ÖAW
Österreichische Akademie der Wissenschaften

Mathematics and Signal Processing in Acoustics
From Theory to Applications

Peter Balazs

Acoustics Research Institute (ARI)
Austrian Academy of Sciences, Vienna

Signal Processing Workshop 2011; Brno, 28.10.2011

Overview:

Mathematics in Acoustics

Peter Balazs

The Acoustics Research Institute

Mathematics in Acoustics

Research Topics

Conclusions

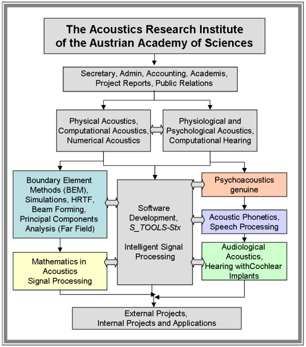
Overview:

Overview:

Overview:

The Acoustics Research Institute

Interdisciplinary approach to application-open fundamental research in acoustics.



22 employees. Part of the Austrian Academy of Sciences.

The Acoustics Research Institute

Work groups:

- Acoustic Phonetics and Phonology

  Speaker identification, Viennese Dialect, Speaker Models, ...
- Audiological Acoustics

  Binaural Hearing, Sound Localization, Cochlea Implants, Psychoacoustic Masking, ...
- Psychoacoustics

  Masking, Computational Auditory Scene Analysis, Pitch, Timbre, ...
- Computational Acoustics

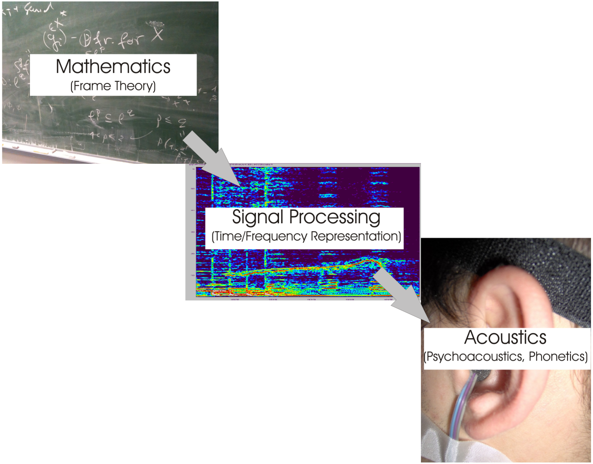
  Finite / Boundary Element Methods, Fast Multipole Method, Modal Analysis, Beam Forming, Acoustic Holography, Noise Barriers, ...
- Software Development

  ST^X, digital signal processing, sound database, ...
- Mathematics and Signal Processing in Acoustics

Mathematics and Signal Processing in Acoustics

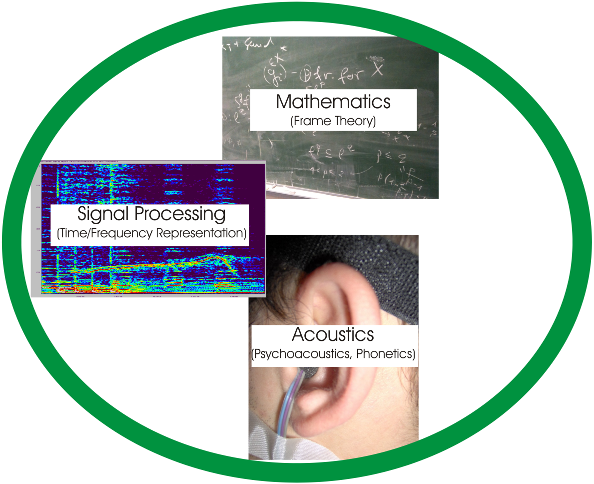
- **Founded 2008:** with a Viennese Science Foundation WWTF Project for 'High Potentials' (*'Frame Multipliers: Theory and Applications in Acoustics - MulAc-'*, March 2008- September 2011)
- **Growing:**
  - Currently 3 Researchers (funded by the Academy of Sciences),
  - 4 this November (FWF Lise Meitner project)
  - and 6 next May (START project *'Frames and Linear Operators for Acoustical Modeling and Parameter Estimation - FLAME-'*, 2012-2018)
- **Goal:**
  Application-oriented mathematics for acoustics with the focus on pluri-disciplinary research and internal, national and international cooperation.

State of the Art:



- Heuristic models
- Lacking precision
- Low efficiency
- Poor control of properties

Our Goal



- Application-oriented mathematics
- True inter-disciplinarity
- Synergy
- Novel methods for acoustics

(some) Research Topics

Frame Multipliers:

Analysis

$\Downarrow$

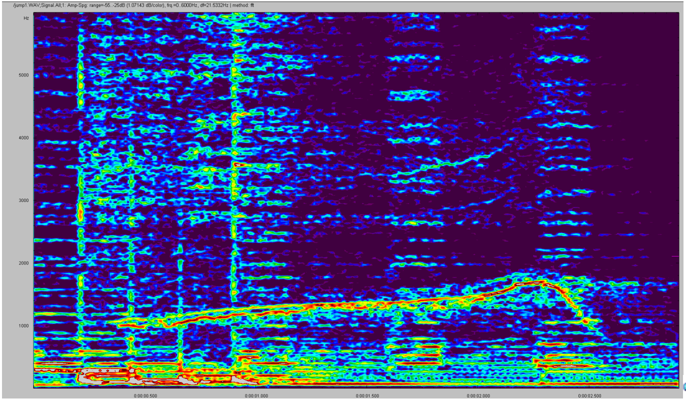
Multiplication

$\Downarrow$

Synthesis

Example for a Multiplier

Original audio file:

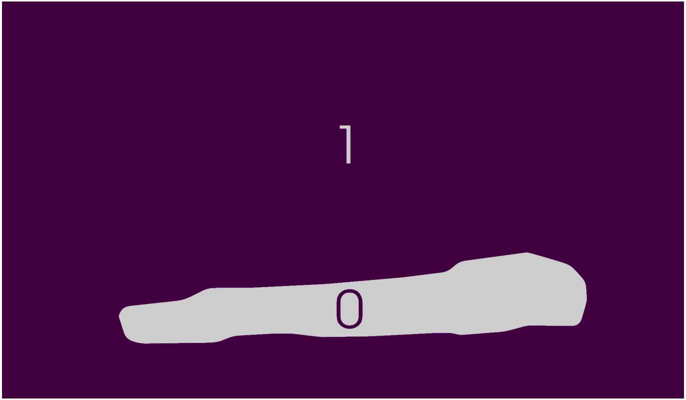


Example for a Multiplier

Mathematics in Acoustics

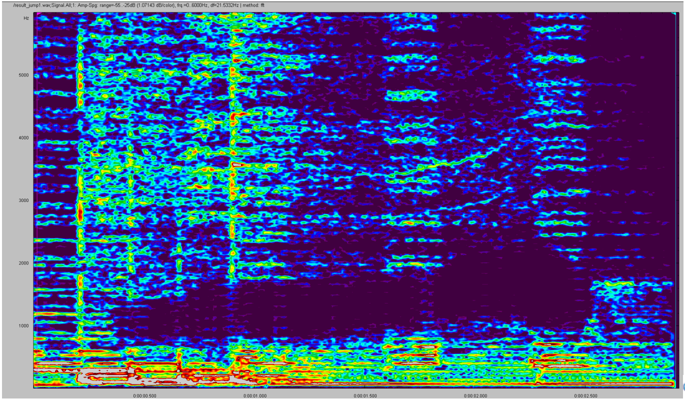
Peter Balazs

Symbol:



Example for a Multiplier

Result of Gabor Multiplier.



Frames

Definition

The sequence $(g_k | k \in K)$ is called a frame for the Hilbert space $\mathcal{H}$, if constants $A, B > 0$ exist, such that

$$A \cdot \|f\|_{\mathcal{H}}^2 \leq \sum_k |\langle f, g_k \rangle|^2 \leq B \cdot \|f\|_{\mathcal{H}}^2 \ \forall f \in \mathcal{H}$$

- Allows non-orthogonal expansion, redundant representation.
- Perfect reconstruction is guaranteed with the 'canonical dual frame' $\tilde{g}_k = S^{-1} g_k$ with S the frame operator (i.e. combined analysis/resynthesis operator).

$$f = \sum \langle f, \tilde{g}_k \rangle g_k = \sum \langle f, g_k \rangle \tilde{g}_k.$$

Frames

Definition

The sequence $(g_k | k \in K)$ is called a frame for the Hilbert space $\mathcal{H}$, if constants $A, B > 0$ exist, such that

$$A \cdot \|f\|_{\mathcal{H}}^2 \leq \sum_k |\langle f, g_k \rangle|^2 \leq B \cdot \|f\|_{\mathcal{H}}^2 \ \forall \ f \in \mathcal{H}$$

- Allows non-orthogonal expansion, redundant representation.
- Perfect reconstruction is guaranteed with the 'canonical dual frame' $\tilde{g}_k = S^{-1} g_k$ with S the frame operator (i.e. combined analysis/resynthesis operator).

$$f = \sum \langle f, \tilde{g}_k \rangle \, g_k = \sum \langle f, g_k \rangle \, \tilde{g}_k.$$

Frame Multipliers: Definition

Definition

Let $(\psi_k)_{k \in K}$, $(\phi_k)_{k \in K}$ be frames in the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Define the operator $\mathbf{M}_{m,(\phi_k),(\psi_k)} : \mathcal{H}_1 \to \mathcal{H}_2$, the frame multiplier, as the operator

$$\mathbf{M}_{m,(\phi_k),(\psi_k)} f = \sum_k m_k \langle f, \psi_k \rangle \phi_k$$

where $m \in l^\infty(K)$ is called the *symbol*.

ONS [Schatten, 1960]; regular Gabor frames [Feichtinger and Nowak, 2003]; general frames resp. Bessel sequences [Balazs, 2007]

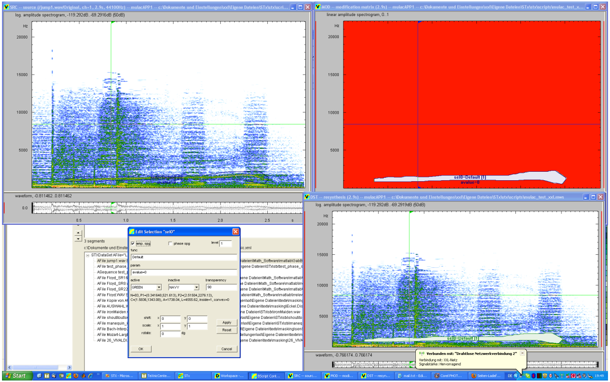
Theorem ([Balazs, 2007])

Let $\mathbf{M} = \mathbf{M}_{m,(\phi_k),(\psi_k)}$ be a frame multiplier for (ψ_k) and (ϕ_k) with the upper frame bounds B and B' respectively. Then

1. *If $m \in l^\infty$, then $\mathbf{M}$ is a well defined bounded operator.* $\|\mathbf{M}\|_{Op} \leq \sqrt{B'}\sqrt{B} \cdot \|m\|_\infty$.
2. $\mathbf{M}^*_{m,(\phi_k),(\psi_k)} = \mathbf{M}_{\overline{m},(\psi_k),(\phi_k)}$. *Therefore if m is real-valued and $\phi_k = \psi_k$ for all k, $\mathbf{M}$ is self-adjoint.*
3. *If $m \in c_0$, $\mathbf{M}$ is compact.*
4. *If $m \in l^1$, $\mathbf{M}$ is a trace class operator with* $\|\mathbf{M}\|_{trace} \leq \sqrt{B'}\sqrt{B} \cdot \|m\|_1$, *and* $tr(\mathbf{M}) = \sum_k m_k \langle \phi_k, \psi_k \rangle$.
5. *If $m \in l^2$, $\mathbf{M}$ is a Hilbert Schmidt operator with* $\|\mathbf{M}\|_{\mathcal{HS}} \leq \sqrt{B'}\sqrt{B} \cdot \|m\|_2$.

Also mathematical results for p-frames in Banach spaces, continuous frames, condition for invertibility for arbitrary sequences, etc...

Numerical results and algorithms



Implementation in STx and MATLAB, in the Linear Time-Frequency Analysis Toolbox (LTFAT) [Soendergaard et al., 2011] (available at Sourceforge) .

Österreichische Akademie der Wissenschaften

Applications in Acoustics: Perceptual Sparsity by Irrelevance

MP3-Player

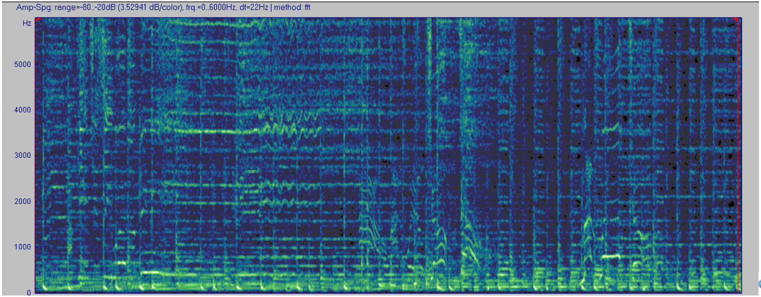
MP3:

- encoding / decoding scheme
- MPEG1/MPEG2 (Layer 3)
- signal processing
- psychoacoustical masking model

Perceptual Sparsity by Irrelevance

Existing algorithm in ST^X: simple model based on simultaneous masking, but effective algorithm!

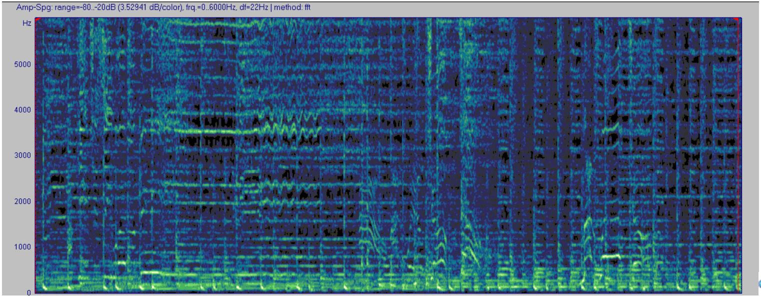
Original audio file



Psychoacoustic Masking

Existing algorithm in STx: simple model based on simultaneous masking, but effective algorithm!

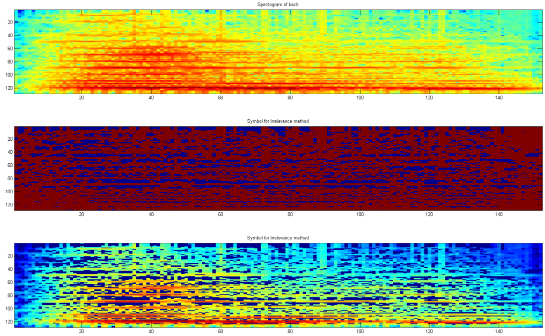
Irrelevance Filter



"Lossy Coding"

Perceptual Sparsity by Irrelevance

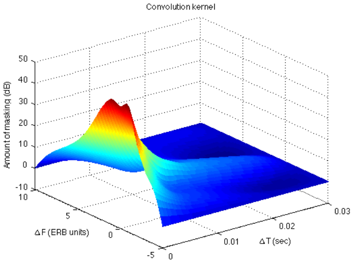
Interpreted as adaptive Gabor multiplier:



Irrelevance by Time Frequency Masking

Extend to True Time-Frequency Model using Multipliers:

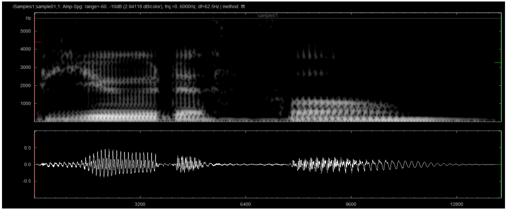
- Base it on non-stationary Gabor transform (NSGT) ([Balazs et al., 2011] $\rightarrow$ see talk by Monika), adapted to perception.
- Use new psychoacoustical data on time-frequency masking [Laback et al., 2011].



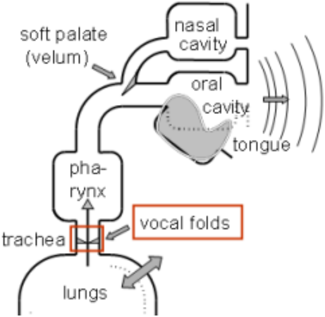
Long-term goal: Develop a new audio coder!

Speech and Speaker Analysis: using a Pole-Zero Model

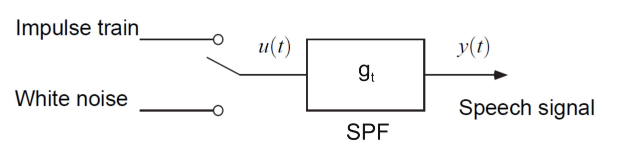
Digital Modeling of Speech



The goal: estimate the vocal tract from speech signal.



Digital Modeling of Speech



Goal: Estimate the speech production filter (SPF) $g(q)$ based on the observation of the speech signal $y(t)$ and the assumption on the excitation signal $u(t)$. See [Rabiner and Schafer, 1978, Markel and Gray, 1976, Quatieri, 2001].

Proposed Method: Estimation of Numerator and Denominator (ND, ND-FFT)

Assumption: Quasi-stationarity

Standard approach: All-Pole Model estimation on linear frequency scale.

Our approach [Marelli and Balazs, 2010]:

- Include Poles and Zeros!
- Take Psychoacoustics into account:
  - Minimize a logarithmic criterion
  - Use perceptual frequency scale
- Use an efficient, iterative algorithm (Quasi Newton-search using BFGS formula [Fletcher, 1987] for inverse of Hessian, analytic formulas for the Jacobian.)

Formant and Anti-Formant Tracking

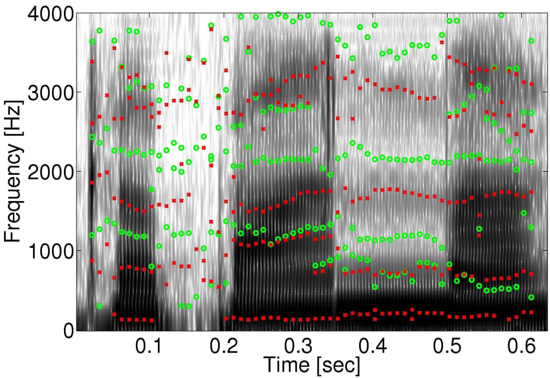


Figure: Poles (x) and zeros (o) of the SPF models estimated from the word "capanna".

Current and Future Work

- Importance for speaker verification (nasal and nasalized sounds).
- Link to physical models.
- Tracking algorithm applied on perception-motivated NSGT.
- Use for speech coding.
- Do a time-variant approach.

See `http://www.kfs.oeaw.ac.at/polezero`.

Improvement of Head-Related-Transfer-Function-Measurements

(see next talk)

Conclusions

Connecting mathematics, signal processing and acoustics leads to

- interesting mathematical results, as well as
- new methods and models for acoustics, as well as their implementation.

From Theory to Applications!

Thank you for your attention!

Questions? Comments?

References: I

Balazs, P. (2007).
Basic definition and properties of Bessel multipliers.
Journal of Mathematical Analysis and Applications, 325(1):571–585.

Balazs, P., Dörfler, M., Holighaus, N., Jaillet, F., and Velasco, G. (2011).
Theory, implementation and applications of nonstationary Gabor frames.
Journal of Computational and Applied Mathematics, to appear:–.

Feichtinger, H. G. and Nowak, K. (2003).
A first survey of Gabor multipliers, chapter 5, pages 99–128.
Birkhäuser Boston.

Fletcher, R. (1987).
Practical methods of optimization.
A Wiley-Interscience Publication. John Wiley & Sons Ltd., Chichester, second edition.

Laback, B., Balazs, P., Necciari, T., Savel, S., Ystad, S., Meunier, S., and Kronland-Martinet, R. (2011).
Additivity of auditory masking for short gaussian-shaped sinusoids.
The Journal of the Acoustical Society of America, 129:888–897.

Marelli, D. and Balazs, P. (2010).
On pole-zero model estimation methods minimizing a logarithmic criterion for speech analysis.
IEEE Transactions on Audio, Speech and Language Processing, 18(2):237–248.

References: II

Markel, J. and Gray, J. A. (1976).
Linear Prediction of Speech.
Springer-Verlag, Berlin Heidelberg.

Quatieri, T. (2001).
Discrete-time speech signal processing: principles and practice.
Prentice Hall Press, Upper Saddle River, NJ, USA.

Rabiner, L. and Schafer, R. (1978).
Digital Processing of Speech Signals.
Prentice Hall, Englewood Cliffs, NJ.

Schatten, R. (1960).
Norm Ideals of Completely Continuous Operators.
Springer Berlin.

Soendergaard, P., Torrésani, B., and Balazs, P. (2011).
The linear time frequency analysis toolbox.
International Journal of Wavelets, Multiresolution and Information Processing, to appear:accepted.