Frames for Psychoacoustics
Erblet transform and perceptual sparsity

Peter Balazs
joint work with T. Necciari, B. Laback, N. Holighaus, D. Stoeva, ...

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

February Fourier Talks 2014

Acoustics Research Institute (ARI)

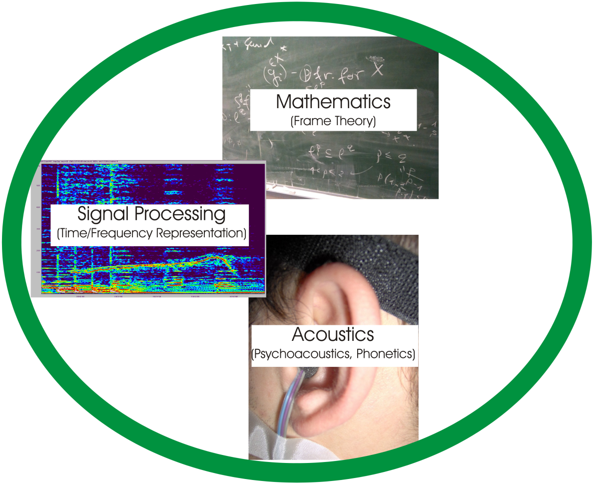
ARI

Interdisciplinary research in acoustics, integrating acoustic phonetics, psychoacoustics and computational physics, based on a solid mathematical background.

Excellence through Synergy

ARI

Advantage of a strong mathematical background

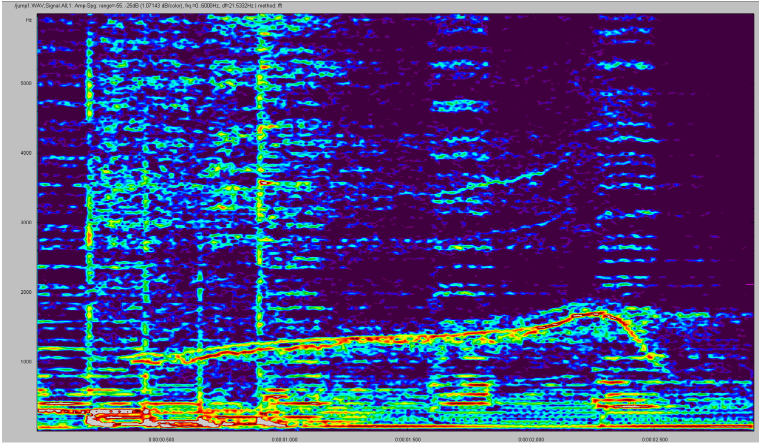


- Application-oriented mathematics
- True inter-disciplinarity
- Synergy
- Novel methods

Overview:

Signal Representations: Time-Frequency Analysis and Frames

Spectrogram



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.

Sampled Version is the Gabor transform:

$$f \mapsto \mathcal{V}_g(f)(a\cdot k, b\cdot l) = \langle f, g_{k,l}\rangle \text{, where } g_{k,l}(t) = g(t-ka)e^{i2\pi lbt}.$$

When is perfect reconstruction possible?

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]

Beautiful abstract mathematical setting:

- Frames = generalization of bases; can be overcomplete, allowing redundant representations. Redundancy
- 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' $\tilde{\psi}_k = S^{-1}\psi_k$

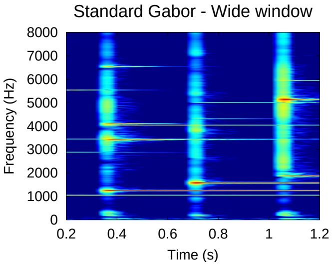
$$f = \sum_k < f, \psi_k > \tilde{\psi}_k = \sum_k < f, \tilde{\psi}_k > \psi_k,$$

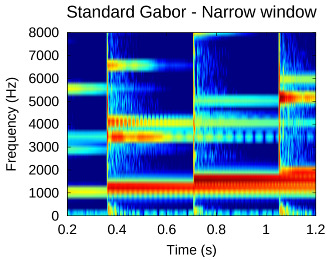
where S is the frame operator $Sf = \sum_k < f, \psi_k > \psi_k$.

Frame Theory:

Non-stationary Gabor transform

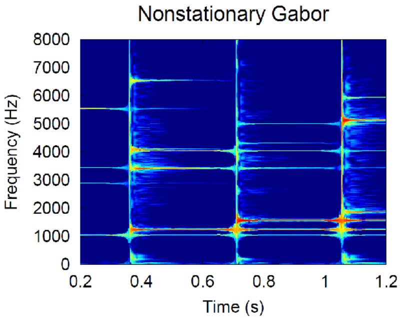
Limitations of Standard Gabor analysis: Quality of representation highly depends on window choice, but optimal window choice is different for different signal components





Non-stationary Gabor Transform

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



Non-stationary Gabor Transform

Given a sequence of windows $(g_n)_{n \in \mathbb{Z}}$ 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:

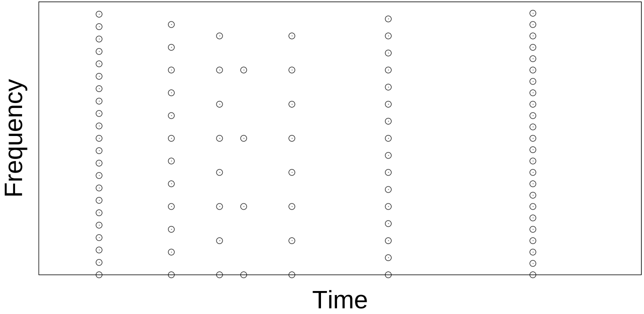
$$g_{m,n}(t) = g_n(t - na_n)e^{i2\pi m b_n t}.$$

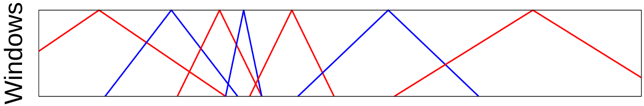
Regular structure in frequency allows FFT implementation.

An analogue construction in the frequency domain allows easy implementation of, e.g. wavelet frames; an invertible CQT [Velasco et al., 2011].

Non-stationary Gabor Transform

Sampling grid example:





Non-stationary Gabor Transform

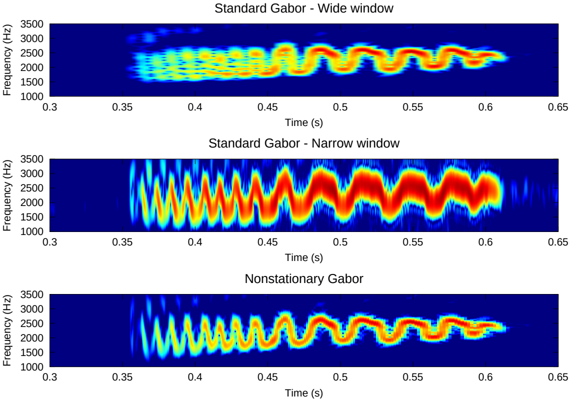
Frame theory allows perfect reconstruction. Particularly efficient in the 'painless' case:

Theorem

For every $n \in \mathbb{Z}$, let the function $g_n \in L^2(\mathbb{R})$ be compactly supported with $\operatorname{supp}(g_n) \subseteq [c_n, d_n]$ 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 $A \leq \sum_n \frac{1}{b_n} |g_n(t - na_n)|^2 \leq B$. In this case, the canonical dual frame has the same structure and is given by:

$$\tilde{g}_{m,n}(t) = \frac{g_n(t)}{\sum_k \frac{1}{b_k} |g_k(t - ka_k)|^2} e^{2\pi i m b_n t}. \tag{1}$$

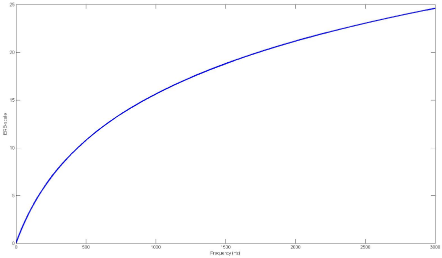
Bird vocalization example:



For an overview of adapted and adaptive time-frequency representations, see [Balazs et al., 2013].

ERBlets

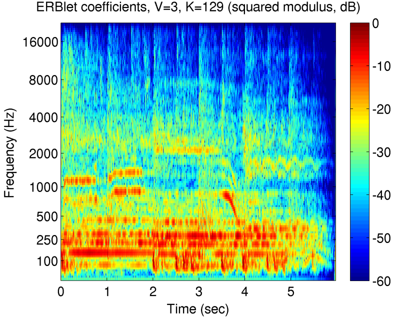
Non-stationary Gabor transform adapted to human auditory perception [Necciari et al., 2013]:



ERB-scale

ERBlets

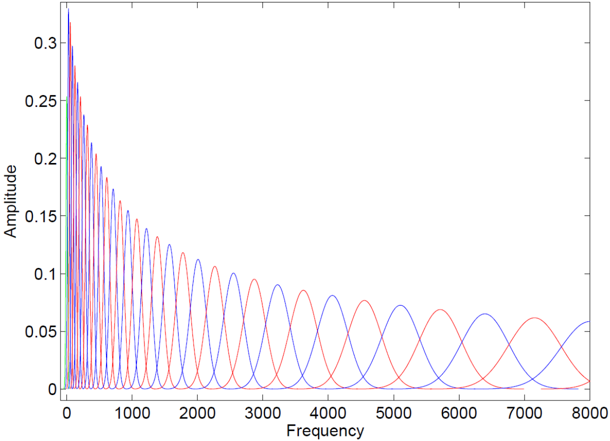
Non-stationary Gabor transform adapted to human auditory perception [Necciari et al., 2013]:



Relative reconstruction error: $< 10^{-15}$.
Implementation in LTFAT [Soendergaard et al., 2012].

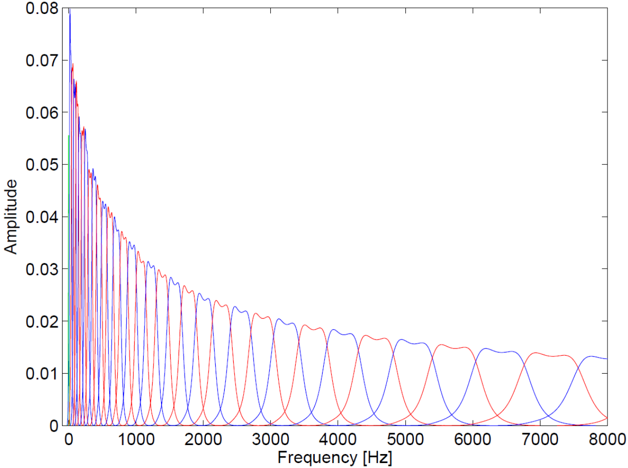
ERBlets

Filterbank:



ERBlets

Dual Filterbank:



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

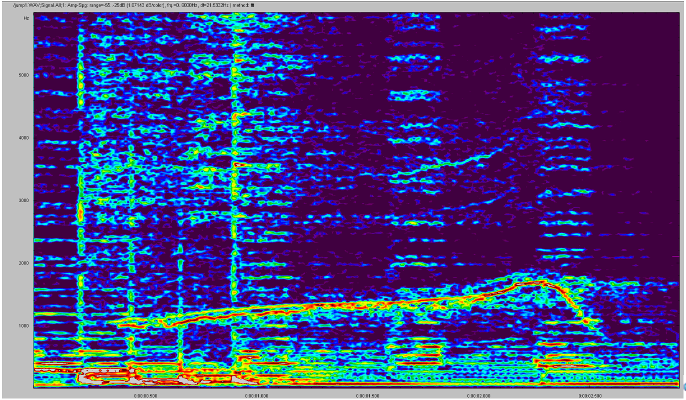
Multipliers

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].

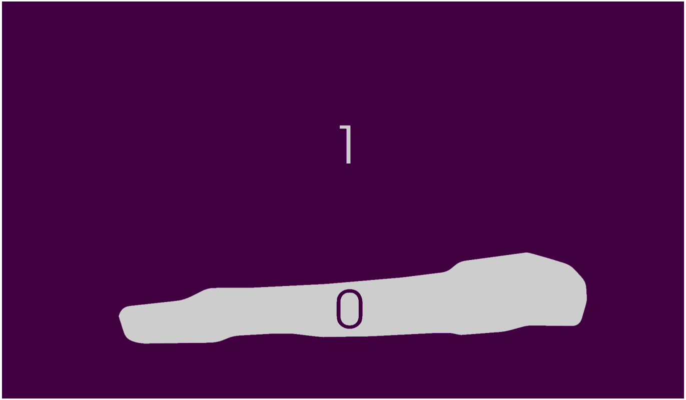
Example for a Multiplier

Original audio file:



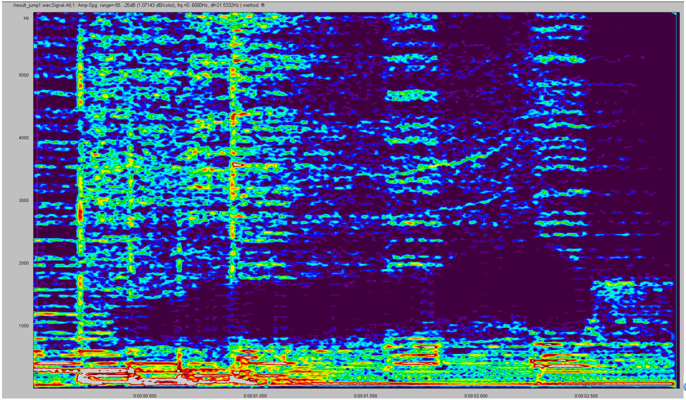
Example for a Multiplier

Symbol:



Example for a Multiplier

Result of Gabor Multiplier.



Frame Multipliers: Definition

Definition ([Balazs, 2007])

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

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

where $m = (m_k)$ is called the symbol.

Generalization of Gabor multipliers [Feichtinger and Nowak, 2003].

We have invested quite some afford into the abstract setting, in particular investigating invertible multipliers, see e.g. [Stoeva and Balazs, 2012].

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

$$M_{m,\Phi,\Psi}^{-1} = M_{1/m,\Psi^\dagger,\Phi^d} = M_{1/m,\Psi^d,\Phi^\dagger}. \tag{2}$$

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

Applications in Acoustics: **Perceptual Sparsity by Irrelevance**

MP3-Player

MP3:

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

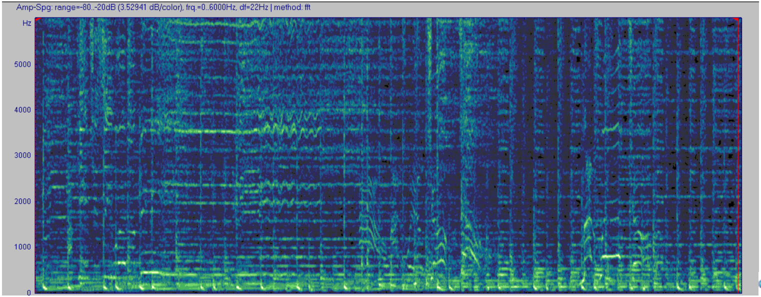
Masking:
presence of one stimulus, the masker, decreases the detectability of another stimulus, the target.

Irrelevance Filter: searches (and deletes) perceptual irrelevant (masked, inaudible) data (in complex signals) using a masking model.

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

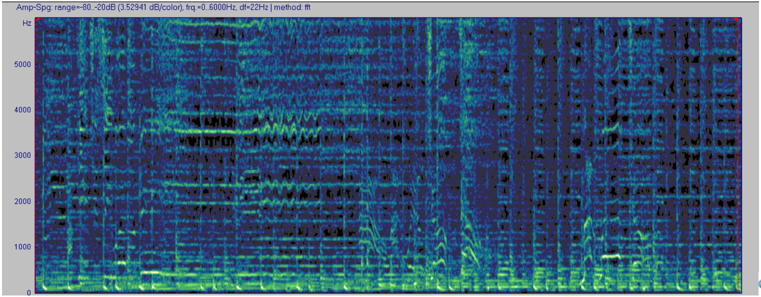
Irrelevance Filter: searches (and deletes) perceptual irrelevant (masked, inaudible) data (in complex signals) using a masking model.

Algorithm in STx:
Original audio file



Algorithm in STx:

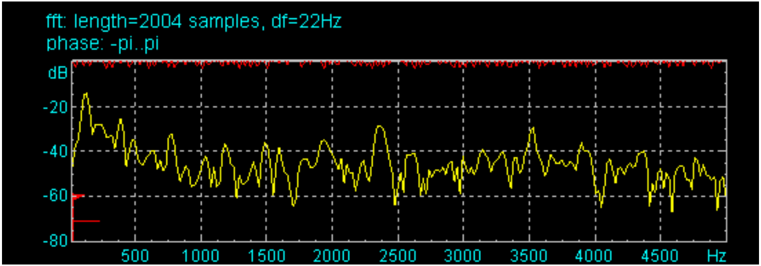
Filtered signal



Residual

Adaptive

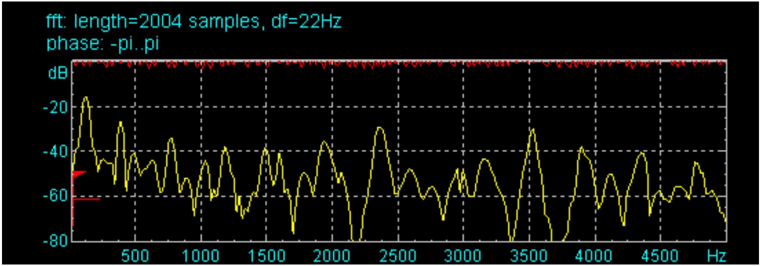
Existing algorithm in STX: Original audio file (Spectrum)



Back

Irrelevance Filter

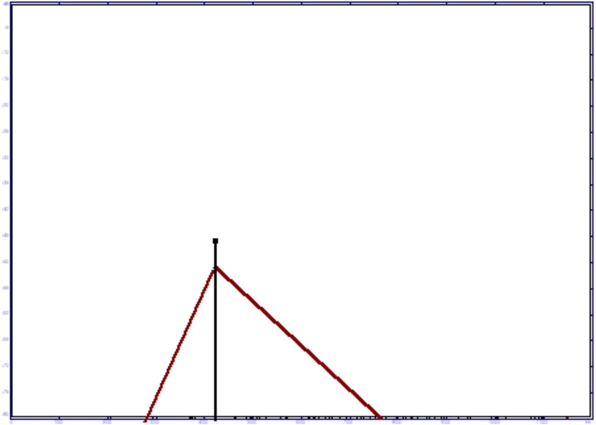
Existing algorithm in STx: Masked signal (Spectrum)



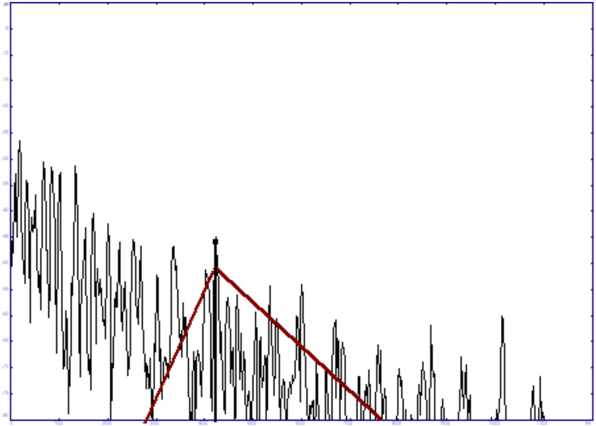
Back

Psychoacoustic Masking : simultaneous masking I

Existing Model, using bark scale

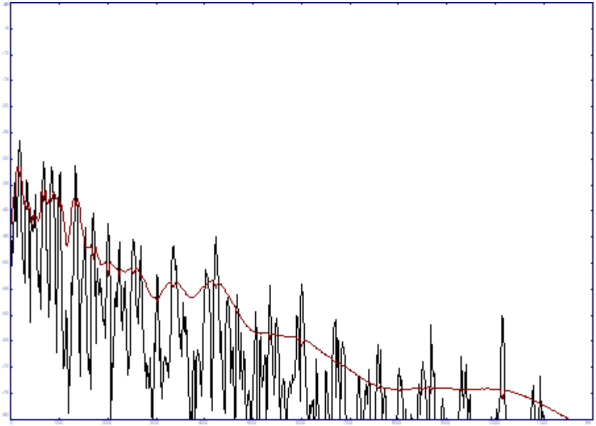


Existing Model, using bark scale



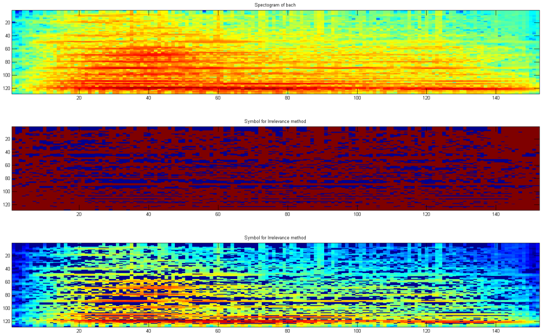
Psychoacoustic Masking : simultaneous masking I

Existing Model, using bark scale



Perceptual Sparsity by Irrelevance

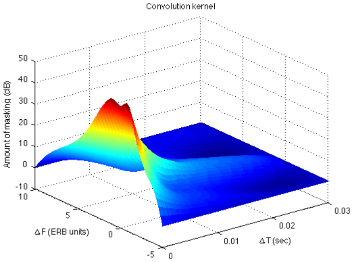
The irrelevance method calculates an adaptive threshold function for each spectra of a Gabor transform. This corresponds to an adaptive Gabor frame multiplier with coefficients in $\{0, 1\}$.



Perceptual Sparsity: Current and Future Work

Extend to True Time-Frequency Model using Multipliers:

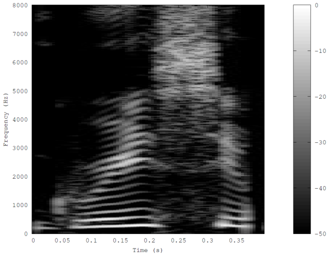
- Base it on ERBlets.
- Use new psychoacoustical data on time-frequency masking [Necciari et al., 2012].

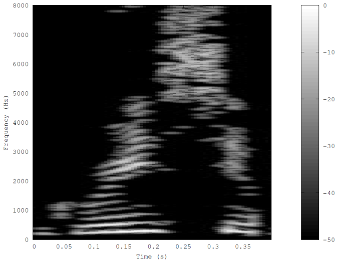


Use this for improved audio codec!

Perceptual Sparsity: Current and Future Work

Perceptual OMP:





OMP reduces to 400 atoms, masking removes another 73.

Frame theory is

- not only a very beautiful abstract setting,
- but also important for applications,
- in particular by linking it to human hearing!

Thank you for your attention!

- http://www.kfs.oeaw.ac.at
- http://magazine.orf.at/alpha/programm/2013/130513.htm

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.

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., Dörfler, M., Kowalski, M., and Torrésani, B. (2013).
Adapted and adaptive linear time-frequency representations: a synthesis point of view.
IEEE Signal Processing Magazine (special issue: Time-Frequency Analysis and Applications), to appear:–.

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

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

References: II

Frames for Psychoacoustics

Peter Balazs

ARI

Frame Theory

Multipliers

Perceptual Sparsity by Irrelevance

Conclusions

References

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.

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 Søndergaard, P. (2013).
The ERBlet transform: An auditory-based time-frequency representation with perfect reconstruction.
In Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), pages 498–502, Vancouver, Canada. IEEE.

Necciari, T., Balazs, P., Kronland-Martinet, R., Ystad, S., Laback, B., Savel, S., and Meunier, S. (2012).
Auditory time-frequency masking: Psychoacoustical data and application to audio representations.
volume 7172 LNCS, pages 146–171.

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

References: III

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.

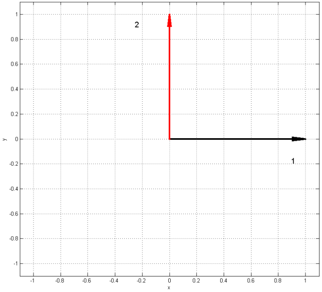
Stoeva, D. T. and Balazs, P. (2012).
Invertibility of multipliers.
Applied and Computational Harmonic Analysis, 33(2):292–299.

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.

Orthonormal Basis (ONB)

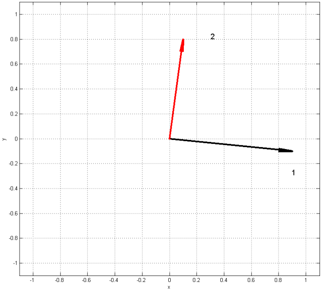
Standard aproach: orthonormal basis.



Problems:

- Perturbation
- Construction
- Error Robustness

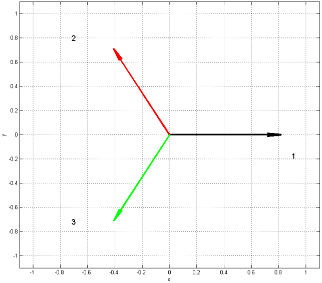
Riesz bases



Problems:

- ~~Perturbation~~
- Construction
- Error Robustness

Alternate approach: introduce redundancy.



Problems:

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

Back