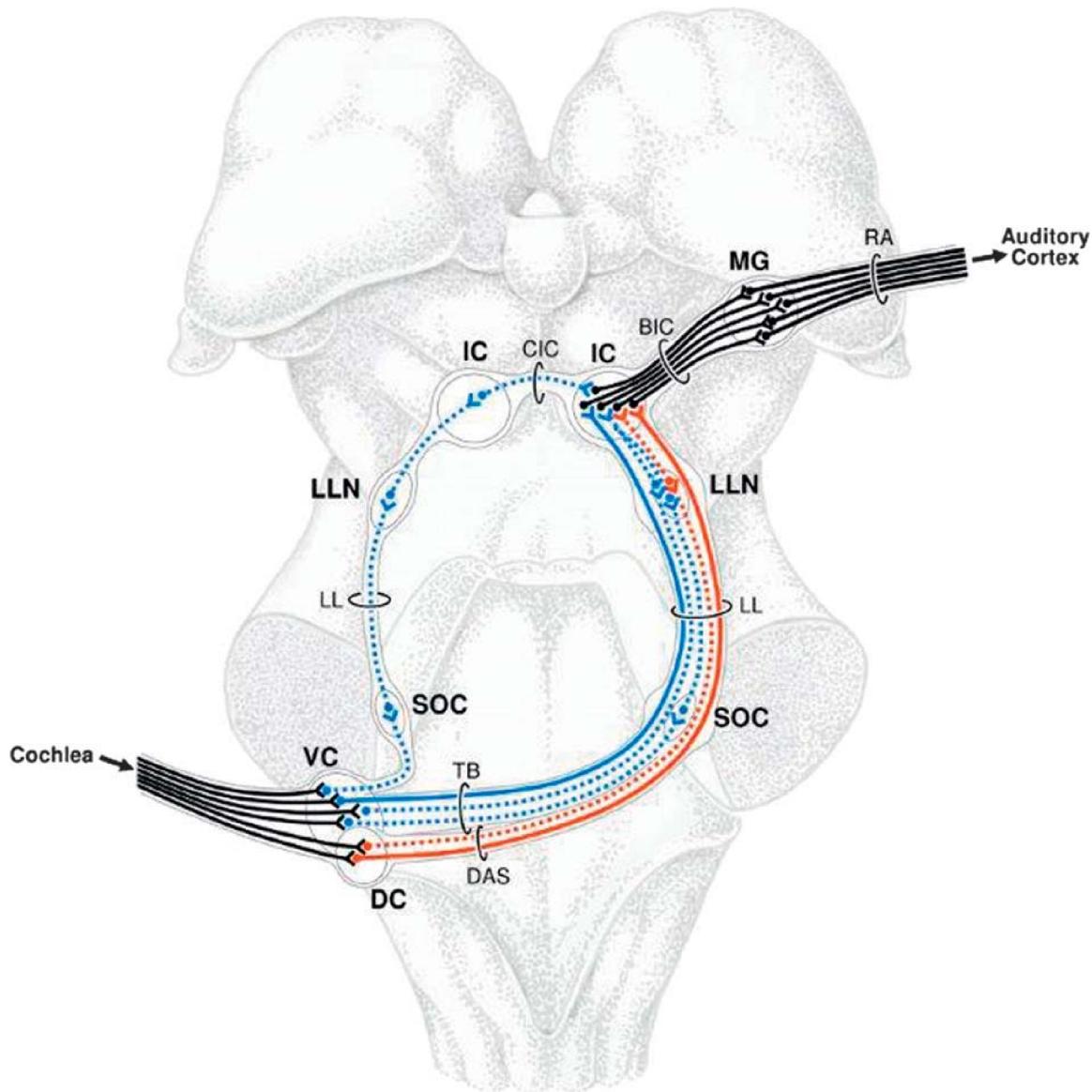


Output at the auditory nerve level



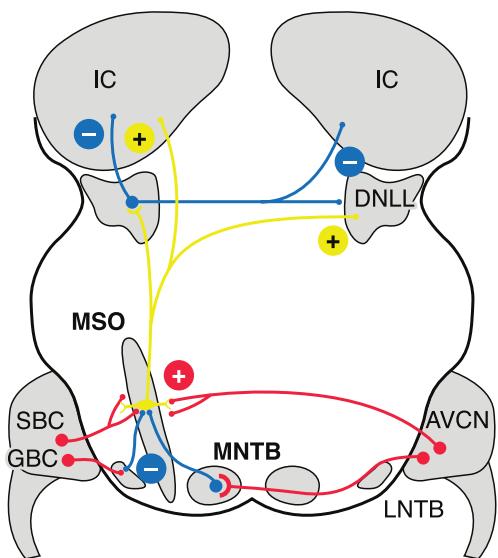


Binaural pathway



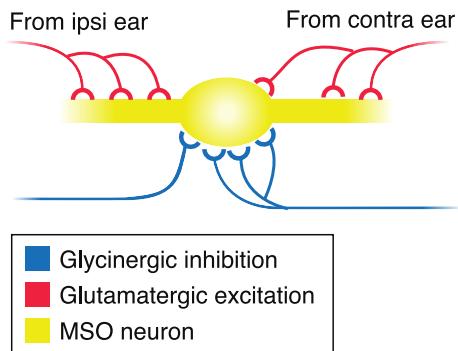
ITD pathway

D

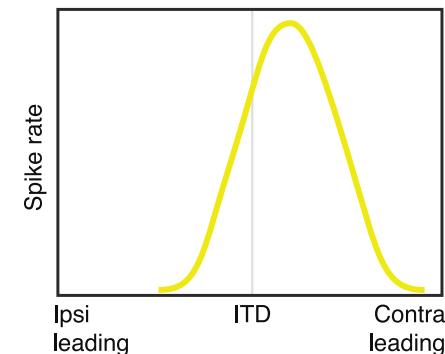


AVCN: Anterio-ventral cochlear nucleus
 DNLL: Dorsal nucleus of the lateral lemniscus
 GBC: Globular bushy cell
 IC: Inferior colliculus
 MNTB: Medial nucleus of the trapezoid body
 MSO: Medial superior olive
 SBC: Spherical bushy cell

E



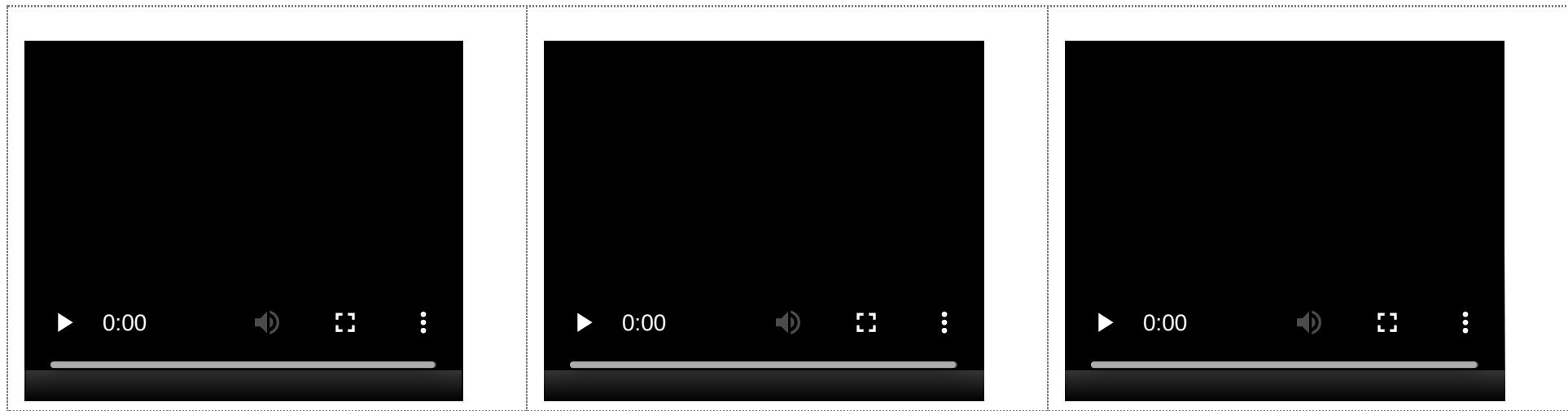
F



(Grothe et al. 2010)

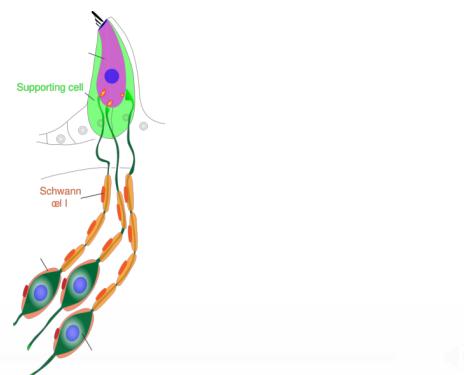
- The initial site of ITD processing is considered to be the MSO.
- MSO are innervated by direct excitatory ipsi- and contra-lateral inputs from spherical bushy cells and inhibitory inputs indirectly originating from the globular bushy cells.

Coincidence detection (MSO)



Hidden Hearing Loss

- Clinically, pure-tone audiometry (PTA) is used to assess the hearing status of human listeners (typically between 125 to 8000Hz). However, some “Normal Hearing (NH)” listeners (5 - 15 % of patients) report great difficulties in understanding speech, particularly in noisy and high reverberation environments, beyond the level that would be predicted from their audiogram (Füllgrabe et al., 2015).
- Animal studies in mice and guinea pigs have shown that noise exposure can lead to 40 - 50 % of loss of fibres synapses with the inner hair cells in the cochlea - **Synaptopathy** (e.g. Kujawa and Liberman, 2015, 2009; Liberman and Kujawa, 2014).
- Histological data seems to confirm this effect in “Normal Hearing” human cochleas (Viana et al., 2015).
- Synaptopathy affect all types of AN fibers: high-SR / low- and medium-SR fibers ratio of about 1:3 (Marmel et al. 2015).
- Noise exposure may reduce the number of synapses between AN fibers and hair cells (Synaptopathy) but also induce **transient loss of cochlear Schwann cells resulting in permanent auditory temporal deficits (Wan and Corfas 2017)**



Binaural hearing and hidden hearing loss

- Binaural hearing plays a critical role when listening speech in noise.
- A deterioration in the ability to encode the temporal fine structure of speech will reduce the ability to encode ITD and speech understanding.
- If noise exposure induces **transient loss of cochlear Schwann cells (Wan and Corfas 2017)**, the ability to encode ITD cues will decrease.

Binaural processing may be more sensitive than monaural measures.

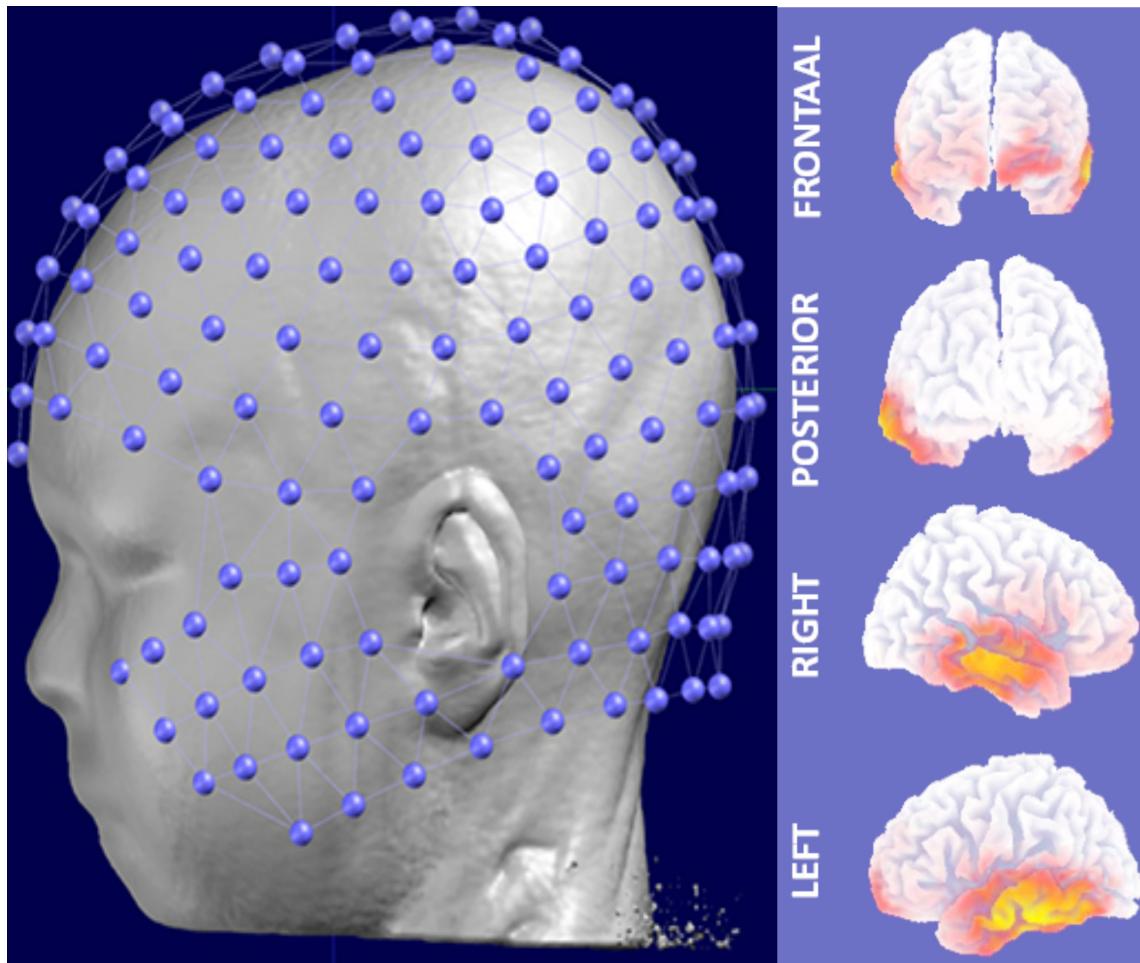
- Processing of binaural stimuli is accomplished via an “**effective multiplication**” of the neural information arising from each monaural channel (Bernstein and Trahiotis 2016). Thus, small monaural neural damage could result in more substantial and perceptually meaningful binaural losses.

effective multiplication of the neural information arising from each monaural channel (e.g., coincidence detection, cross-correlation). Within this rubric, one can imagine that even a small loss of the monaural neural units that provide inputs from frequency-matched left-right cochlear regions could result in a much larger loss within the central, binaural processor. To illustrate, let us assume that, without loss, the neural inputs from 100 left-right pairs of monaural nerve fibers represent inputs for binaural interaction. One would expect that if the neural information were perfectly interaurally synchronized, then 100 multiplications of the monaural events would result in 100 outputs after binaural interaction. Now, let us assume that 20% of the inputs from each monaural channel are removed in a manner that is effectively random with respect to left and right ears. Then, one would expect, everything else being equal, that the information from each of the 80 independent monaural neural units would yield, on average, 64 outputs after binaural interaction. That is, under this scenario, a loss of 20% of the neural information stemming from each ear, could result in a loss of 36% centrally. Following this reasoning, one could understand how small monaural neural losses, including some not revealed via typical audiometric screening, could result in more substantial and perceptually meaningful binaural losses.

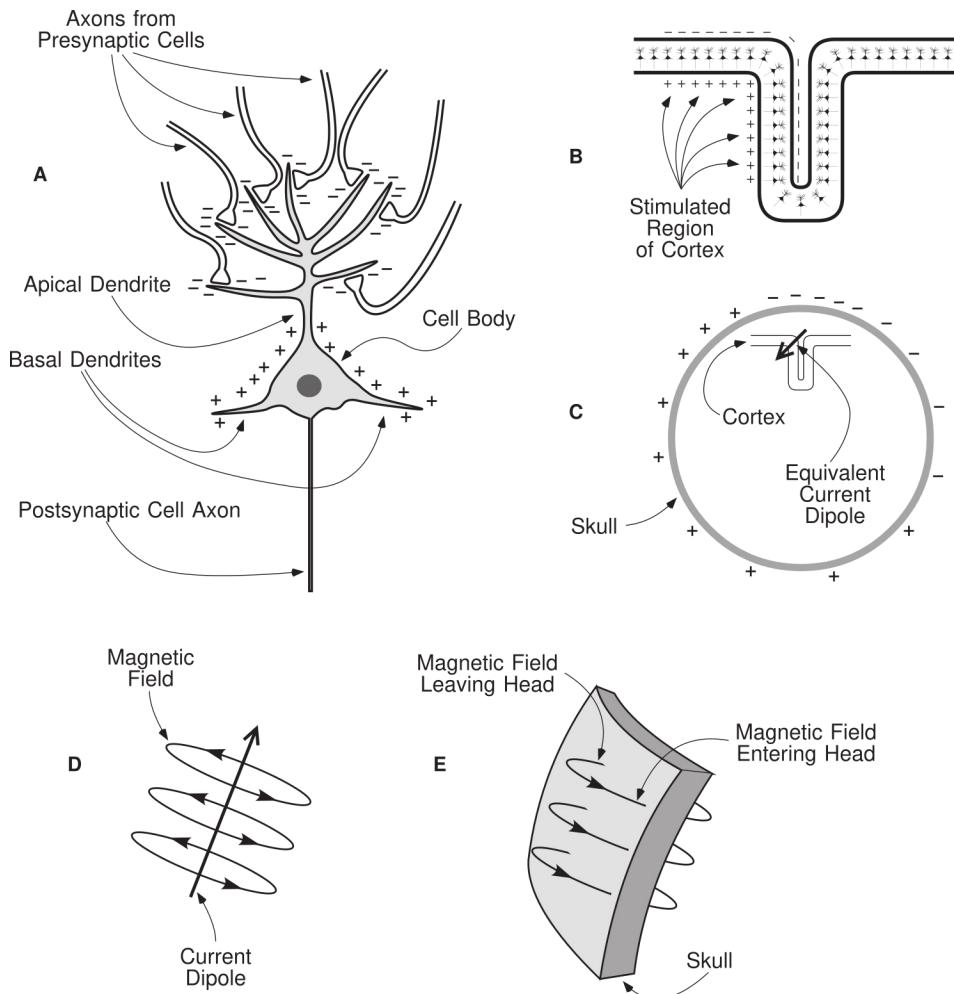
Objective detection of binaural processing in humans

Electroencephalogram (EEG)

- EEG allows to record electrical activity of the brain by means of electrodes placed along the scalp.

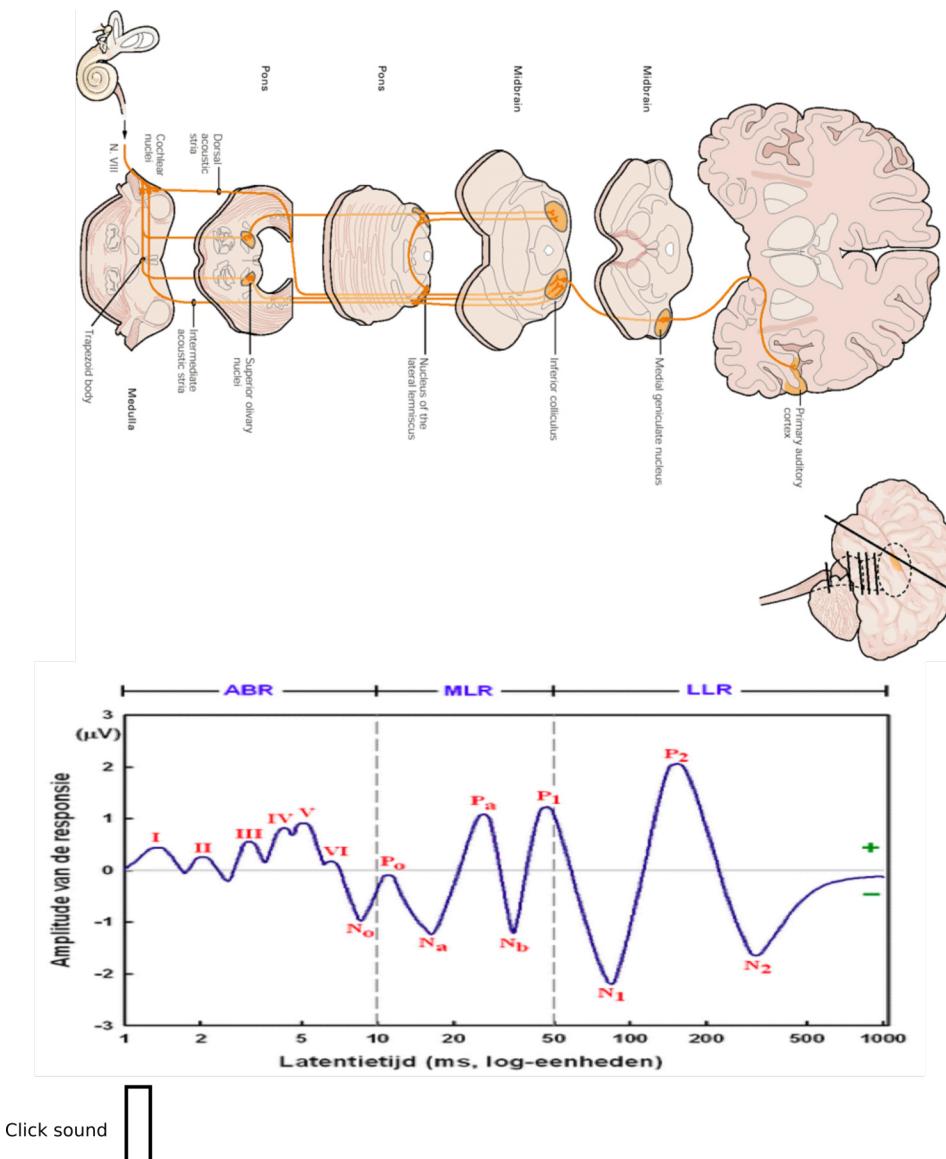


EEG principle

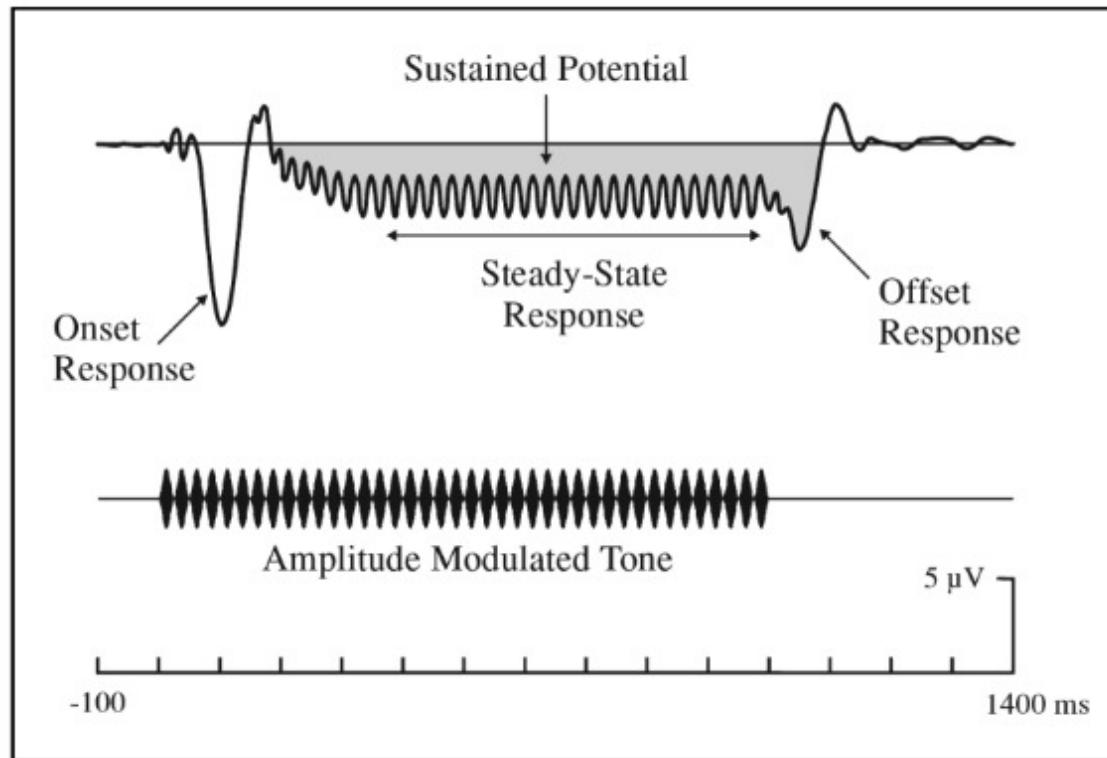


(From Luck 2005)

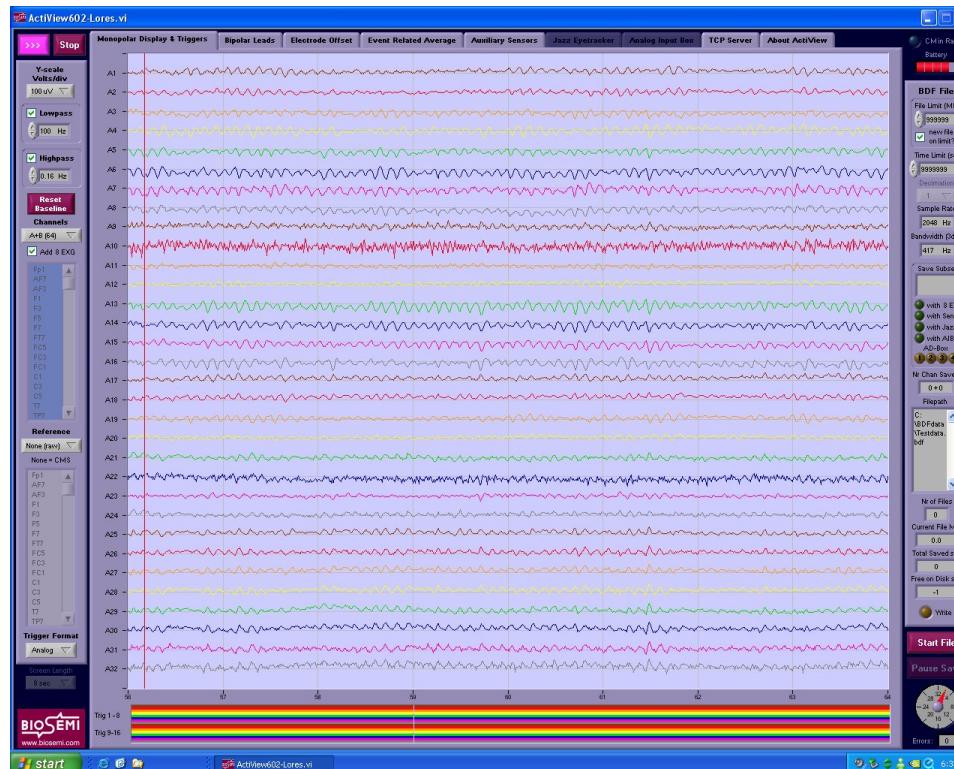
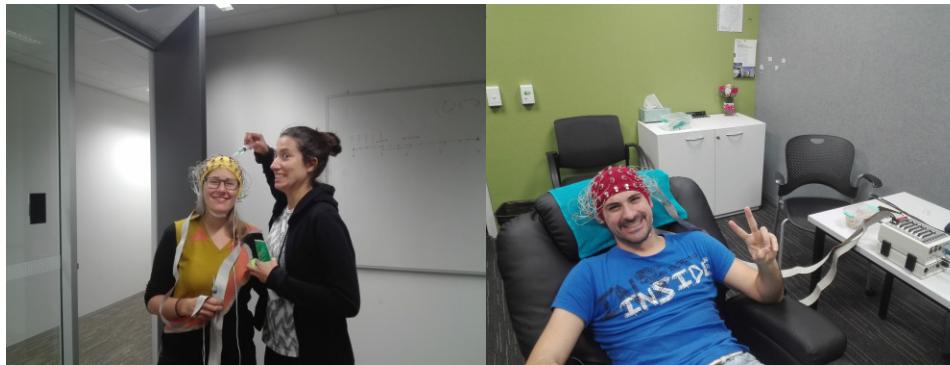
Auditory evoked potentials



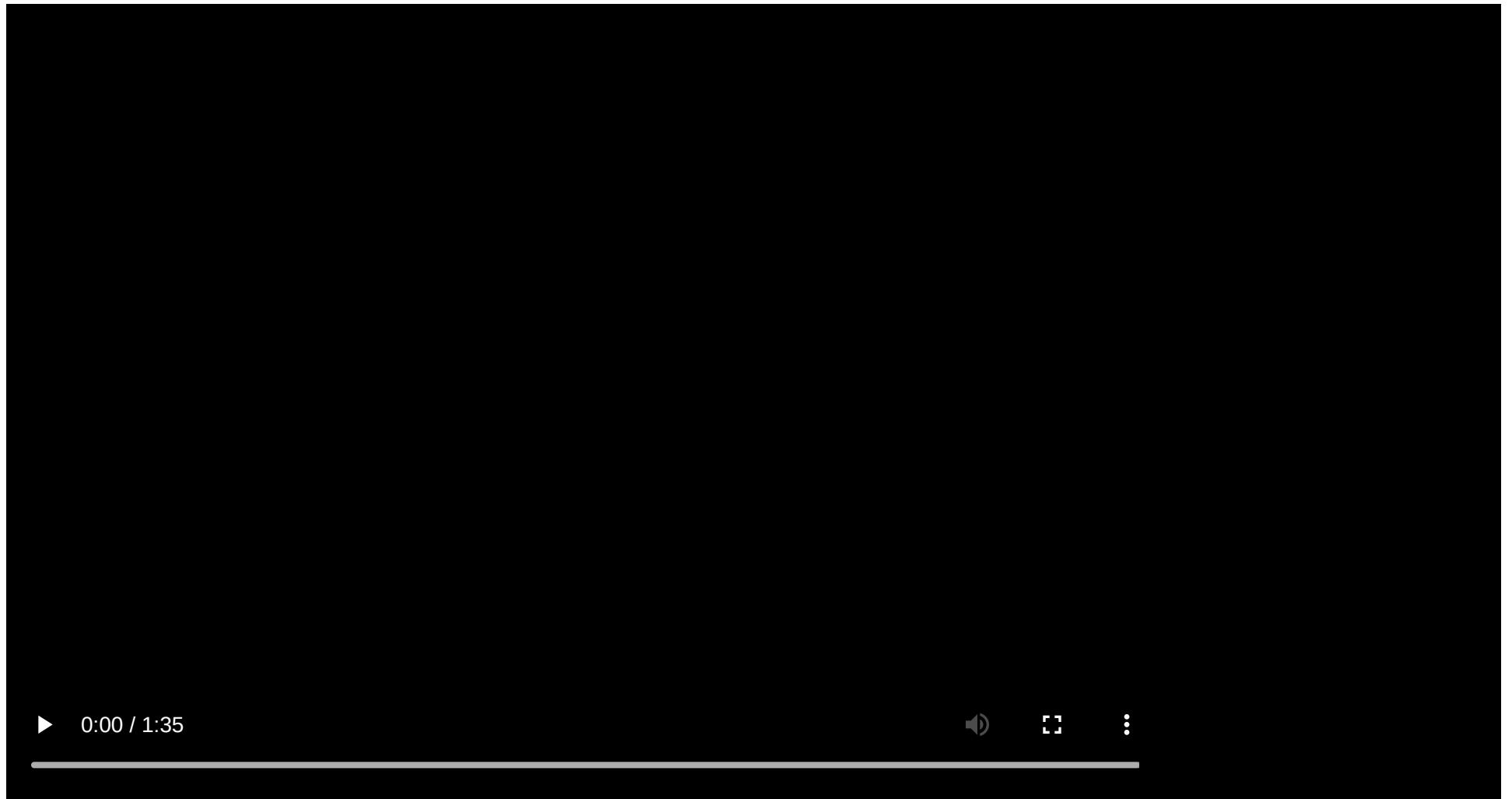
Transient and sustained response



Lab settings



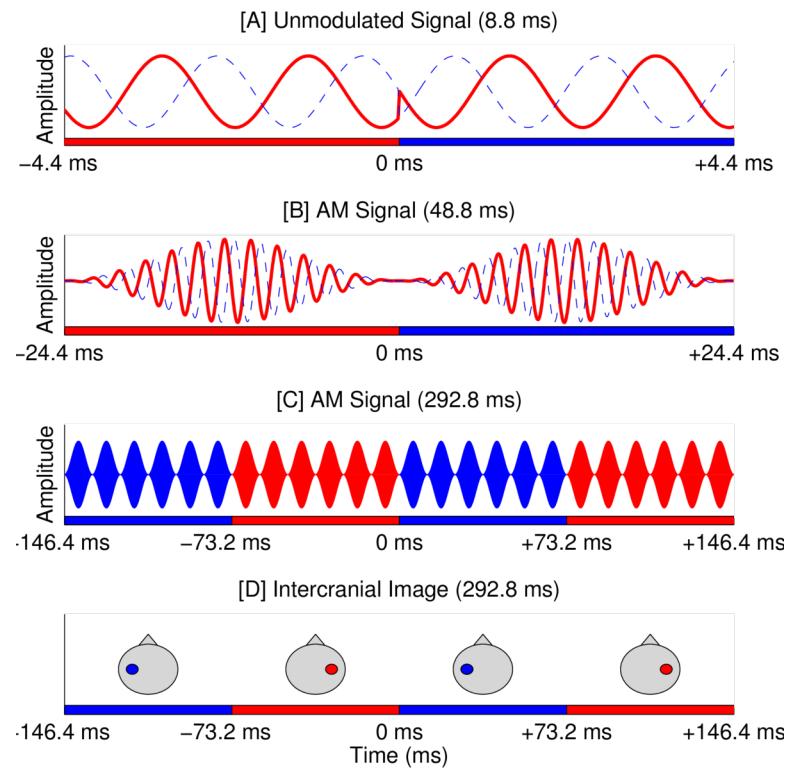
Realtime processing



Objective detection of binaural processing

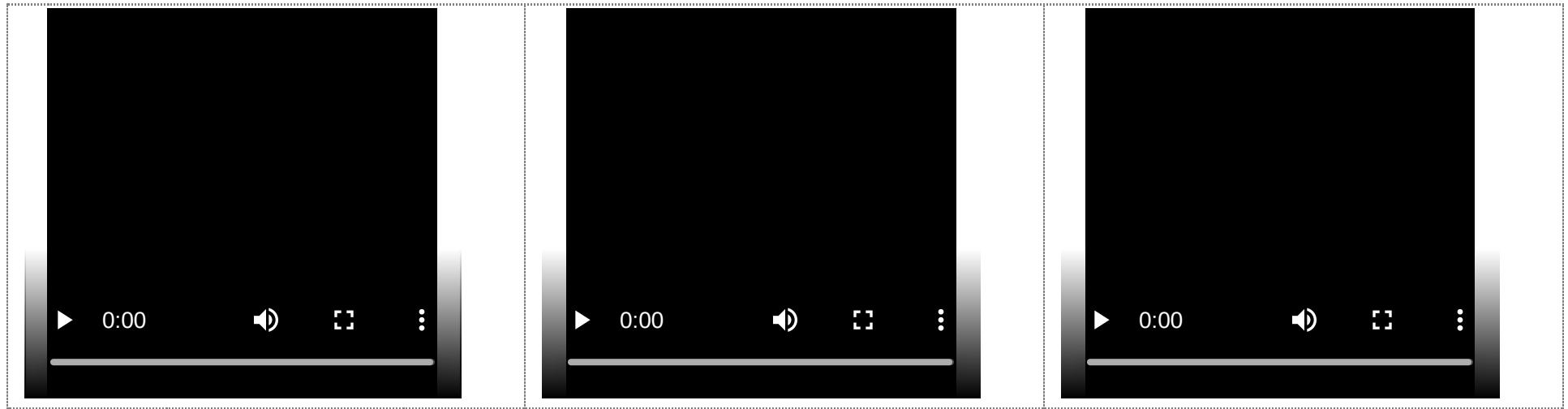
Interaural-phase modulation following-responses (IPM-FR)

- Objective measures of binaural processing can be obtained by using stimuli where the temporal fine structure is manipulated so that the perceived location of the sound image changes periodically (e.g. 6.7 Hz) over time.



(Undurraga et al. 2016)

- The size of the interaural phase difference (relative ITD with respect to the frequency of the sound) will determine the lateralization of the sound.

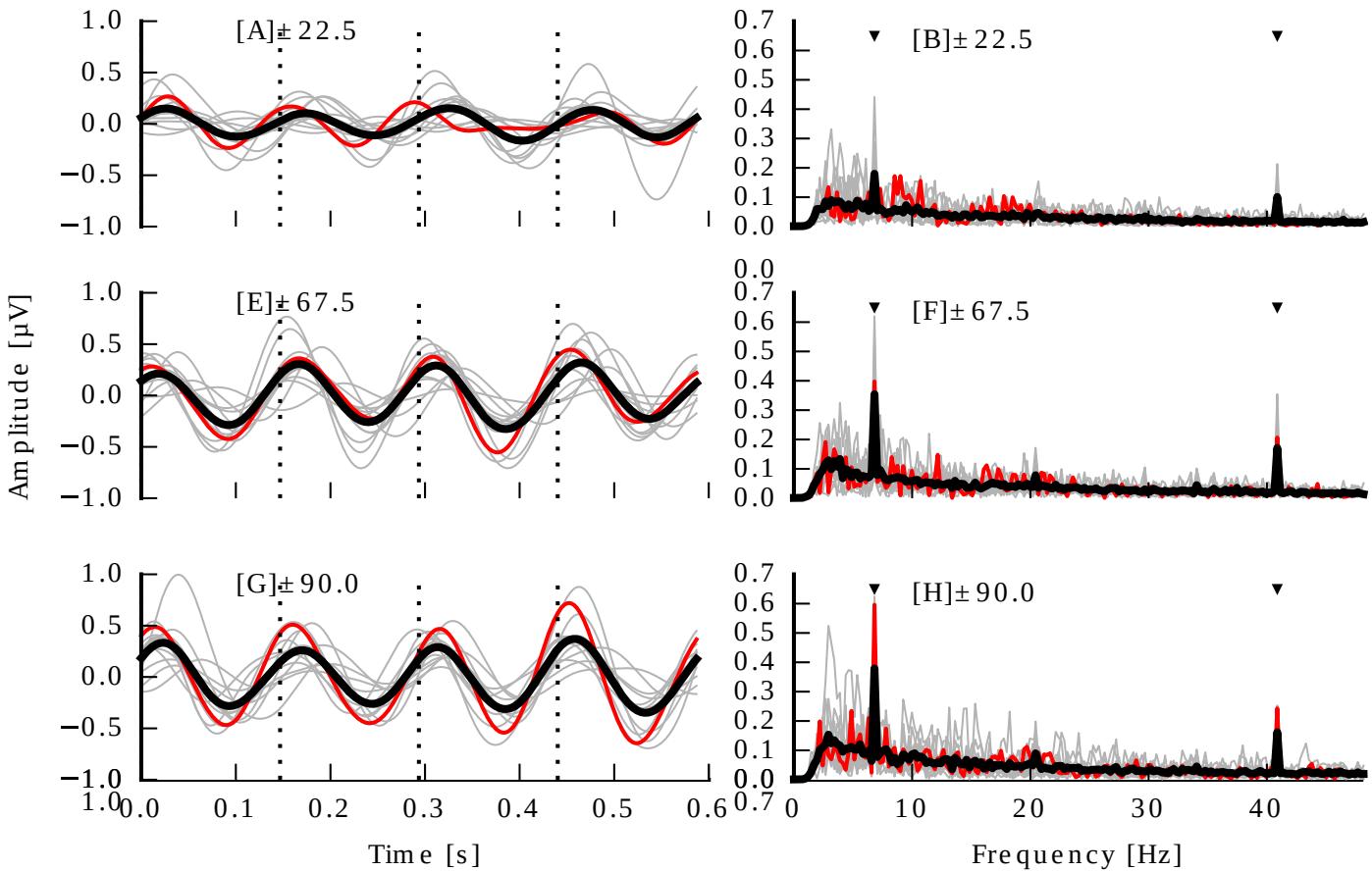
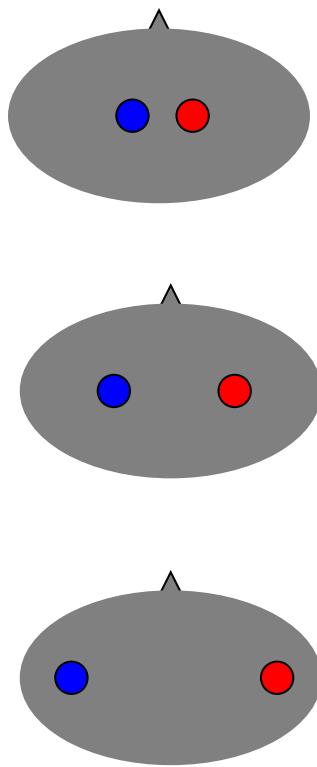


EEG processing

All data was processed with “**pyeeg-python**” (Python 3.7)

- Data referenced to Cz and down-sampled to 1024 samples per second
- Poor electrode were automatically detected and removed
- Eye blink artifacts were removed using a template matching suppression method (Valderrama et al., 2018)
- Data filtered using a FIR Kaiser filter (–65 dB ripple and 1 Hz transition between pass and stop band)
- Epochs were sorted and de-noised using spatial filtering (de Cheveigné and Simon, 2008)
- Epochs averaged using a weighted averaging method (Don and Elberling, 1994)
- Frequency response (FFT of 4255 points at 0.24 Hz resolution) and tested using Hotelling’s T-squared test Picton et al. (1987) and Picton et al. (2003)

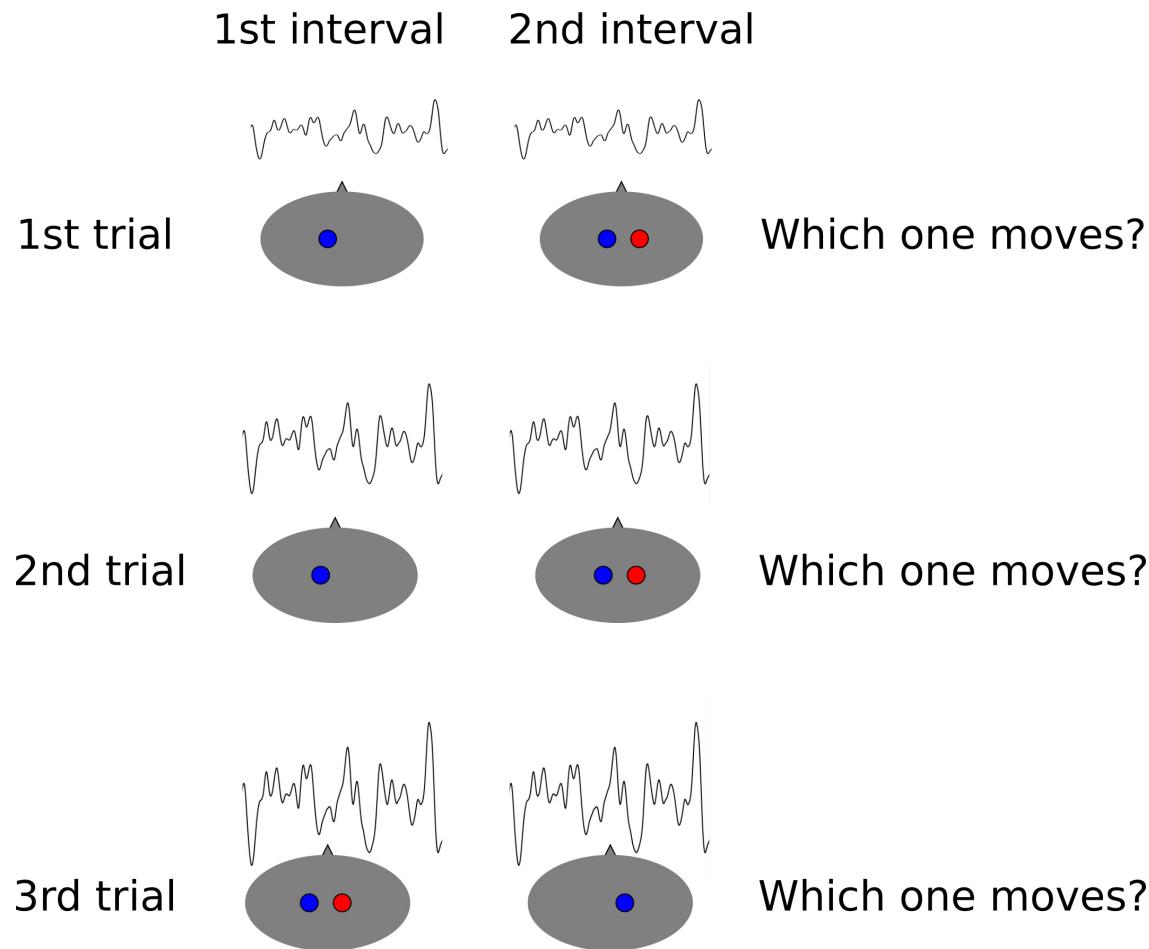
Results



- By switching the “sound image” from left to right at 6.7 Hz, a strong steady-state response is evoked at that particular frequency
 - IPM-FRs are larger when the stimuli causes a strong lateralization percept.

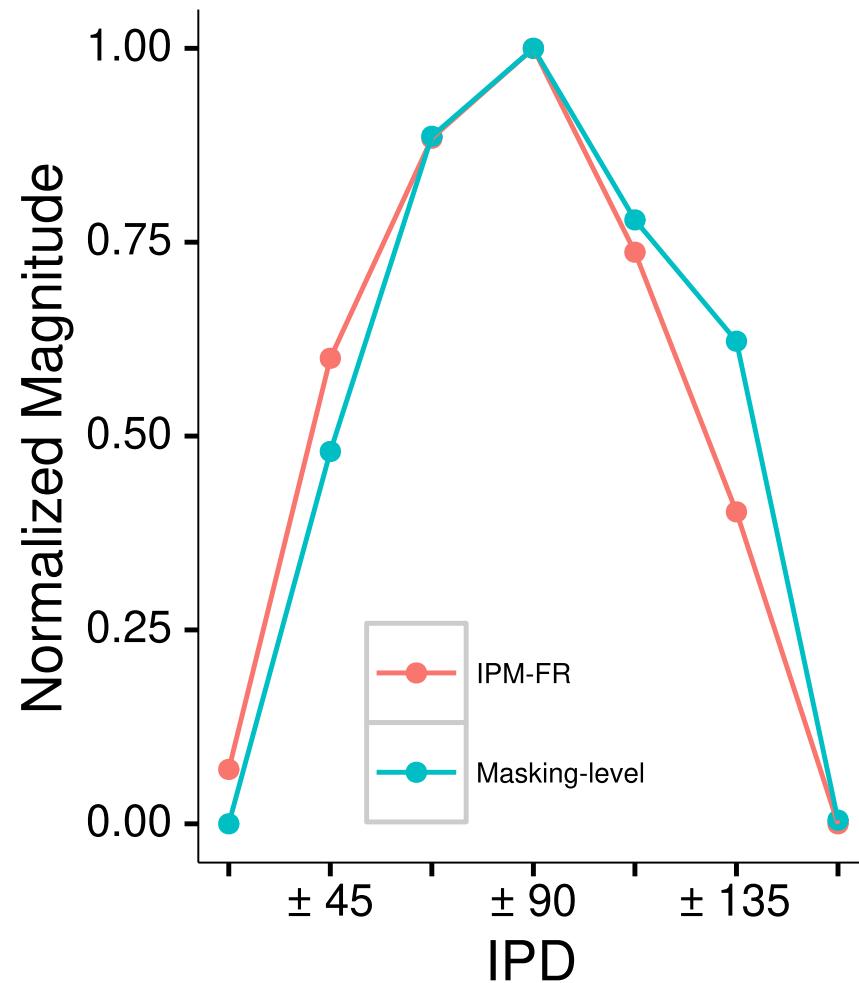
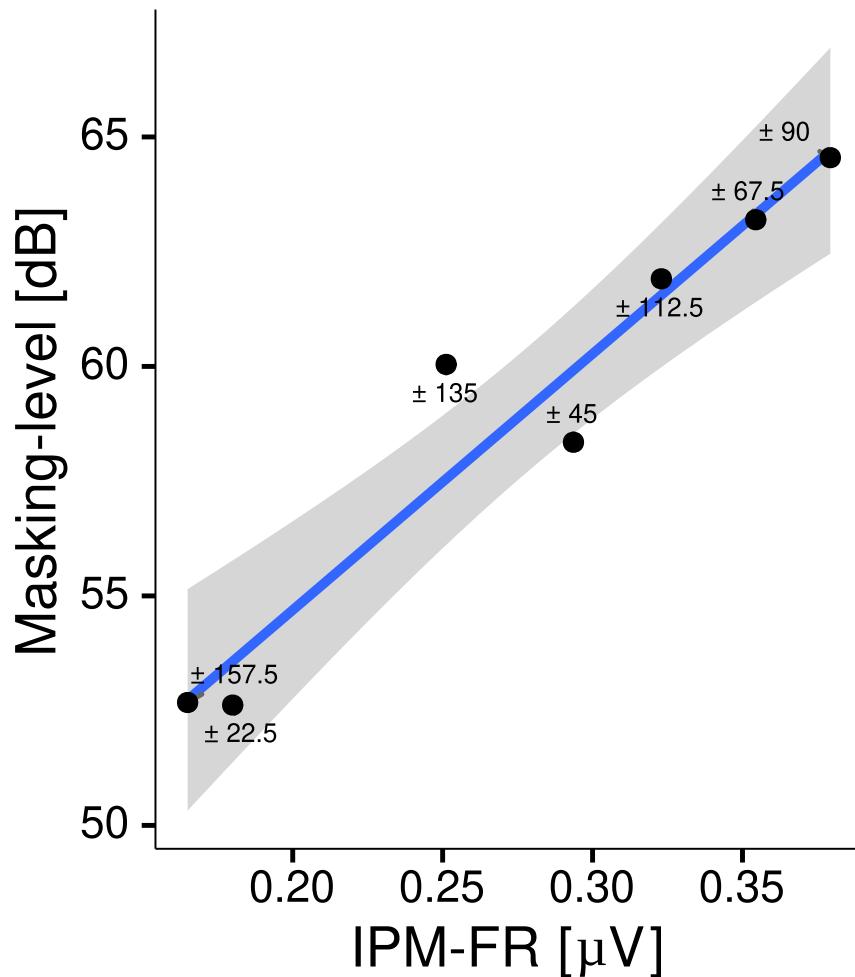
How well do objective IPM-FR and behavioural measurements correlate with each other?

Behavioural task



The level of the masking noise to perform at chance is determined and compared and then compared with brain responses (IPM-FR)

Behavioural vs. Objective measures

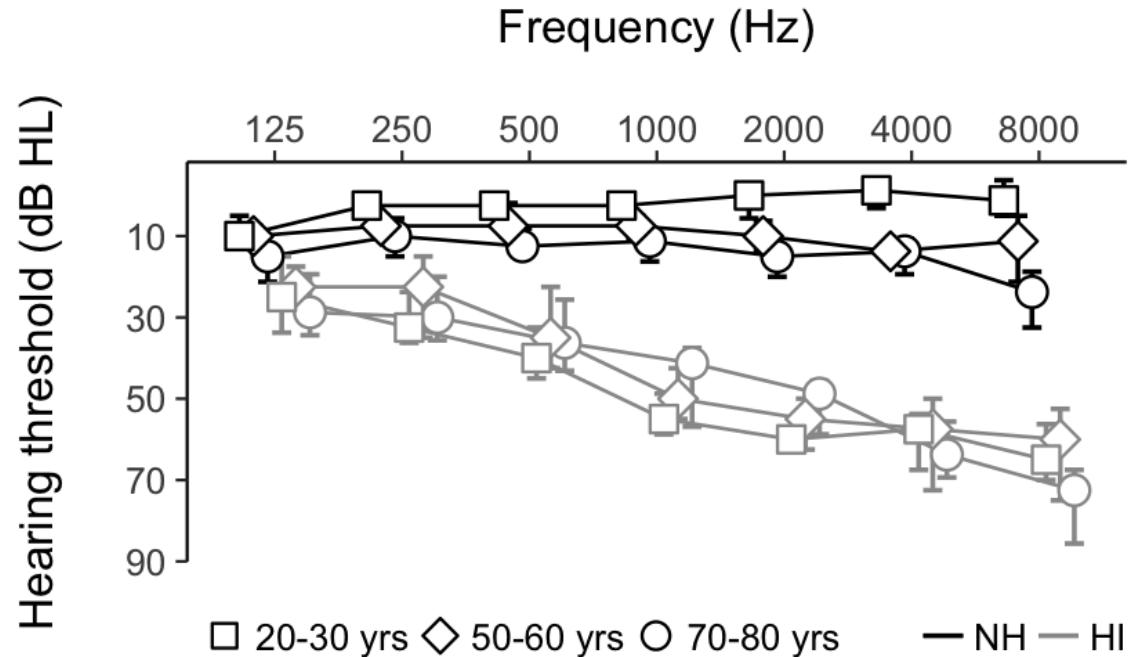


- The correlation between objective and behavioural measurements is excellent ($r = 0.96$)

The effect of age and hearing loss in objective binaural measures

Aims

- Does the IPM-FR differ between healthy and impaired hearing listeners?

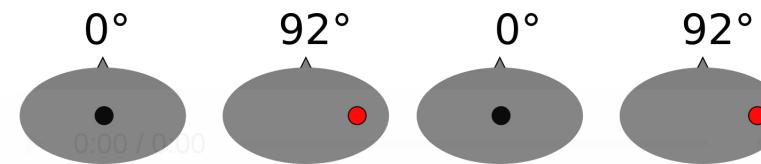
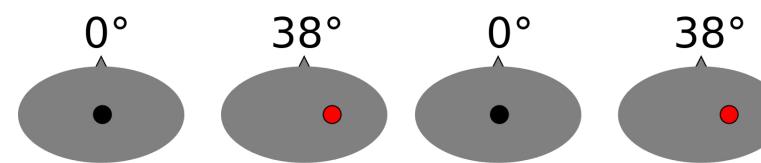
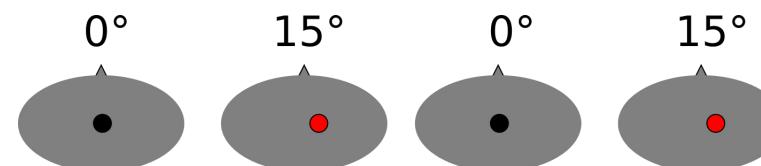
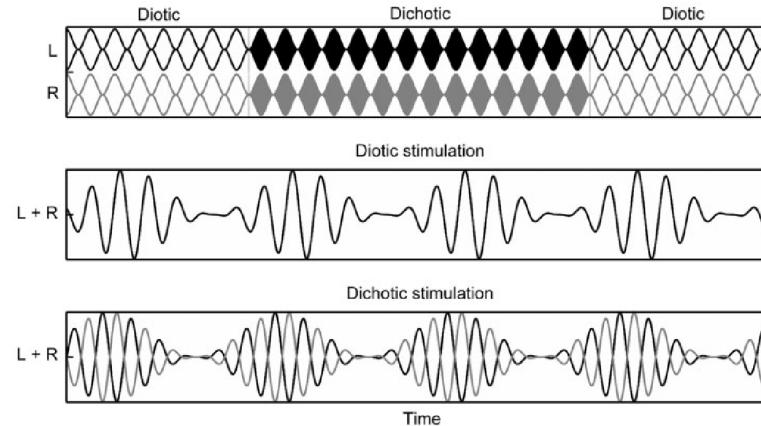


Vercammen et al. (2018)

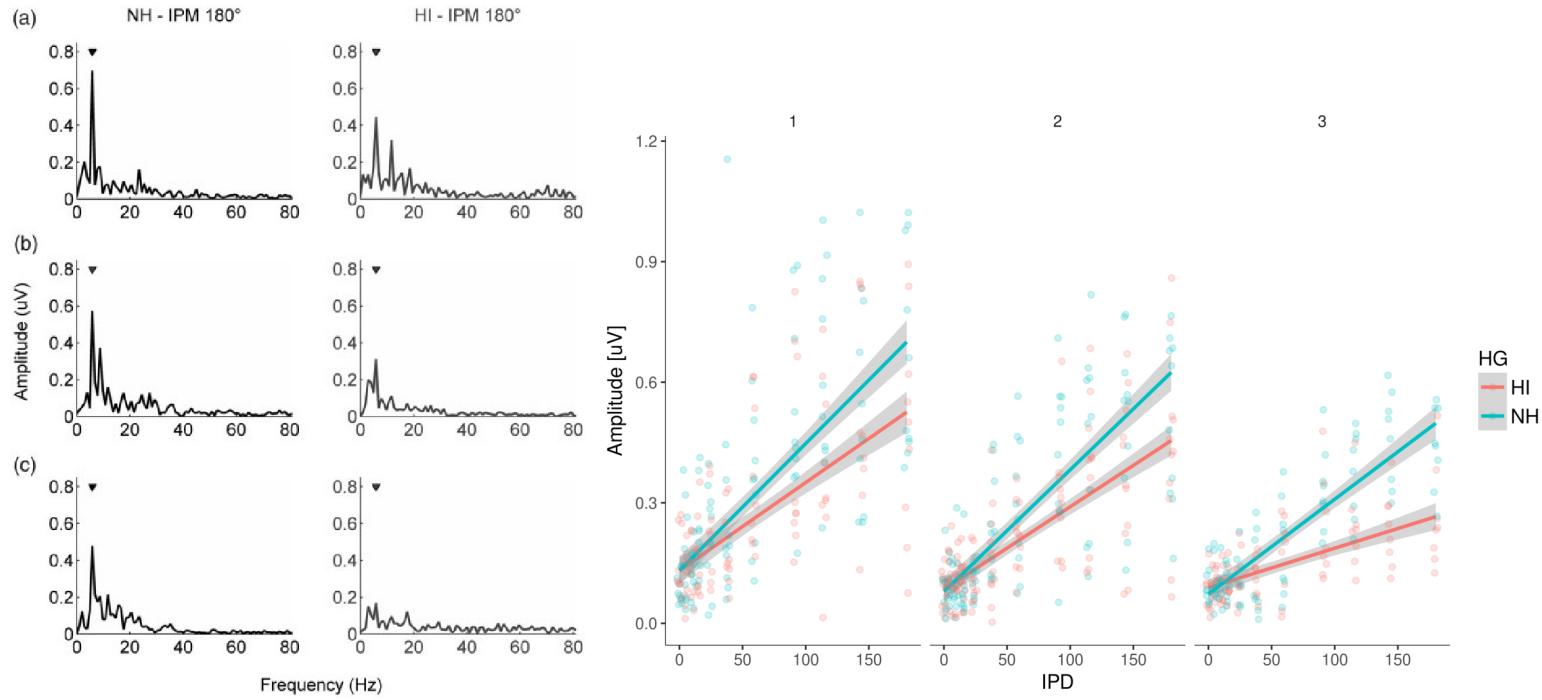
- Interaural difference in hearing thresholds near the stimulation frequency (500 Hz) did not exceed 10 dB HL.
- The stimulus level was 65 dB SPL for normal listeners.
- The stimulus level for hearing impaired listeners was adjusted using visual analogue scale (“inaudible”, “very soft”, “soft”, “comfortable”, “loud”, “very loud”, and “uncomfortably loud”). The subjective loudness was adjusted to match that of normal listeners.

Stimuli

Stimuli were amplitude modulated tones (500 Hz) using IPDs ranging from 0 (diotic) to 180 (dichotic).



Results



- Binaural processing of low frequency sounds is significantly deteriorated by age and hearing status.
- A larger IPM-FR dynamic range was associated with lower (better) IPD discrimination.

Speech in noise and binaural processing

Aims

Does binaural brain response relate to speech-in-noise?

- Here we investigate the relation between neural coding of temporal processing using objective (Electroencephalogram (EEG)) and behavioural (speech-in-noise listening) binaural measures in NH listeners.

Preliminary Subjects Assessment

- 23 Participants took part in this study

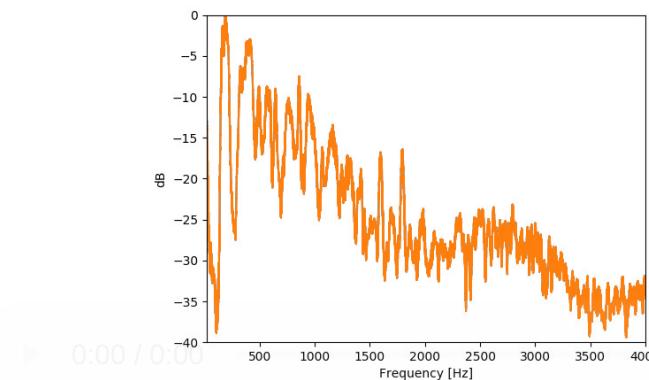
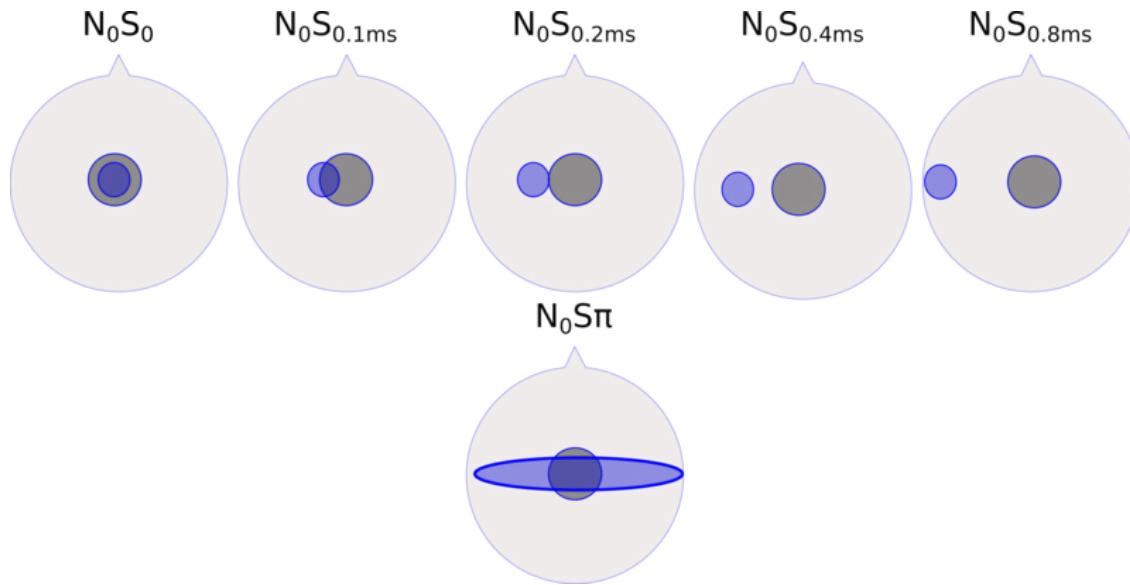
All participants undergone:

- Otoscopy (to ensure integrity of external ear and tympanic membrane)
- Pure tone audiometry (hearing threshold between 250 Hz – 8 kHz)
- Immittance test with acoustics reflex (to ensure integrity of middle ear and lower brainstem pathways)
- Distortion product otoacoustic emissions between 500 Hz – 10 kHz (DPOAEs, to assess outer hair cell functioning)

Behavioural Measures

Digits in noise test

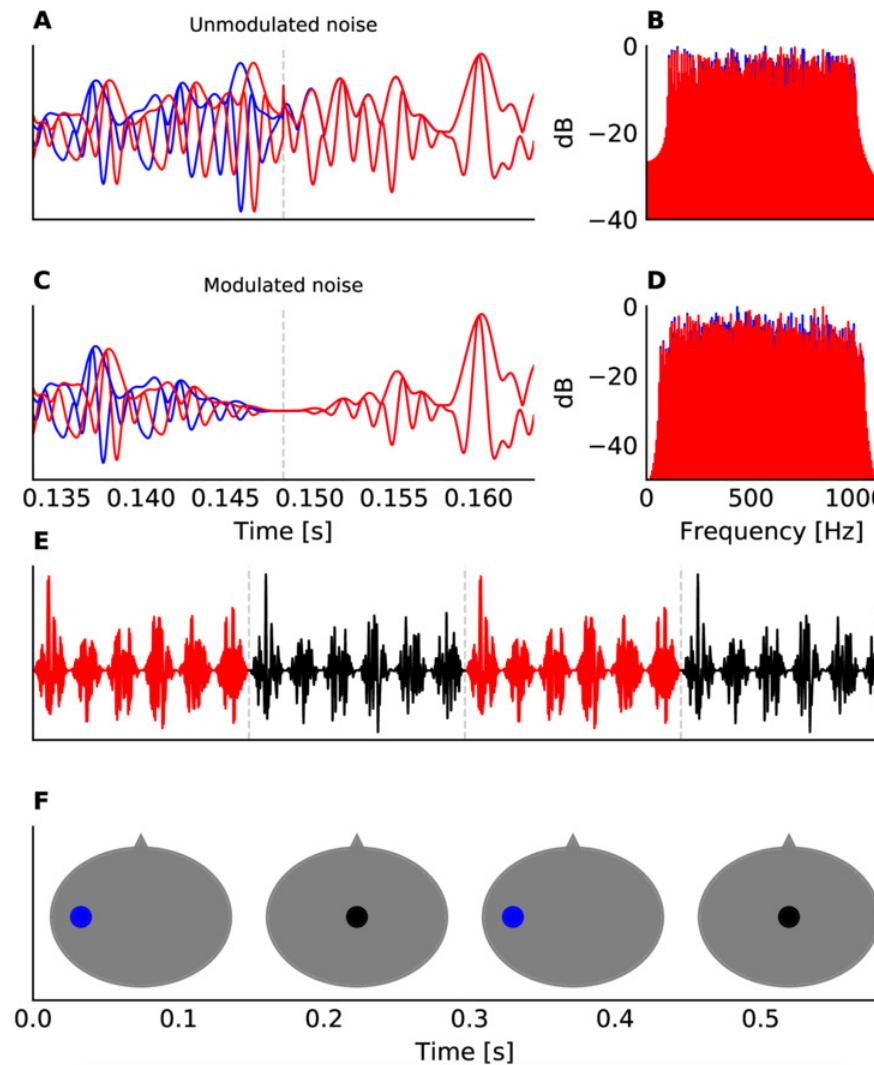
- Adaptive staircase procedure (2-down/1-up) with variable adaptive step (Leek, 2001; Denys et al. 2019)
- Three randomly chosen digits were presented in background speech shaped noise.



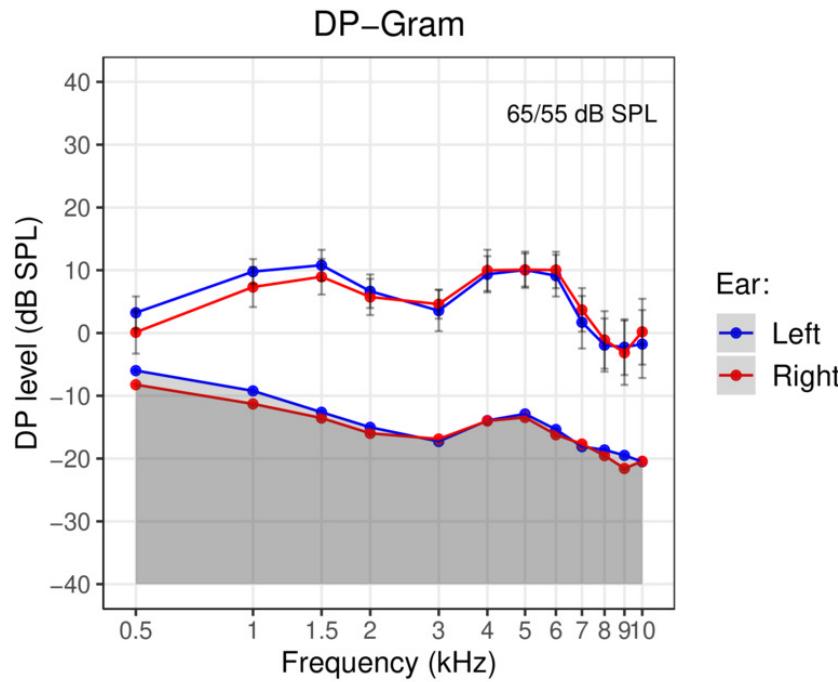
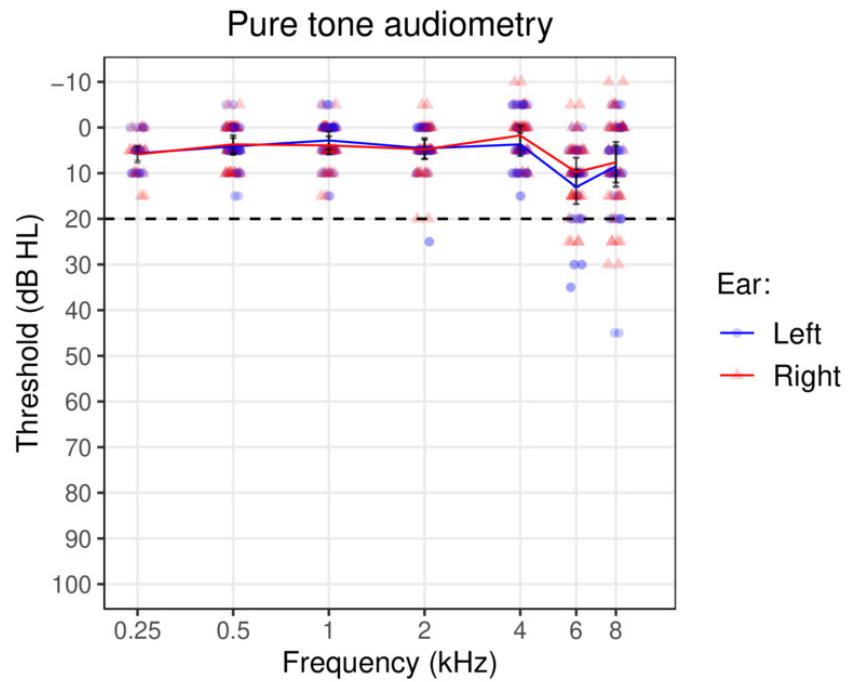
- Initial step size: 6 dB first two reversals; 3 dB next two reversals, and 2 dB last six reversals.
- Noise was presented diotically (i.e. identical signal in both ears) at 65 dB SPL.
- Digits were presented using several **ITDs: 0, 0.1, 0.2, 0.4, 0.8 ms**. In addition, digits were also presented having the opposite polarity in both ears - **antiphasic**.
- 3 runs were obtained per ITD (starting SNR at 10 dB and -30 dB).
- All were presented in random order.

EEG recordings

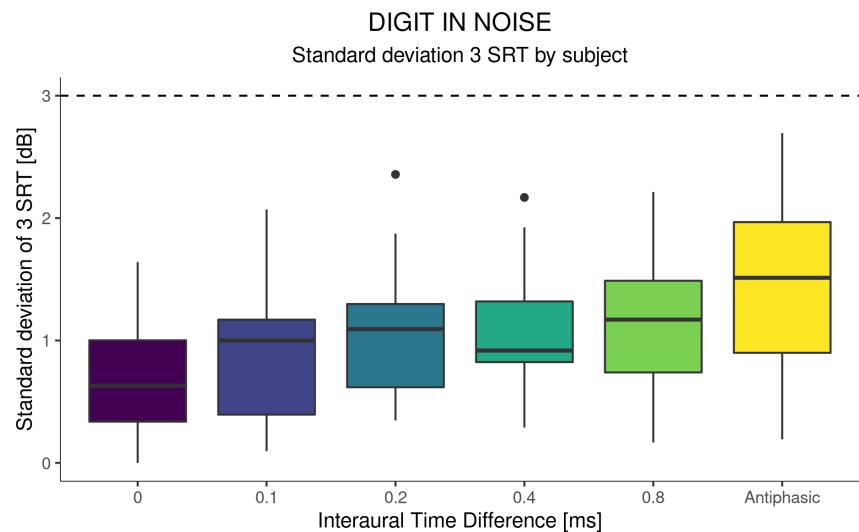
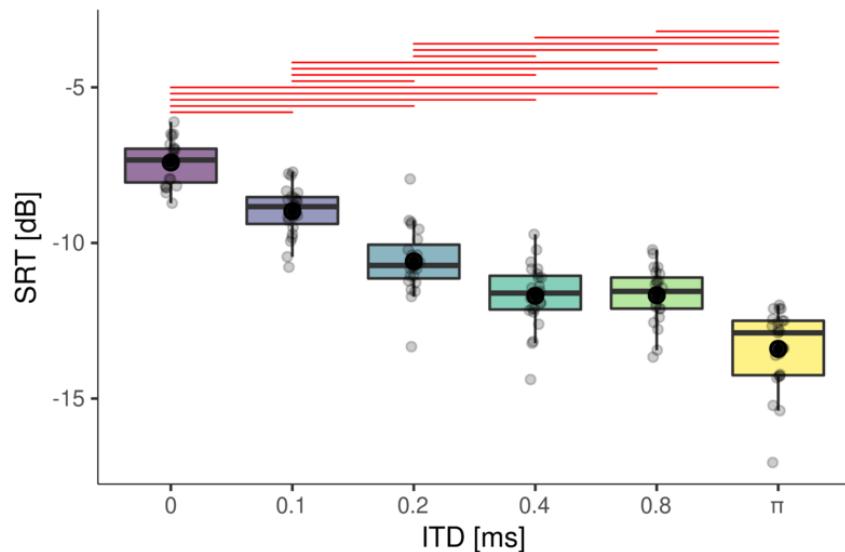
- Amplitude modulated (80 Hz) bandpass noise (100 - 1000 Hz) at 65 dB SPL.
- Interaural time modulations (ITM) presented at a rate of 6.7 Hz
- ITM consisted of: 0/0 (diotic), 0/0.1 ms, 0/0.2 ms, 0/0.4 ms, 0/0.8 ms, and antiphasic condition.



Results

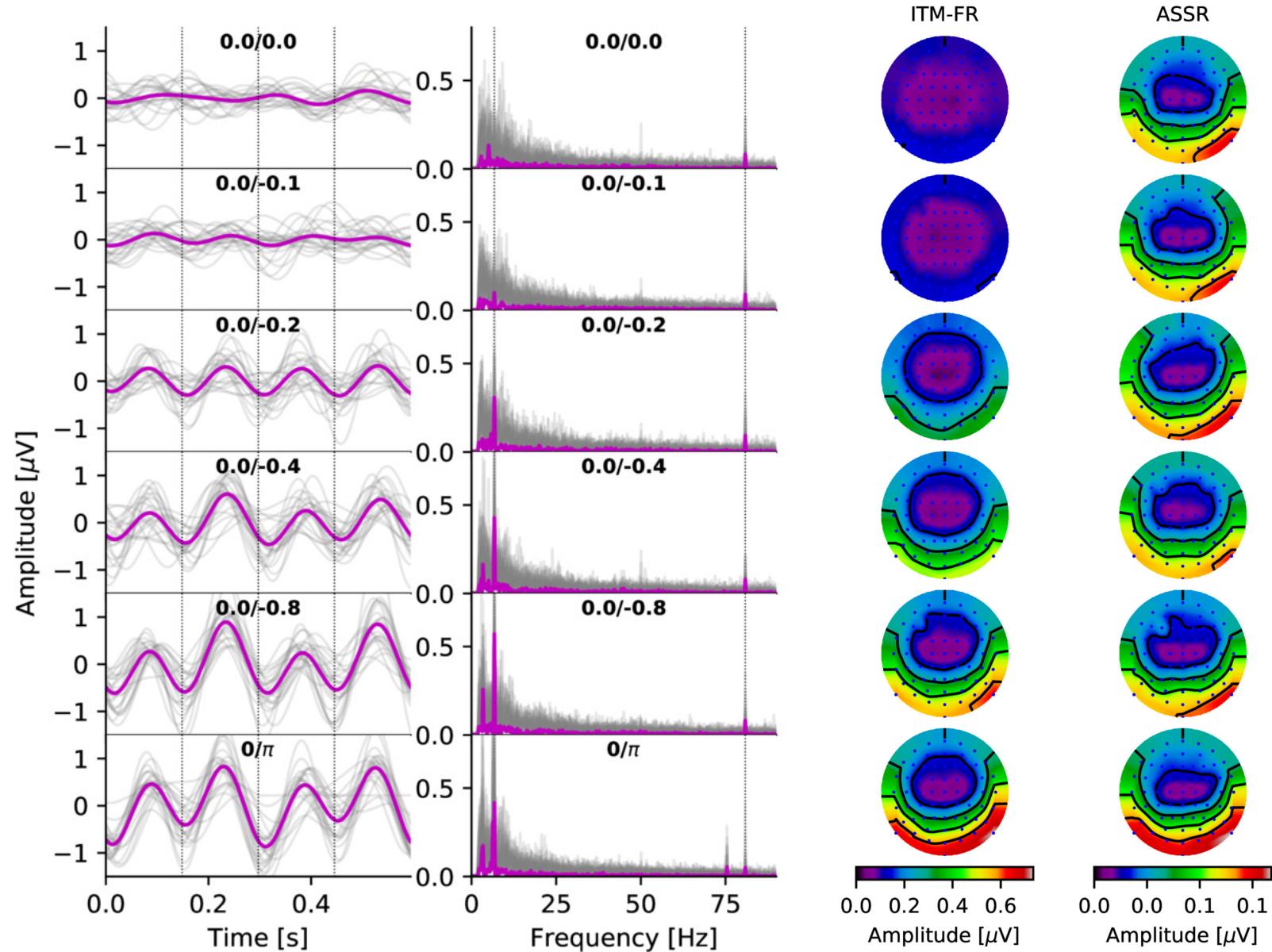


Speech reception thresholds

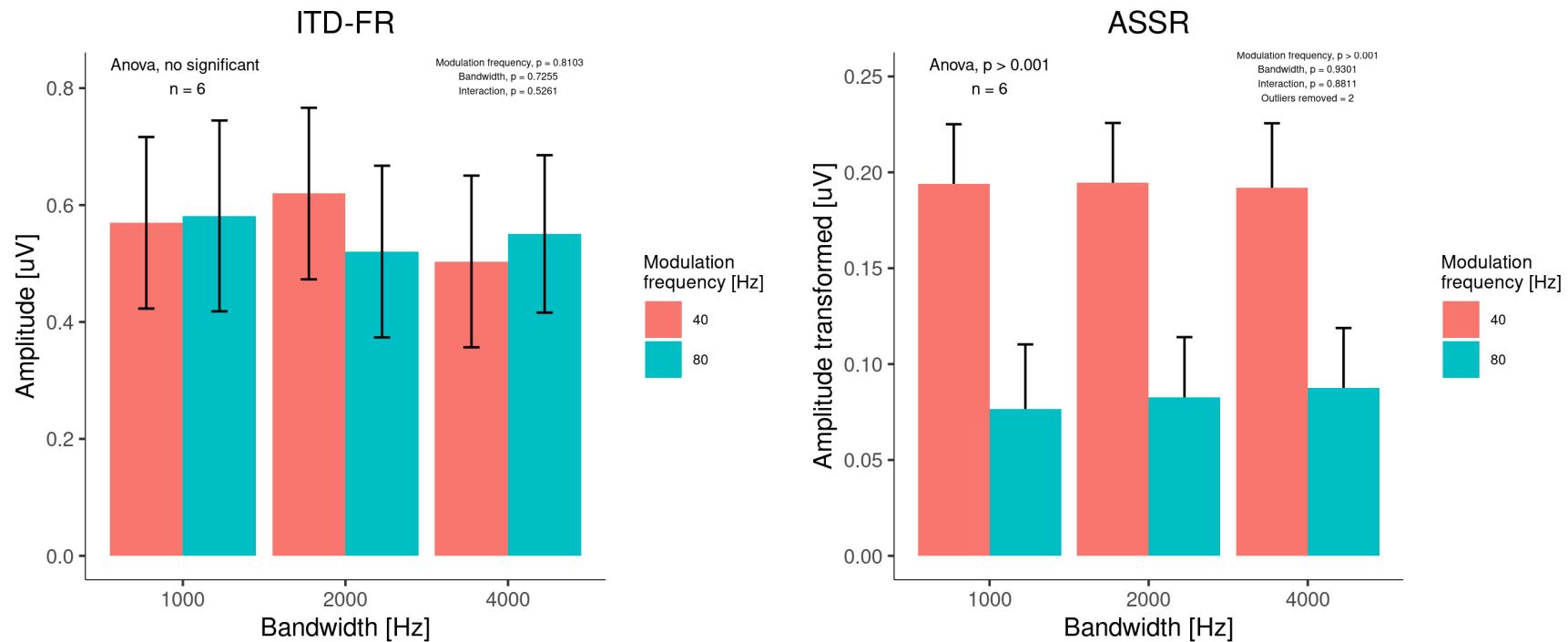


- No effect of presentation was no significant ($F(5, 352.6) = 1.7, p = 0.14$)
- Trial number had a significant effect (difference < **0.88 dB**) ($F(2,370.4) = 22.5, p < 0.001$)
- SRT improvement with ITD ranged between 4.1 and 8.8 dB ($F(5, 110) = 181.68, p < 0.001$)
- SRT standard deviation increased with ITD ($F(5, 105) = 3.7317, p = 0.004$). Mean SD: 0.7 (0 ITD) to 1.47 dB (Antiphasic).

EEG responses

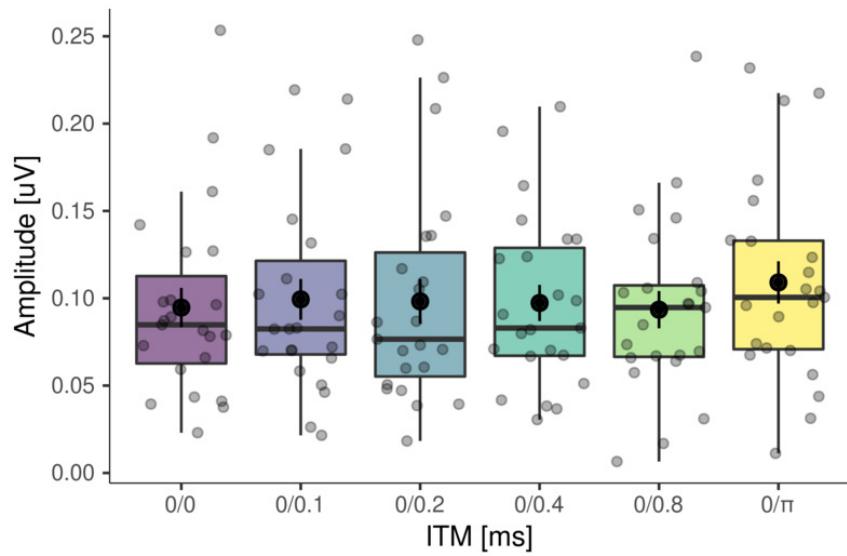
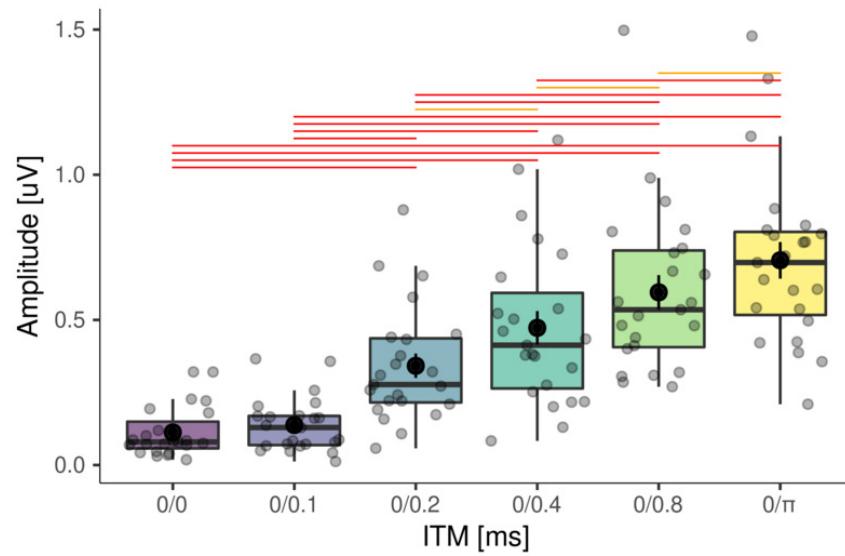


EEG vs stimulus bandwidth



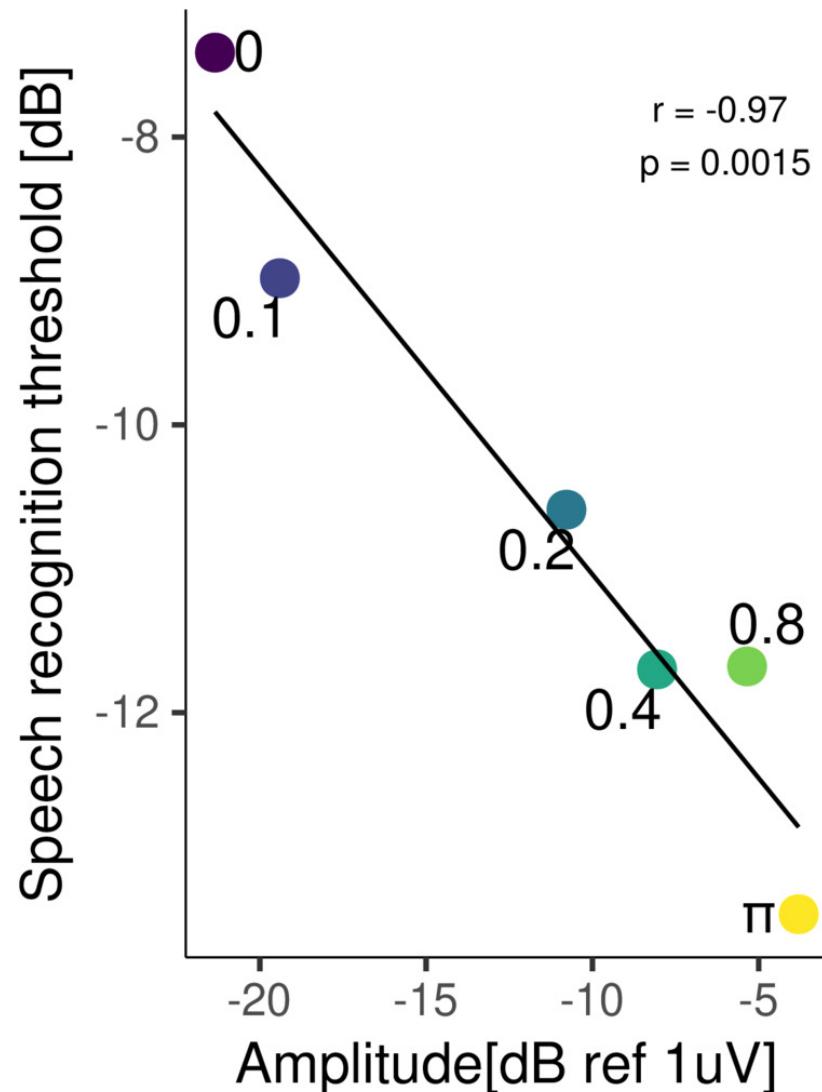
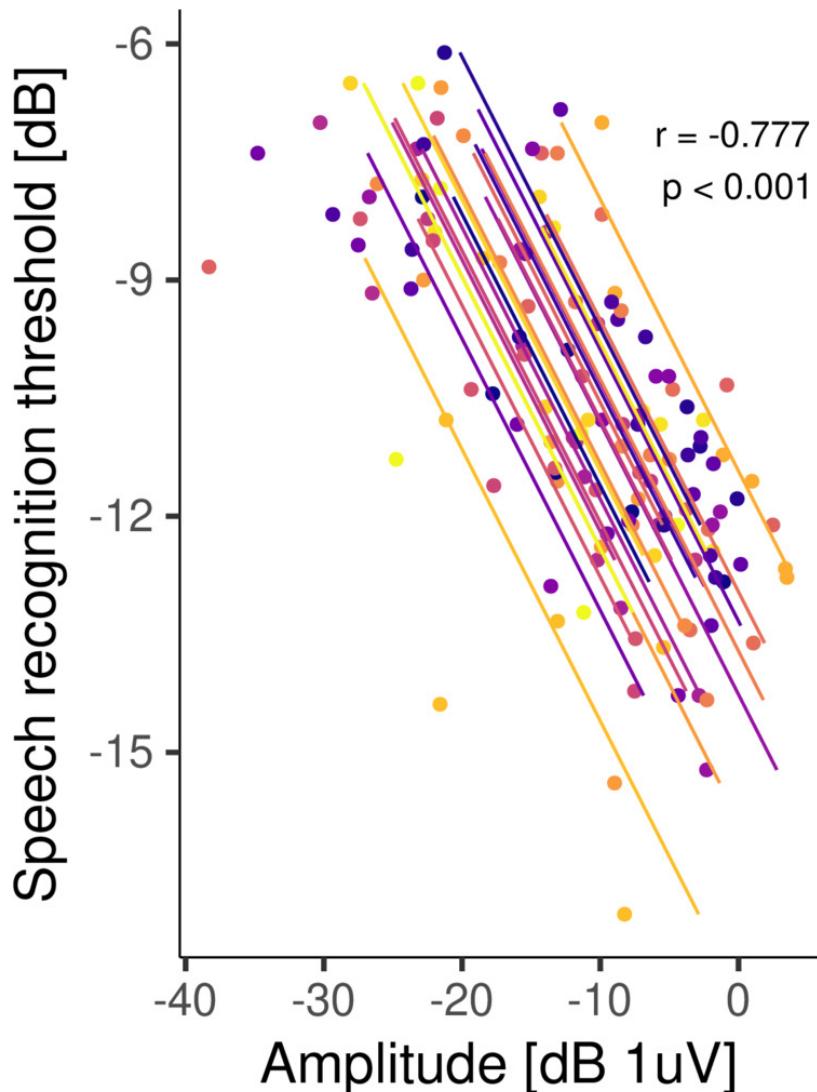
- ITD-FR not affected by stimulus bandwidth, suggesting that response is mostly elicited by energy below 1 kHz.

ITM-FR and ASSR



- ITD-FR significantly affected by ITDs ($F(5, 110) = 43.5, p << 0.001$)
- ASSR showed no significant effects ($F(5, 110) = 1.92, p = 0.09$)

SRT vs EEG



- ITM-FR amplitude, SNR or global field power (GFP) correlate significantly with SRT.
- Excellent correlation at group level $p < 0.001$
- Good correlations at subject level: **19 (AMP) - 21 (GFP) significant correlations**

Conclusions

General observations

- Interaural level and, in particular, interaural time differences conveyed by low frequencies (< 1500 Hz) are critical for speech understanding and segregation in the cocktail party problem (e.g. Swaminathan et al. 2016)
- Hearing-impaired listeners typically find noisy environments disproportionately difficult for understanding speech, whether they wear a hearing aid or not (Killion 1997; Moore 1998).
- Similar problems are encountered by users of cochlear implants, again owing to the limited frequency resolution available with implants (Clark 2003).
- Binaural release from masking may improve detection thresholds up to 12 dB for multiple speech interferers (Jones and Litovsky, 2011), and facilitates source segregation provided that streaming can build up and natural onset cues are present (Drennan, Gatehouse, and Lever, 2003).
- Segregation is always better for the combination of both ITDs and ILDs cues (Culling, Hawley, and Litovsky 2004)

Neurophysiological measures

Objective measures of binaural processing

- agree with behavioural measures of IPD processing
- show a **reduced amplitude in hearing impaired and elder listeners.**
- **showed great sensitivity to ITDs** and had an excellent correlation with SRTs (across several different metrics) across and withing participants.

All the above suggests that objective and behavioural measures of binaural processing may be more sensitive than monaural measures to unveil hidden hearing loss.

Acknowledgments

- Australian Research Council [project number FL160100108]
- Thanks to you for “listening”

