State of art and Challenges in Improving Speech Intelligibility in Hearing Impaired People

Stefan Launer, Lyon, January 2011

Phonak AG, Stäfa, CH

Content

- Speech intelligibility in complex listening environments for hearing impaired persons
- Noise reduction technologies in hearing instruments
  - De-reverberation
  - Single microphone technology
  - Multi-microphone technology
  - FM systems
- Results
- Challenges

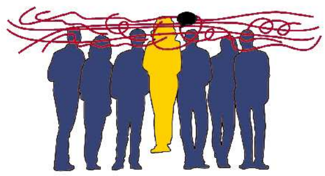
Speech Intelligibility in Noise???
⇔ *Speech Intelligibility in Complex Listening Conditions!!*

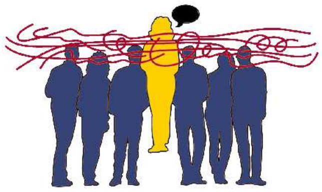
- Different types of interfering sources
- Different spatial arrangements of sources and interferers
  - Dynamic…
- Room acoustics
  - Reverberation
  - Distance

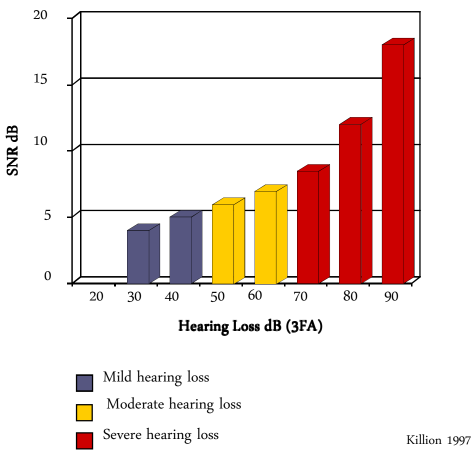
Speech Intelligibility in Noise???
⇔ *Speech Intelligibility in Complex Listening Conditions!!*

- Test methodology
  - Speech tests: short sentences, words, phonemes
    - target from front, static
  - White noise… from the back
  - Anechoic environment
  - Lab / real life results
- Speech intelligibility
- Listening effort

Speech Intelligibility in Noise







Physical structure of interfering signal has a strong impact on speech intelligibility

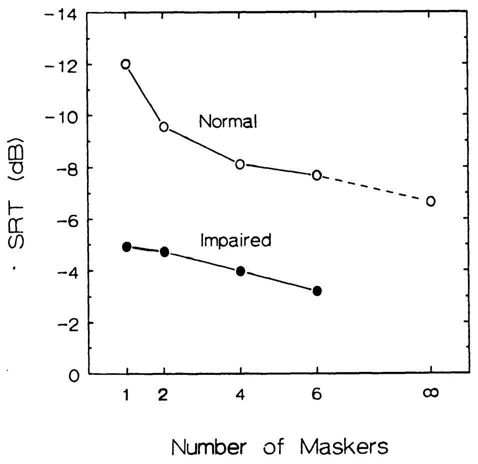
Introducing ....

- spectral dips: SH 3 - 4 dB SRT NH: 9 - 15 dB
- temporal dips: SH 1 - 2 dB SRT NH: 6 - 7 dB
- combination of both: SH 4 - 5 dB SRT NH: 15 - 20 dB

- ... improves speech intelligbility a lot for normal-hearing subjects, much less so for hearing impaired subjects !

Peters, Moore and Baer 1998, JASA

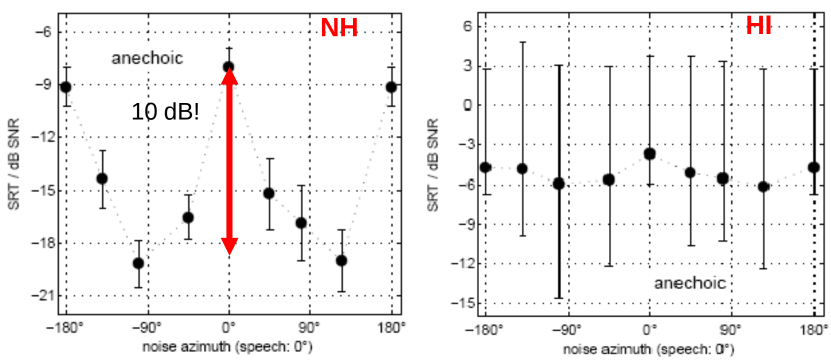
Speech Intelligibility in Multi-talker Environment



- Speech intelligibility as a function of interfering talkers

Fig. 2,
Bronhorst and Plomp, JASA 1992

Spatial Release from Masking – Anechoic Chamber

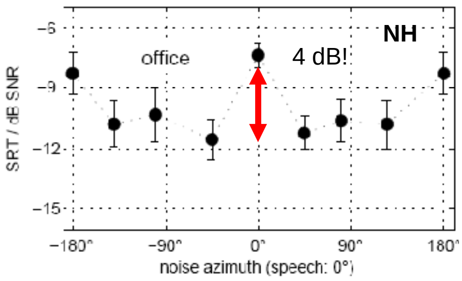


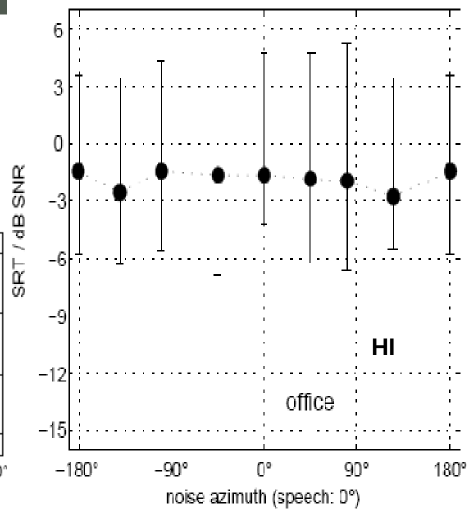
Beutelmann & Brand, JASA 2006

Spatial Release from Masking - Office

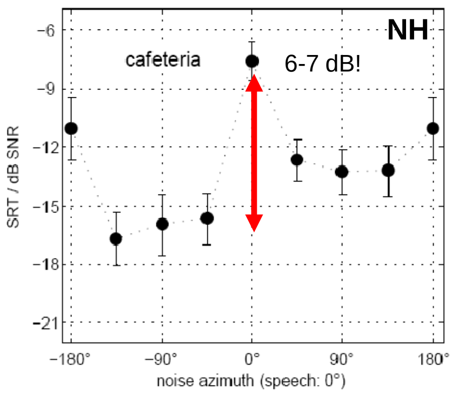
- Spatial release reduced by reverberation

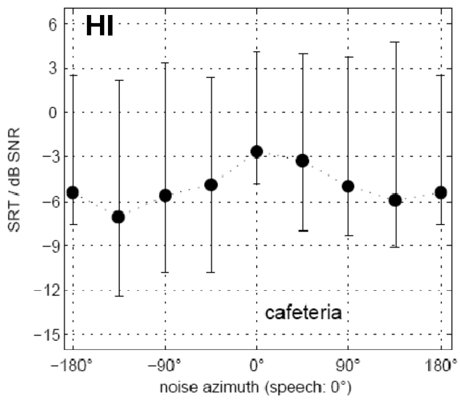
Beutelmann & Brand, JASA 2006





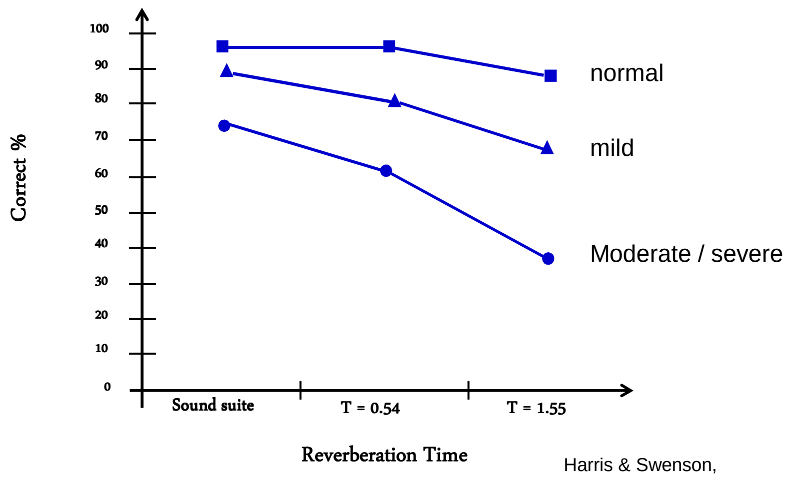
Spatial Release from Masking - Cafeteria





Beutelmann & Brand, JASA 2006

Speech intelligibility in reverberant environments

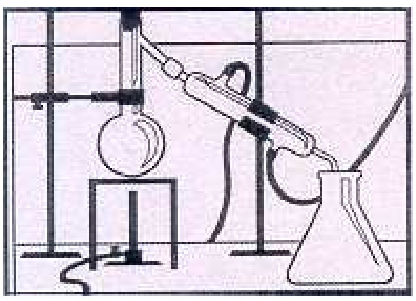


Harris & Swenson,
Audiology 1990, p. 314-321

How to mix a Speech in Noise Cocktail

directional microphones

noise
canceling



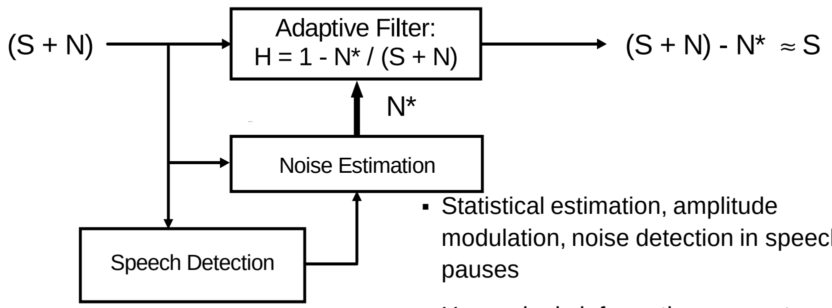
Speech in
noise cocktail

Objectives for a hearing instrument:

- Speech intelligibility improvement!!!!
- Ease of listening, listening effort, listening comfort

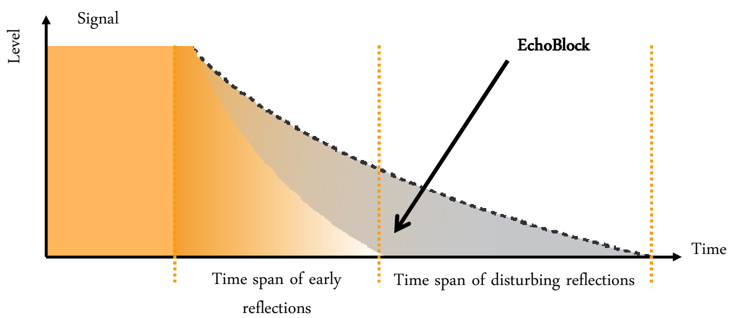
Noise Reduction Using a Single Microphone

- Single Microphone Noise-Cancellers: in principal estimate the noise and subtract it from the noisy signal.



- Statistical estimation, amplitude modulation, noise detection in speech pauses
- Use a single information source to separate two signals

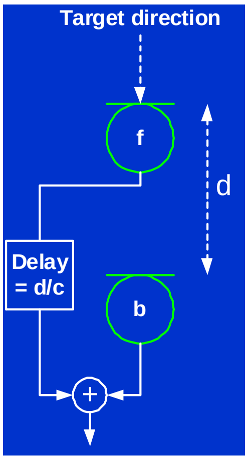
Reverberation Canceller – Reduces the smearing effects by de-blurring the speech signal



Single Microphone Noise Reduction - Summary

- This technique performs well eliminating stationary noises like a fan or in a car, etc.
  - Reverberation: very reverberant rooms
- Speech like noises can't be suppressed without degrading speech quality at the same time.
- ... ease of listening: improving listening comfort
  - reduction of perceived noisiness
  - less annoyance
- Improvement of speech intelligibility ???
  - Sound quality is a trade off…

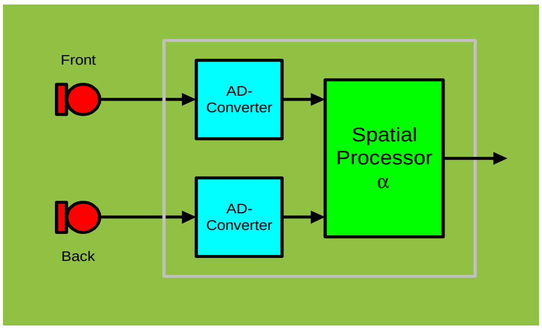
Delay & Sum - Technique



- The acoustical signal is picked up at two different locations by the front and the back microphones
- The signal from the back is delayed
- The signals from both microphones are summed
- Depending on delay - different directions are attenuated

Digital Adaptive Directional Microphones

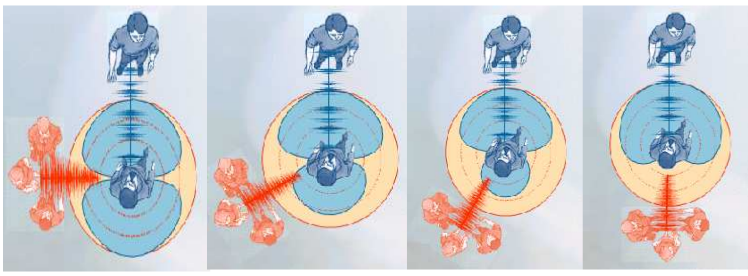
- Adaptive: minimize output energy of the two microphones



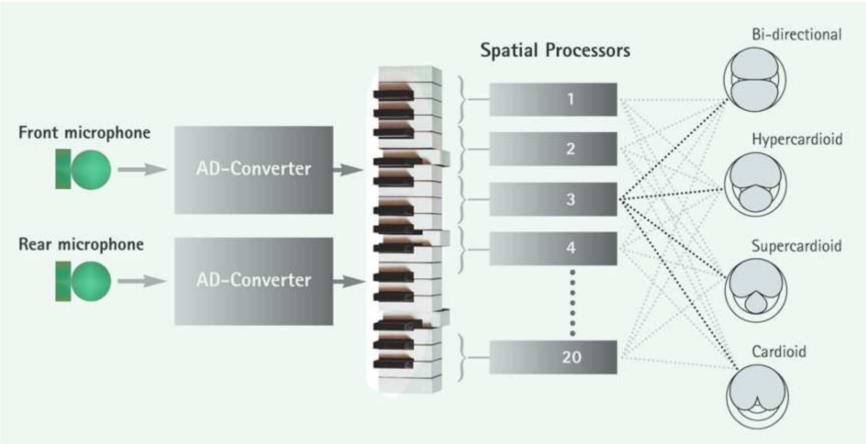
The spatial weighting factor (a) is continuously adapted, the Directivity Index hereby optimized.

Digital Adaptive Directional Microphones

- Amplify sounds from front
- Adaptively attenuates strongest noise source

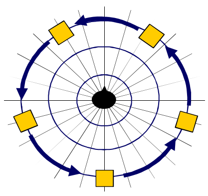


Frequency Specific Beamforming
Directivity in each frequency band



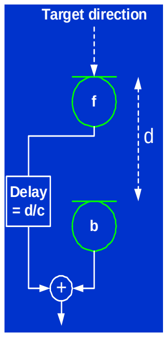
Directional Microphones: Potential and Limitations

- Significant speech intelligibility benefit compared to omnidirectional systems in complex listening conditions
  - from side & asymmetric
  - diffuse
  - moving noises
  - reverberant environments & larger distances
- Lab results: 3-6 dB improvements

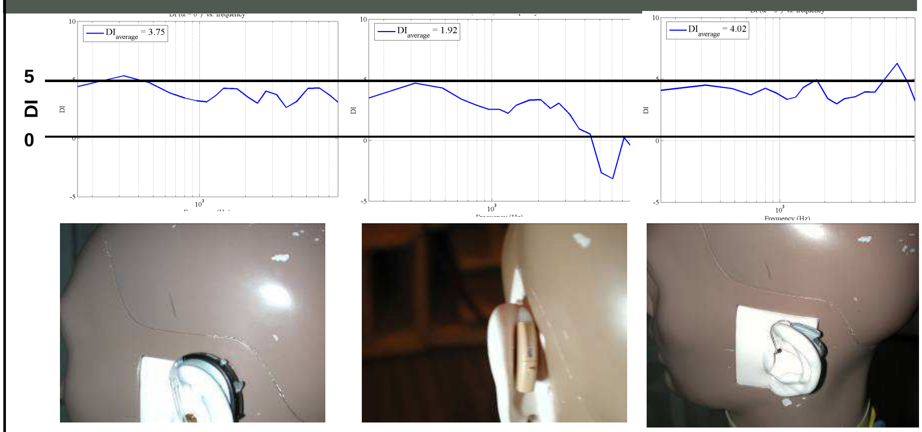


Directional Microphones: Potential and Limitations

- Positioning on head
- Microphone mismatch, ageing etc
- More than two microphones
  - Noise floor
  - Narrow beam pattern acceptable?
- Size constraint: low frequency roll-off
- Computational complexity



Directivity Index for Different Products Styles and Placements



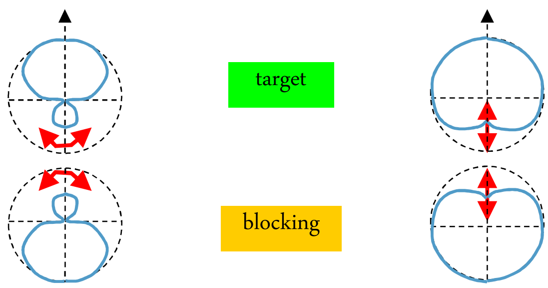
Factors causing BF mismatch

The beamformer performance in our current products can be limited due to level and phase mismatch caused by the following factors:

time invariant	time variant
Microphone production mismatch	Microphone ageing
HI assembly	HI repairing
Clean W&W variability	W&W pollution
Customer individual head/pinna shape	Non-idealities of current adaptive level matching block
Device geometry: ITEs and microBTEs have unfavorable mic positions	Customer HI positioning variance

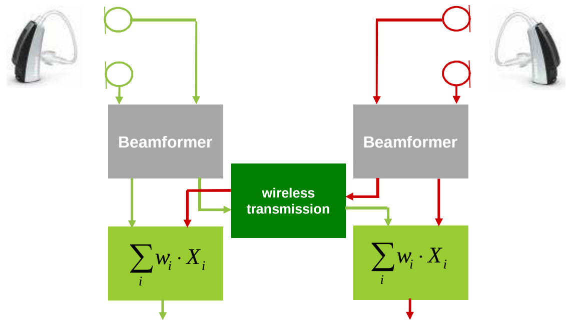
Effects of Microphone/BF mismatch

- microphone phase deviation
  - → rotated null direction
- microphone magnitude deviation
  - → reduced suppression



Binaural Directional Microphones

Improving directivity by linear combination of monaural directional microphone outputs



Maximum SNR improvement: 3 dB

Test setup

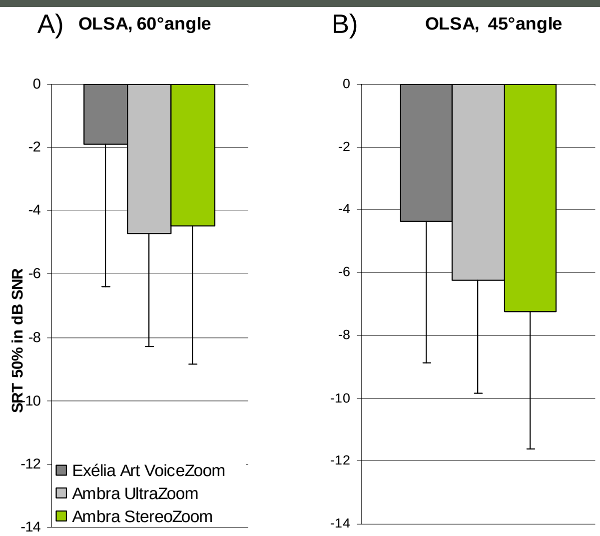
- Subjects
- 20 adults
- Moderate - moderately-severe hearing loss
- Exélia Art and Ambra microP BTE

- Algorithms
- Excelia
- Ambra UltraZoom (monaural)
- Ambra StereoZoom (binaural)

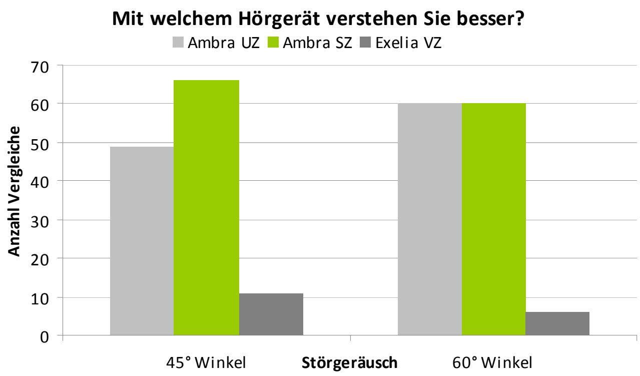
- Test setup
- OLSA: speech intelligibility in noise
- Listening effort scaling
- Paired comparison

Binaural Beamforming

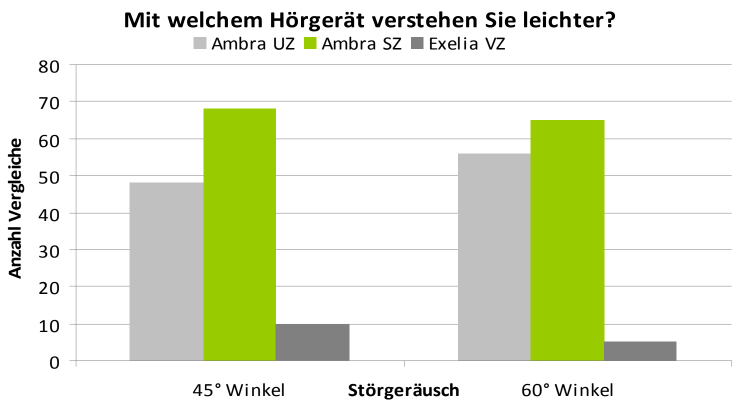
OLSA



Paired Comparison – Subjective Speech Intelligibility



Paired Comparison – Subjective Listening Effort

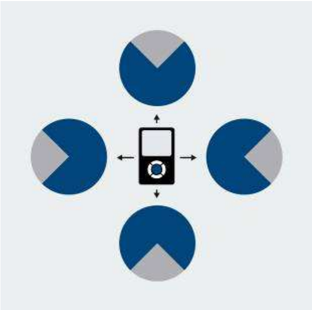


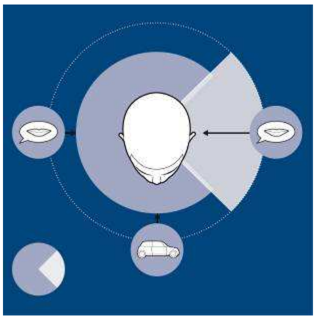
User-Steered directionality

- Traditional beamforming systems focus only to the front
- Speech signals do not always come from the front and facing the speaker is not always possible
  - Car, restaurants, small groups
- ZoomControl, accessible through myPilot, allows Exélia wearers to select in which direction to focus hearing

Listen to the side: User-Steered directionality

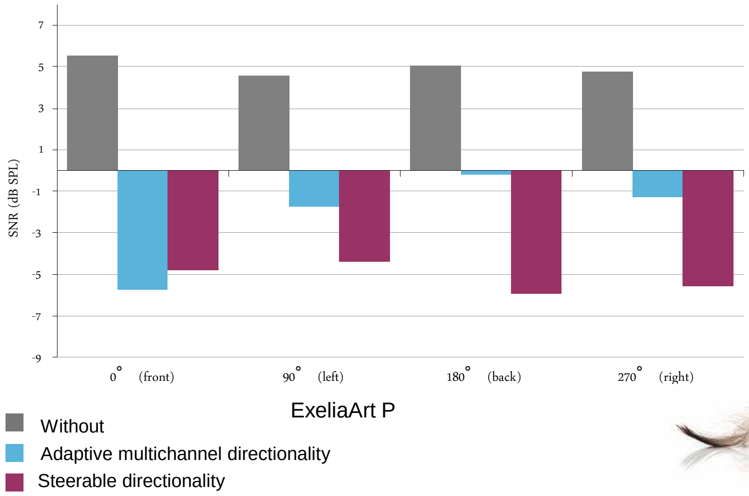
- Uses four-microphone network of full bandwidth binaural instruments
- Broadband audio data transfer between devices focuses hearing in one specific direction, while suppressing all signals in other directions



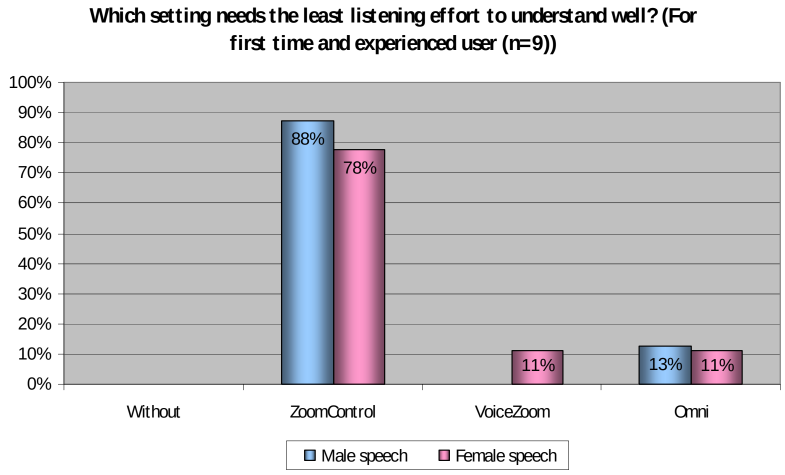


User-Steered directionality

User-Steered directionality

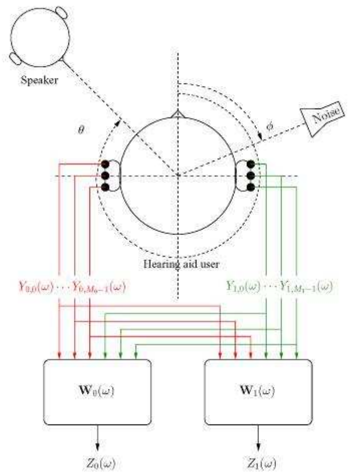


Subjective Evaluation – Listening Effort

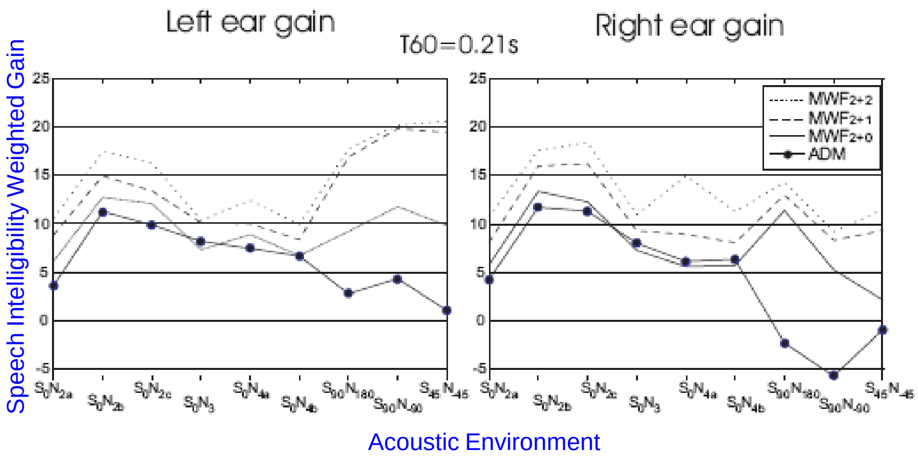


Binaural noise reduction techniques

- Different types of algorithms
- Beam former:
  spatial information, timing difference
- Binaural Wiener Filter
- Blind source separation:
  statistical information estimating
  room transfer function
- Auditory processing schemes

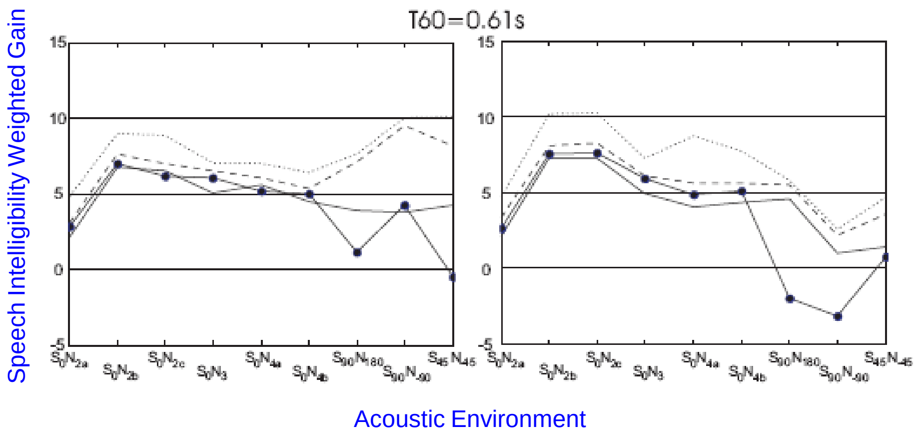


BWF: Speech Intelligibility Weighted Gain



PhD Thesis van den Bogaert 2008

BWF: Speech Intelligibility Weighted Gain



PhD Thesis van den Bogaert 2008

Binaural Beam Forming / Noise Reduction

- No stereo output signal => loss of spatial sensation / localization
  - Artificially re-introduce that by “split-directionality”
  - Mixing in part of the original signal at the output
- Narrower beam width
  - How narrow should the beam be (head movement!)?
- Complex environments
  - Dynamic -> target tracking, target identification
  - Reverberation & distance

- Expected improvements: specific situation, no generic solution
  - Single /few strong interfering source, frontal hemisphere
  - Environments with little reverberation

Technical constraints

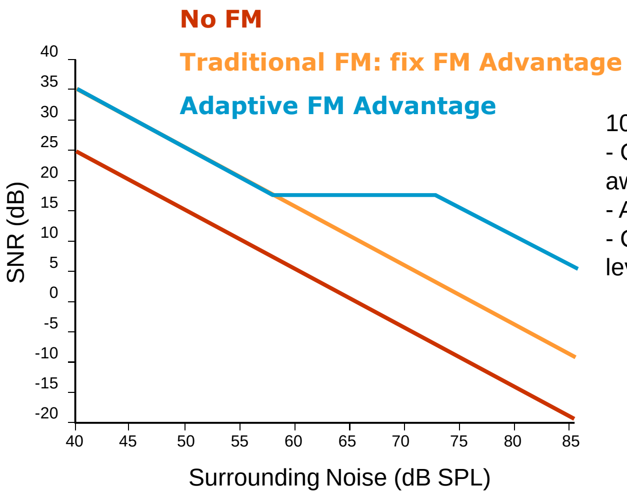
- Delay over the link
- Clock jtter
- Noise floor, signal degradation
- Microphone calibration (amplitude and phase)

Earlevel FM

Modern FM Technology

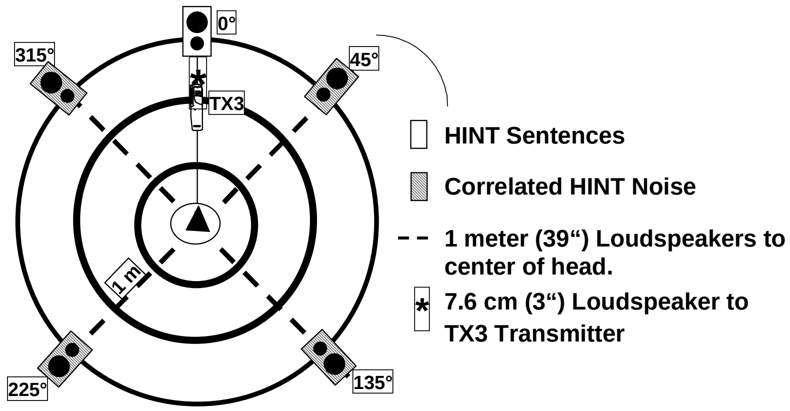
- Dynamic Speech Extraction
  - Automatic FM advantage: Adjusts the FM gain depending on the environmental noise level
  - Surrounding Noise Compensation
  - Voice Activity Detector
- Multi-talker networks: New team teaching concept using up to 10 transmitters

SNR at ear level for different technologies



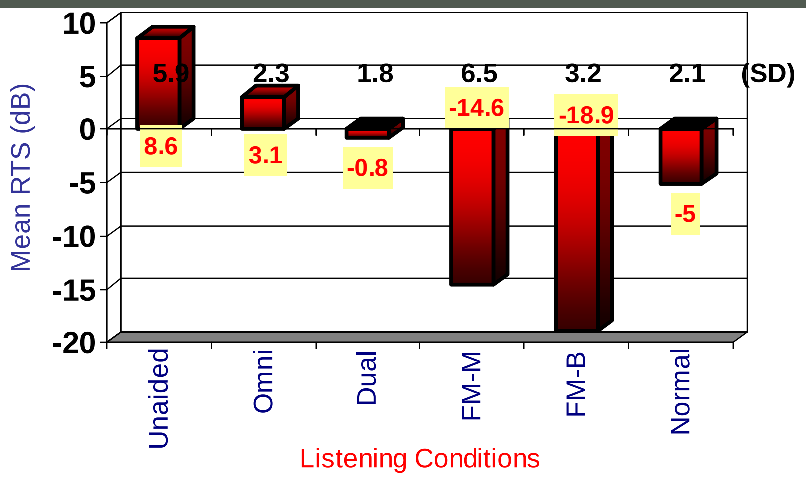
10 dB FM advantage:
- Good environmental awareness
- Audibility of the own voice
- Compromise at high noise levels

Fieldstudy with 48 adults



Source: Valente, 2002

Speech Intelligibility Threshold



Source: Valente, 2002

Auditory Scene Analysis / Hearing Instrument Processing

Auditory Processing	**Hearing Instrument Processing**
Bottom up / top down	Bottom up
No delay constraint, no real-time processing Higher resolution signal analysis Much higher computational power => Stream segregation & source formation: works on several different time scales	**Delay constraint**, real-time processing comp. power constraint - limited signal analysis, spectro-temporal resolution
No signal reconstruction! ⇒ Perceptual attenuation, focus attention, suppression of neuronal activity	**Signal reconstruction** & signal modification: amplification & attenuation / filtering -> "distortions"
Channel: full information capacity	Channel with limited information capacity
A priori knowledge, "situational knowledge" - other sensory modalities - "world knowledge", models of sources ⇔ fill in information…	**Retrospective analysis** Dynamic aspects head / source movement …
Attention control: Target signal identification and tracking, switching back and forth between objects, overcoming salient sources	Target signal – assumption: in front…

Conclusion

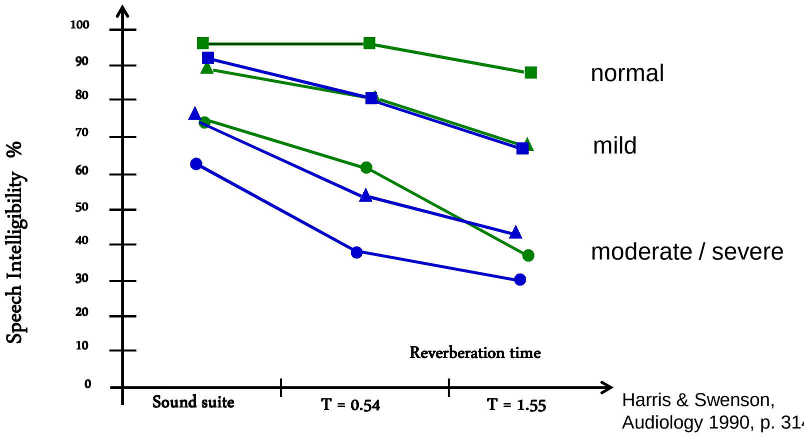
- Hearing instruments offer several algorithms to improve speech intelligibility in complex listening environments
  - Algorithms based mainly on
- Speech intelligibility in complex listening environments remains a huge challenge
  - Reverberation and distance
  - Dynamic target selection and tracking
  - Technical limitations
- Realistic test setups and test procedures

Questions

- Speech intelligibility: how much is top-down driven versus bottom-up processing?
- Speech intelligibility: how fast is it really??
  - How much information do we infer at the end of a „sentence“?
- Which cues (pitch, temporal fine structure, location,….) are the essential ones, does it depend on situation?
  - How does the auditory system pick the relevant one??
- How do we achieve „perceptual constancy“ – voices in real life always sounds the same, (almost) independent of environment?

Thank you…!!!

Speech Intelligibility in reverberant environments



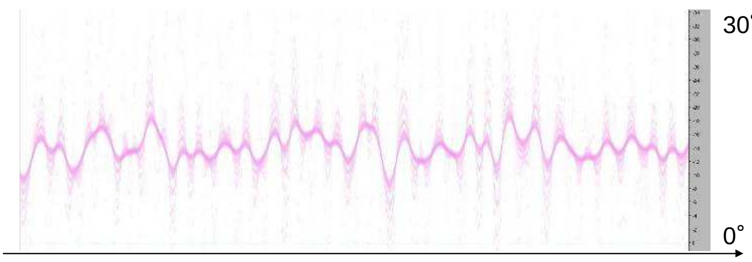
Harris & Swenson, Audiology 1990, p. 314-321

Binaural processing - audio delay

- Group delay:
  - Is mainly determined by Radio bandwidth, ADC, CODEC, buffering (Error correction)
- Delay shall be deterministic and constant
- For binaural audio processing the link delay adds to the other signal processing delay ie. FFT block processing, ADC.
- Overall system delay should be less than 10 ms (Stone & Moore 2005, …)
- Audiosignals + control data:
  - Some more delay for Gain control is acceptable (Hohmann 2009)

Jitter – examples: 800 Hz pure tone

Acoustic delay from head dimension: typ. 500 µs for ear distance
Normal hearing minimum audible angle: a few µs
Jitter should be smaller than 20 µs RMS
-> allows for binaural beamforming without significant localization errors



$$T_{jitter} = \frac{phasediff}{360° \cdot freq} \leq 30\mu s$$