The Sensitivity Matrix

Integrating Perception into the Flexcode Project

Jan Plasberg

Flexcode Seminar Lannion
June 6, 2007

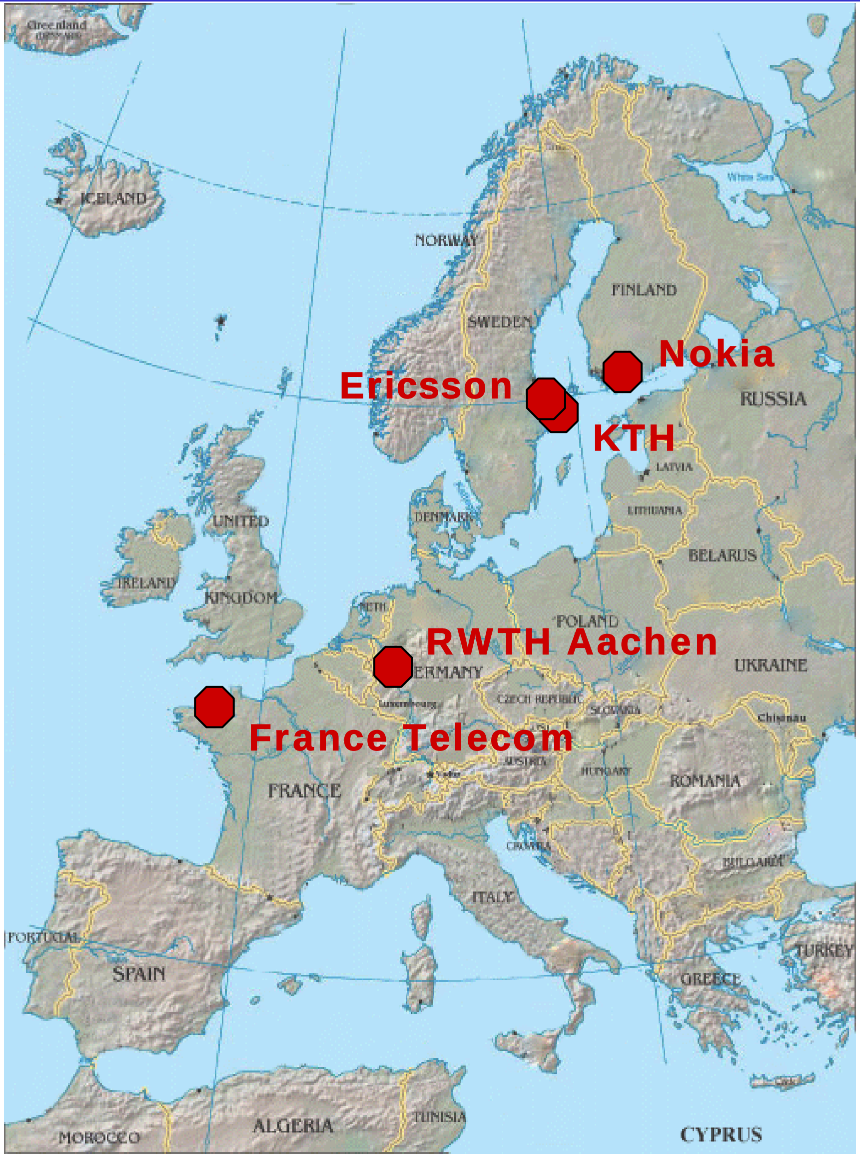
FlexCode

Outline

- Flexcode in a Nutshell
- The Sensitivity Matrix
- Coding with the Sensitivity Matrix
- Open Issues

FlexCode

Who?

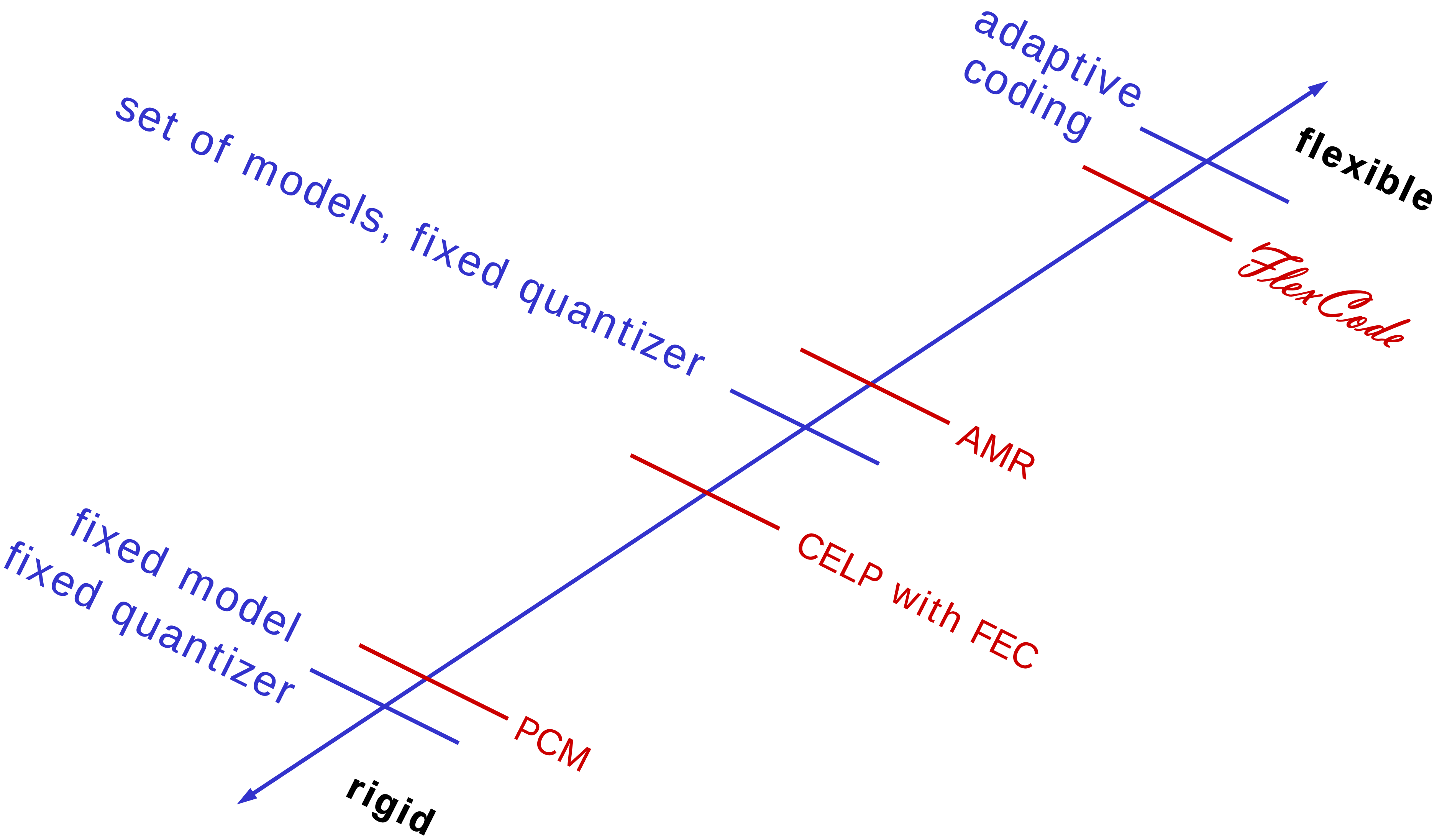


The Problem

- Heterogeneity of networks increasing
- Networks inherently variable (mobile users)

- But:
  - Coders not designed for specific environment
  - Coders inflexible (codebooks and FEC)
  - Feedback channel underutilized

Adaptation and Coding



The Tools

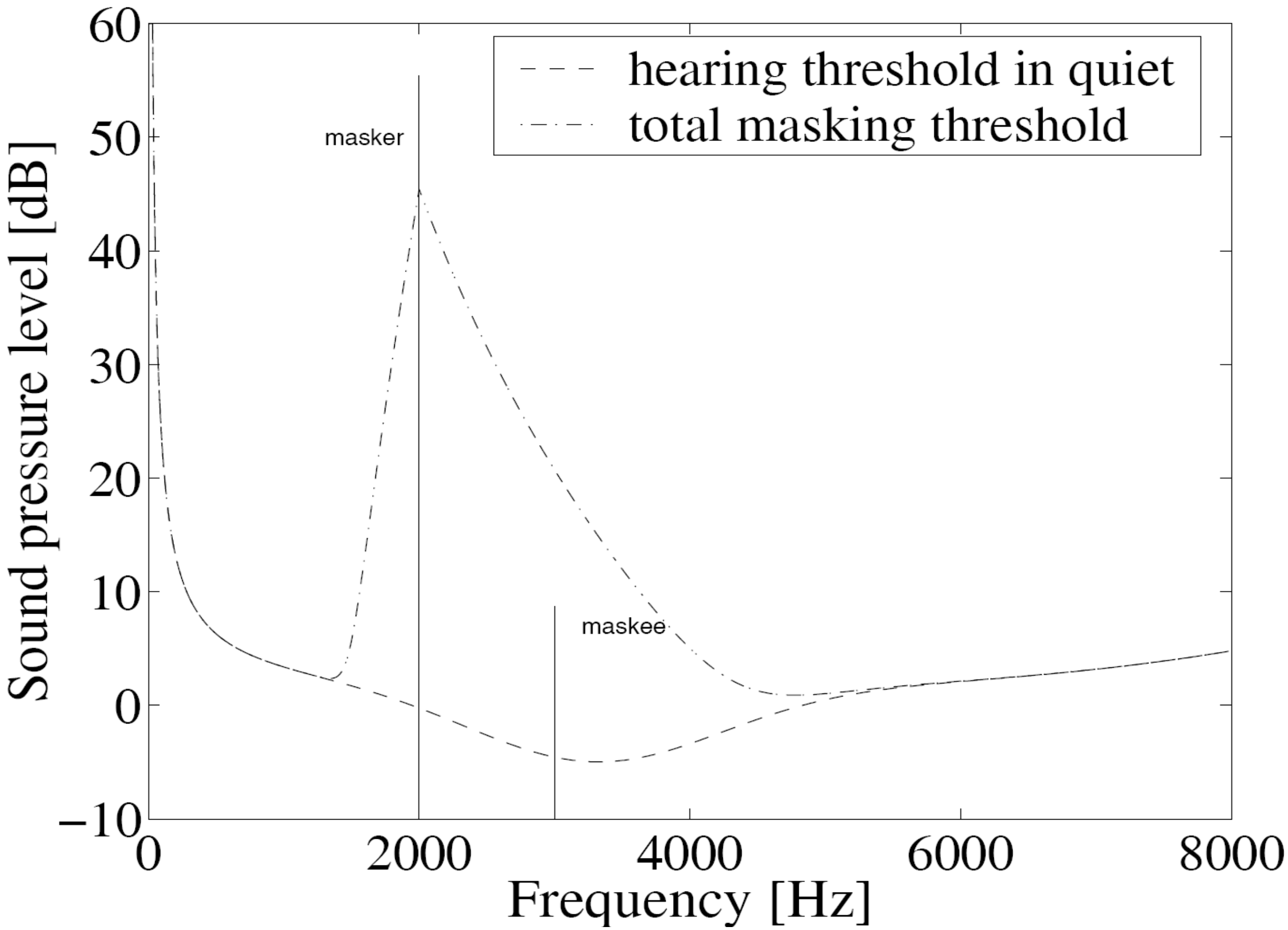
- Tools include
  - Models of source, channel, receiver
  - High-rate quantization theory
  - Multiple description coding (MDC)
  - Iterative source-channel decoding
  - Distortion measures using the sensitivity matrix

Outline

- Flexcode in a Nutshell
- The Sensitivity Matrix
  - Why, what and how?
  - What can it tell us?
- Coding with the Sensitivity Matrix
- Open Issues

Human Auditory Masking

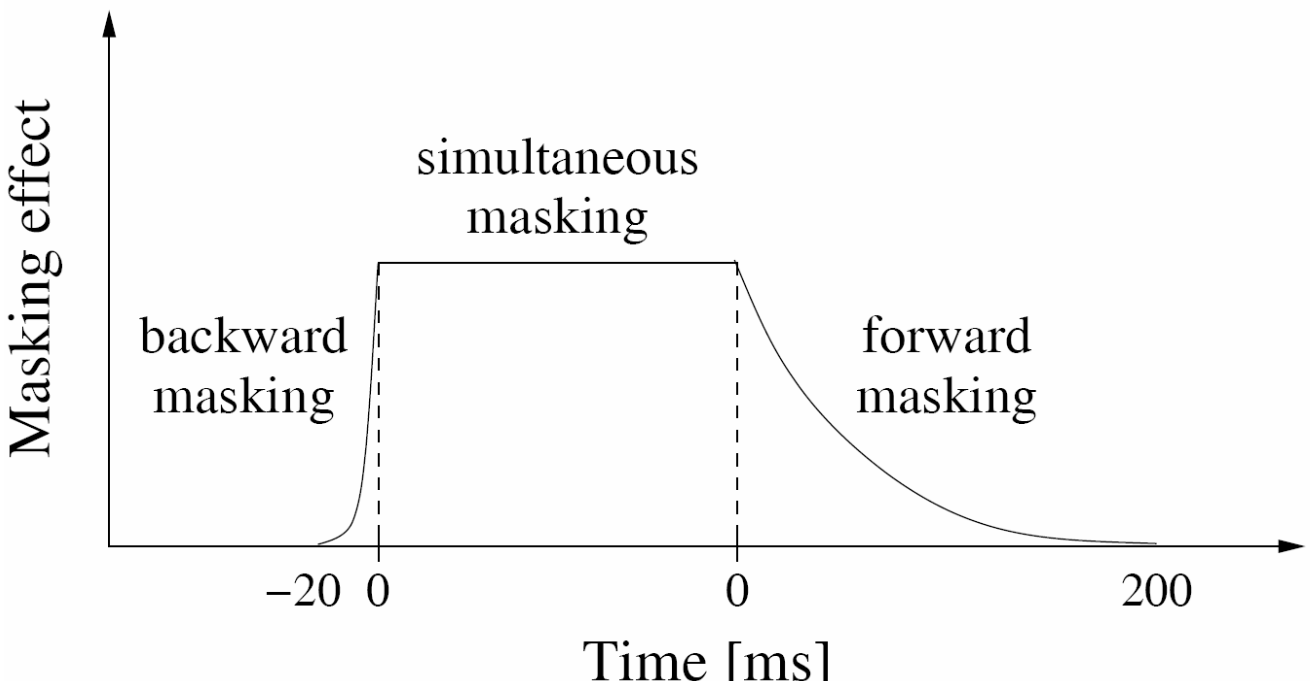
- Simultaneous masking (frequency masking)



FlexCode

Human Auditory Masking

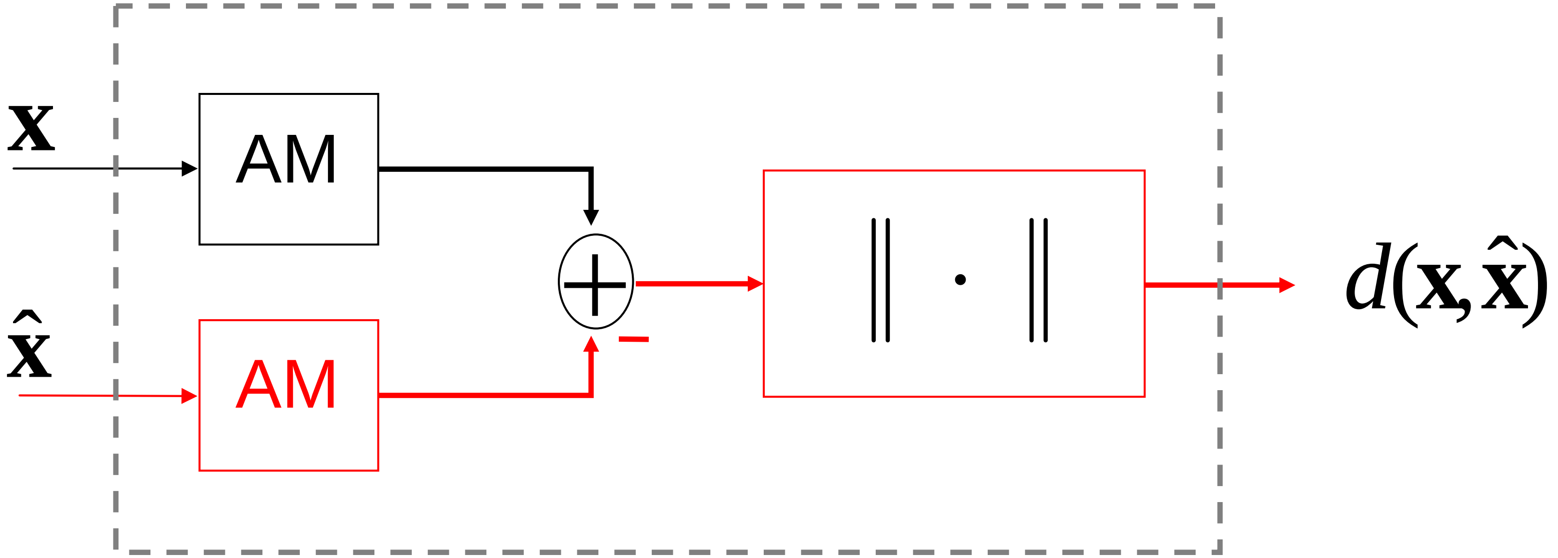
- Non-simultaneous masking (time masking)



- Not accounted for by simple auditory models
- Coding standards use (heuristic) work-arounds

Motivation

- Merge the worlds of auditory modeling and audio coding
- Advanced auditory models too complex for coding



perceptual distortion criterion

The Sensitivity Matrix

- Sensitivity Matrix **M** from Taylor expansion

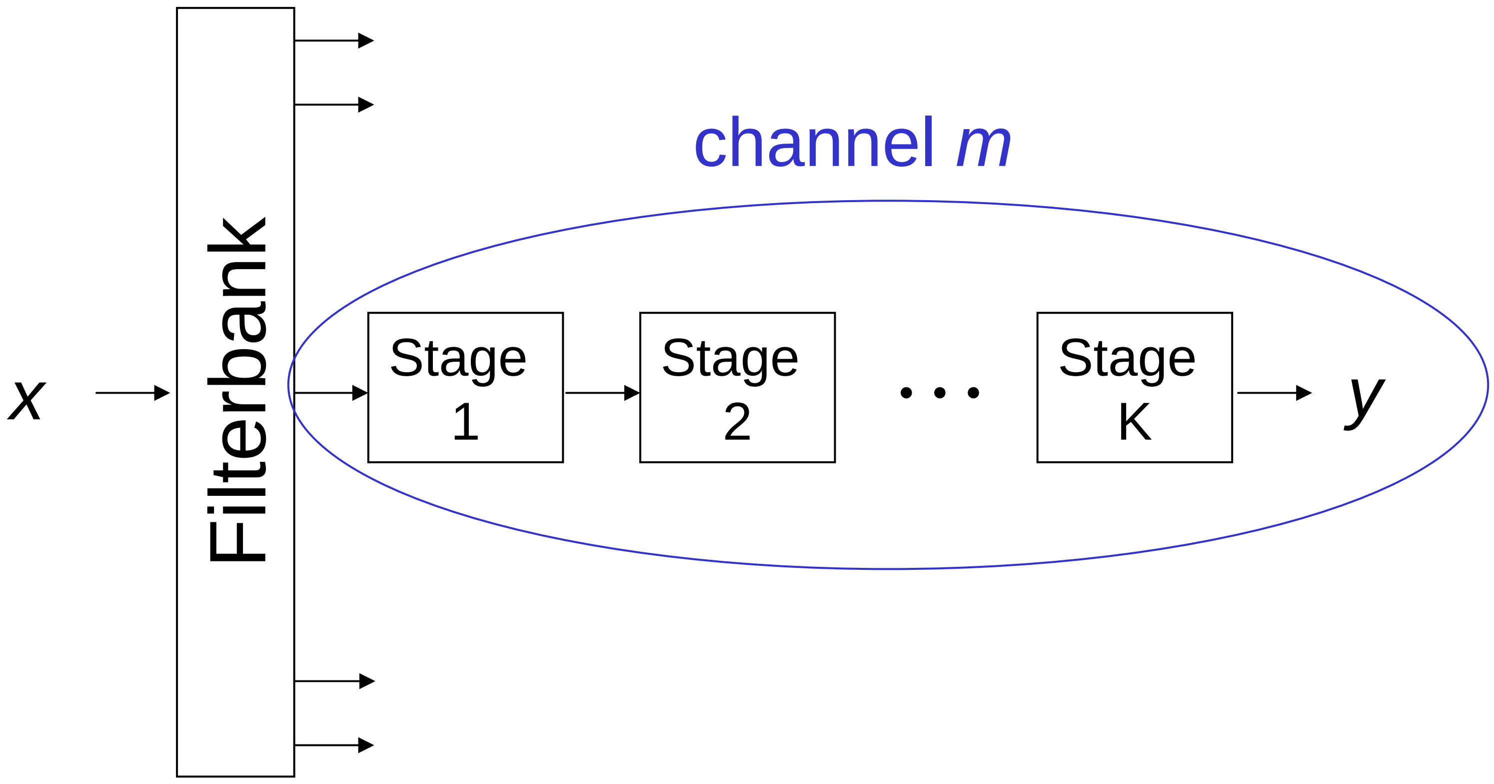
$$d(\mathbf{x},\hat{\mathbf{x}}) \approx \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^H \mathbf{M}_x(\mathbf{x}) (\mathbf{x}-\hat{\mathbf{x}}), \quad d(\mathbf{x},\hat{\mathbf{x}}) \to 0$$

$$[\mathbf{M}_x(\mathbf{x})]_{i,j} = \left. \frac{\partial^2 d(\mathbf{x},\hat{\mathbf{x}})}{\partial \hat{x}_i \partial \hat{x}_j} \right|_{\hat{\mathbf{x}}=\mathbf{x}}$$

- **Complexity** problem solved!
- **Analysis** of model properties using linear algebra
  - Masking curve
  - Subspace-based analysis

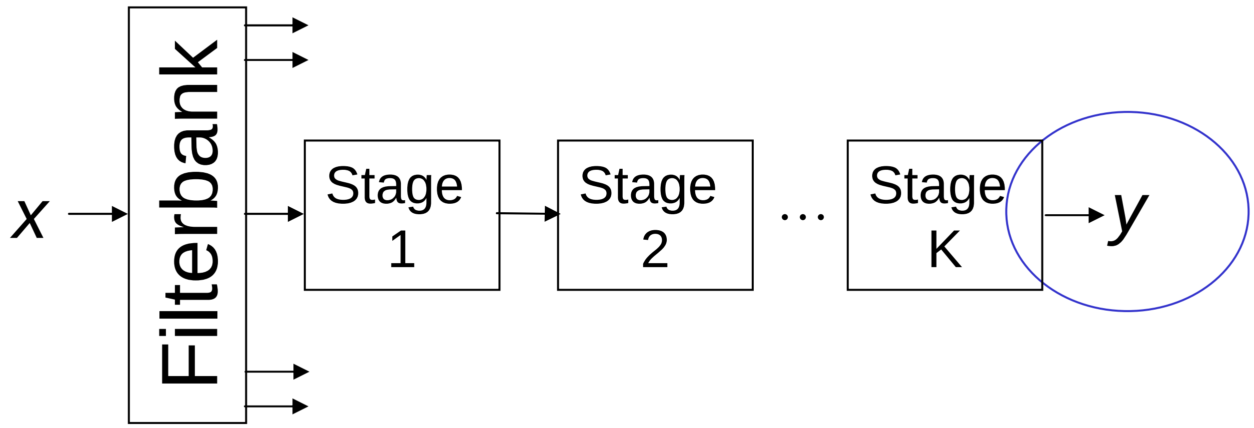
Application to an Auditory Model

- A typical auditory model (here: the Dau model)



- Distortion is sum of per-channel distortions

Application to an Auditory Model

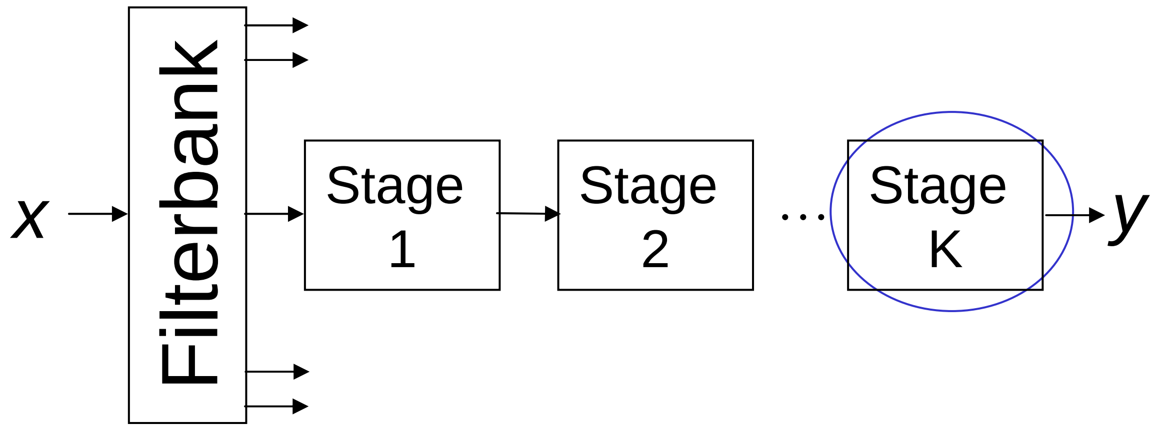


- Find sensitivity matrix for model output *y* per channel *m*

$$d^{(m)}(\mathbf{y},\hat{\mathbf{y}}) = \frac{1}{2}(\mathbf{y}-\hat{\mathbf{y}})^H \ \mathbf{M}_y^{(m)}(\mathbf{y}) \ (\mathbf{y}-\hat{\mathbf{y}})$$

here $\mathbf{M}_y^{(m)}(\mathbf{y}) = 2\mathbf{I}$

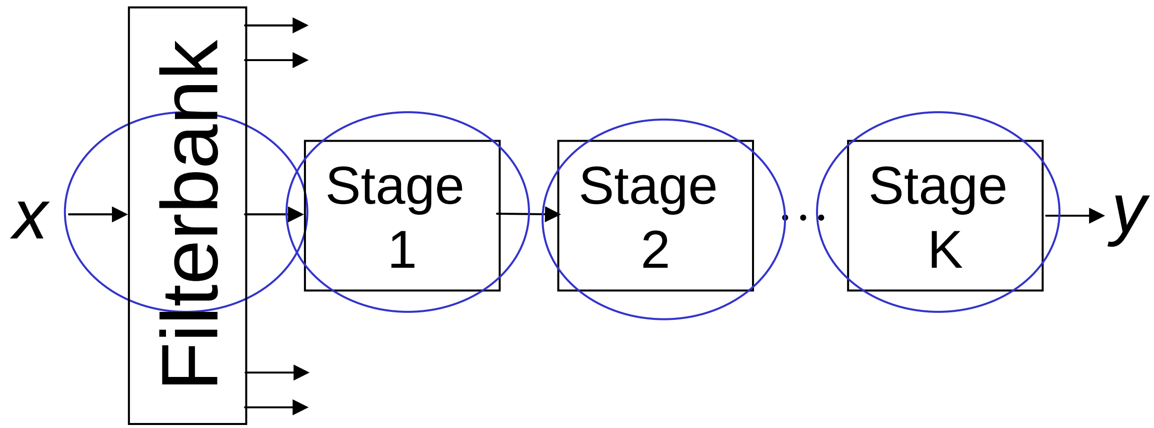
Application to an Auditory Model



- Each model stage linearized by respective Jacobian

$$\partial \mathbf{z} \longrightarrow \boxed{\mathbf{J}_z(\mathbf{z})} \longrightarrow \partial \mathbf{y}$$

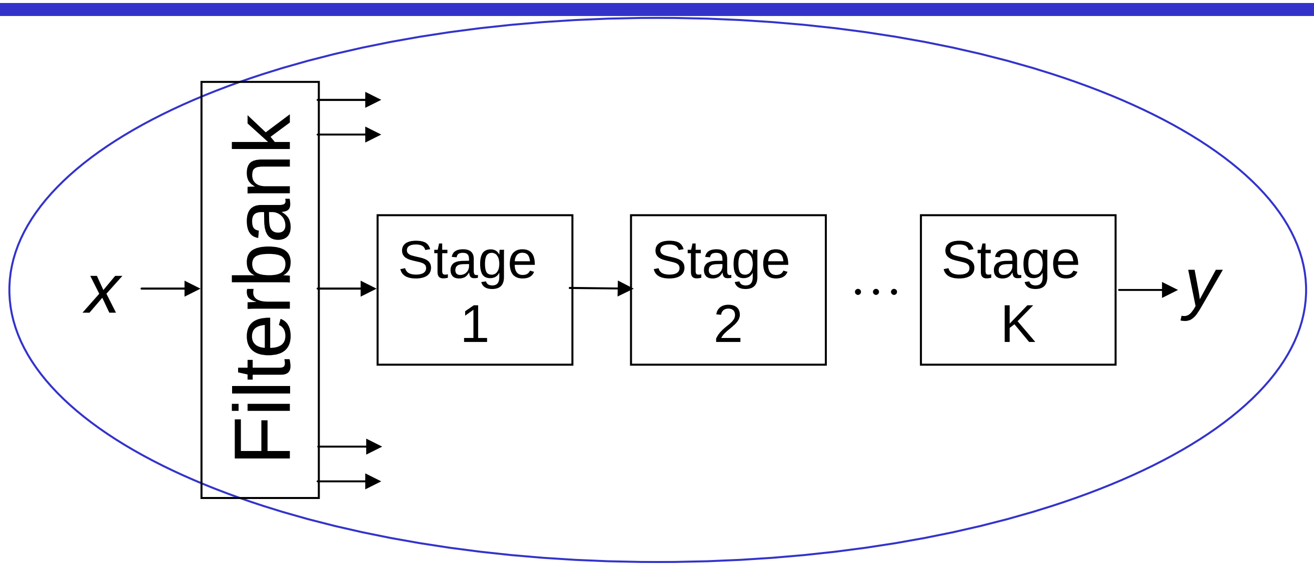
Application to an Auditory Model



- Get sensitivity matrix for channel ***m*** in input signal domain from chain rule;

$$\mathbf{M}_x^{(m)}(x) = 2\left[\prod_k \mathbf{J}_k^{(m)}\right]^H \prod_k \mathbf{J}_k^{(m)}$$

Application to an Auditory Model

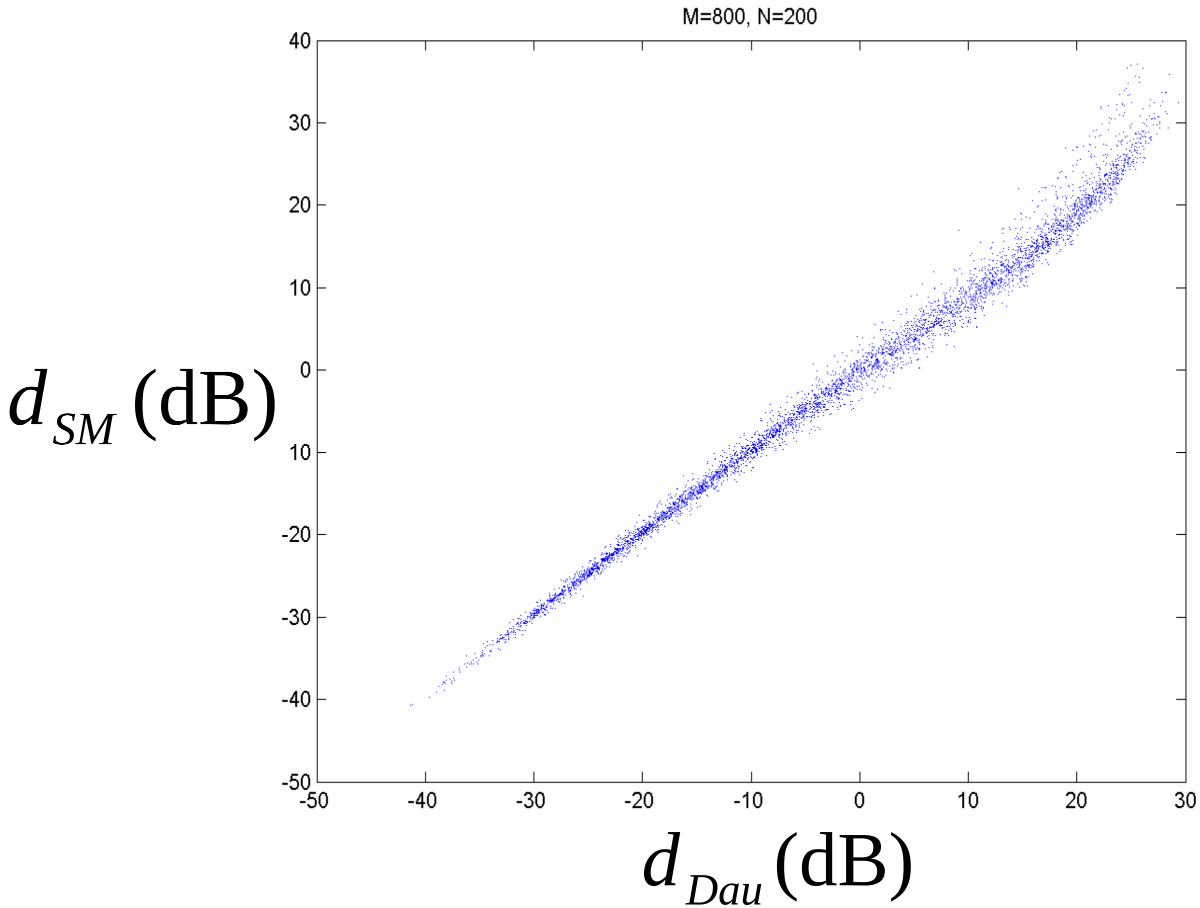


- Sensitivity Matrix is sum of per-channel matrices

$$\mathbf{M}_x(\mathbf{x}) = \sum_m \mathbf{M}_x^{(m)}(\mathbf{x}) = 2\sum_m \left[\prod_k \mathbf{J}_k^{(m)}\right]^H \prod_k \mathbf{J}_k^{(m)}$$

Validity of the Approximation

- Narrow-band music + white noise
  at different SNR (10-60 dB), correlation > 0.9



→ **coding possible**

Analysis: Fourier Transform

- Sensitivity matrix $\mathbf{M}_X(\mathbf{X})$ for frequency domain vectors, with $\mathbf{F}$ = DFT in matrix notation

$$(\mathbf{x}-\hat{\mathbf{x}}) = \mathbf{F}^H\ (\mathbf{X}-\hat{\mathbf{X}})$$

$$\mathbf{M}_X(\mathbf{X}) = \mathbf{F}\,\mathbf{M}_x(\mathbf{x})\,\mathbf{F}^H$$

Analysis: Masking Curve

- Assume masking threshold is at $d(\mathbf{X},\hat{\mathbf{X}})=1$
- Masking curve at frequency *i:*
  *g*ain g_i for unit distortion in DFT bin *i*

$$\mathbf{u}_i = \begin{bmatrix}0 & \cdots & 1 & \cdots & 0\end{bmatrix}^H$$

pos. *i*

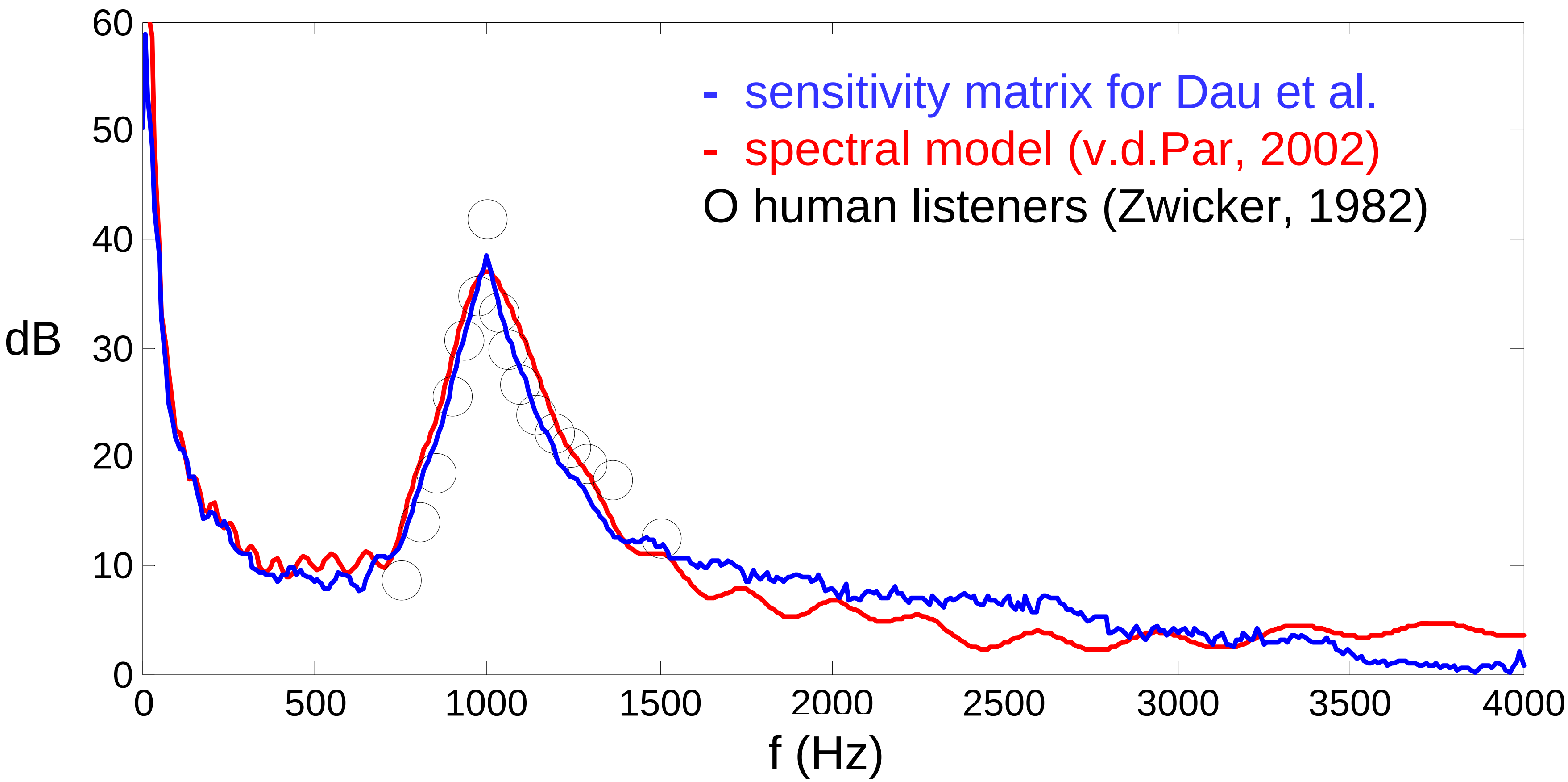
to approach masking threshold

$$1 = \frac{1}{2}\mathbf{u}_i^H \mathbf{M}_X(\mathbf{X})\,\mathbf{u}_i\, g_i$$

$$\Leftrightarrow \quad g_i = \frac{2}{\left[\mathbf{M}_X(\mathbf{X})\right]_{i,i}}$$

Example 1: Simultaneous Masking

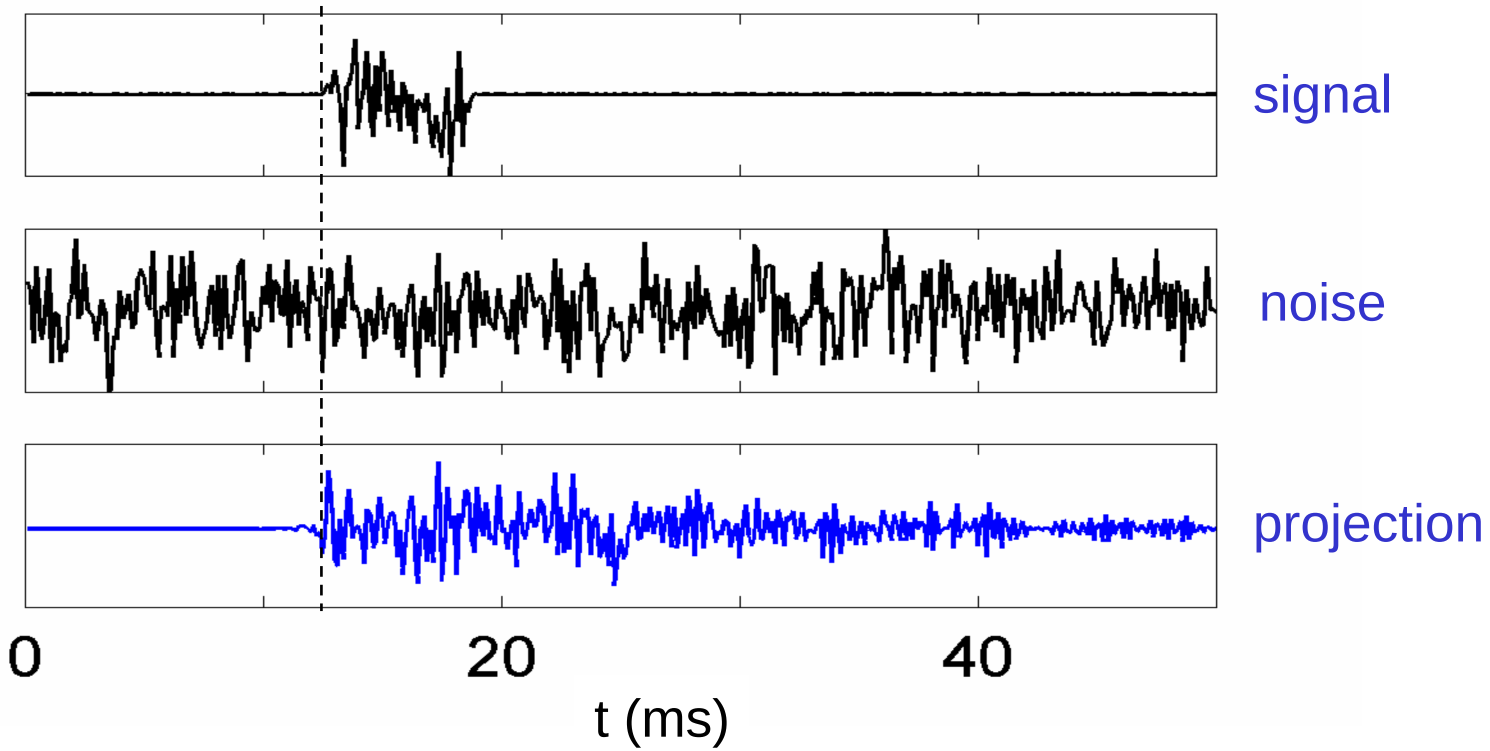
- Masking curve for 50 dB sinusoid at 1000 Hz



FlexCode

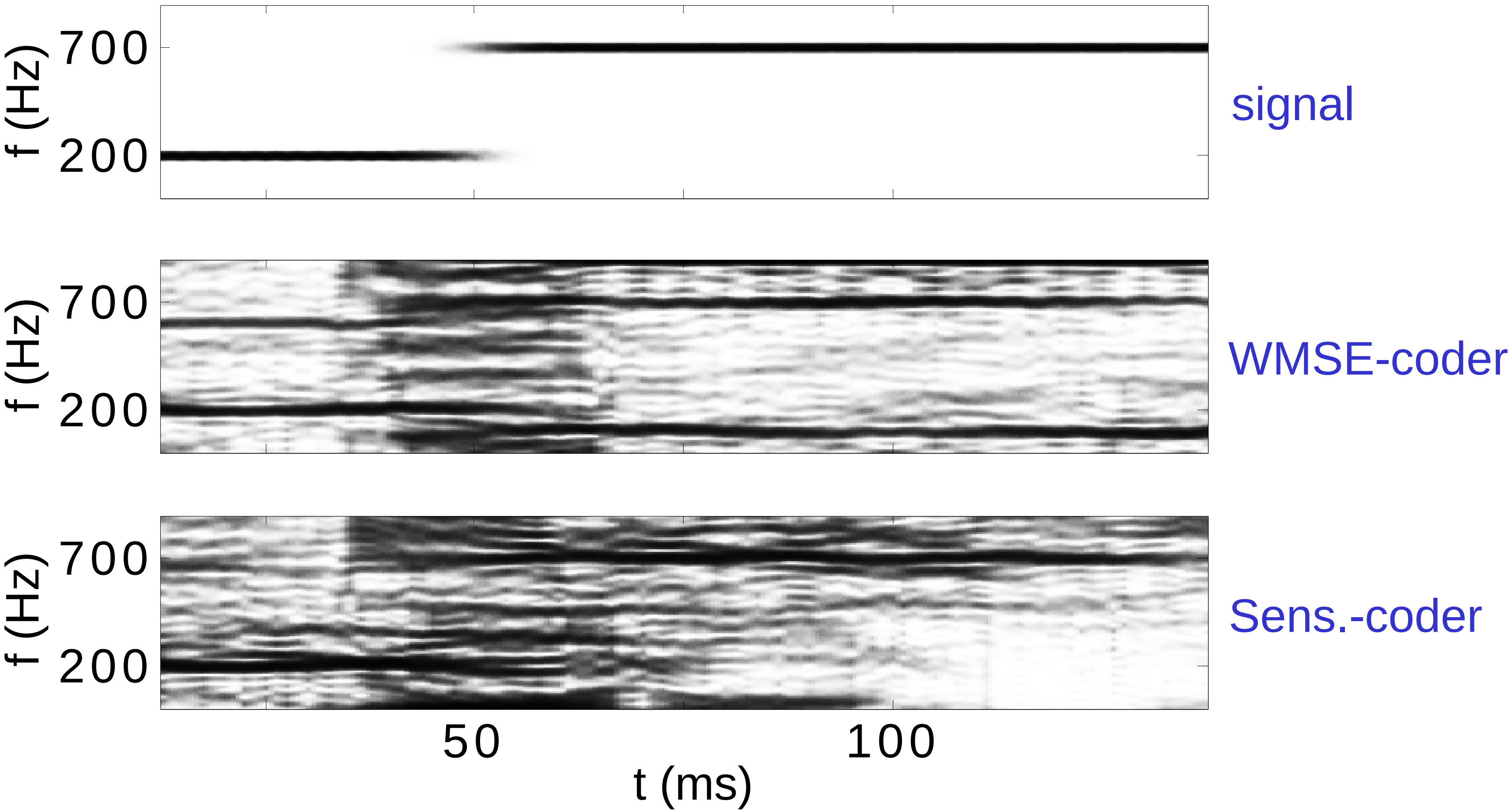
Example 2: Non-Simultaneous Masking

- Subspace analysis of **M**: low sensitivity space



A Spectro-Temporal Experiment

- Spectrograms for two sinusoids (200 Hz and 700 Hz) coded with an APCM coder



Outline

- Flexcode in a Nutshell
- The Sensitivity Matrix
- Coding with the Sensitivity Matrix
- Open Issues

FlexCode

Problem

- Sensitivity Matrix generally not diagonal in typical coding domains (e.g., time, frequency, etc.)
  - distortion not single-letter:

$$d(\mathbf{x},\hat{\mathbf{x}}) \approx \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^H \ \mathbf{M}_x(\mathbf{x}) \ \ (\mathbf{x}-\hat{\mathbf{x}})$$

  - no analytical solution known for optimal coding
  - so far only trained VQ available (undoable for large dimensions)!

Solution Outline

- Sensitivity Matrix diagonal in perceptual domain
  - transform signal vectors into perceptual domain
  - (W)MSE-optimal coding in perceptual domain results in minimum perceptual distortion
- Optimal Transform:

$$\mathbf{F}'^{\,T}_{opt}(\mathbf{x})\mathbf{F}'_{opt}(\mathbf{x}) = c\mathbf{M}_x(\mathbf{x})$$

Making it Practical

- Problem: optimal transform depends on ***x***
  - not known at decoder!
- Solution: Parameterize transform as DCT + diagonal weights

$$\mathbf{F}'(\mathbf{x}) = \mathbf{W}_f \mathbf{T}_{DCT} \mathbf{W}_t$$

  - combines DCT coding gain with perceptual transform!
  - Model-driven approach to temporal noise shaping (TNS)

Examples

- Original
- Freq. weights $\mathbf{F}'(\mathbf{x}) = \mathbf{W}_f \mathbf{T}_{DCT}$
- Time-freq. weights $\mathbf{F}'(\mathbf{x}) = \mathbf{W}_f \mathbf{T}_{DCT} \mathbf{W}_t$

Outline

- Flexcode in a Nutshell
- The Sensitivity Matrix
- Coding with the Sensitivity Matrix
- Open Issues

FlexCode

Complexity

- Main complexity lies here:

$$\mathbf{M}(\mathbf{x}) = 2\sum_{m}\left[\prod_{k}\mathbf{J}_{k}^{(m)}\right]^{H}\prod_{k}\mathbf{J}_{k}^{(m)}$$

- Can we make multiplications ‘sparse’?
- Can we reduce the size of the matrices?

Conclusion

- Using auditory models in coding is possible
- General method
- Direct applicability in AbS structures (CELP)
- Solution outline for transform coding

The End

Thank you!