Frame Theory and its Acoustical Applications

Peter Balazs
(joint work with a lot of people, many of them are in the room!)

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

Modern Methods of Time-Frequency Analysis II, ESI
Time-frequency methods for the applied sciences,
04.12.2012

Overview:

Overview:

Overview:

1. Acoustics Research Institute
2. Frame Theory
   - Time-Frequency representation
   - Semi-Frames
   - Non-stationary Gabor Transform
3. Frame Multipliers
4. Applications
   - Perceptual Sparsity by Irrelevance
   - Acoustic System Estimation
5. Conclusions

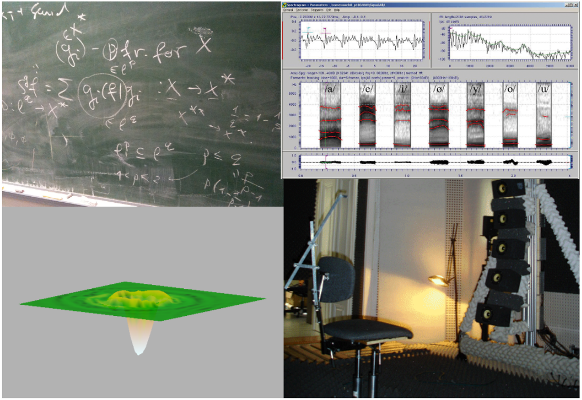
Österreichische Akademie der Wissenschaften

Overview:

Overview:

Acoustics Research Institute

Part of the Austrian Academy of Sciences, approx. 30 employees.
Work-Groups:



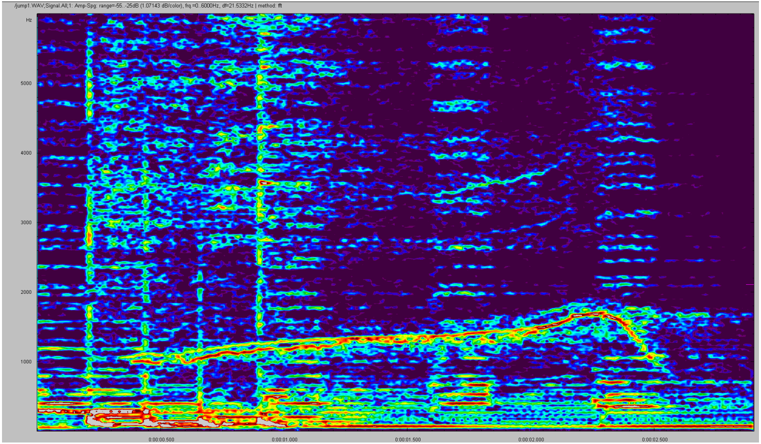
Acoustics Research Institute

Part of the Austrian Academy of Sciences, approx. 30 employees.
Work-Groups:

- Acoustic Phonetics
- Physical and Computational Acoustics
- Psychoacoustics and Experimental Audiology
- Mathematics and Signal Processing in Acoustics

Signal Representations: Time-Frequency Analysis and Filterbanks

Spectogram


/jump1.WAV;Signal.All;1: Amp-Spg: range=-55..-25dB (1.07143 dB/color), frq.=0..6000Hz, df=21.5332Hz | method: fft
Hz
5000
4000
3000
2000
1000
0:00:00.500
0:00:01.000
0:00:01.500
0:00:02.000
0:00:02.500

Short Time Fourier Transformation (STFT)

Definition (see e.g [Gröchenig, 2001])

Let $f,g \neq 0$ in $L^2\left(\mathbb{R}^d\right)$, then we call

$$\mathcal{V}_g f(\tau,\omega) = \int\limits_{\mathbb{R}^d} f(x)\overline{g(x-\tau)}e^{-2\pi i \omega x}dx\,.$$

the Short Time Fourier Transformation (STFT) of the signal f with the window g.

$$\mathcal{V}_g(f)(\tau,\omega) = \mathcal{F}\left(f\cdot\overline{T_\tau g}\right).$$

Sampled Version: Gabor transform : $\tau = a\cdot k$, $\omega = b \cdot l$ with $k,l \in \mathbb{Z}$.

$$f \mapsto \mathcal{V}_g(f)(a\cdot k, b\cdot l).$$

When is perfect reconstruction possible?

Short Time Fourier Transformation (STFT)

Definition (see e.g [Gröchenig, 2001])

Let $f,g \neq 0$ in $L^2\left(\mathbb{R}^d\right)$, then we call

$$\mathcal{V}_g f(\tau,\omega) = \int\limits_{\mathbb{R}^d} f(x)\overline{g(x-\tau)}e^{-2\pi i \omega x} dx \,.$$

the Short Time Fourier Transformation (STFT) of the signal f with the window g.

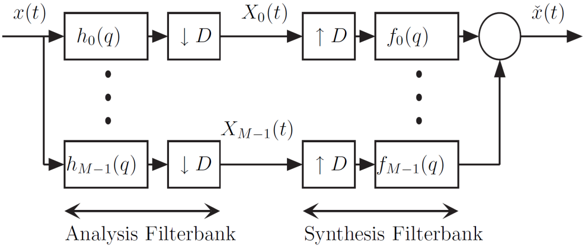
$$\mathcal{V}_g(f)(\tau,\omega) = \mathcal{F}\left(f \cdot \overline{T_\tau g}\right).$$

Sampled Version: Gabor transform : $\tau = a \cdot k$, $\omega = b \cdot l$ with $k,l \in \mathbb{Z}$.

$$f \mapsto \mathcal{V}_g(f)(a \cdot k, b \cdot l).$$

When is perfect reconstruction possible?

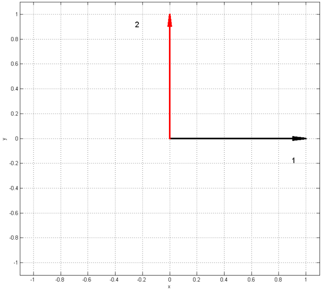
Filterbank



When is perfect reconstruction possible?

Orthonormal Basis (ONB)

Standard aproach: orthonormal basis.

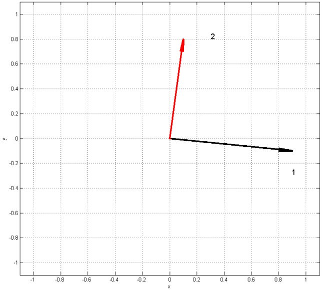


Problems:

- Perturbation
- Construction
- Error Robustness

Riesz bases

Riesz bases

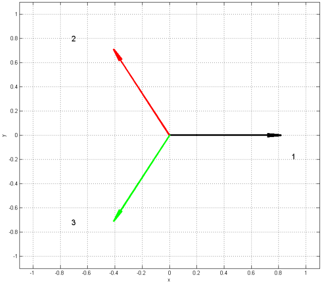


Problems:

- ~~Perturbation~~
- Construction
- Error Robustness

Frames

Alternate approach: introduce redundancy.



Problems:

- ~~Perturbation~~
- ~~Construction~~
- ~~Error Robustness~~

Frames

Definition

The (countable) sequence $\Psi = (\psi_k | k \in K)$ is called a frame for the Hilbert space $\mathcal{H}$ if constants $A > 0$ and $B < \infty$ exist such that

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

[Duffin and Schaeffer, 1952, Daubechies et al., 1986]

Frames

Beautiful abstract mathematical setting:

- Frames = generalization of bases; can be overcomplete, allowing redundant representations.
- Frame inequality = generalization of Parseval's condition.
- Active field of research in mathematics!

Interesting for applications:

- **Much more freedom.** Finding and constructing frames can be easier and faster.
  Some advantageous side constraints can **only** be fulfilled for frames.
- Perfect reconstruction is guaranteed with the 'canonical dual frame' $\tilde{\psi}_k = S^{-1}\psi_k$

$$f = \sum_k < f, \psi_k > \tilde{\psi}_k = \sum_k < f, \tilde{\psi}_k > \psi_k.$$

Semi-Frames

Sometimes even the concept of frames is too restrictive, in the sense that one cannot satisfy both frame bounds simultaneously.

Ψ is an upper semi-frame [Antoine and Balazs, 2011] if there exists a constant $\mathsf{B} < \infty$ such that

$$0 < \sum_{k \in \Gamma} |\langle \psi_k, f \rangle|^2 \leq \mathsf{B} \|f\|^2, \forall f \in \mathcal{H}, \ f \neq 0.$$

Theorem

For all $f \in R_D$, we have the reconstruction formula

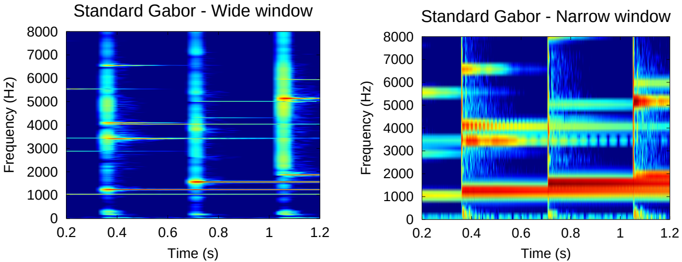
$$f = \sum_k \left[G^{-1} \left(\langle f, \psi_k \rangle_{\mathcal{H}} \right) \right] \psi_k \tag{1}$$

with unconditional convergence.

Non-stationary Gabor transform

Non-stationary Gabor transform

Limitations of Standard Gabor analysis: Quality of representation highly depends on window choice, but optimal window choice is different for different signal components



Our proposition [Jaillet, 2005, Balazs et al., 2011]: simple extension to reduce this limitation by using window evolving over time.

Non-stationary Gabor transform

Given a sequence of windows $(\gamma_n)_{n\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ and sequences of real numbers $(a_n)_{n\in\mathbb{Z}}$ and $(b_n)_{n\in\mathbb{Z}}$, the non-stationary Gabor transform (NSGT) elements are defined, for $(m,n)\in\mathbb{Z}^2$, by:

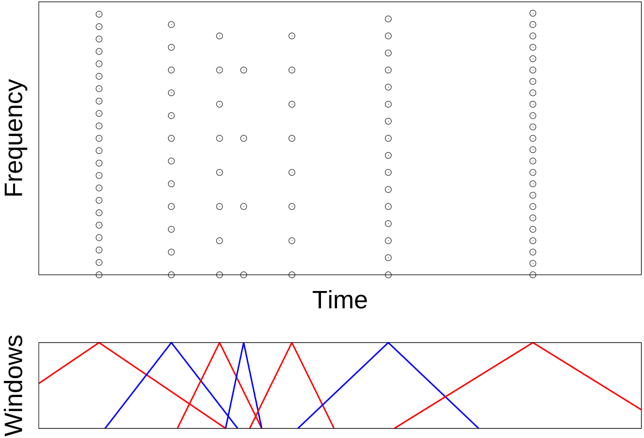
$$\gamma_{m,n}(t) = \gamma_n(t - na_n)e^{i2\pi mb_n t} = M_{mb_n} T_{na_n} \gamma_n.$$

Regular structure in frequency allows FFT implementation.

A analogue construction in the frequency domain allows easy implementation of, e.g. wavelet frames; an invertible CQT [Velasco et al., 2011]; or a filterbank adapted to human auditory perception (see talk by Thibaud Necciari).

Non-stationary Gabor transform

Sampling grid example:



Non-stationary Gabor transform

Frame theory allows perfect reconstruction. Especially easy and fast in the 'painless' case:

Theorem

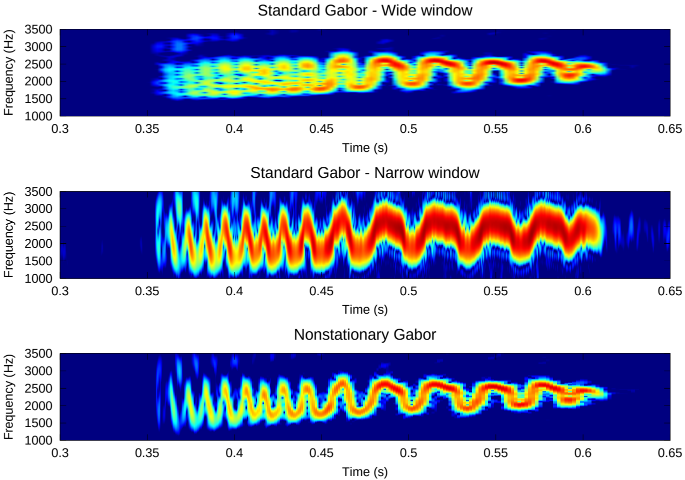
For every $n \in \mathbb{Z}$, let the function $\gamma_n \in L^2(\mathbb{R})$ be compactly supported with $\operatorname{supp}(\gamma_n) \subseteq [c_n + na_n, d_n + na_n]$ such that $d_n - c_n \leq \frac{1}{b_n}$. the system of functions $g_{m,n}$ forms a frame for $L^2(\mathbb{R})$ if and only if there exists $A > 0$ and $B < \infty$, such that $A \leq \sum_n \frac{1}{b_n} |\gamma_n(t - na_n)|^2 \leq B$. In this case, the canonical dual frame has the same structure and is given by:

$$\tilde{\gamma}_{m,n}(t) = \frac{\gamma_n(t)}{\sum_k \frac{1}{b_k} |\gamma_k(t - ka_k)|^2} e^{2\pi i m b_n t}. \quad (2)$$

For approaches and algorithm beyond the painless case see [Necciari et al., 2013] (and talk by Thibaud).

Non-stationary Gabor transform

Bird vocalization example:



What is a

Frame Multiplier:

Analysis

$\Downarrow$

Multiplication

$\Downarrow$

Synthesis

What is a

Frame Multiplier:

Analysis

$\Downarrow$

Multiplication

$\Downarrow$

Synthesis

What is a

Frame Multiplier:

Analysis

$\Downarrow$

Multiplication

$\Downarrow$

Synthesis

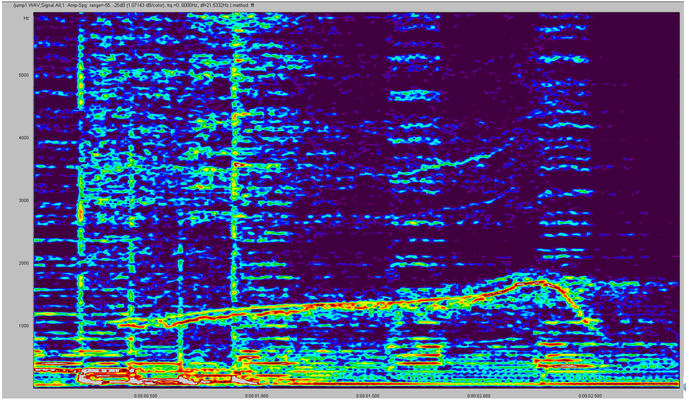
Multipliers

Those are operators, that are of utmost importance in

- Mathematics, where they are used for the diagonalization of operators [Schatten, 1960].
- Physics, where they are a link between classical and quantum mechanics, so called quantization operators [Ali et al., 2000].
- Signal Processing, where they are a particular way to implement time-variant filters [Matz and Hlawatsch, 2002].
- Acoustics, where those time-frequency filters are used in several fields, for example in Computational Auditory Scene Analysis [Wang and Brown, 2006].

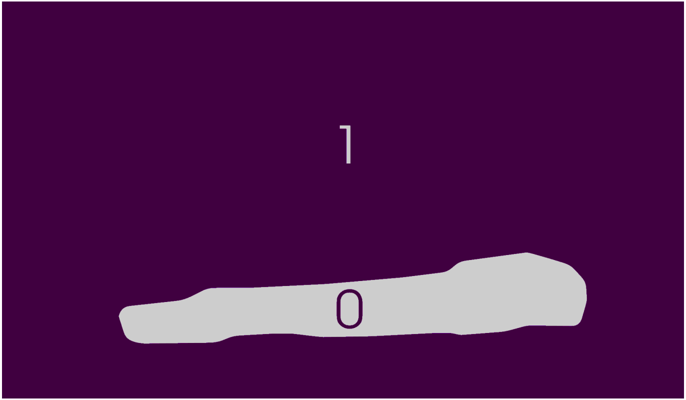
Example for a Multiplier

Original audio file:



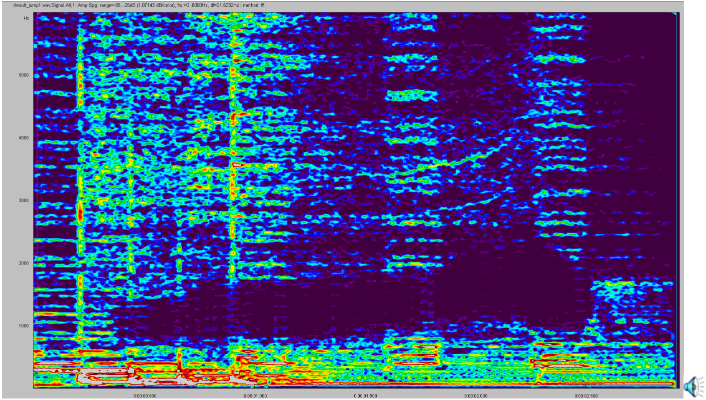
Example for a Multiplier

Symbol:



Example for a Multiplier

Result of Gabor Multiplier.



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.

Generalization of Gabor multipliers [Feichtinger and Nowak, 2003] to the general frame case [Balazs, 2007].

Fundamental Research in the Theory of Multipliers

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* $\|\mathbf{M}\|_{\mathcal{HS}} \leq \sqrt{B'}\sqrt{B} \cdot \|m\|_2$.

Unconditionally Convergence and Invertibility of Frame Multipliers

We put quite some effort into the investigation of invertible multipliers. A quite surprising result is the following:

Theorem (B., Stoeva; submitted)

Let Φ and Ψ be frames for $\mathcal{H}$, and let m be semi-normalized. Let $\mathbf{M}_{m,\Phi,\Psi}$ be invertible. Then there exist a dual frame $\Phi^\dagger$ of Φ and a dual frame $\Psi^\dagger$ of Ψ, so that for any dual frame Φ^d of Φ and any dual frame Ψ^d of Ψ we have

$$M_{m,\Phi,\Psi}^{-1} = M_{1/m,\Psi^\dagger,\Phi^d} = M_{1/m,\Psi^d,\Phi^\dagger}. \tag{3}$$

The frames $\Psi^\dagger$ are uniquely determined.

Unconditionally Convergence and Invertibility of Frame Multipliers

We also found sufficient conditions, when multipliers are invertible. In this case, formulas for $\mathbf{M}^{-1}_{(m_n),(\phi_n),(\psi_n)}$ are determined. For example in the following case:

Proposition ([Stoeva and Balazs, 2012])

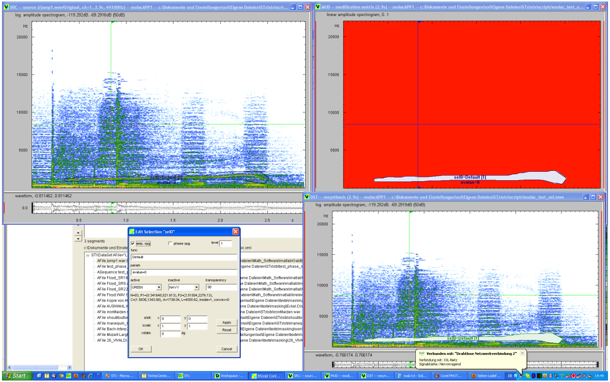
Let $\Phi = (\phi_k)$ *be a frame for* $\mathcal{H}$. *Assume that* $\exists\, \mu \in [0, \frac{A_\Phi^2}{B_\Phi})$ *such that* $\sum |\langle f, m_n\psi_n - \phi_n\rangle|^2 \leq \mu \|f\|^2, \quad \forall\ f \in \mathcal{H}$. *Then* $m\Psi$ *is a frame for* $\mathcal{H}$, *the multipliers* $\mathbf{M}_{\overline{m},\Phi,\Psi}$ *and* $\mathbf{M}_{m,\Psi,\Phi}$ *are invertible on* $\mathcal{H}$ *and*

$$\frac{1}{B_\Phi + \sqrt{\mu B_\Phi}} \|h\| \leq \|\mathbf{M}^{-1} h\| \leq \frac{1}{A_\Phi - \sqrt{\mu B_\Phi}} \|h\|, \ \forall h \in \mathcal{H}, \qquad (4)$$

$$\mathbf{M}^{-1} = \sum_{k=0}^{\infty} [S_\Phi^{-1}(S_\Phi - \mathbf{M})]^k S_\Phi^{-1} \qquad (5)$$

where $\mathbf{M}$ *denotes any one of* $\mathbf{M}_{\overline{m},\Phi,\Psi}$ *and* $\mathbf{M}_{m,\Psi,\Phi}$.

Numerical results and algorithms



Implementation in STx and MATLAB, in the Linear Time-Frequency Analysis Toolbox (LTFAT) [Soendergaard et al., 2012] (available at Sourceforge, see talk by Peter Søndergaard.) .

Applications in Acoustics: Perceptual Sparsity by Irrelevance

MP3-Player

MP3:

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

Psychoacoustic Masking: introduction

Masking:
presence of one stimulus, the masker, decreases the response to another stimulus, the target.

Irrelevance Filter: searches (and deletes) perceptional irrelevant data (in complex signals) using a masking model.

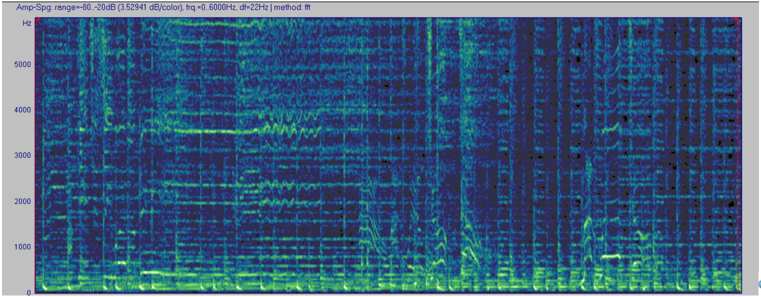
Psychoacoustic Masking: introduction

Masking:
presence of one stimulus, the masker, decreases the response to another stimulus, the target.

Irrelevance Filter: searches (and deletes) perceptional irrelevant data (in complex signals) using a masking model.

Perceptual Sparsity by Irrelevance

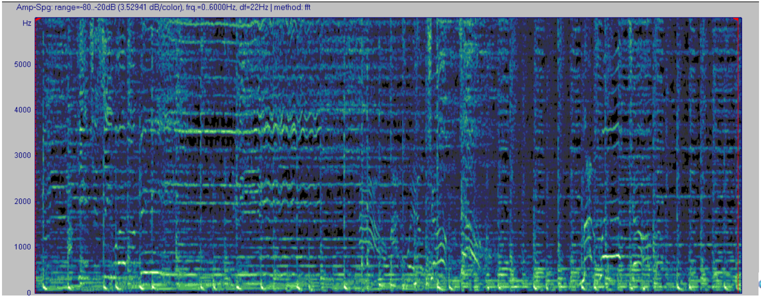
Algorithm in STx:
Original audio file



Psychoacoustic Masking

Algorithm in STx:

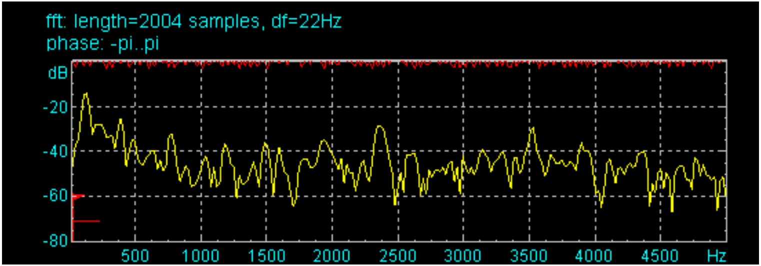
Filtered signal



"Lossy Coding"

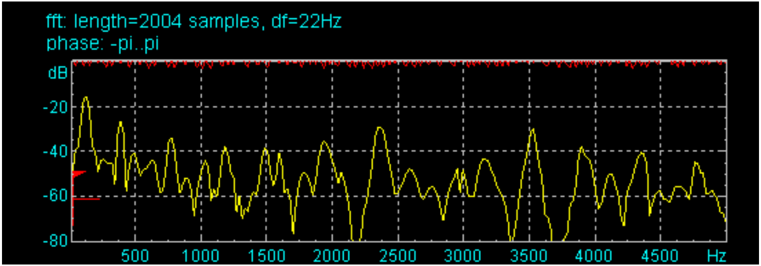
Irrelevance Filter

Existing algorithm in ST^X: Original audio file (Spectrum)



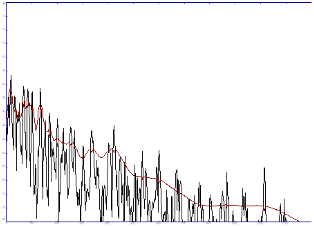
Irrelevance Filter

Existing algorithm in STx: Masked signal (Spectrum)



Psychoacoustic Masking

[Balazs et al., 2010] models simultaneous frequency masking and uses single spectra.

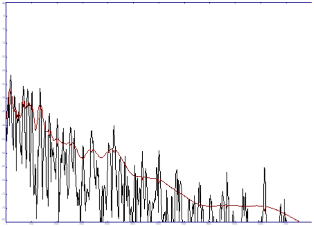


The irrelevance method calculates an adaptive threshold function for each spectra of a Gabor transform. This corresponds to an adaptive Gabor frame multiplier with coefficients in $\{0, 1\}$.
With frame theory some properties are explained:

- perfect reconstruction,
- time frequency concentration (spread),
- smoothness and
- numerical efficiency.

Psychoacoustic Masking

[Balazs et al., 2010] models simultaneous frequency masking and uses single spectra.

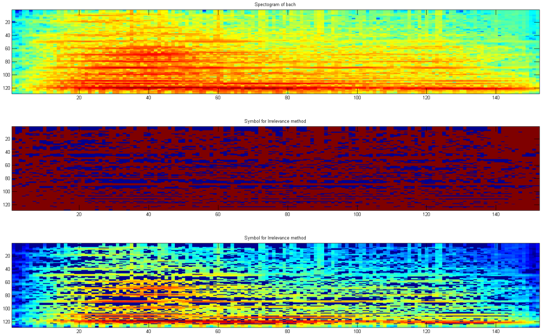


The irrelevance method calculates an adaptive threshold function for each spectra of a Gabor transform. This corresponds to an adaptive Gabor frame multiplier with coefficients in $\{0, 1\}$.
With frame theory some properties are explained:

- perfect reconstruction,
- time frequency concentration (spread),
- smoothness and
- numerical efficiency.

Perceptual Sparsity by Irrelevance

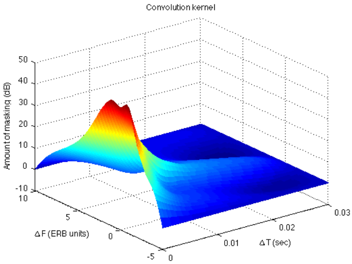
Interpreted as adaptive Gabor frame multiplier:



Irrelevance by Time Frequency Masking

Extend to True Time-Frequency Model using Multipliers:

- Base it on ERBlets (NSGT adapted to perception) (see talb by Thibaud).
- Use new psychoacoustical data on time-frequency masking [Laback et al., 2011].



Applications in Acoustics:

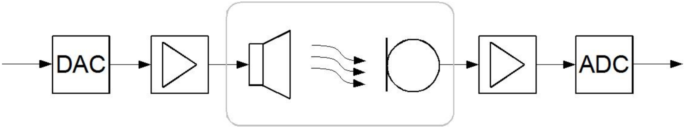
Acoustic System Estimation

Acoustic System Estimation

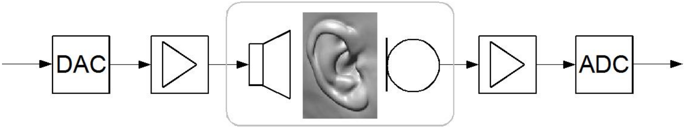
Measurement of Head Related Transfer Functions (HRTFs)

Acoustic System Estimation

Electro-acoustic signal path:
weakly non-linear, time invariant systems (PA, Speakers)

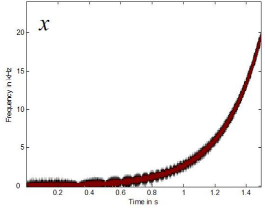


with head-movement weakly non-linear, time variant system
But the interesting part is the HRTF: an LTI system!

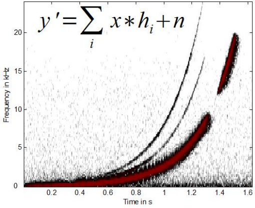


Acoustic System Estimation

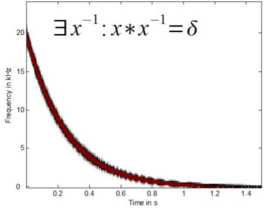
Input

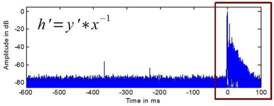


Output



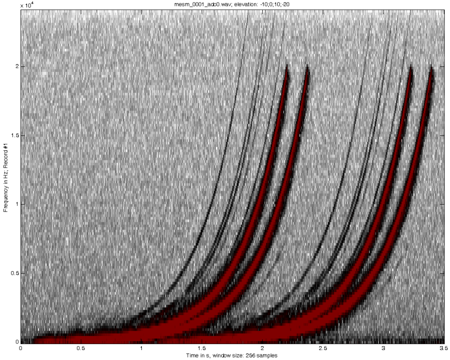
Deconvolution





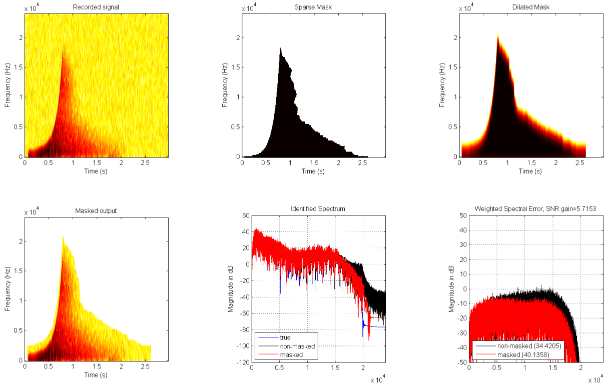
Acoustic System Estimation

Measurement of Head Related Transfer Functions (HRTFs) by the Multiple Exponential Sweeps Method (MESM) [Majdak et al., 2007]



Speed up measurement by factor of four.

Time-Frequency Denoising [Majdak et al., 2011]:



Conclusions

Frames and Frame Multipliers allow

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

From Theory to Applications!

Thank you for your attention!

Questions? Comments?

References: I

Ali, S. T., Antoine, J.-P., and Gazeau, J.-P. (2000).
Coherent States, Wavelets and Their Generalization.
Graduate Texts in Contemporary Physics. Springer New York.

Antoine, J.-P. and Balazs, P. (2011).
Frames and semi-frames.
Journal of Physcis A: Mathematical and Theoretical, 44:205201.
submitted.

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, 236(6):1481–1496.

Balazs, P., Laback, B., Eckel, G., and Deutsch, W. A. (2010).
Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking.
IEEE Transactions on Audio, Speech and Language Processing, 18(1):34–49.

Daubechies, I., Grossmann, A., and Meyer, Y. (1986).
Painless non-orthogonal expansions.
J. Math. Phys., 27:1271–1283.

References: II

Duffin, R. J. and Schaeffer, A. C. (1952).
A class of nonharmonic Fourier series.
Trans. Amer. Math. Soc., 72:341–366.

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

Gröchenig, K. (2001).
Foundations of Time-Frequency Analysis.
Birkhäuser Boston.

Jaillet, F. (2005).
Représentation et traitement temps-fréquence des signaux audio numériques pour des applications de design sonore.
PhD thesis, Université de la Méditerranée - Aix-Marseille II.

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.

Majdak, P., Balazs, P., Kreuzer, W., and Dörfler, M. (2011).
A time-frequency method for increasing the signal-to-noise ratio in system identification with exponential sweeps.
In Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011, Prag.

References: III

Majdak, P., Balazs, P., and Laback, B. (2007).
Multiple exponential sweep method for fast measurement of head related transferfunctions.
Journal of the Audio Engineering Society, 55(7/8):623–637.

Matz, G. and Hlawatsch, F. (2002).
Linear Time-Frequency Filters: On-line Algorithms and Applications, chapter 6 in 'Application in Time-Frequency Signal Processing', pages 205–271.
eds. A. Papandreou-Suppappola, Boca Raton (FL): CRC Press.

Necciari, T., Balazs, P., Holighaus, N., and Soendergaard, P. (2013).
The erblet transform: An auditory-based time-frequency representation with perfect reconstruction.
In ICASSP 2013.

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

Soendergaard, P., Torrésani, B., and Balazs, P. (2012).
The linear time frequency analysis toolbox.
International Journal of Wavelets, Multiresolution and Information Processing, 10(4):1250032.

Stoeva, D. T. and Balazs, P. (2012).
Invertibility of multipliers.
Applied and Computational Harmonic Analysis, 33(2):292–299.

References: IV

Velasco, G. A., Holighaus, N., Dörfler, M., and Grill, T. (2011).
Constructing an invertible constant-Q transform with non-stationary Gabor frames.
volume Paris. AudioMiner;Locatif.

Wang, D. and Brown, G. J. (2006).
Computational Auditory Scene Analysis: Principles, Algorithms, and Applications.
Wiley-IEEE Press.