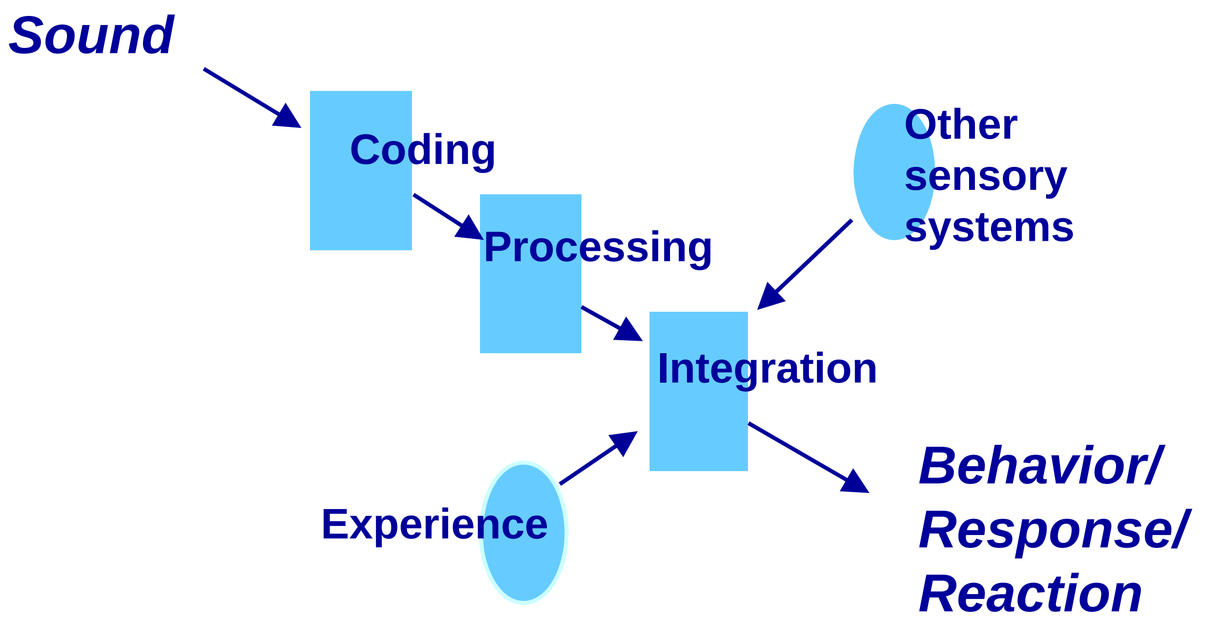
Mechanisms and models of normal and impaired hearing

Torsten Dau

Hearing Systems
Department of Electrical Engineering
Technical University of Denmark

Hearing: Information processing system

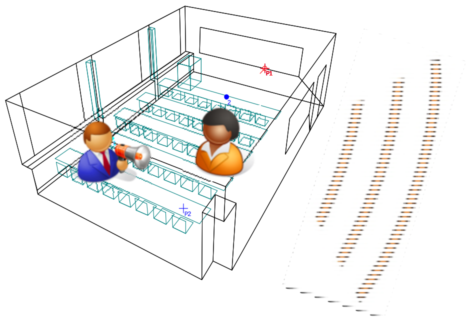


Challenges:

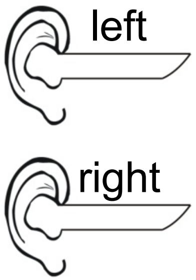
- Representation and processing of signal information in the auditory system; neural correlates of perception.
- Modelling auditory signal processing and perception.
- Integration of processing strategies in technical and clinical applications.

Various challenges

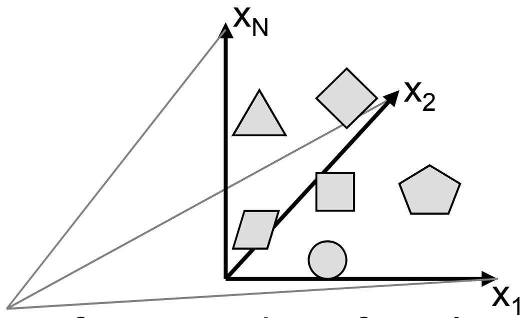
Physical domain



Complex sound field



1-dim.
input signal



features: time, freq, level, location, modulation, …

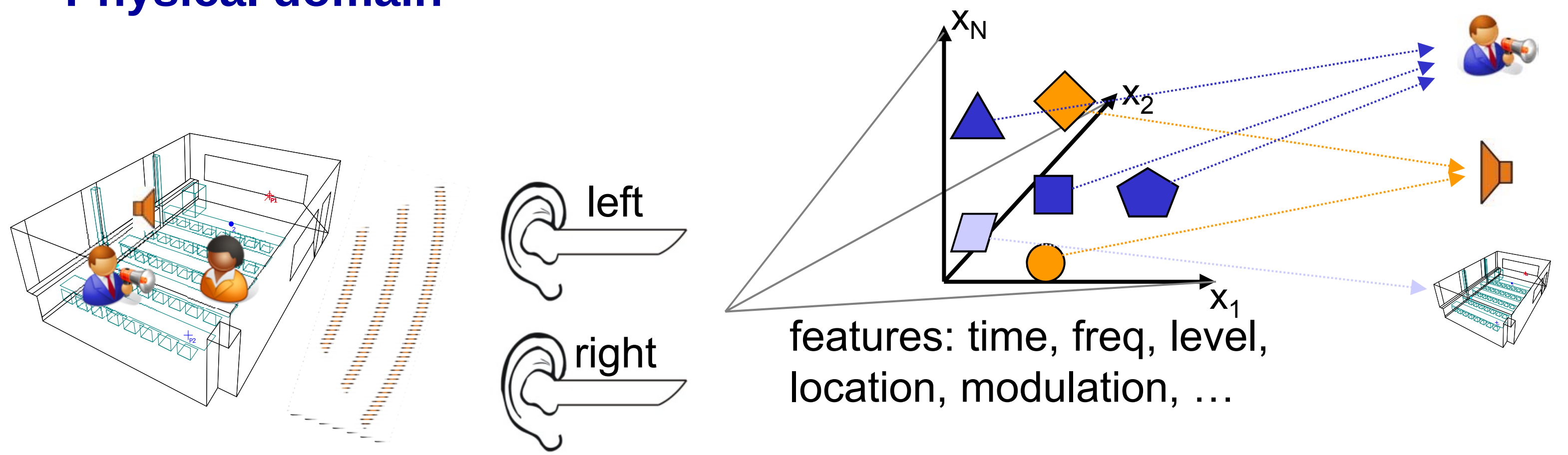
Coding

Multi-dimensional feature representation

Various challenges

Psychophysical domain

Physical domain



- intelligibility
- loudness
- …
- pitch
- distance
- …
- room size
- envelopment
- …

Complex sound field

1-dim. input signal

Coding
Multi-dimensional feature representation

Decoding
Segmentation, feature binding

Outcome measure

Some "hot topics":

- Solving the cocktail-party problem
- Solving the dynamic range problem
- Coding of spatial sounds in humans
- Neural correlates of learning and attention

The “cocktail-party” problem

In ”cocktail-party” situations, normally-hearing listeners effortlessly segregate different sources. No artificial system performs anywhere near as well.

The "cocktail-party" problem

Key problem: Hearing-impaired people have difficulty with speech communication (even with hearing aids) when background noise is present.

- What gets lost in the impaired system (besides sensitivity)?
- How can we compensate for the deficits with hearing instruments?
- What are the properties that makes the intact auditory system so special?
- Why is the intact auditory system so robust in challenging situations?

Can auditory models help?

Main goals:

- To represent the results from experiments within one framework
- To explain the functioning of the system

Specifically:

– Models can help generate hypotheses that can be explicitely stated.

– Models can help determine how a deficit in one or more components affects the operation of the system.

– Models can illustrate how complex a problem is.

Classes of models: Conceptual, biophysical, physiological, mathematical, computational, … perceptual models, depending on which aspects are considered.

Example

Physical

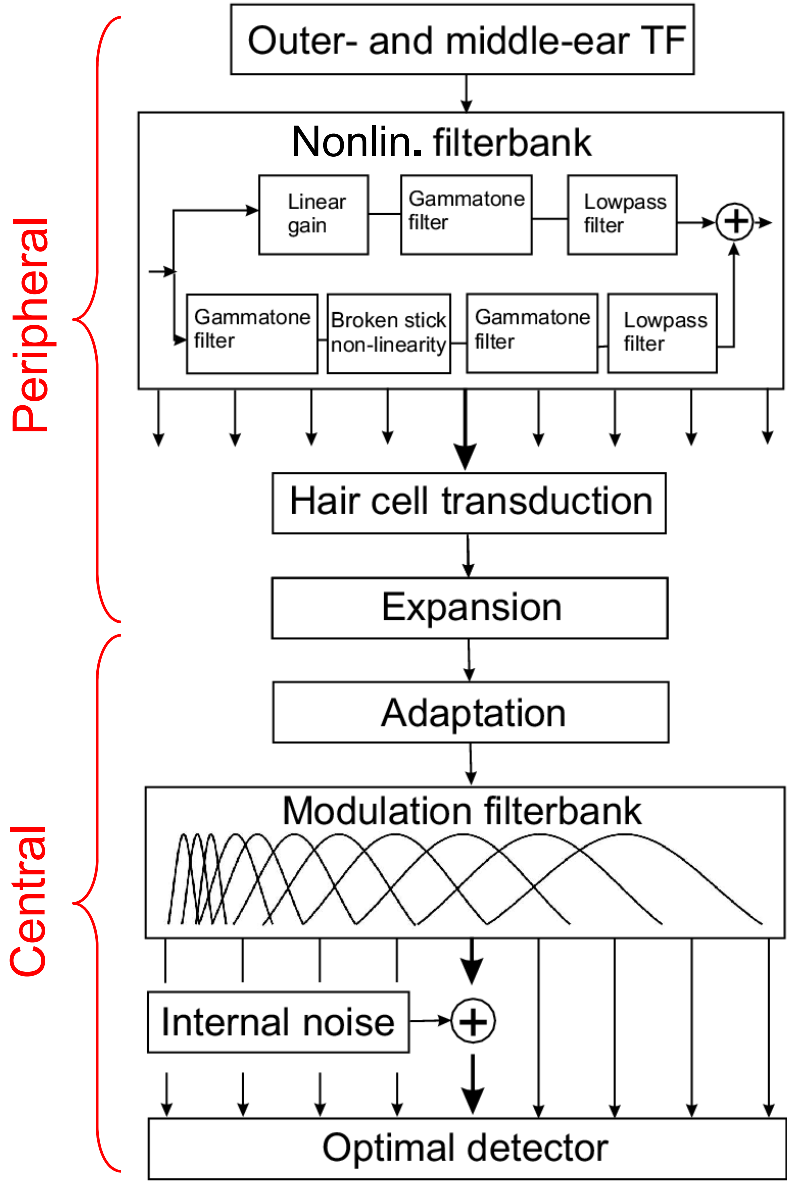
Φ ⟶ Observer ⟶ Ψ

Psychophysical

(e.g.: $I, f, \sigma_I, \ldots$)

(e.g.: L, *pitch*, *jnds*, ...)

Example for a perception model
(from an engineering point of view)



Model focusses on limitations in resolution rather than predictions of sensations.

Focus on the simulation of perception data (inspired by physiology)

Computational auditory signal processing and perception model CASP

Jepsen *et al.* (2008) (based on Dau *et al.*, 1997)

HEARING SYSTEMS

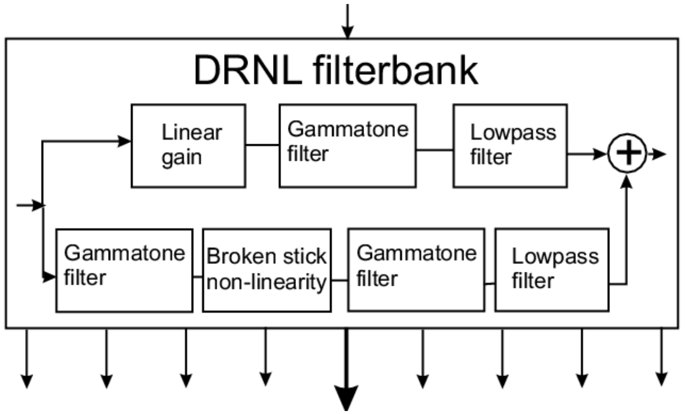
Overview

- Key aspects of nonlinear cochlear processing for auditory perception
- Across-channel processing and coincidence detection
- Adaptation: Steady-state compression and dynamic contrast enhancement
- Processing of temporal and spectral modulations and consequences for speech perception
- Computational auditory scene analysis: An approach based on coherence

Nonlinear cochlear filtering

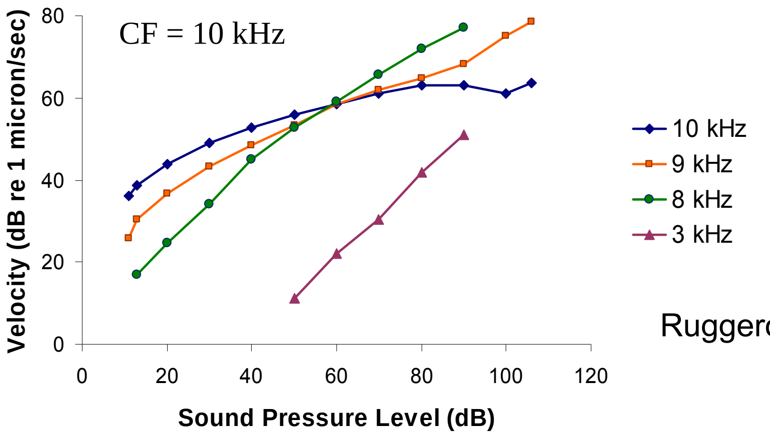
Dual **R**esonance **N**on**l**inear (DRNL) filtering

Meddis *et al.* (2001)

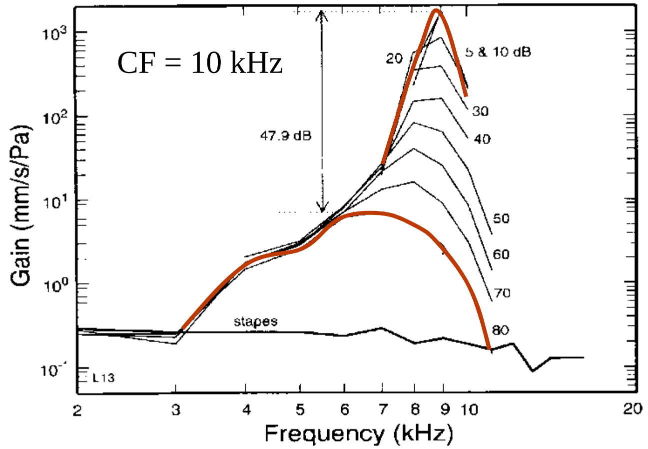


Alternative models:

e.g., Heinz *et al*. (2001); Zilany and Bruce (2006); Irino and Patterson (2007), …



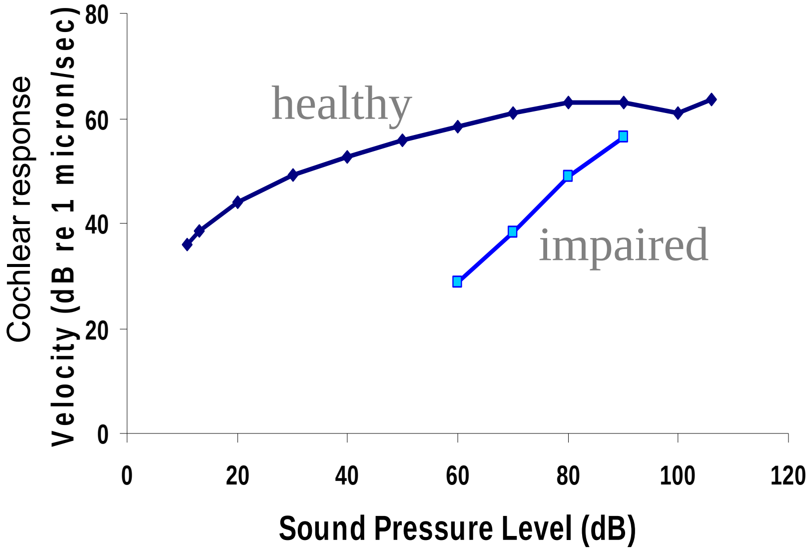
Ruggero *et al.* (2001)

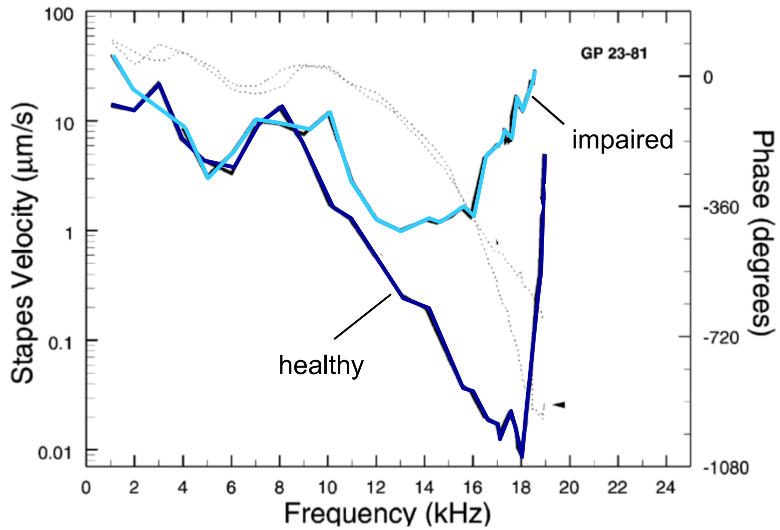


- Amplification and compression only at the characteristic frequency (CF).
- Linear response properties for off-freq. stimulation.

- Frequency selectivity is level dependent.
- Largest gain at low stimulation levels

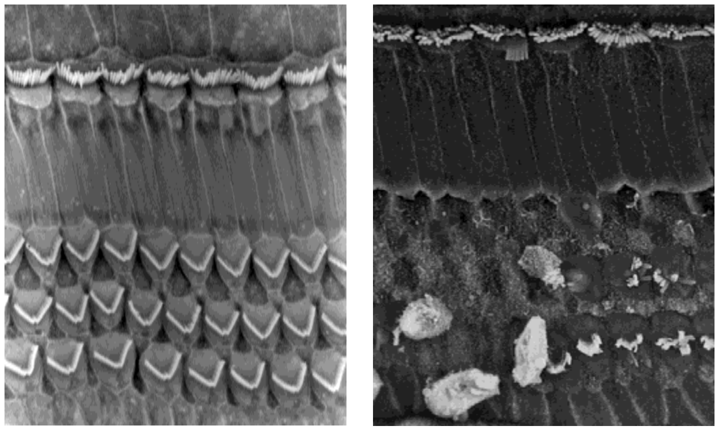
Cochlear damage





IHC

OHC



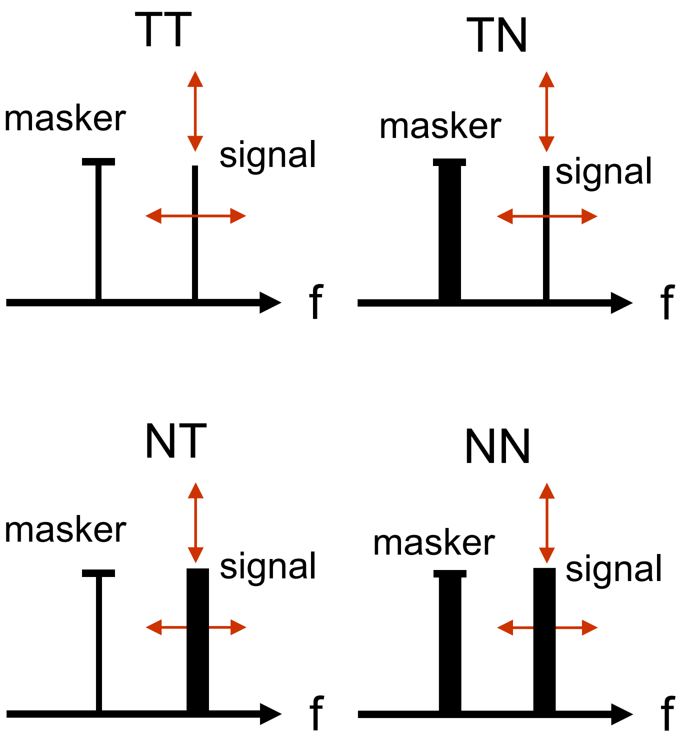
healthy　　impaired

Simulation of loss of frequency selectivity (broadening factor of 3) but no threshold elevation.

original　　processed

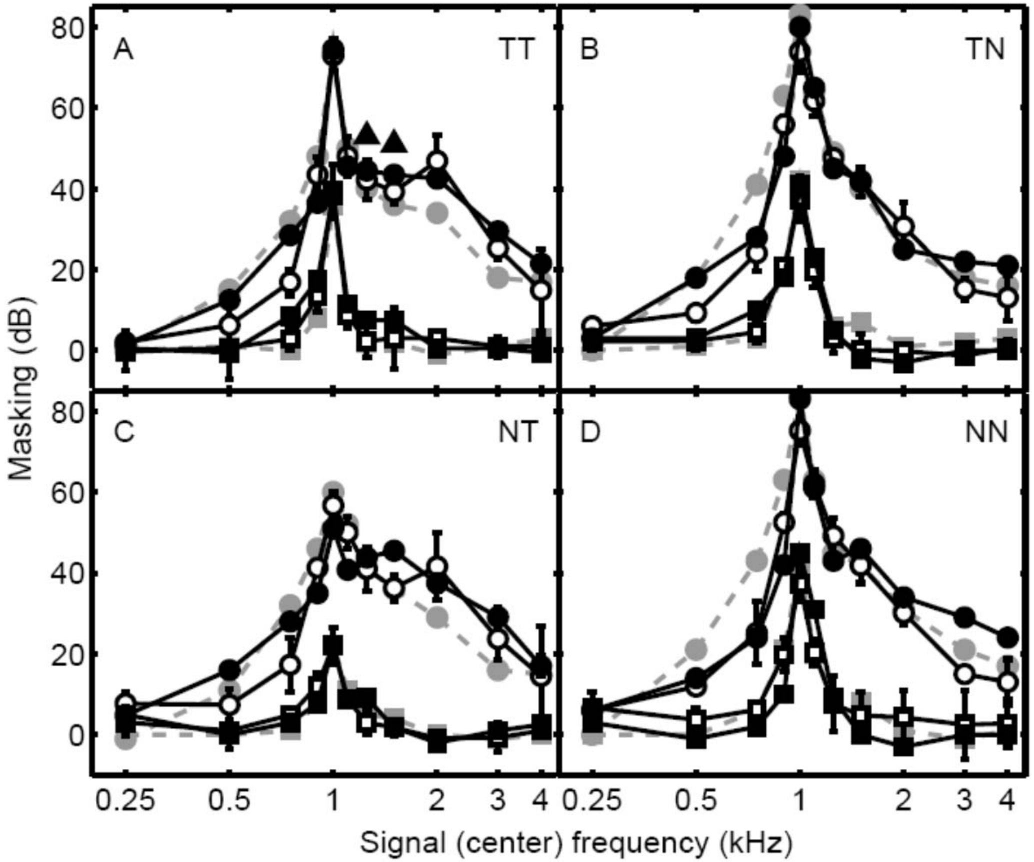
Moore (1995)

Spectral masking patterns (NH)



Masker level: 45 or 85 dB SPL

Jepsen *et al.* (2008)

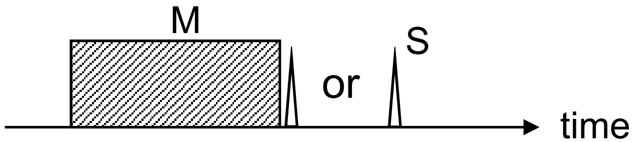


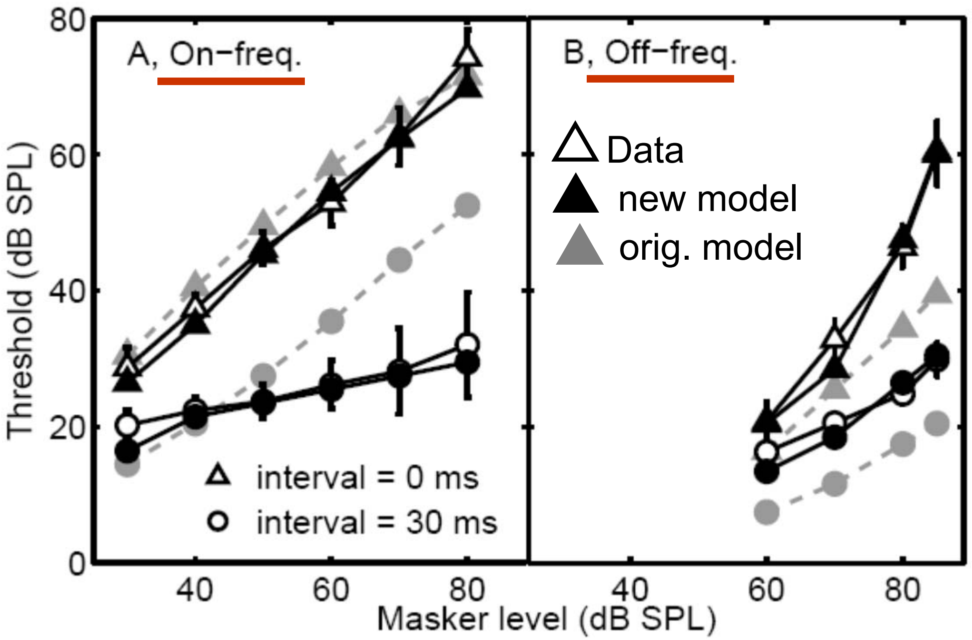
Data (Moore *et al.*, 1998)

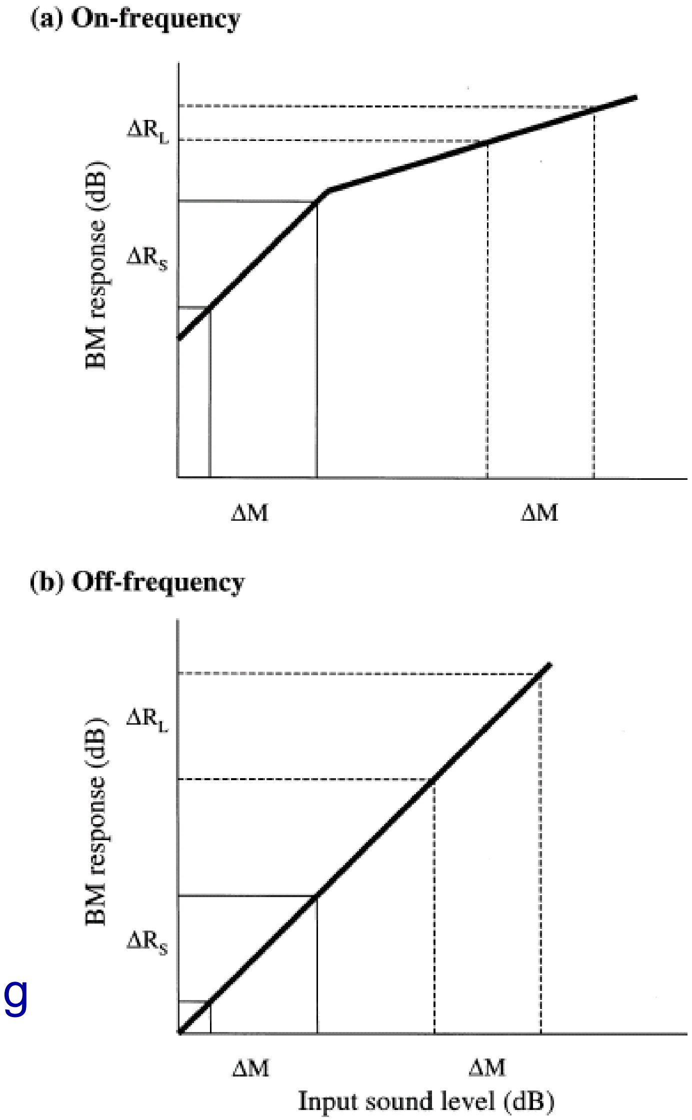
Model

Original model (with *linear* BM)

On- versus off-frequency forward masking (NH)



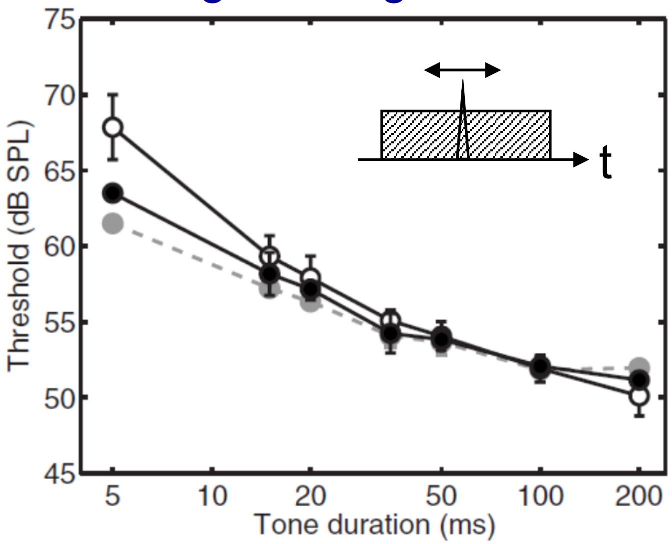




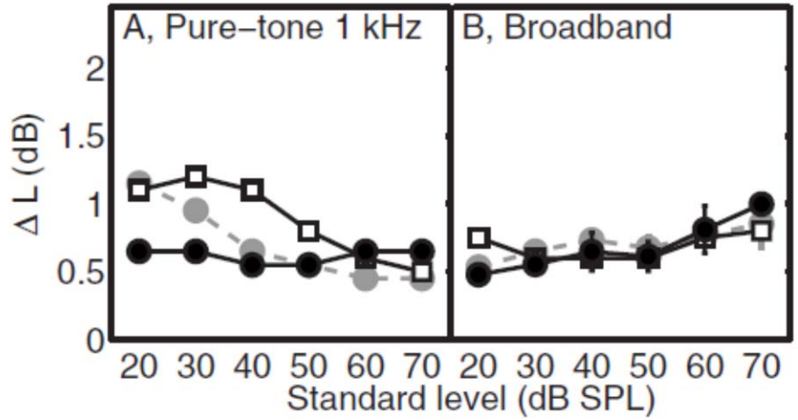
⇒ Model accounts for effects of BM processing on forward masking (NH).

Intensity discrimination, signal integration and AM detection (NH)

Signal integration

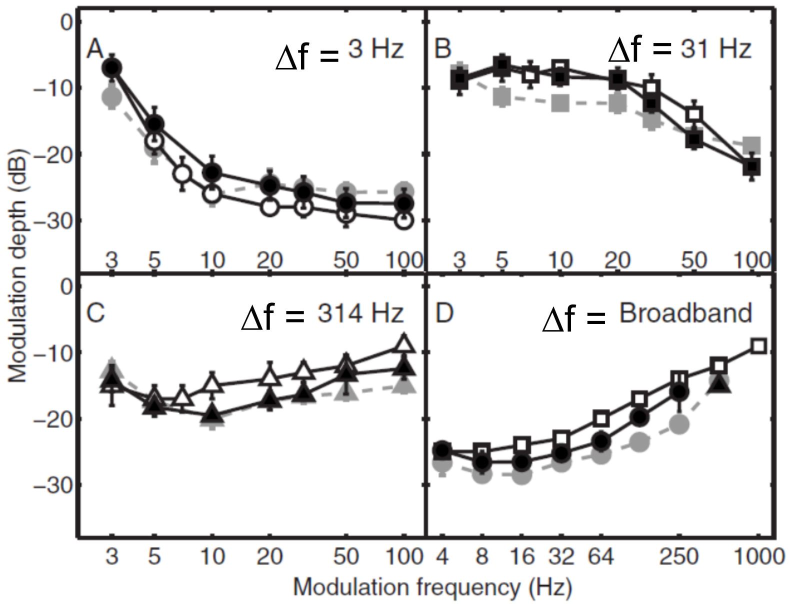


Intensity discrimination



… and modulation detection

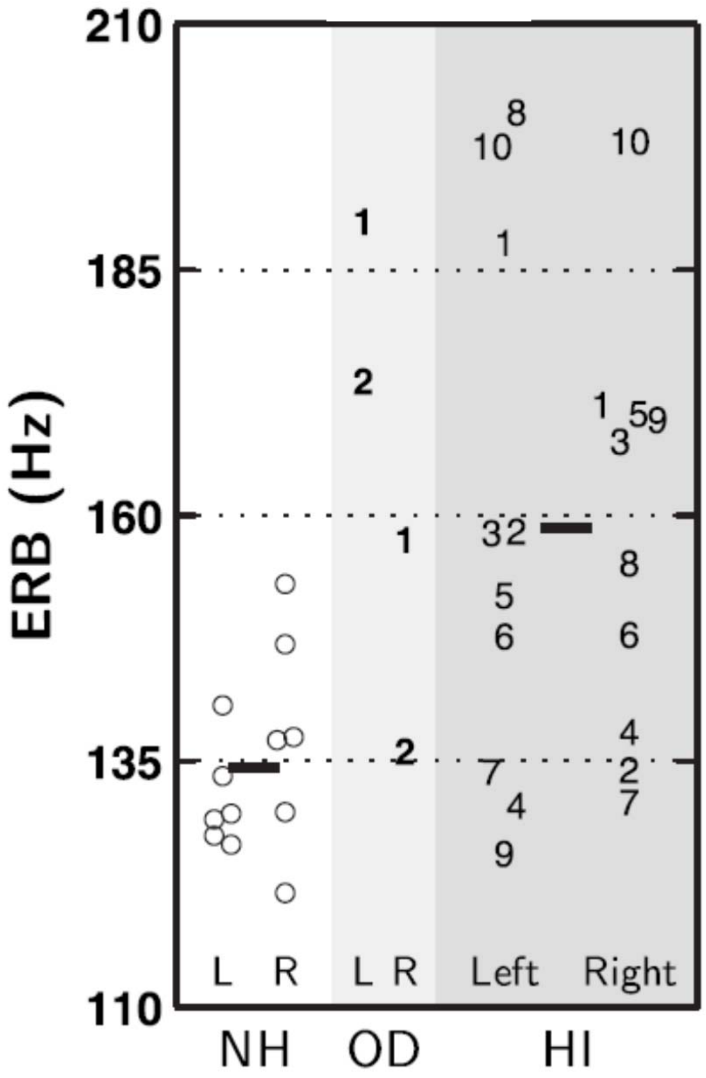
Modulation transfer functions



… resulting from the properties of optimal detector, adaptation stage and modulation filterbank

Jepsen *et al*. (2008)

Modelling variability in the data?



So far so nice. Preprocessing has also been successful as front end in certain applications
- *for mean NH listeners.*

However, a model is missing that accounts for the variability in the data (particularly in HI listeners).

(Example: Frequency selectivity @750 Hz in listeners with normal audiogram at low frequencies.)

1. step: Characterisation of HI ("auditory profile").
2. step: Prediction of individual HI ("basic" functions).
3. step: Evaluation in other tasks (e.g. speech).

Strelcyk and Dau (2009)

Modelling individual hearing impairment?

Measures of basic auditory functions provide information about individual deficits

⇒ Modifications in the model (based on auditory profile)

Peripheral stages:

– Reduced compression

– Change of frequency selectivity

– Reduced sensitivity

– Degraded fine-structure processing

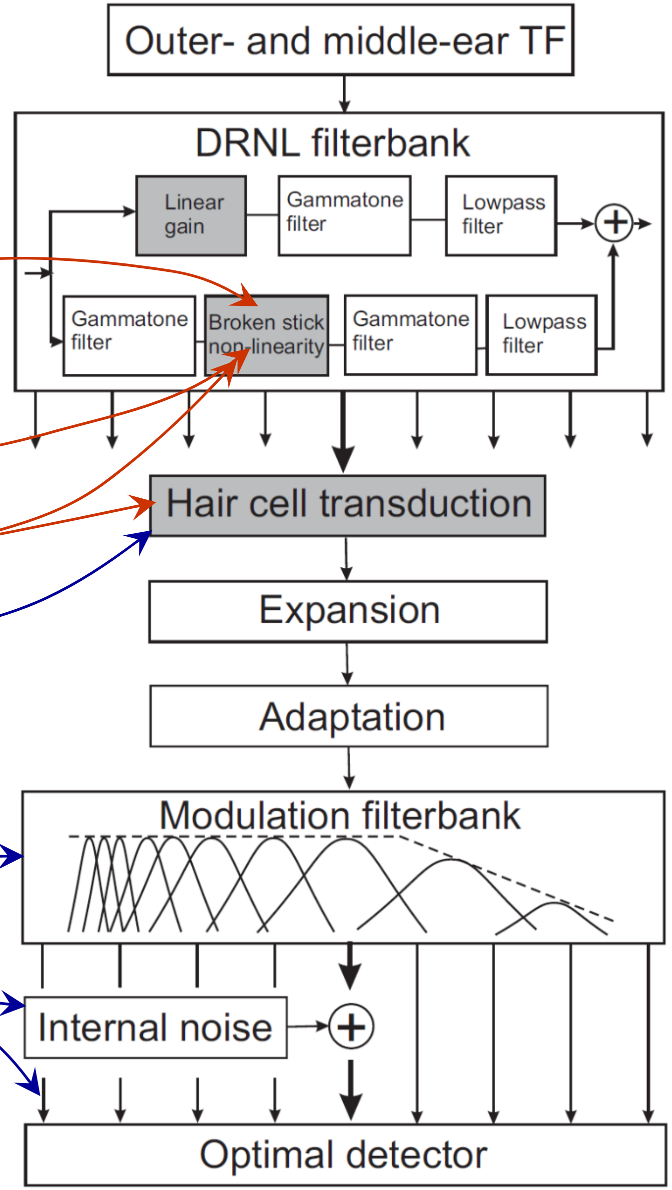
– …

Later stages:

– Deficits in modulation coding

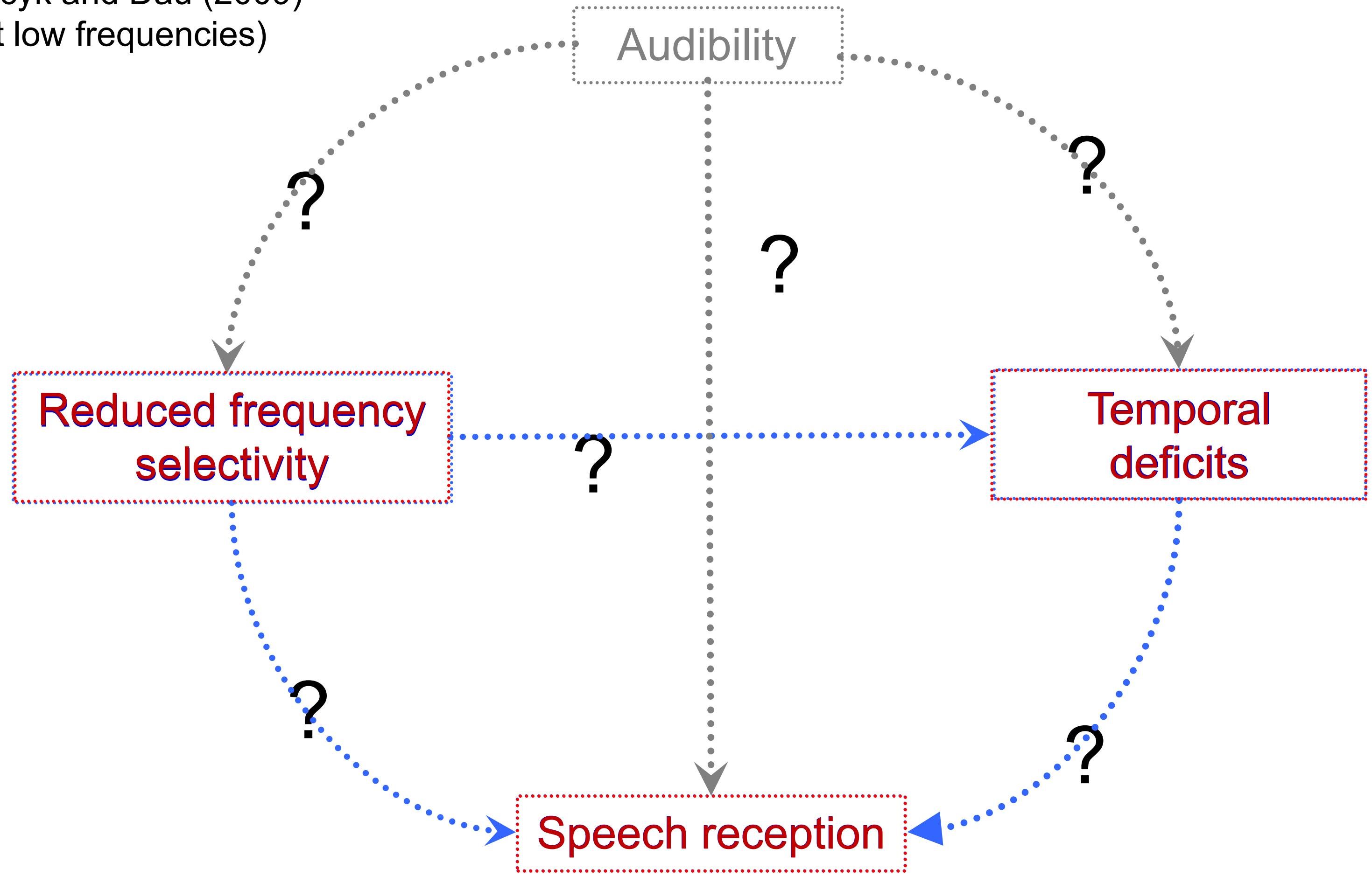
– ”Sub-optimal” processing

– Central/cognitive deficits?



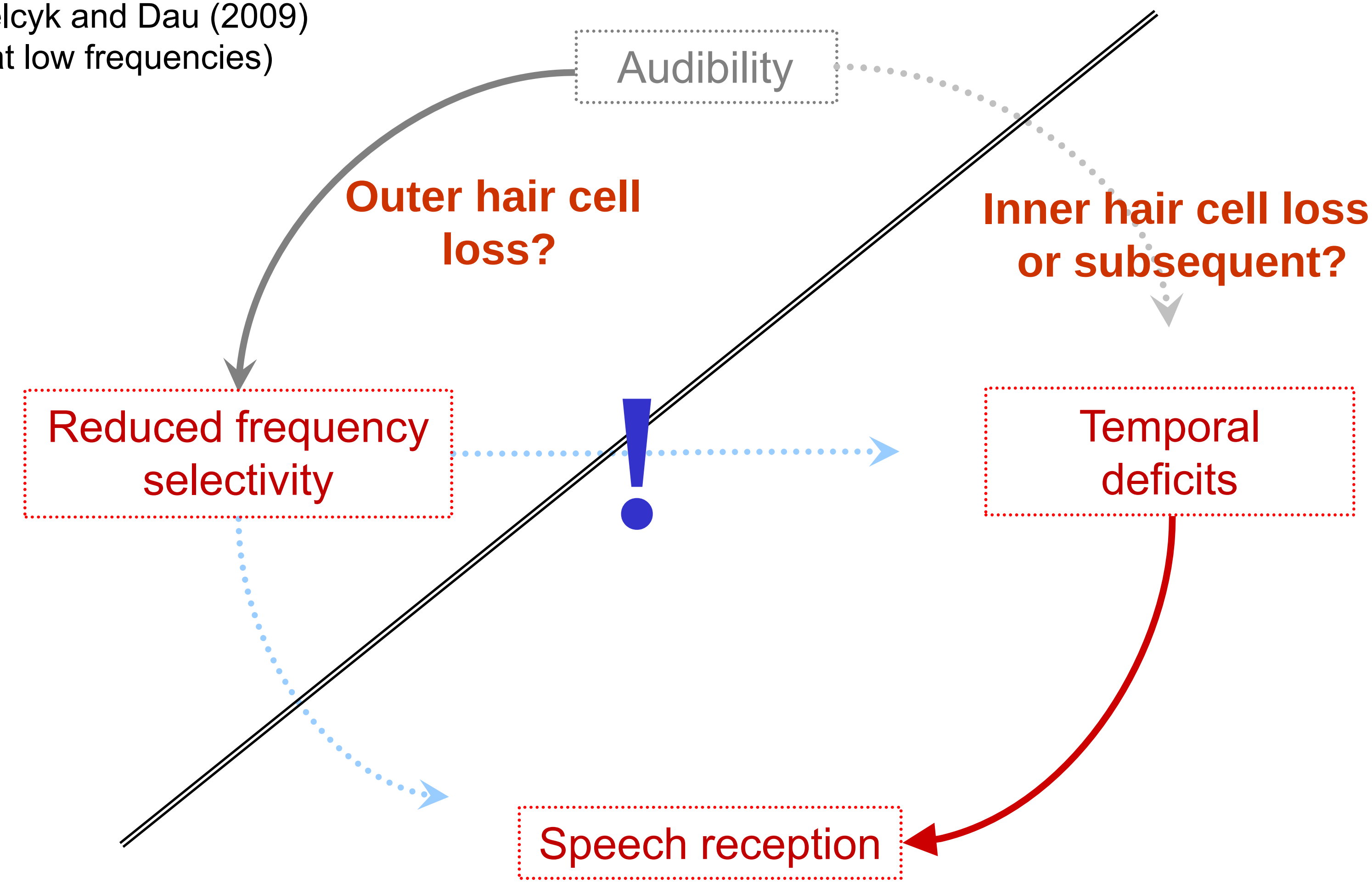
Relations between functions in impaired hearing

Strelcyk and Dau (2009)
(at low frequencies)



Relations between functions in impaired hearing

Strelcyk and Dau (2009)
(at low frequencies)



in two-talker and lateralized noise,
but not in modulated noise
⇒ Important for spatial segregation and talker separation?

HEARING SYSTEMS

Overview

- Key aspects of nonlinear cochlear processing for auditory perception
- Across-channel processing and coincidence detection
- Adaptation: Steady-state compression and dynamic contrast enhancement
- Processing of temporal and spectral modulations
- Computational auditory scene analysis: An approach based on coherence

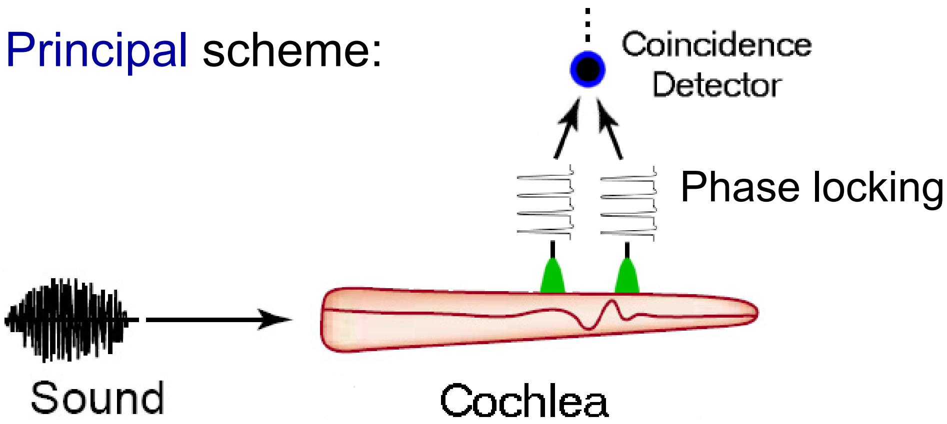
Across-channel processing

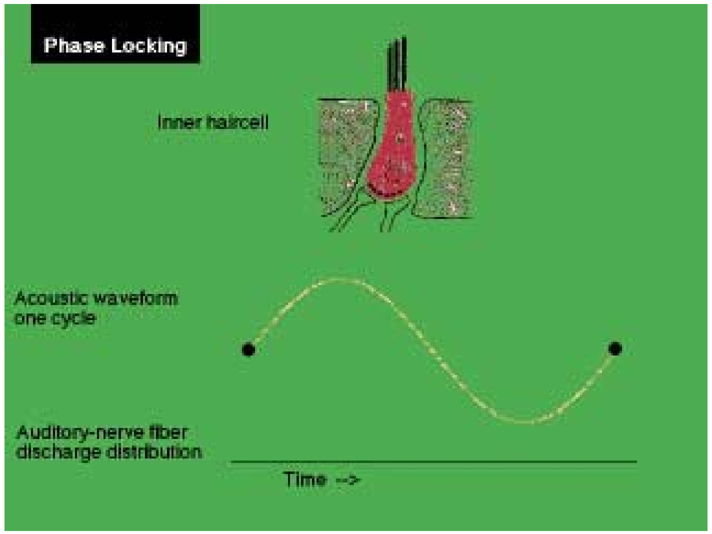
Reduced frequency selectivity thus represents only one impairment factor.
Probably not sufficient to explain the major problems in noisy environments.

Hypothesis: Across-channel processing is important for robust signal encoding (Loeb *et al*., 1981; Carney *et al*., 2002).

Spatio-temporal models: Idea: To extract information from the spatio-temporal pattern of cochlear activity.

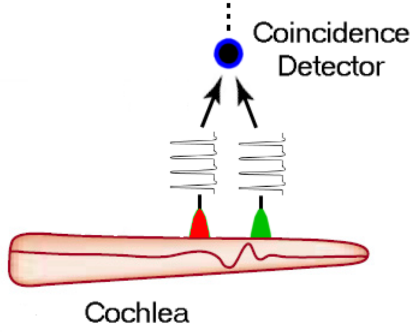
Principal scheme:

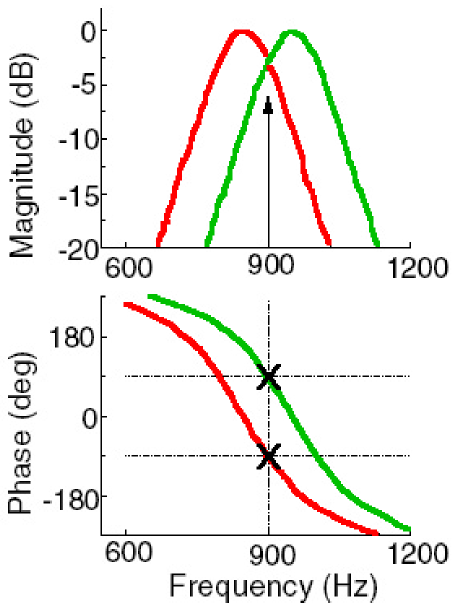


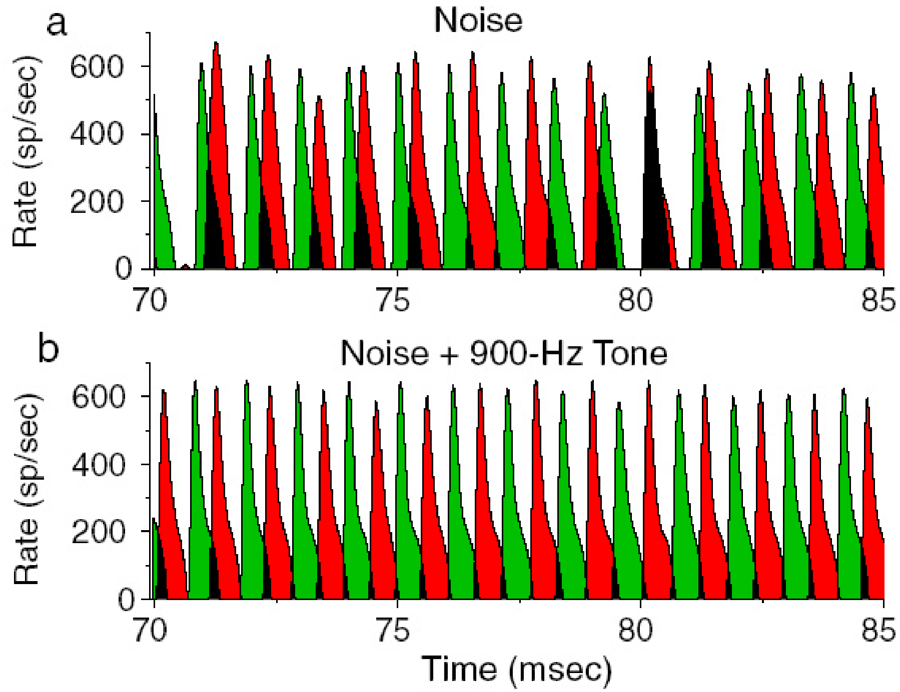


Across-channel model for signal-in-noise detection

Detection of formants and tones in noise (e.g., Deng & Geisler, 1987; Carney *et al.*, 2002)







Presence of the signal (e.g., formant): Reduction in the rate of cross-frequency coincidence.

Response pattern is very robust to overall level changes (as long as the 3 ”components” are intact).

Cochlear damage and across-channel processing

Typical consequences of cochlear damage:

1. Reduced frequency selectivity as a consequence of loss or reduction of compression (outer hair cells)

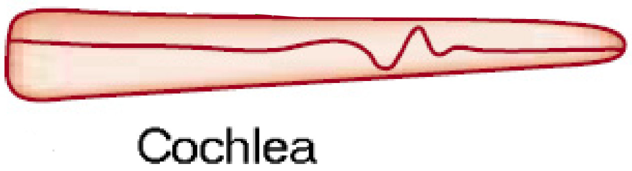
2. Deterioration of the encoding of temporal fine structure, by:

i) Reduced precision of phase locking in AN fibers

ii) Reduced number of cochlear hair cells → reduction of converging inputs

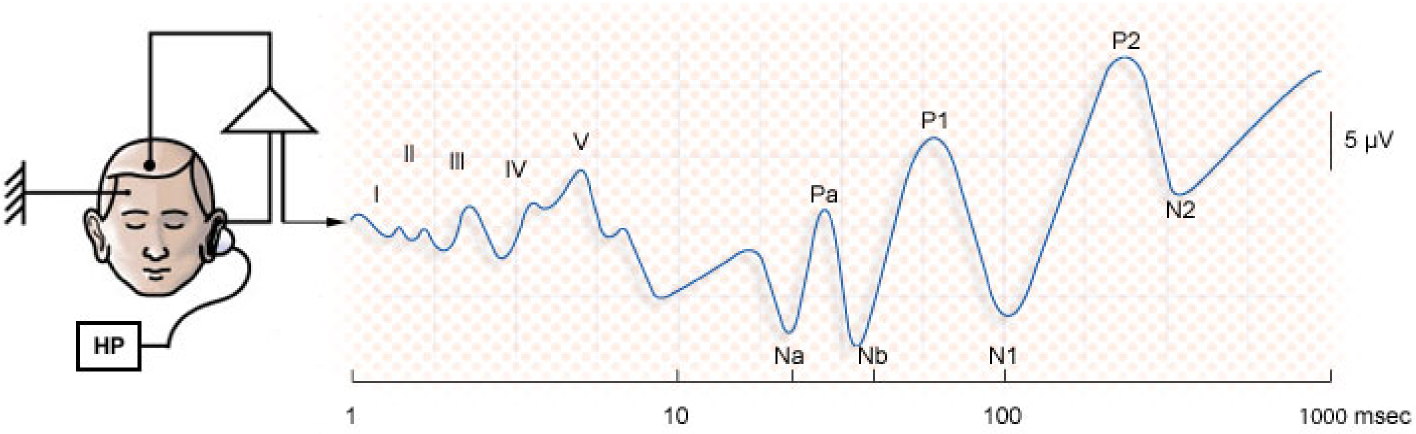
iii) Loss of coincidence detectors

Effect on the output of spatio-temporal processing:



How can we measure phase locking in humans?

Evoked potentials

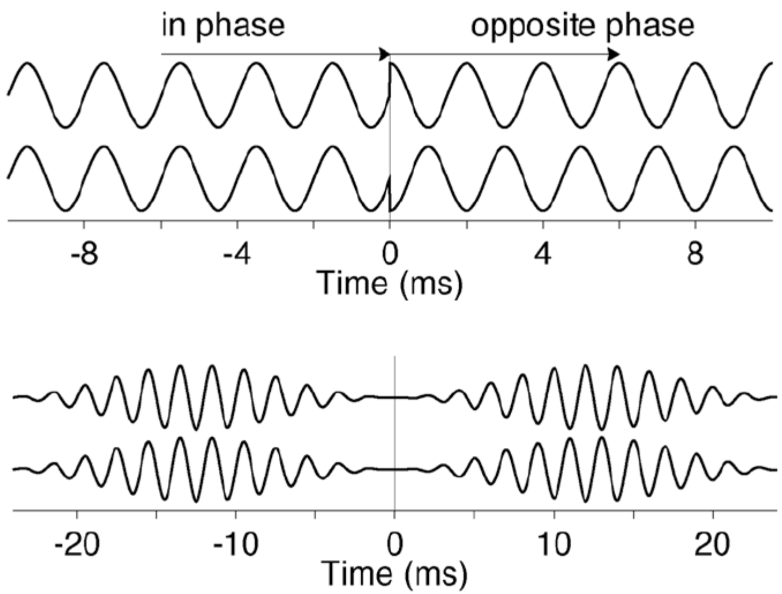


Ross *et al*. (2007)

Sound stimulation

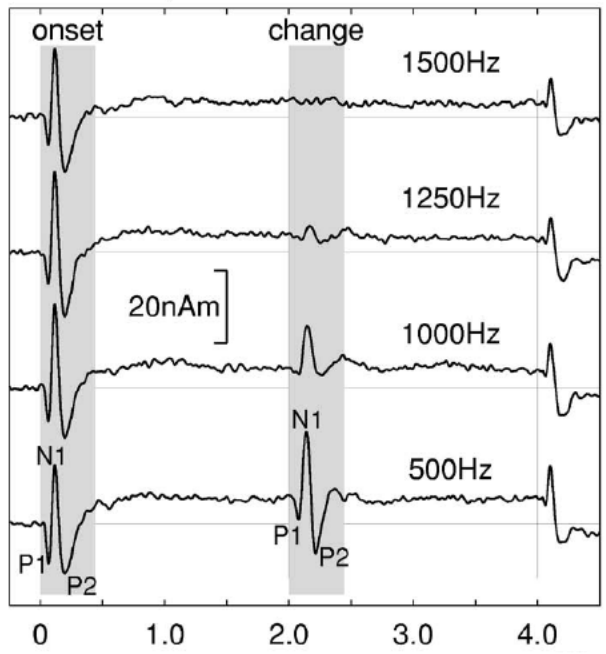
Left ear:

Right ear:



⇒

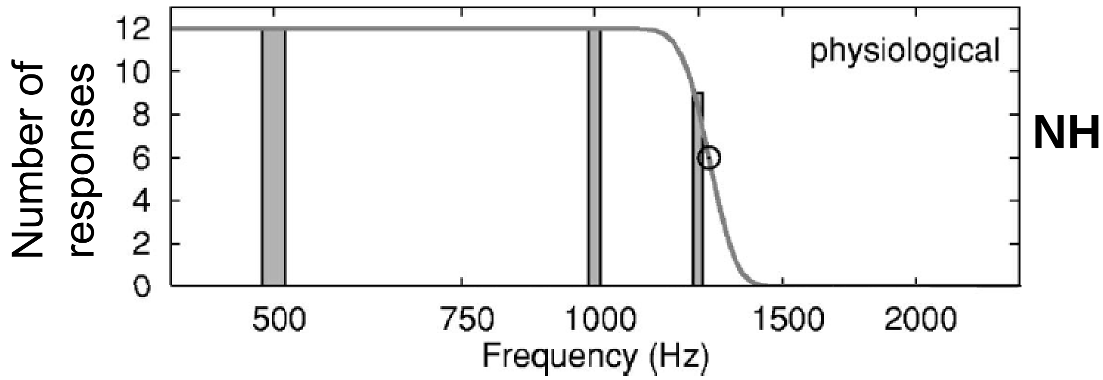
Responses



Phase locking in normal versus impaired hearing

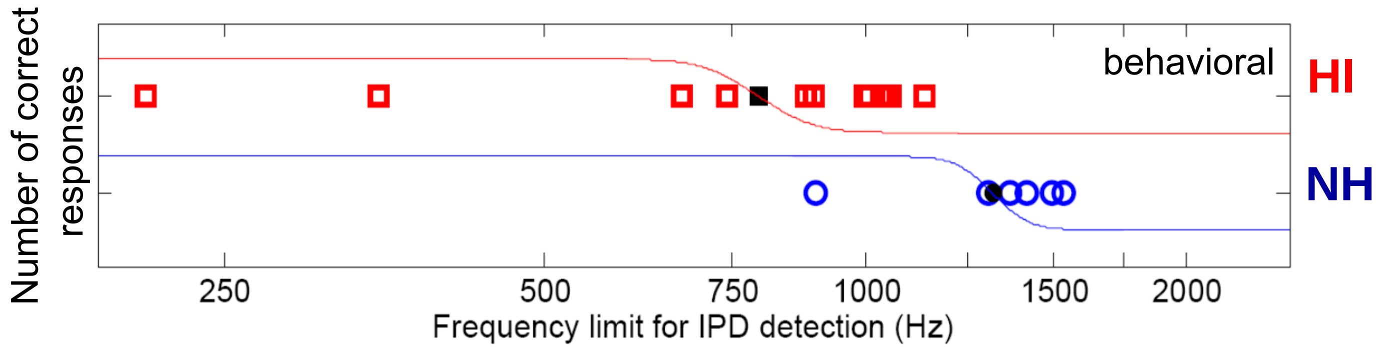
Physiological results

Ross *et al.* (2007)
(MEG study)



Behavioral results

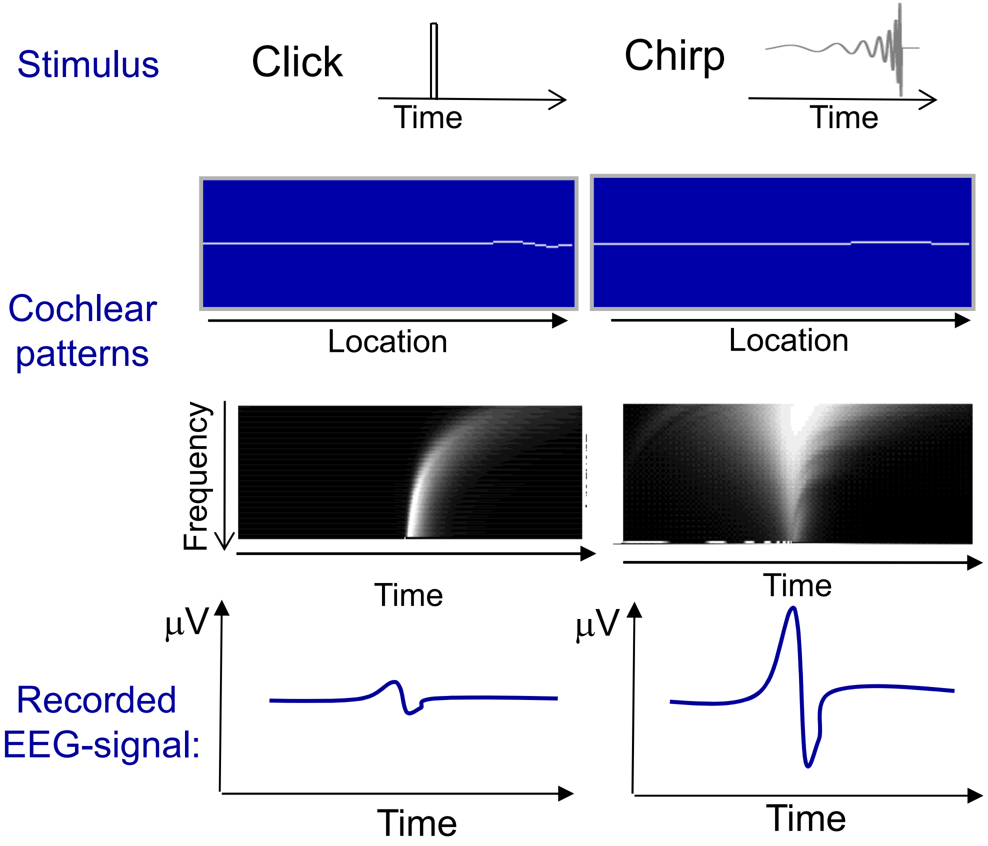
Santurette and Dau (2009)



⇒ Good correspondence between physiological and behavioral estimates

⇒ Most hearing-impaired listeners show a lower frequency limit - either due to degraded monaural phase locking or deficits in the ”binaural operator”.

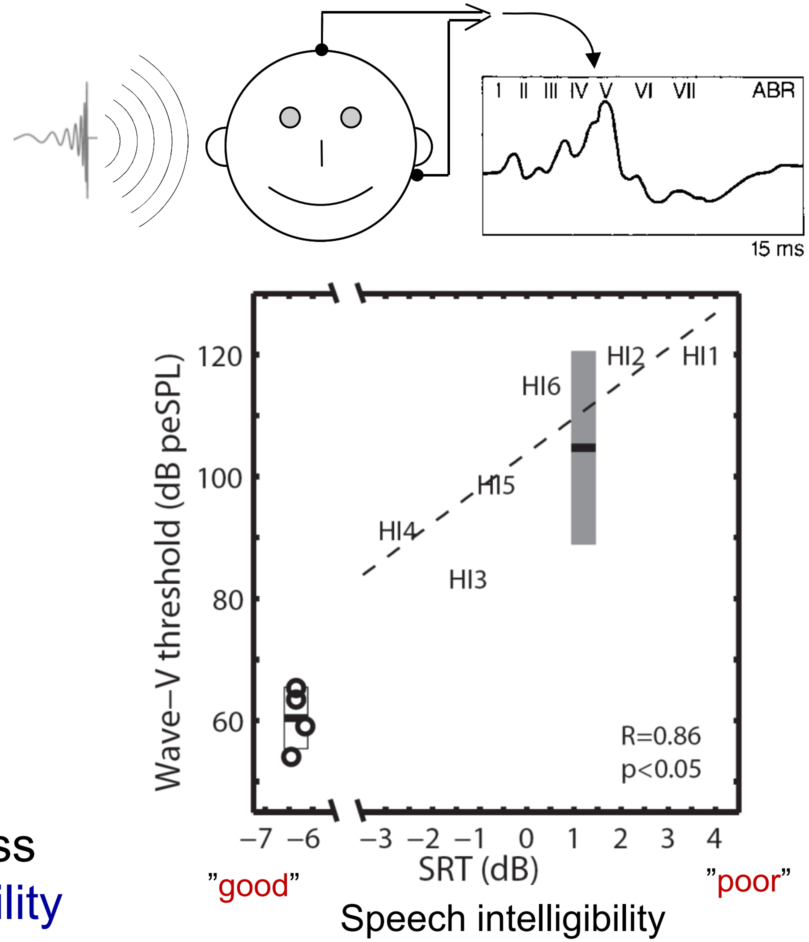
Relation between speech intelligibility and measures of temporal processing



Dau *et al*. (2000); Harte *et al.* (2010)

⇒ The amount of neural synchronization across frequeny is correlated with speech intelligibility (but not with audibility).

Auditory-evoked potentials (ABR)



Papakonstantinou *et al*. (2011)

HEARING SYSTEMS

Overview

- Key aspects of nonlinear cochlear processing for auditory perception
- Across-channel processing and coincidence detection
- Adaptation: Steady-state compression and dynamic contrast enhancement
- Processing of temporal and spectral modulations and consequences for speech perception
- Computational auditory scene analysis: An approach based on coherence

Level adaptation

We can hear sounds extending over a huge range of sound levels (of 120 dB).

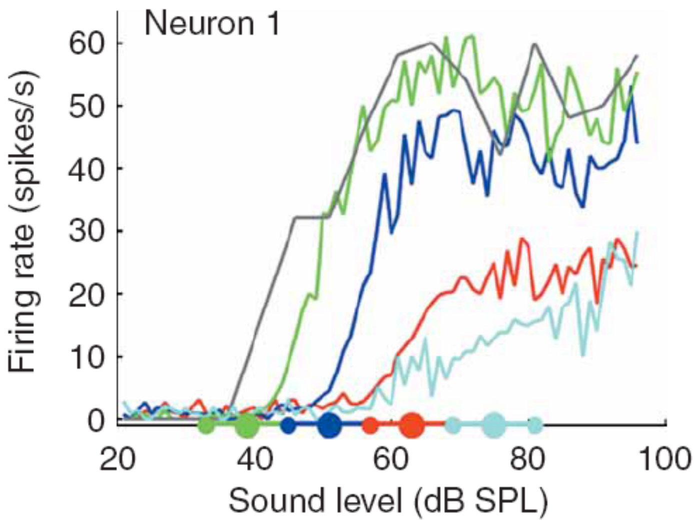
At the same time, we can hear level changes of about 1 dB across entire level range.

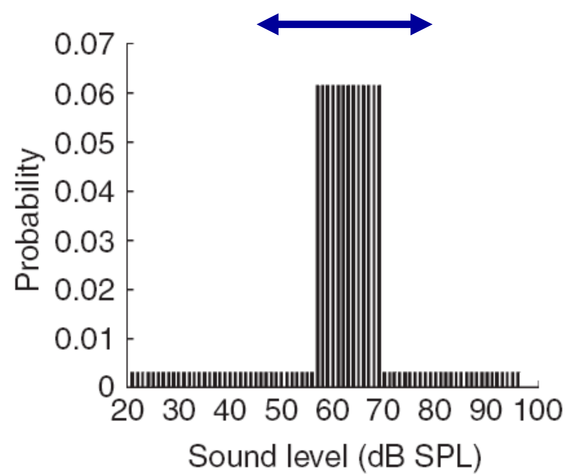
However: The neural dynamic range of individual neurons is very limited as studied extensively in the cochlea (auditory nerve).

Which mechanisms exist that extend the range of coding?

Possible solution:

Dean *et al.* (2005)

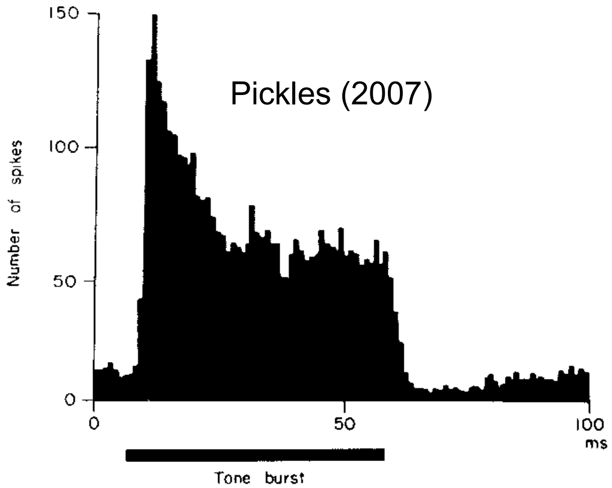




⇒ Adaptive processes of neurons throughout the auditory system (here: brainstem).

Dynamic adaptation

Dynamic changes: Temporal pattern of adaptation is similar throughout auditory pathway but ''time constants'' change from *ms* to *s*.

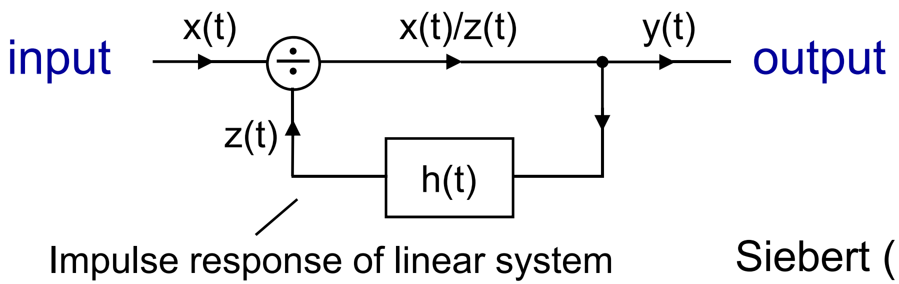


Firing patterns of many neurons show a form of adaptation to a sudden change in stimulus level.

Auditory nerve: rapid adaptation

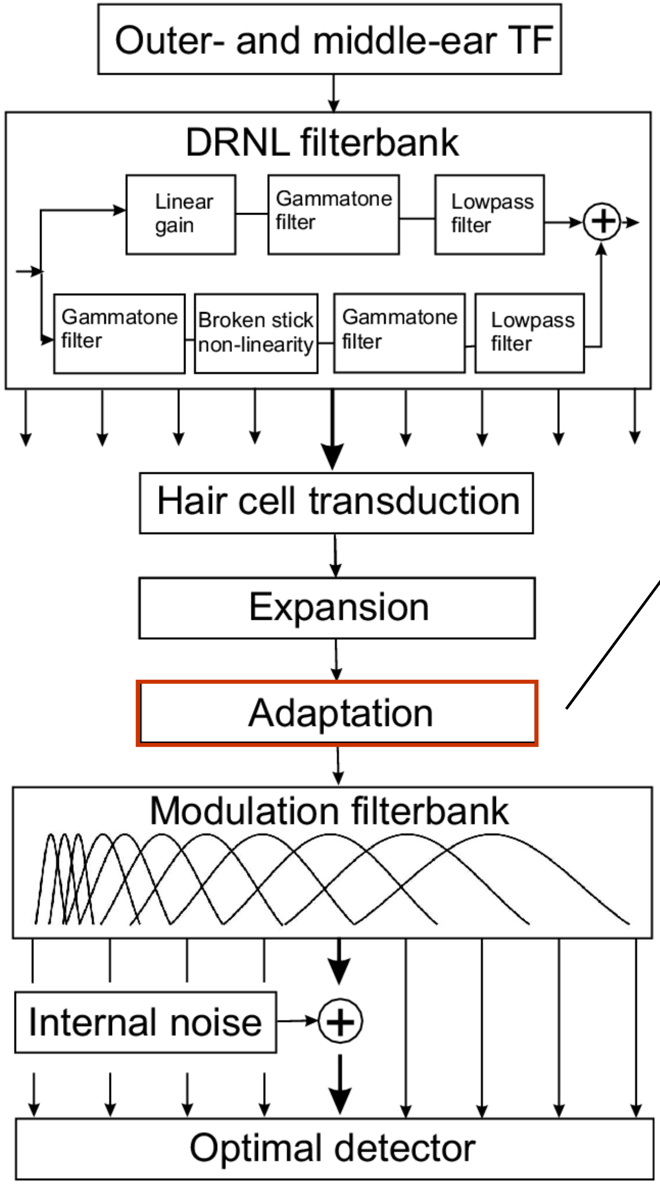
Phenomenological model of adaptation:

⇒ Steady-state compression and contrast enhancement



Siebert (1968); Dau *et al.*(1996)

Model including an adaptation circuit



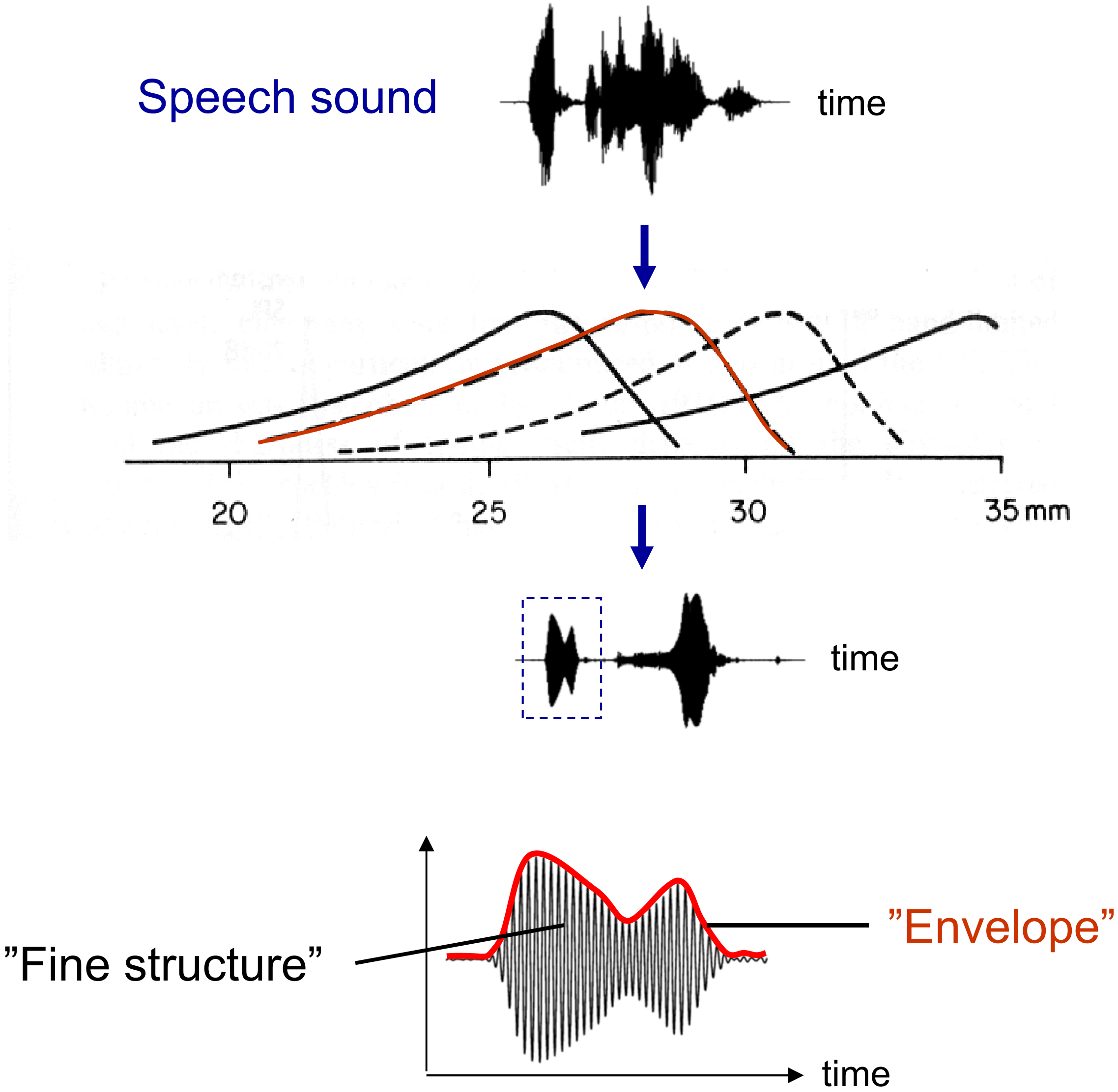
- \+ Simple circuit that accounts for large variety of behavioral data
- \+ Provides robust internal representation in model applications
- – No explanation of the mechanisms underlying adaptation

Jepsen et al. (2008)
(based on Dau *et al.*, 1997)

Overview

- Key aspects of nonlinear cochlear processing for auditory perception
- Across-channel processing and coincidence detection
- Adaptation: Steady-state compression and dynamic contrast enhancement
- Processing of temporal and spectral modulations and consequences for speech perception
- Computational auditory scene analysis: An approach based on coherence

How is the envelope coded in the auditory system?

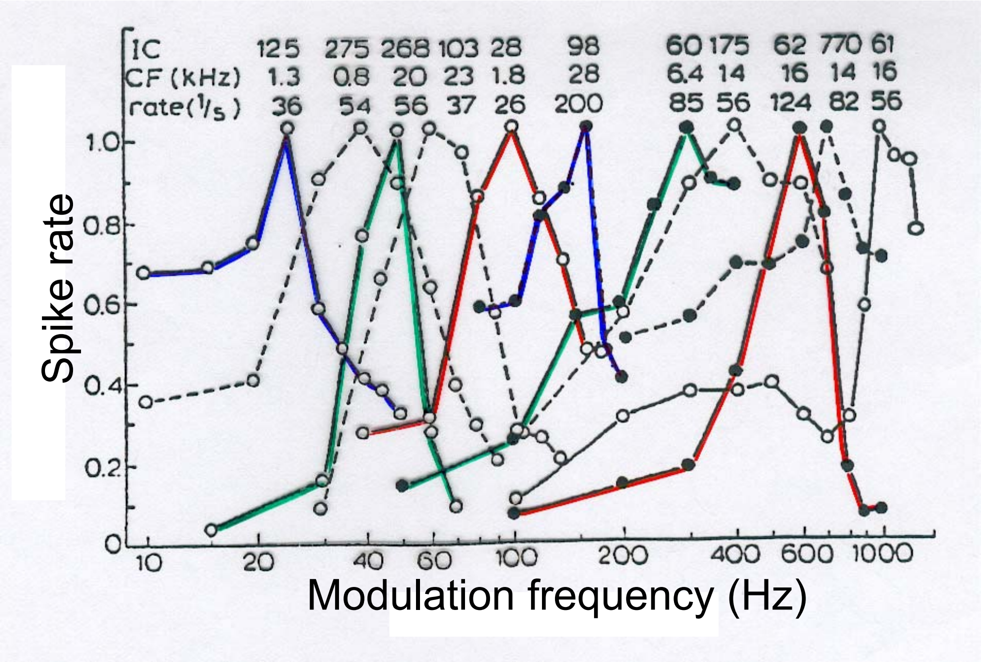


Modulation selectivity

Evidence from physiological and perceptual data: Decomposition of the temporal envelope at the output of each cochlear filter.

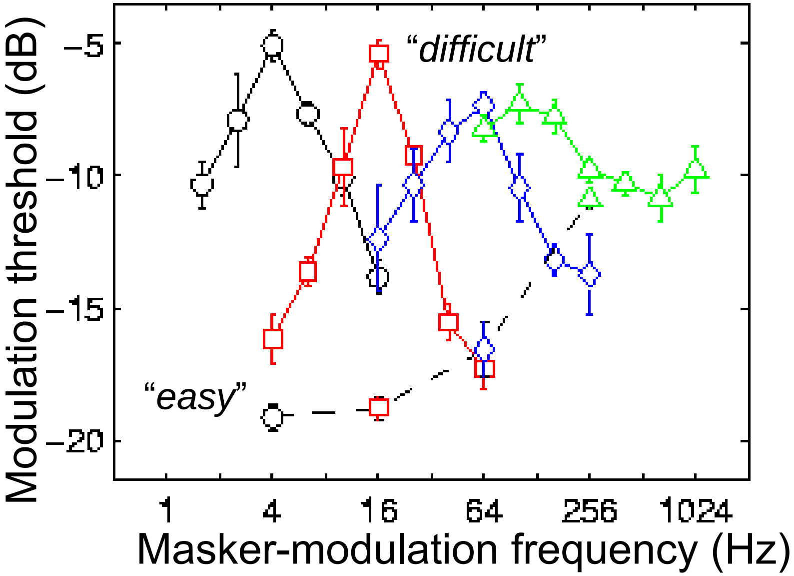
Langner (1992)

Physiological data



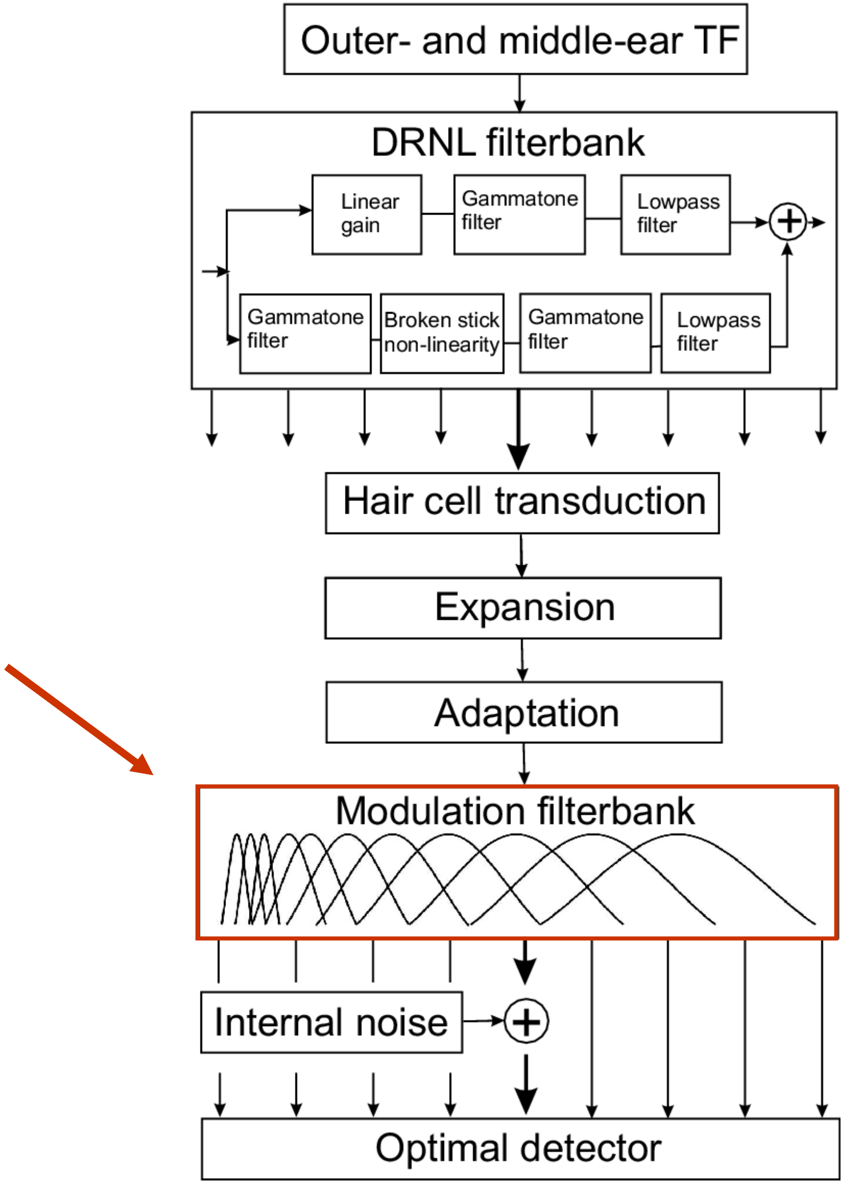
Dau *et al*. (1997);
Ewert & Dau (2000)

Psychophysical data



⇒ Modulation frequency-to-place transformation in the brain.

(1-D) Modulation filterbank model



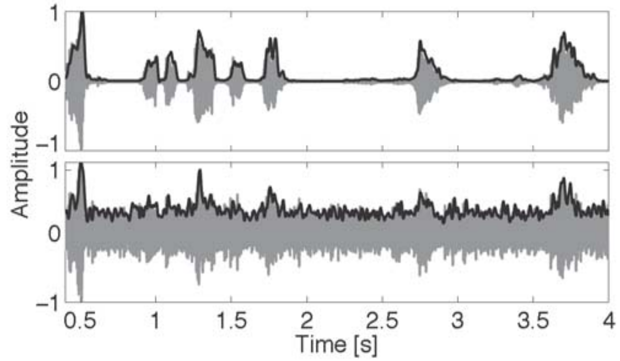
Speech perception:

- Modulation filterbank consistent with concept of speech transmission index (STI) (e.g., Houtgast and Steeneken, 1985).
- RASTA algorithm (Hermansky, 1994) in speech recognition systems: filters out "irrelevant" temporal modulations.

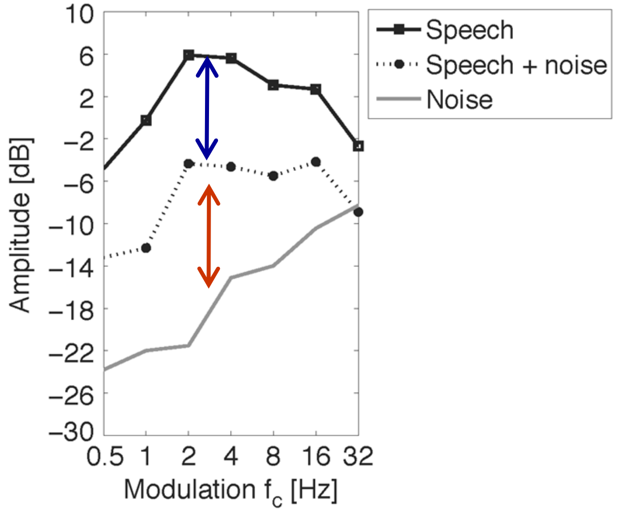
Dau *et al*. (1997); Jepsen *et al*. (2008)

Speech intelligibility prediction: The STI concept

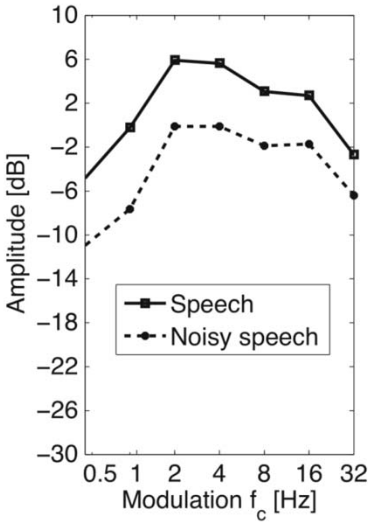
Speech signal
(1/3 oct. filtered @ 2kHz)



Modulation spectra



No processing



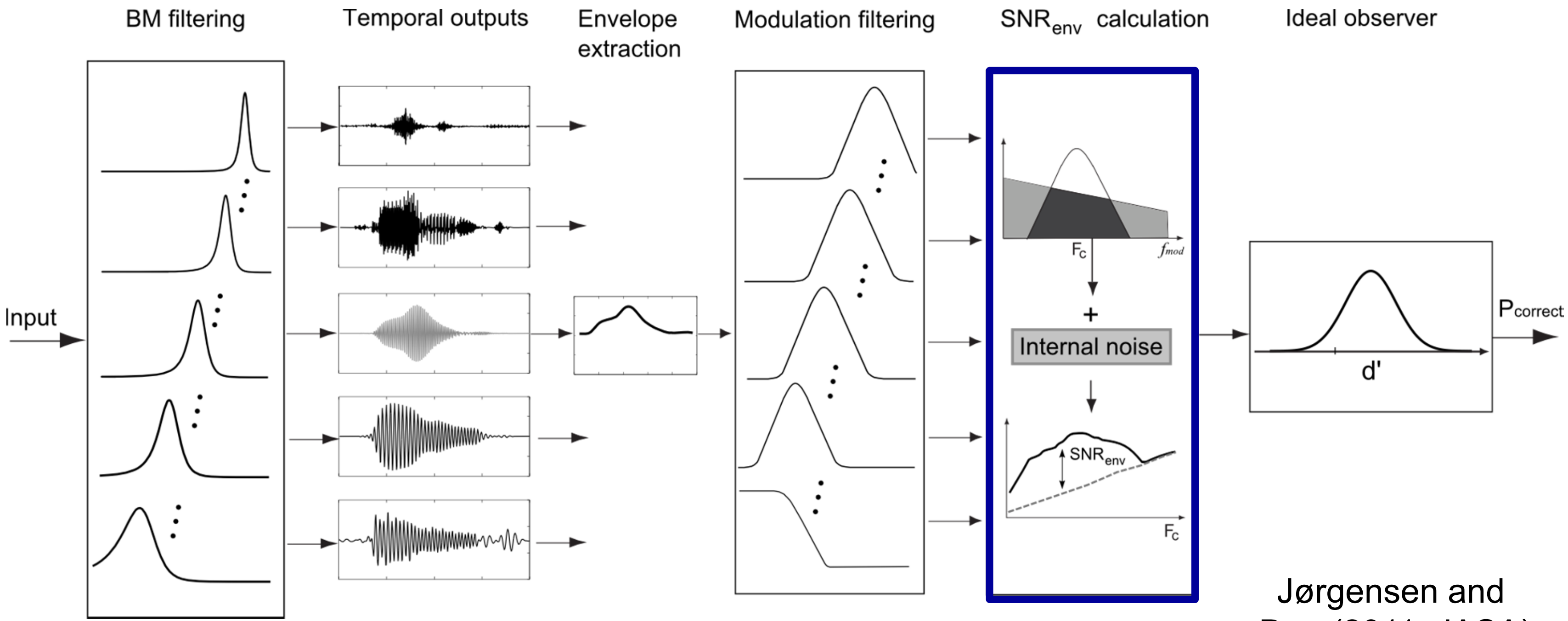
- STI accounts for effects of additive noise.
- Noise reduction (via spectral subtraction) *increases* the SNR (in the audio domain) and the STI.

$\Rightarrow$ Prediction of *increase* in intelligibility.

- However, data typically show a decreased speech intelligibility.

$\Rightarrow$ Noise reduction *paradox*

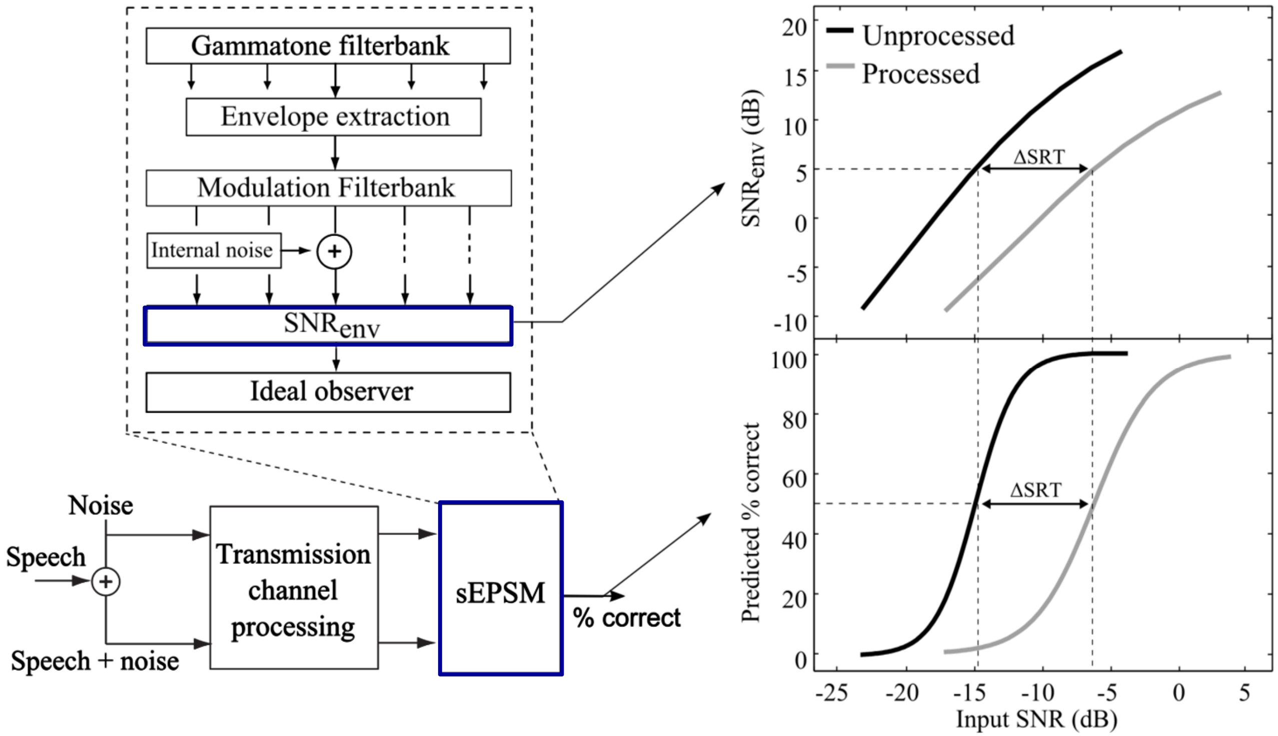
Speech-based envelope power spectrum model (sEPSM)

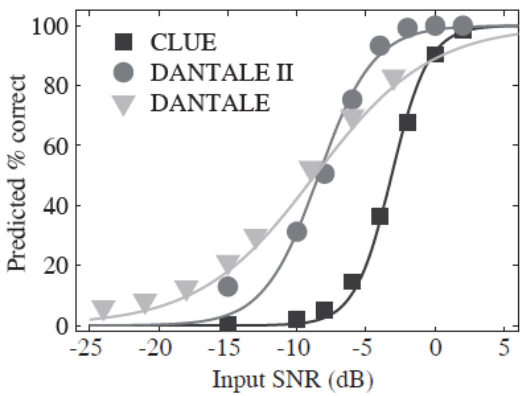


Jørgensen and Dau (2011; JASA)

- Based on the EPSM (Ewert and Dau, 2000) used for prediction of modulation detection and masking.
- Key component: Metric based on the signal-to-noise ratio in the envelope domain (SNR_{env}).

Components of the framework





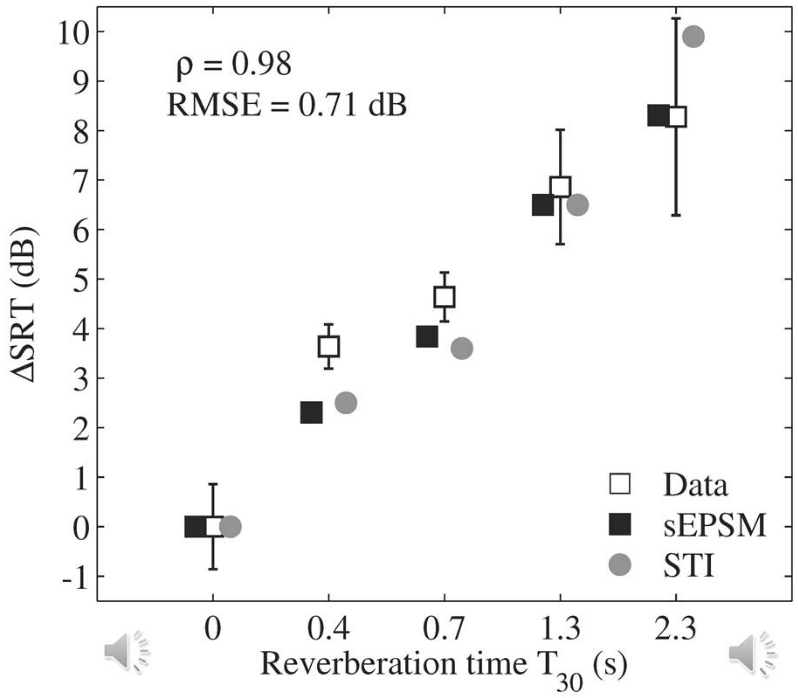
The *ideal observer* makes assumptions about the response alternatives and redundancy (*m*, σ) of a given speech material ⇒ shape of psych. function.

However, the ”key” measure affected by the transmission channel is considered to be SNR_{env}.

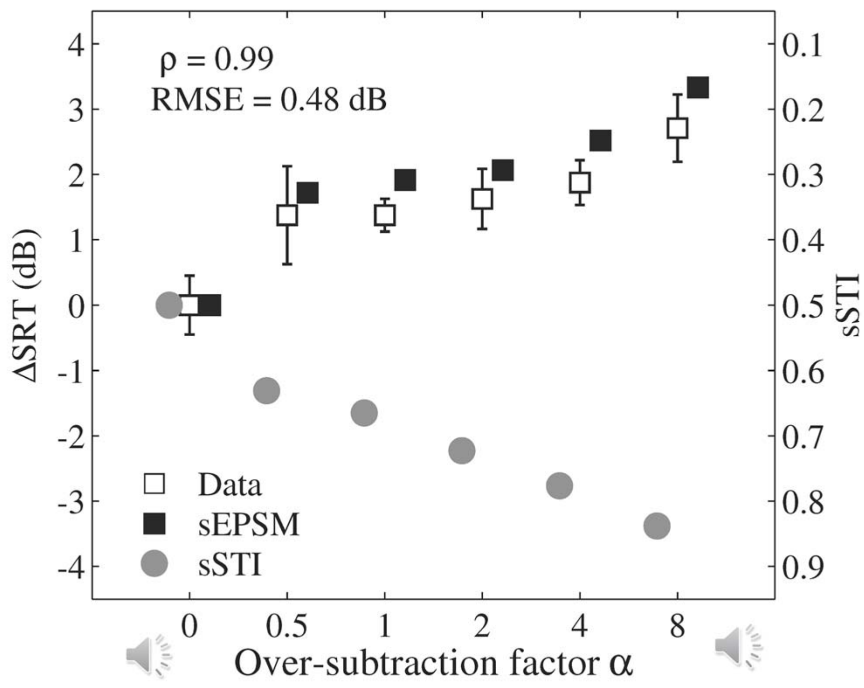
Data and simulations

HEARING SYSTEMS

Reverberation



Spectral subtraction

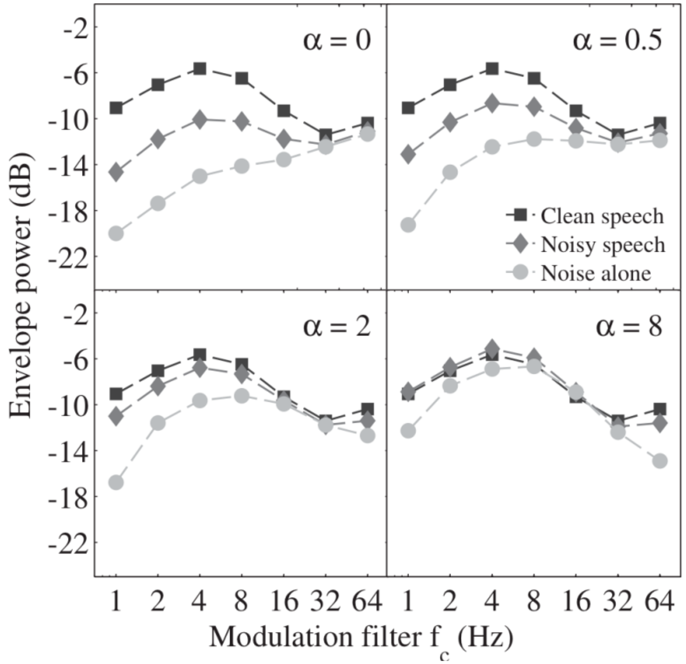


⇒ In conditions of reverberation, STI and sEPSM perform similarly (and successfully).

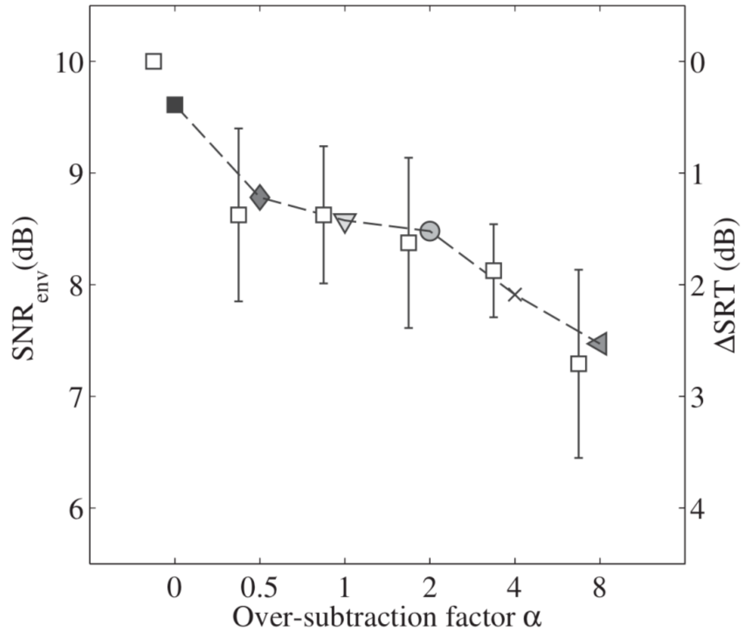
⇒ In conditions of spectral subtraction, the sEPSM accounts for the data while the STI fails completely (Jørgensen and Dau, 2011).

Model analysis
(for spectral subtraction)

Modulation spectra @ 1 kHz



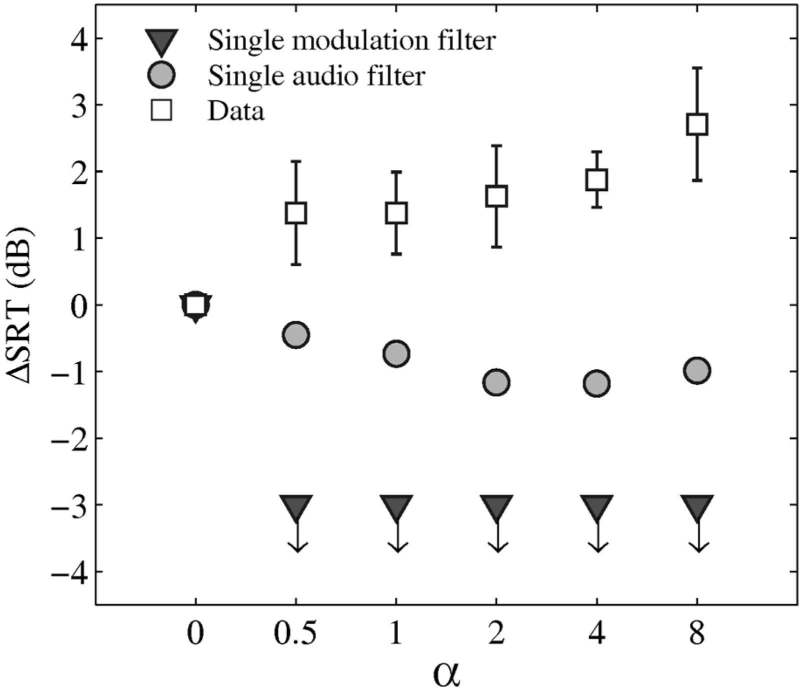
SNR$_{env}$(α)



- No contributions from modulations ≥ 32 Hz (consistent with earlier work).
- The envelope power of the noisy speech increases with α; However, the envelope power of the (estimated) noise floor increases more strongly with α.

⇒ Thus, SNR$_{env}$ decreases with α as does the measured speech intelligibility.

Further model analysis



Is frequency selectivity in audio and envelope frequency domain critical?

Additional simulations with:

– one ”broad” auditory filter

– a 150-Hz modulation LP filter

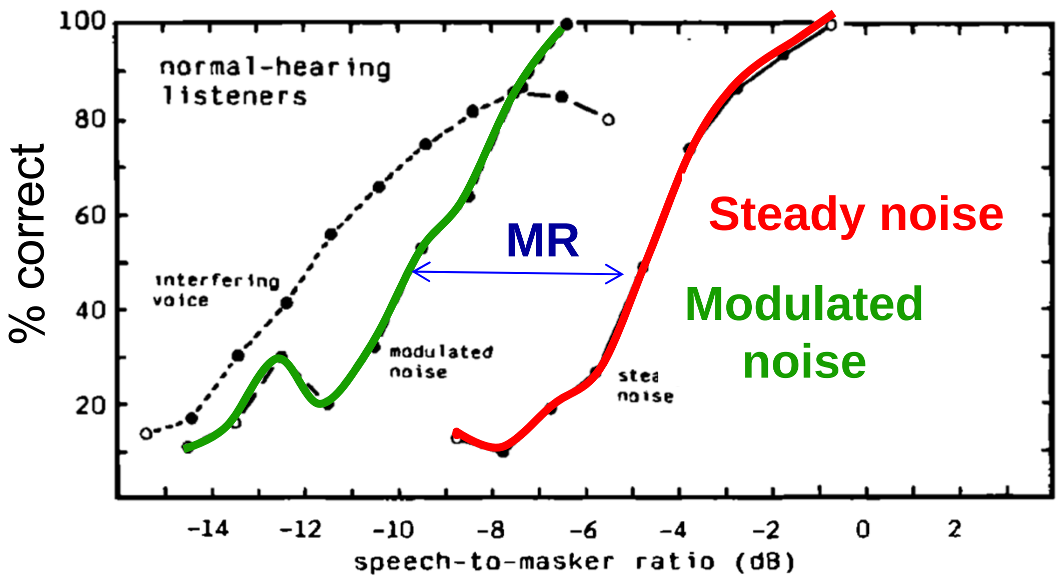
In both cases, the modified model fails to account for the data.

⇒ The integration of SNR_{env} information *after* frequency-selective processing (in both domains) is crucial for speech-intelligibility prediction.

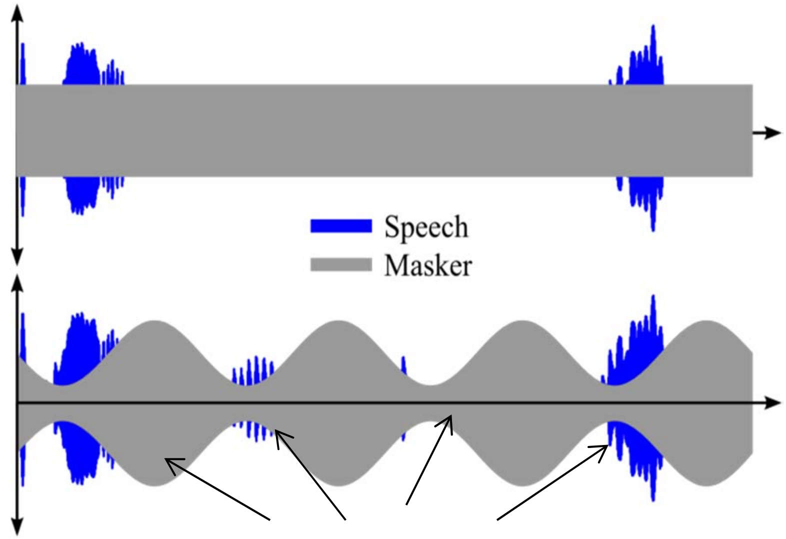
However, the model does not reflect, which modulations contribute at which time and (carrier) frequency to speech intelligibility.

Speech masking release in modulated interferers

Festen and Plomp (1990)



Speech masking release



increased SNR_{env}

⇓

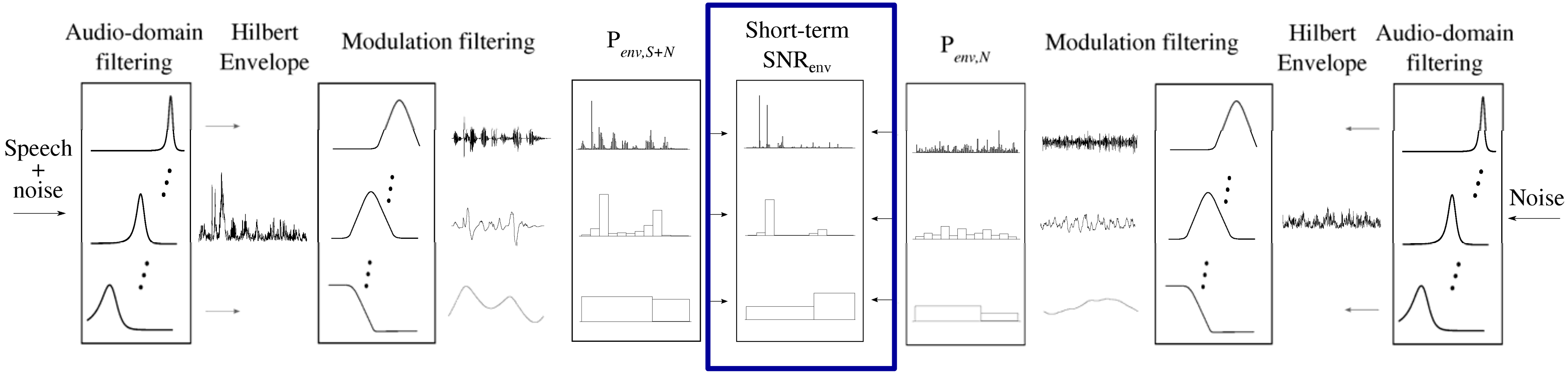
Speech perception in fluctuating noise enhanced compared to a stationary noise interferer

Large masking release (MR) observed due to the ability to ”listen in the dips”.

Hypothesis: The SNR_{env} might be increased in the dips.

”Short-term” SNR_{env} calculation required

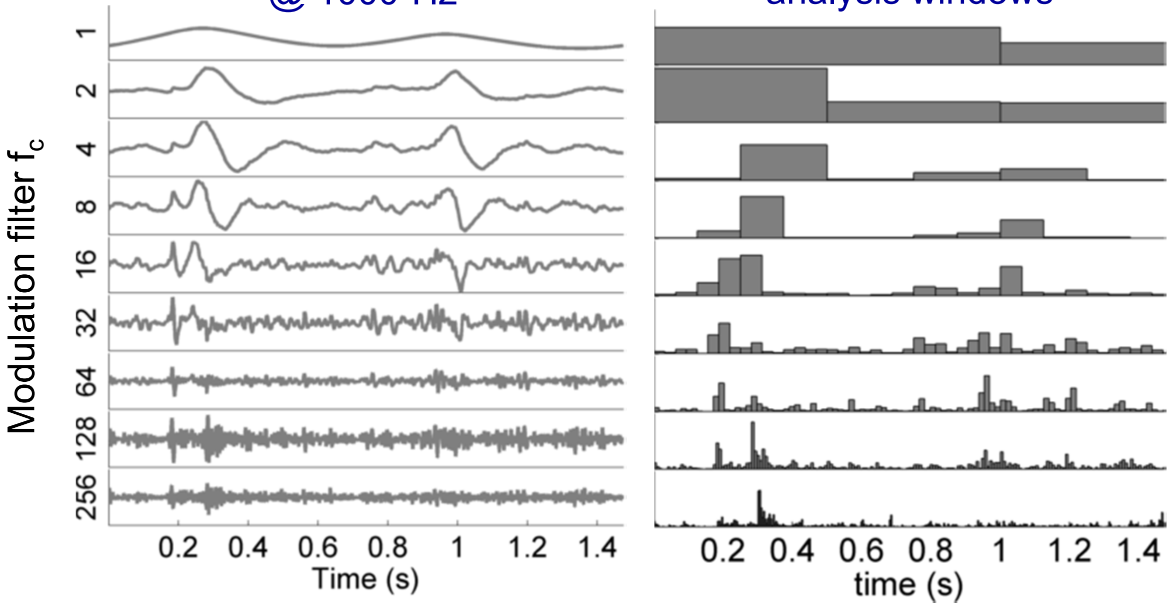
“Multi-resolution” sEPSM



(Jørgensen, Ewert and Dau) JASA 2013

Internal representation @ 1000-Hz

Channel-dependent analysis windows



Envelope power calculation

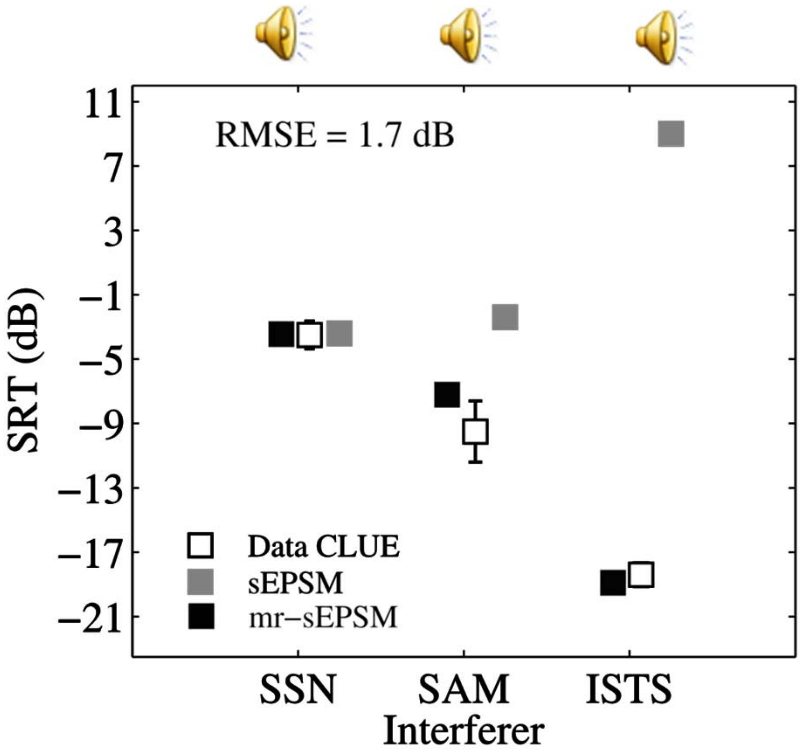
(with $\tau \sim \frac{1}{\Delta f}$)

Modulated-noise and speech-like interferers

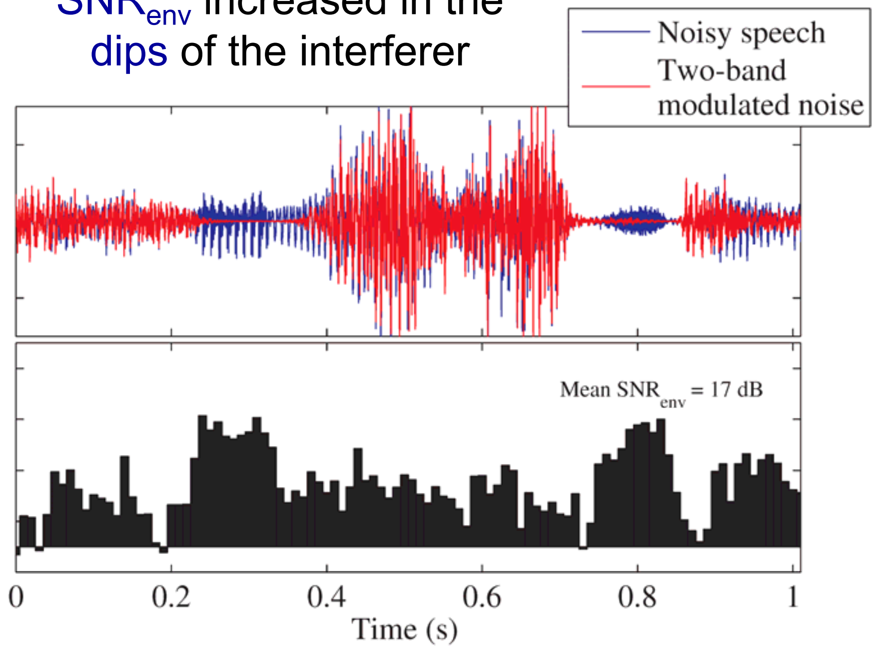
Three conditions: SSN - as the stationary reference condition

SAM - 8-Hz modulated noise (e.g., Festen, 1987)

ISTS - International speech test signal (Holube *et al*., 2010).



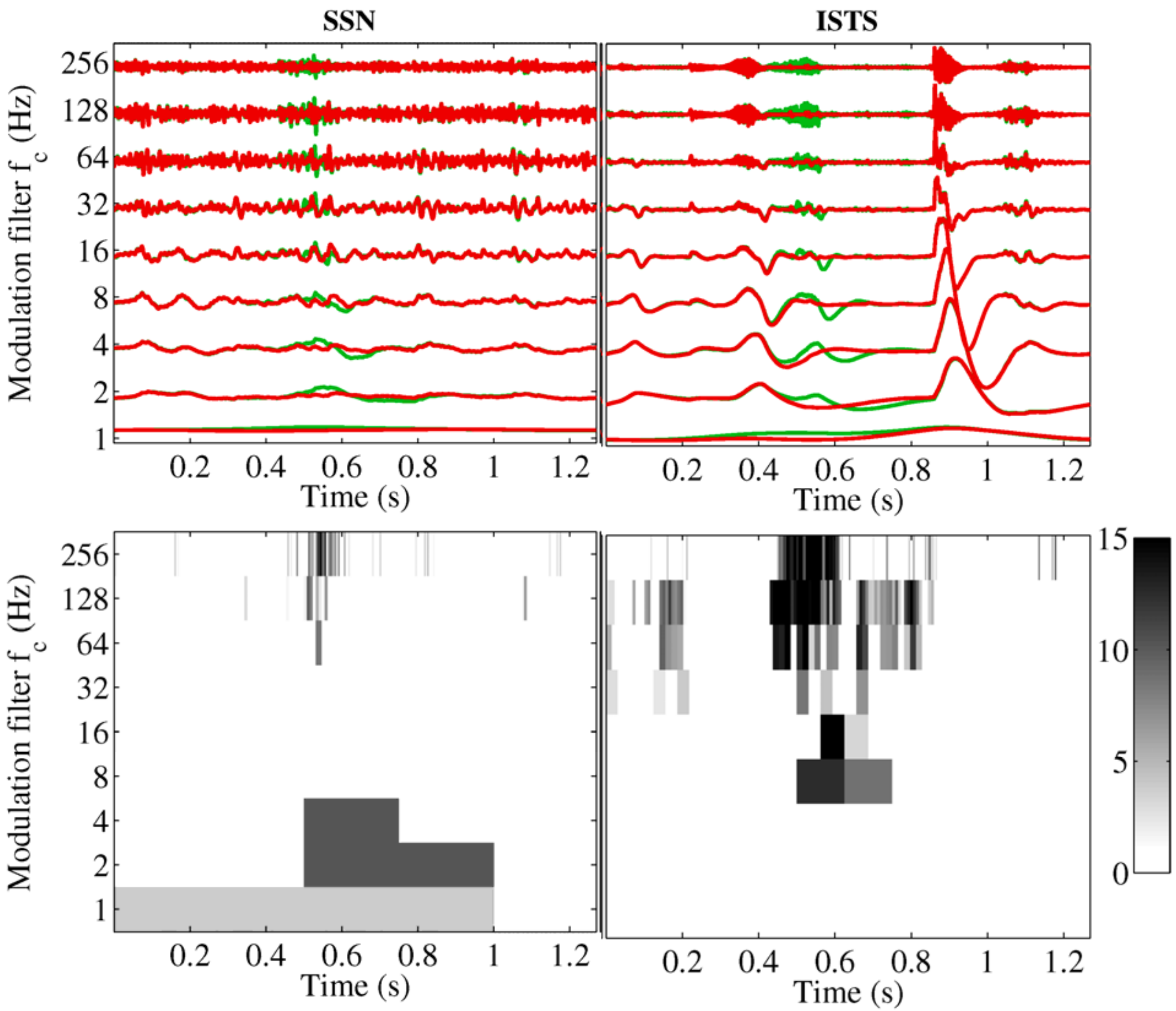
SNR_{env} increased in the dips of the interferer



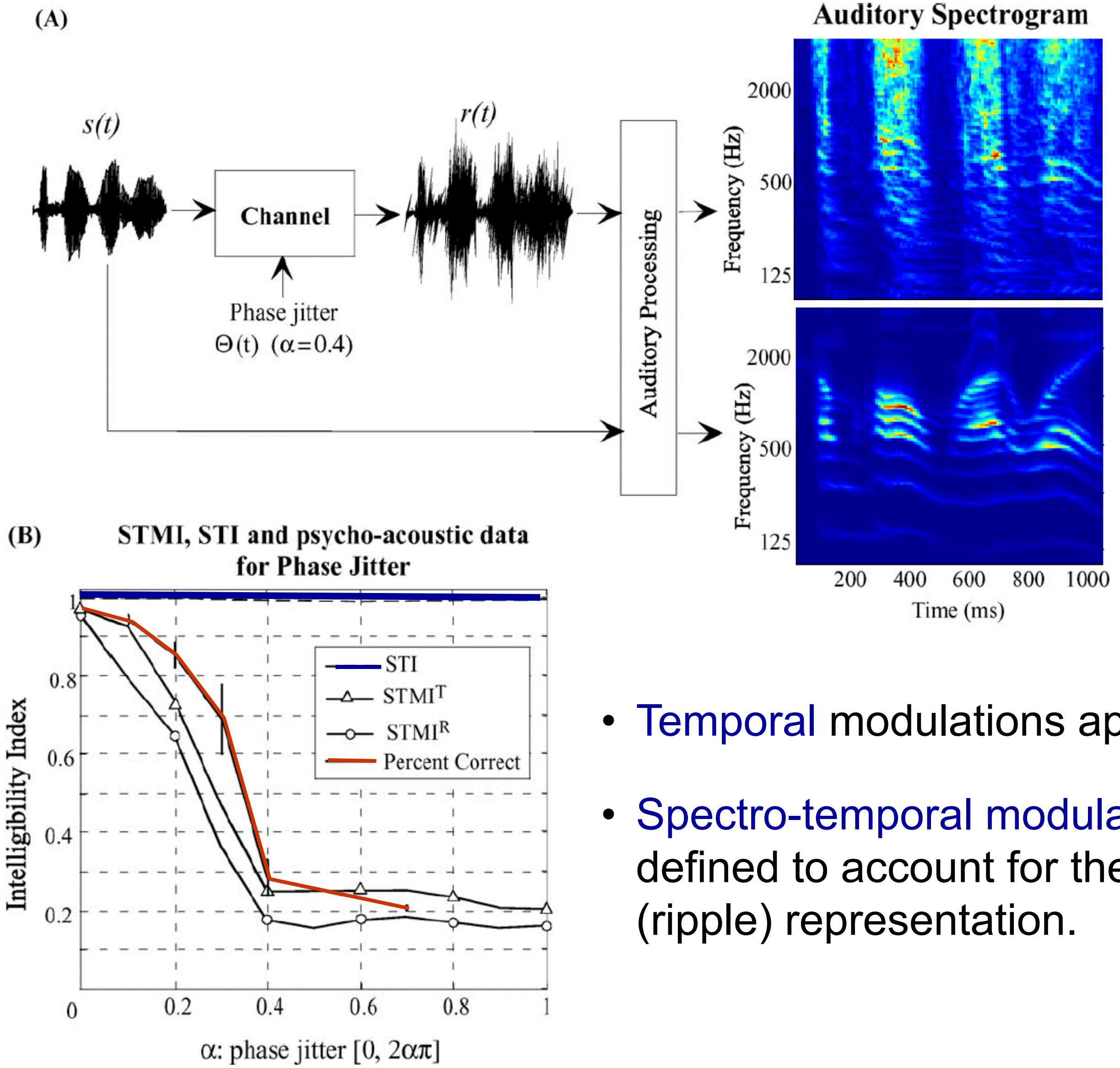
In fact, the simulations suggest that high-frequency modulations (>30 Hz) contribute effectively to speech intelligibility in the case of the SAM and ISTS interferers.

The role of fast modulations

Model-output from audio-filter @1 kHz



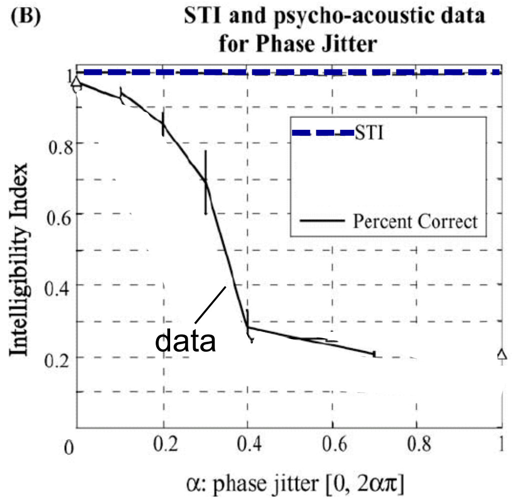
Another challenge:
Phase jitter (nonlinear) distortion



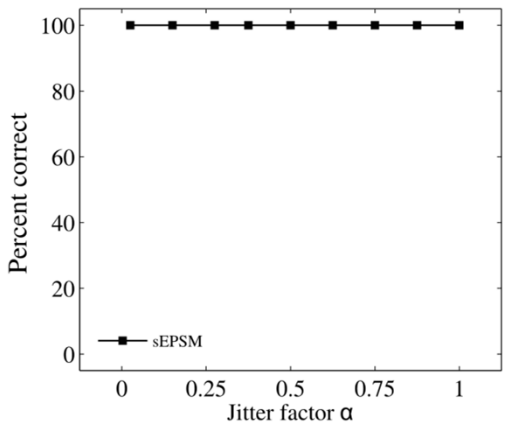
Elhilali *et al.* (2003)

- Temporal modulations appear not to be sufficient.
- Spectro-temporal modulation index (STMI) was defined to account for the degraded spectral (ripple) representation.

Phase jitter distortion



sEPSM predictions

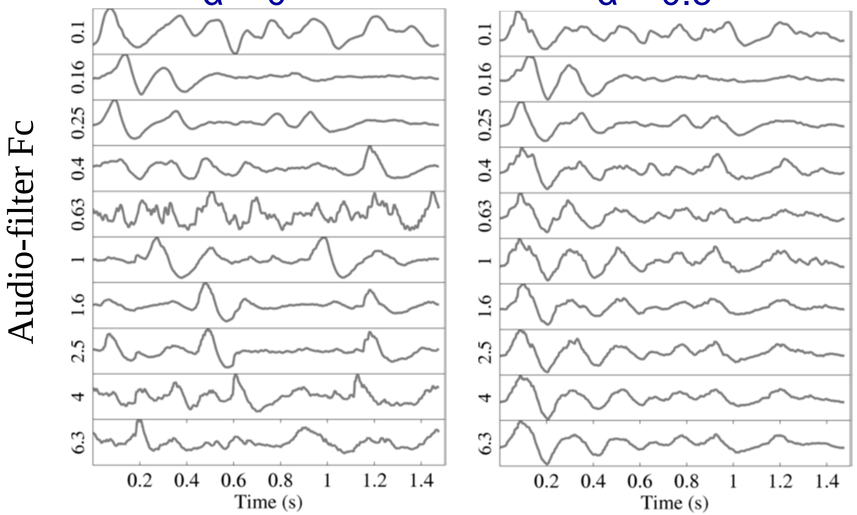


As the STI, the sEPSM fails in this condition.

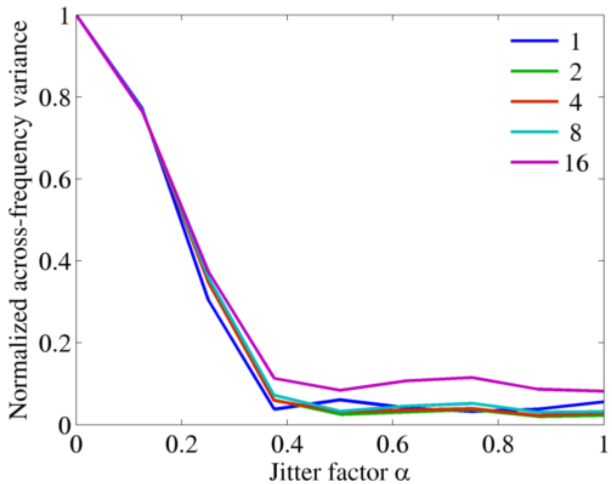
Both models are insensitive to across-frequency distortions

Internal representation
at output of 4-Hz modulation filter

α = 0 α = 0.5



⇒ Across audio-frequency variance decreases with increasing phase jitter



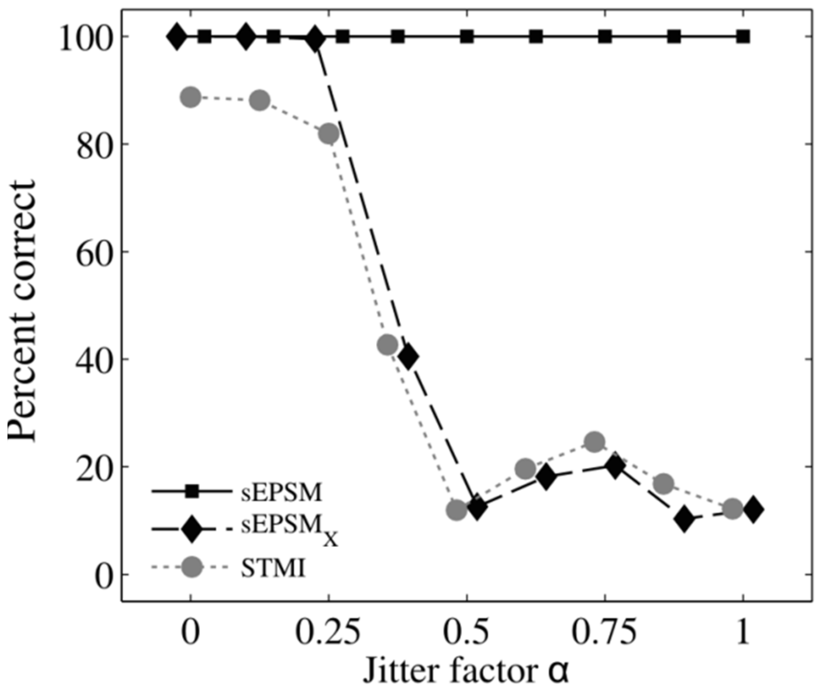
Modulation-filter CF

Across-frequency variance

Weigthing of SNR_{env} by across-channel variance (Chabot-Leclerc *et al.*, 2015):

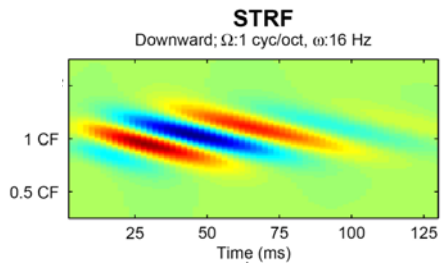
⇒ $sEPSM_x$ predictions (incl. weighting) provide good agreement with measured data

⇒ Similar results as with STMI predictions by Elhilali *et al*. (2003)



- Conceptually related to correlation of neural activity across sensory channels that has been proposed in connection to auditory streaming (Elhilali *et al*., 2009) and CMR (Piechowiak *et al*., 2007).

- Different from STRF concept on which STMI is based.



Recent focus points

Current focus:

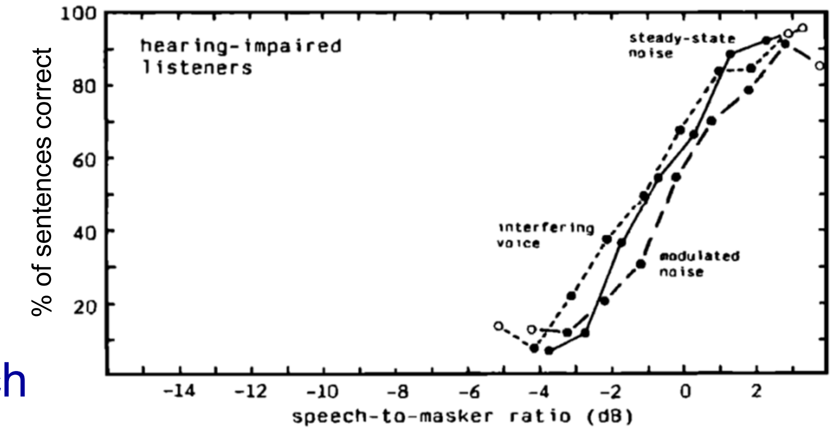
- Expansion towards two ears: Prediction of spatial release from masking due to ”true” binaural unmasking vs ”better-ear” listening (Chabot-Leclerc *et al*., 2016).
- Combination of modulation-based preprocessing with correlation-based decision metric (Iborra *et al*., 2017). (Back to the template-matching approach)?
- Prediction of consequences of hearing loss on speech intelligibility (e.g., consequences of IHC vs. OHC loss).
- Analysis of ”distortion vs attenuation” component of a hearing loss (Plomp, 1986) in a modeling framework.
- …

Implications and applications

Hearing-impaired listeners often show highly reduced speech masking release.

Fast envelope fluctuations may be inaudible or distorted.

Interesting input for models of impaired speech perception and the evaluation of hearing aids.



Festen and Plomp (1993)

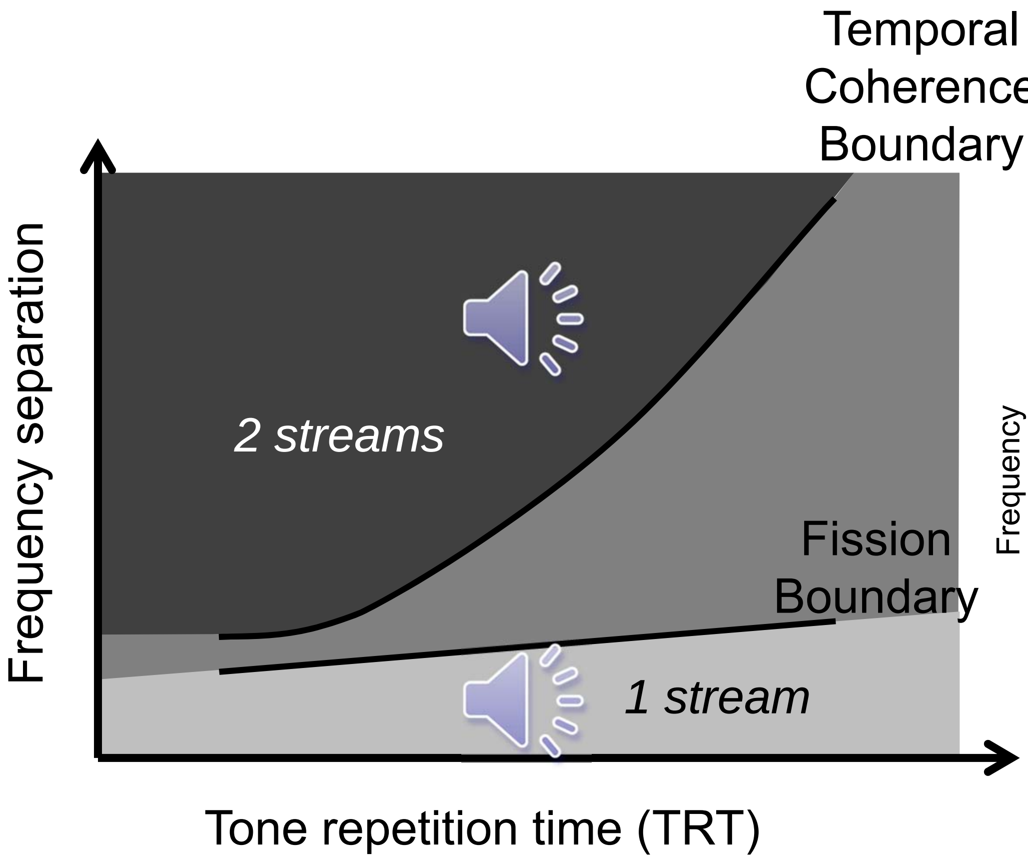
The model is based on the concept of modulation masking. It even accounts for conditons with interfering talkers – often associated with ”informational masking”.

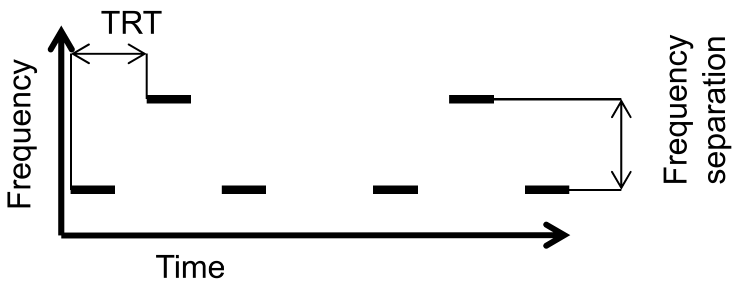
However, the model fails if it does not have *a priori* information about the signal and the masker. It cannot provide stream segregation.

Unfortunately, stream segregation is one of the major challenges of hearing-impaired people.

Towards a model of stream segregation

Example of stream segregation due to frequency separation



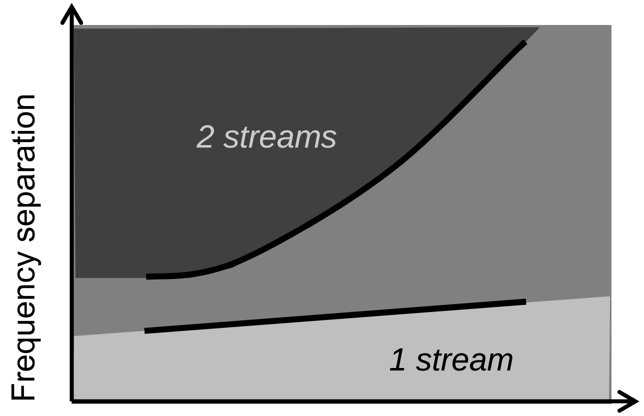


(van Noorden, 1975)

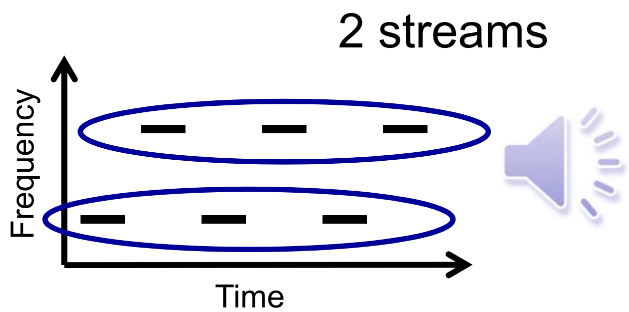
Grouping due to synchrony

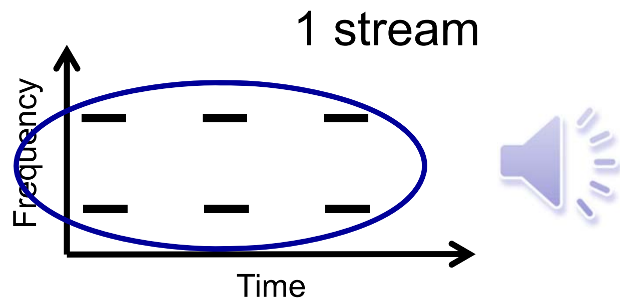
- Tones with a sufficiently large frequency separation always split into separate streams (van Noorden, 1975).

- Unless the tones are synchronized! Then they merge into a single stream despite tonotopic separation (Bregman, 1990; Elhilali *et al*., 2009).



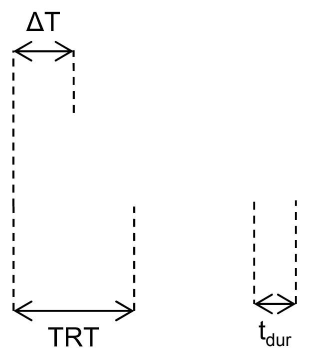
(van Noorden, 1975)

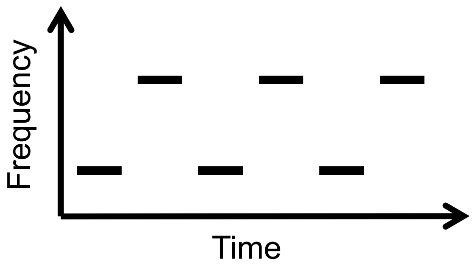




Grouping due to synchrony

- How strict is the "synchrony" grouping mechanism?
- Can the tones be slightly asynchronous ($\Delta T \neq 0$) and still be fused?
- Experiment investigating the influence of:
  - Tone repetition time (TRT)
  - Tone duration (t_{dur})

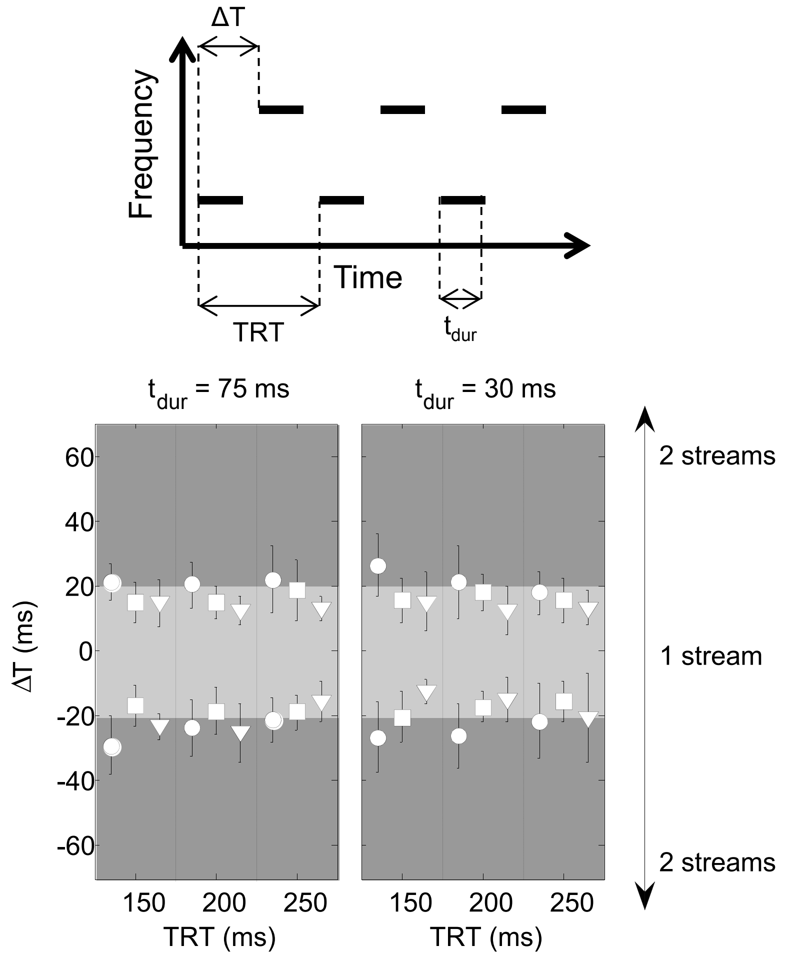




Grouping due to synchrony

Results:

- The tones *can* fuse together without perfect synchrony ($\Delta T = 0$).
- Fusion occurs if the asynchrony is less than ~ 20 ms.
- No significant effect of TRT and t_{dur}.

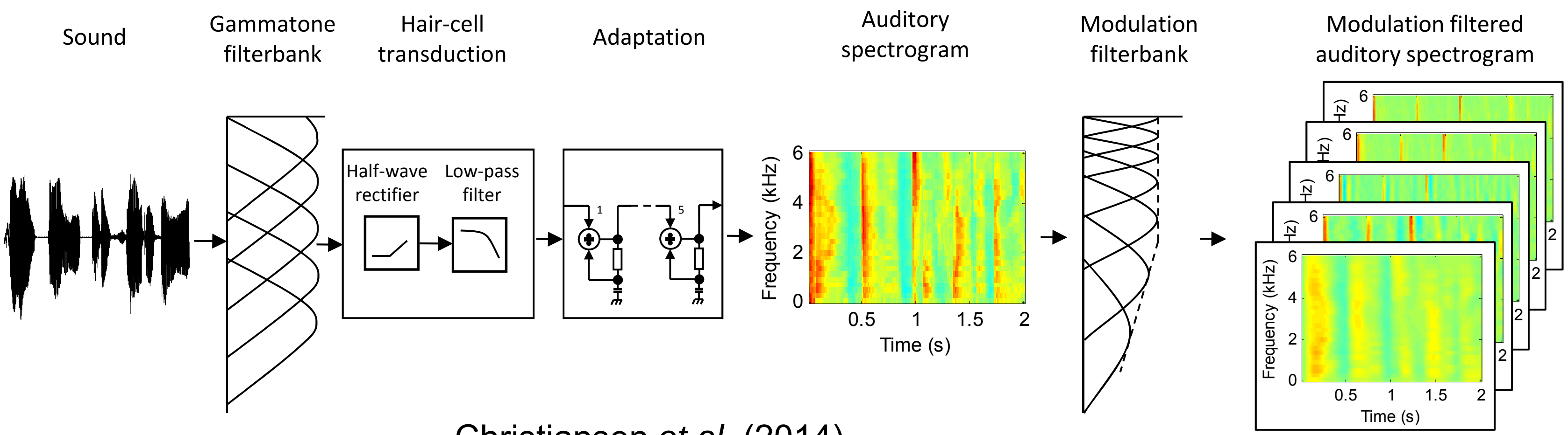


(Christiansen *et al*., 2014; JASA)

Modelling auditory stream segregation

Step 1: Decomposition of acoustic stimuli into a collection of sensory elements (following the concepts of Bregman, 1990):

- Using a physiologically inspired model of the auditory periphery, CASP (e.g., Dau *et al*., 1997).



Christiansen *et al*. (2014)

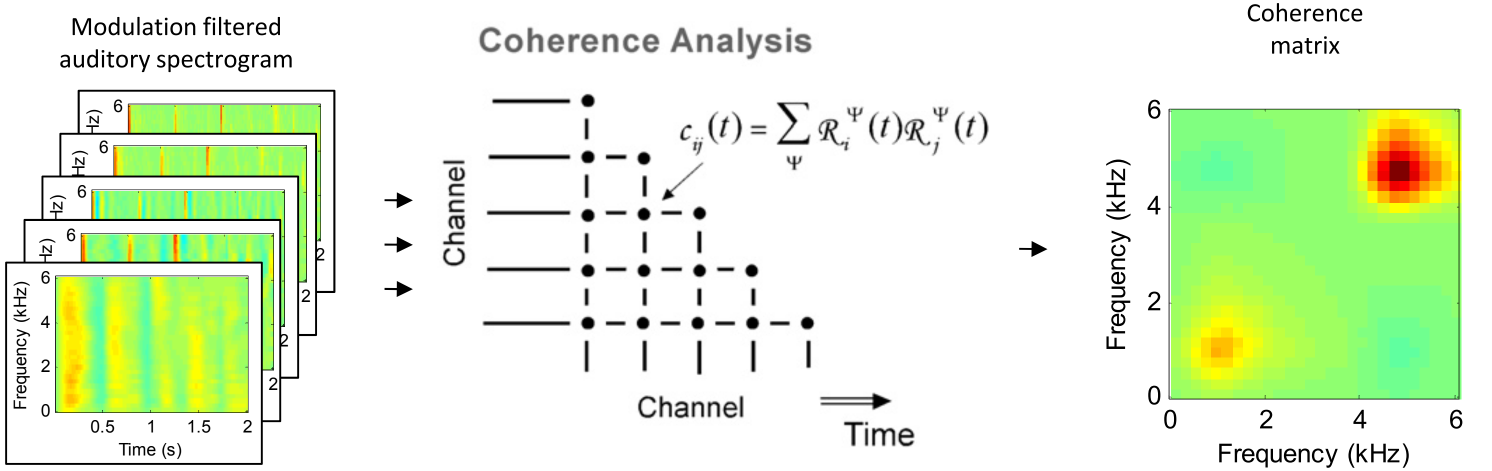
- The model has earlier been evaluated in various conditions of spectro-temporal masking in the auditory system.

Modelling auditory stream segregation

Step 2: Grouping of sensory elements:

- Based on synchrony of neural activity (including conditions with distant spectral components).

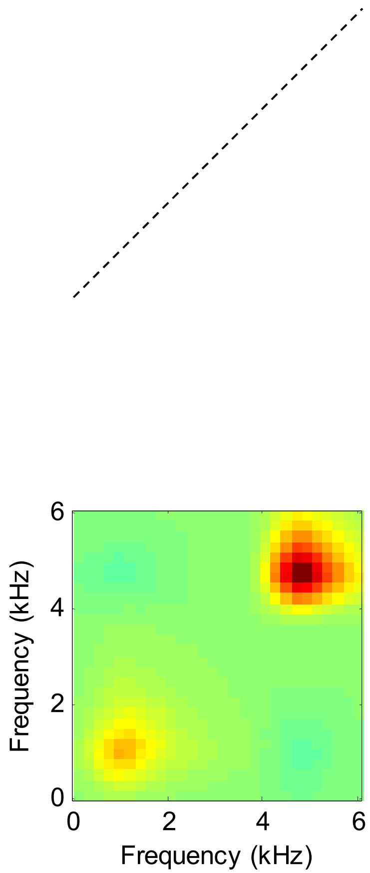
- Utilizing a correlation process across tonotopic channels, e.g., a ”temporal coherence analysis” similar to Elhilali *et al.* (2009).



- Correlation between each pair of frequency channels ⇒ Dynamic coherence matrix C that evolves over time.

Analysis of model output

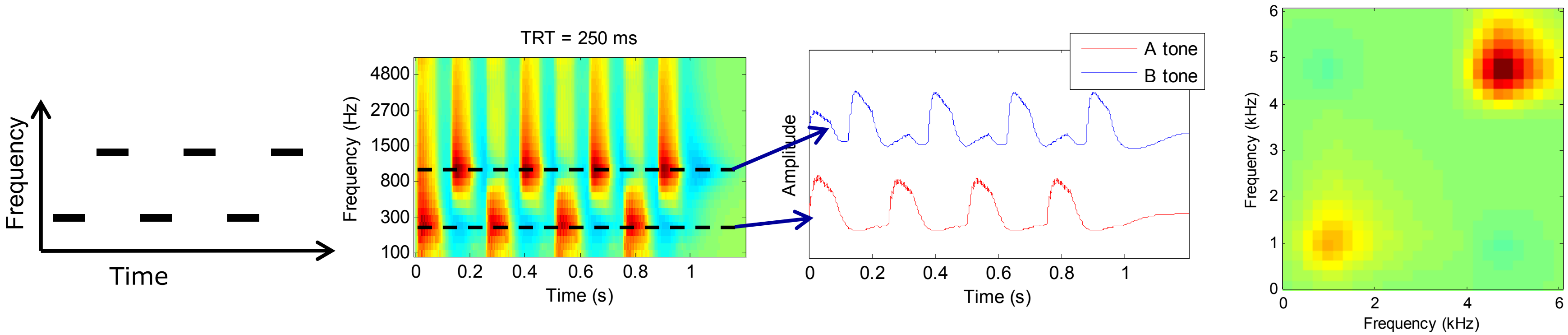
- Diagonal entries of the matrix shows the correlation of a given peripheral channel with itself.
- Off-diagonal entries reflect correlation across separate channels.
- To quantify the coherence matrix, an eigenvalue decomposition is performed.
- The ratio of the second largest to the largest eigenvalue (λ_2/λ_1) shows the strength of the ”two-stream percept”.



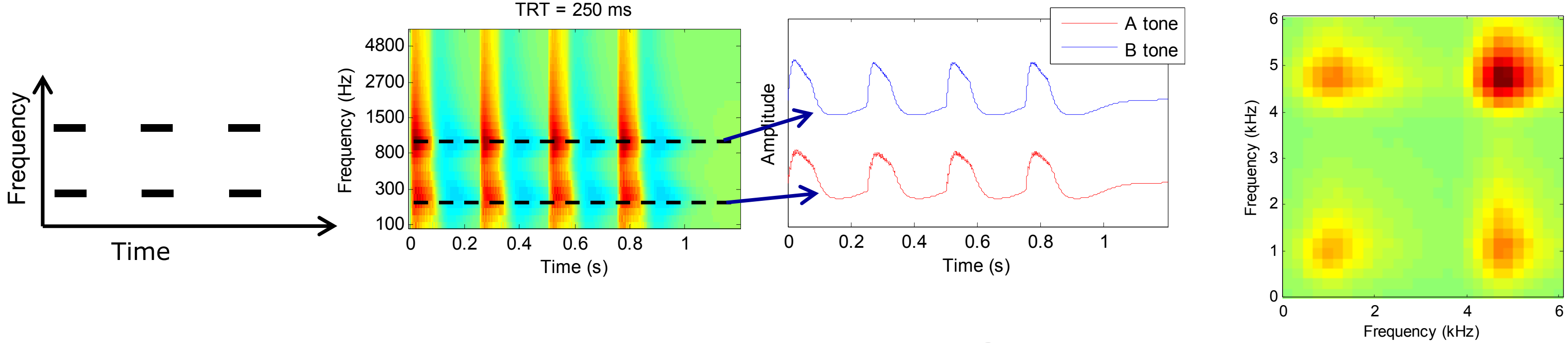
$\frac{\lambda_2}{\lambda_1} \approx 0$ => Only 1 stream

$\frac{\lambda_2}{\lambda_1} \gg 0$ => At least 2 streams

Synchrony simulation



$$\frac{\lambda_2}{\lambda_1} \approx 0.46 \gg 0$$ => 2 streams

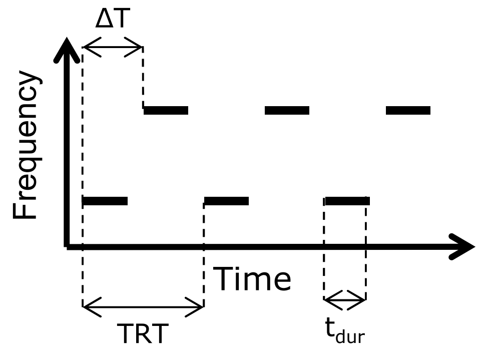


$$\frac{\lambda_2}{\lambda_1} \approx 0.08 \approx 0$$ => 1 stream

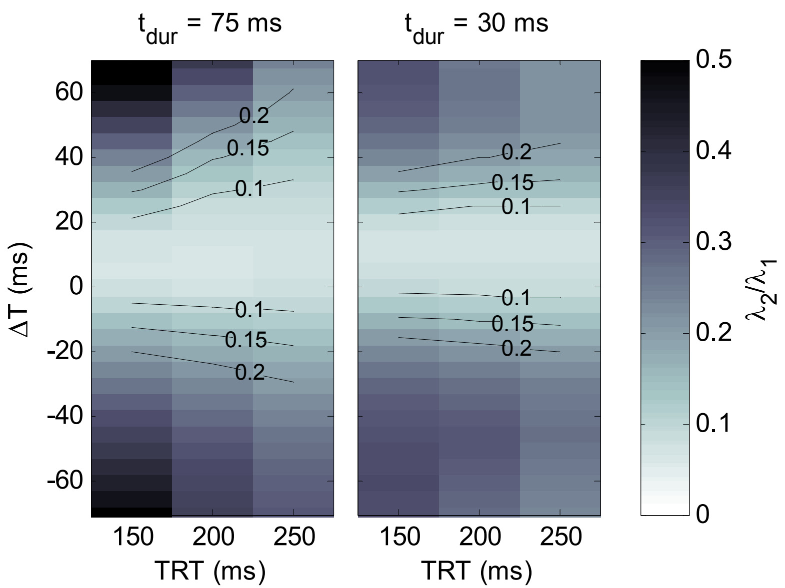
Synchrony simulation

Applying the model on the same experimental setup as used in the psychoacoustic experiment:

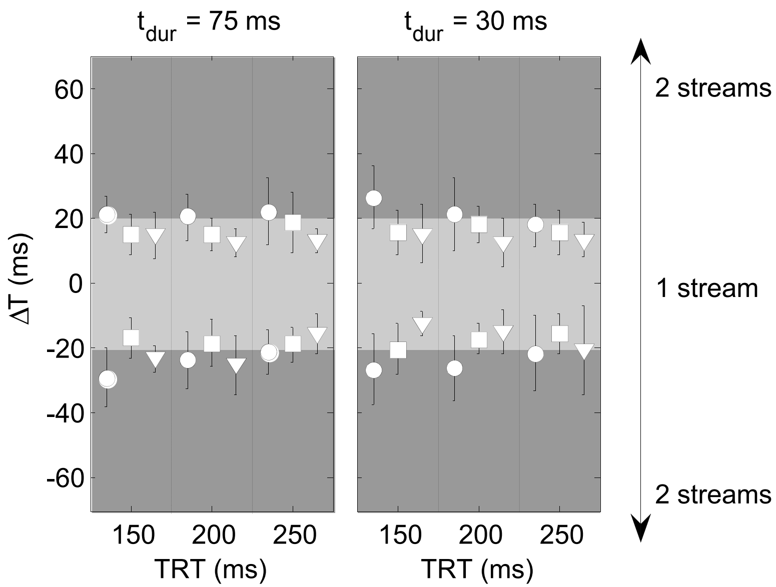
- Similar overall behaviour
- However, the model shows some dependency on TRT



Model



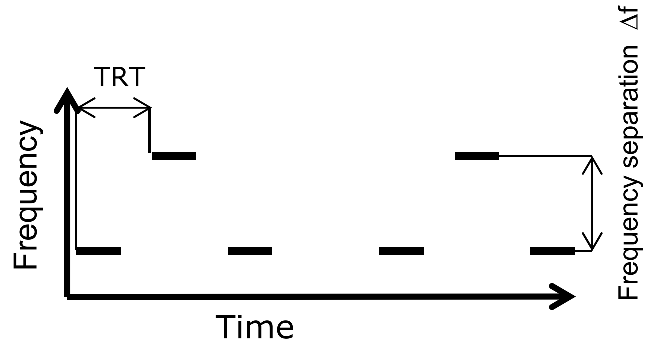
Data

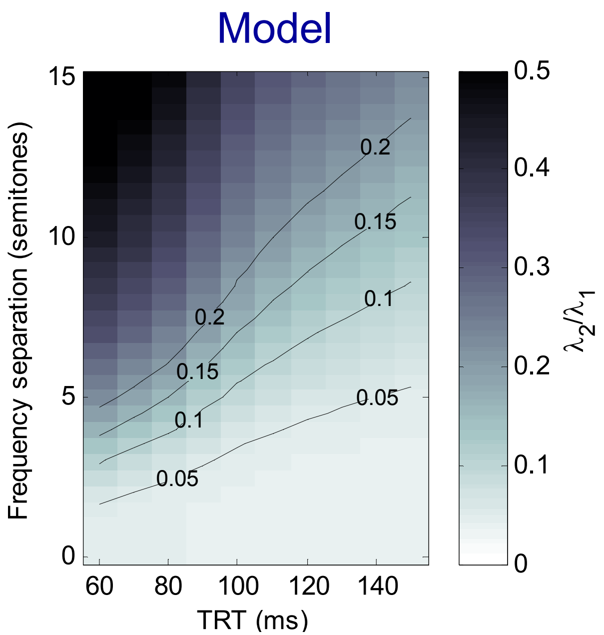


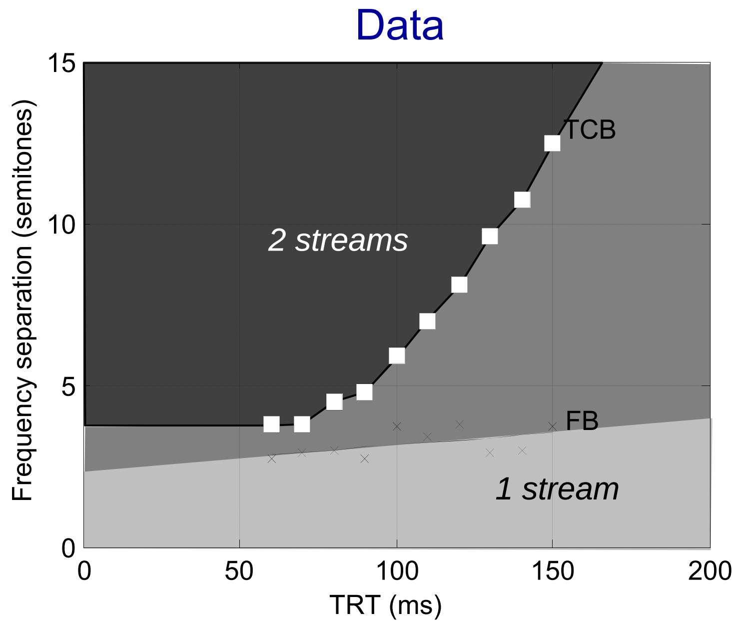
Van Noorden simulation

Applying the model to the classical van Noorden stimuli:

$\Rightarrow$ Model accounts for the dependency of segregation/fusion on TRT and Δf.



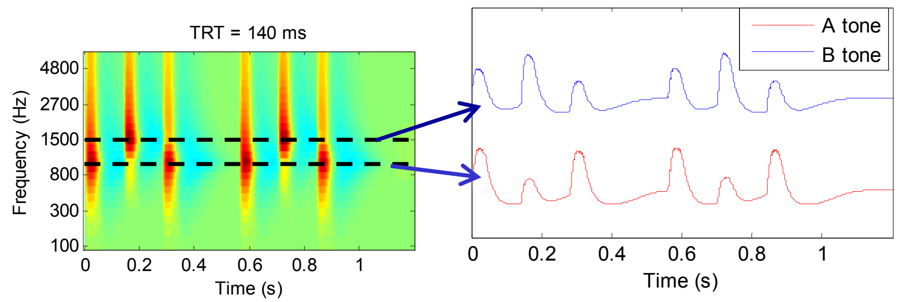




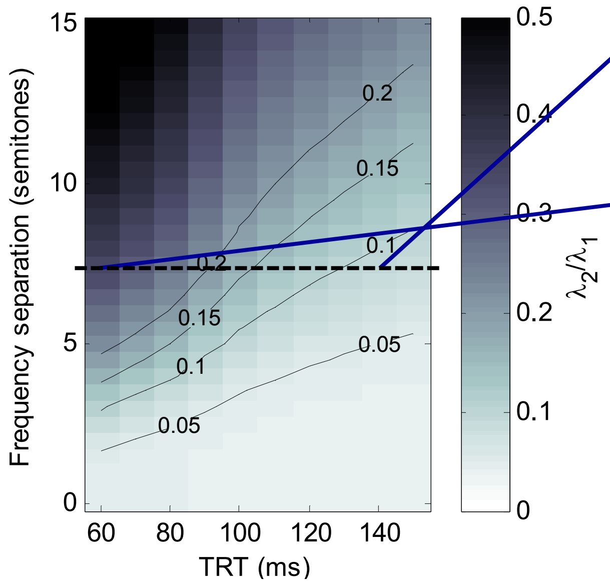
(Christiansen *et al*., 2014; JASA)

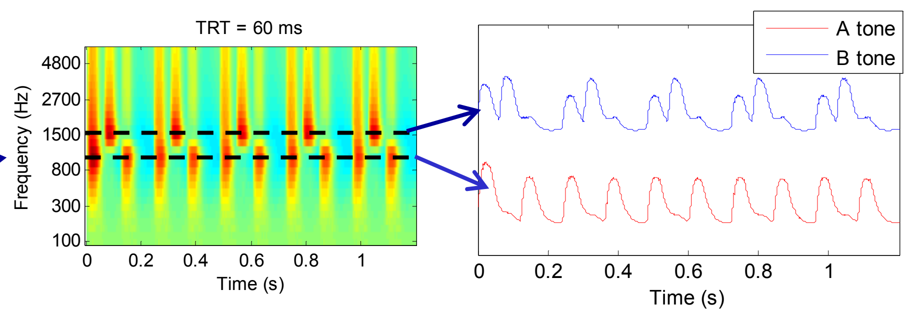
Van Noorden simulation

- Spread of excitation causes a high cross-frequency correlation
- For short TRTs: forward masking reduces spread of excitation
  - ➢ reduced cross-frequency correlation



High correlation → One stream





Low correlation → Two streams

Consistent with physiological studies (e.g., Bee and Klump, 2005).

Implications and applications

- Temporal coherence may be the organizing principle behind primitive stream segregation.
- However, this requires an appropriate preprocessing (i.e., realistic frequency-selective filtering in the cochlea; an adaptive process accounting for forward masking and onset enhancement, and a modulation filterbank).
- The concept of coherence may be generalizable to other sensory channels (e.g., binaural processing).
- A model of auditory stream segregation might be useful for:
    - Source separation algorithms (ideally performing as well as NH listeners)
    - Classification of the number of sources in complex acoustic scenes
    - Evaluation of hearing-aid processing (e.g., does the processing “corrupt” acoustic cues necessary for stream segregation?)

Overall discussion

- Modeling can be helpful to test specific hypotheses. It allows to quantify the effects of individual components in the framework.
- The presented examples have highlighted several features that seem important for robust auditory signal analysis.
- Despite the different methods and ”outcome measures” illustrated here, similar features and processes were found to be essential.
- Combination of approaches might be interesting; e.g. to model target-interferer confusion in stream segregation could help interpret the speech-in-noise problem for the hearing impaired.
- Some of the model insights might be useful for applications:
  - mostly regarding objective evaluations of the effects of hearing-instrument processing (“analysis approach”).
  - and maybe less in terms of applying the signal-processing directly in compensation strategies