RUHR-UNIVERSITÄT BOCHUM

RUB

Noise Reduction for Hearing Aids:
Enabling Communication in Adverse Conditions

Rainer Martin

October 22, 2013

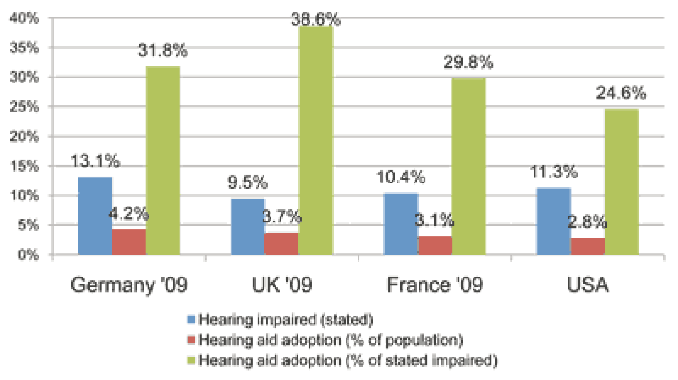
Institute of
Communication Acoustics

Communication in Adverse Acoustic Conditions

Source: http://cdsweb.cern.ch, accessed on Oct 28, 2012

Hearing Loss Prevalence Data

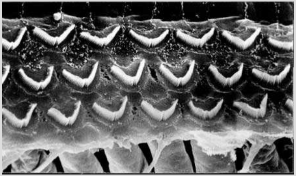
Hearing loss prevalence and hearing aid adoption rates, based on stated hearing loss on the screening survey.



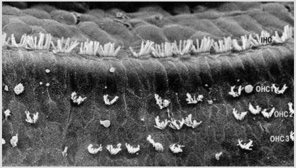
Source: http://www.hearingreview.com/issues/articles/2011-02-01.asp

Hearing Loss and its Consequences I

healthy outer hair cells



damaged outer hair cells



Source: http://dontlosethemusic.co.nz, accessed on Sept 16, 2013

- Increase of the threshold of hearing
  - soft sounds are not heard anymore
  - speech intelligibility (even without additional noise) is insufficient
  - compensation via strong amplification (up to 70 dB) without exceeding the loudness discomfort level (LDL)
  - target amplification is derived from fitting rules.

Hearing Loss and its Consequences II

- Reduction of spectral and/or temporal resolution in the inner ear
  - speech sounds are loud enough but not intelligible
  - speech communication in noisy environments is severely degraded
  - direct compensation of these effects is not possible

- Speech enhancement / noise reduction pre-processing is very important for successful rehabilitation!

Hearing Aids

Sources: Siemens Audiologische Technik, Oticon, varibel

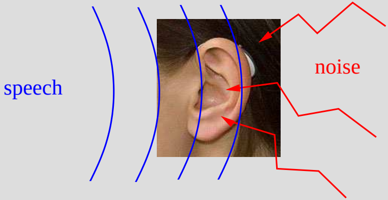
Open-Fit Hearing Aids

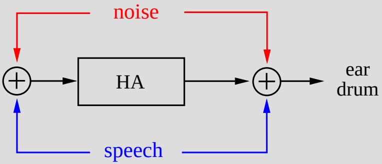
- Open-fit devices
  - are best for mild to moderate hearing loss with good residual hearing at low frequencies,
  - are comfortable to wear,
  - improve own voice reproduction,
  - require powerful feedback cancellation.

Source: www.lloydhearingaid.com

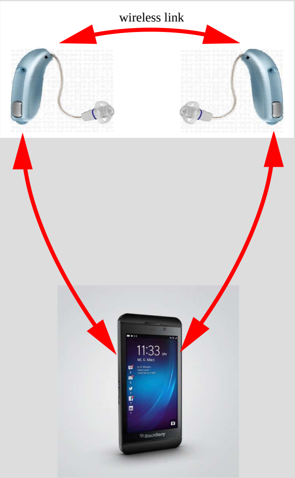
Open-Fit Signal Model

- Open-fit devices
  - require very short processing latency,
  - may be less effective in high levels of ambient noise
  - a case for active noise control?
    see [Dalga and Doclo 2013].





Wireless Connectivity



- Binaural link for the exchange of settings and parameters
- Full audio bandwidth is desired
- Audio streaming via wireless relay
- Streaming directly from a smartphone to hearing aids
- Full bi-directional signal transmission using sensors and computational power of the smartphone

Challenges

Sources: Blackberry, Nokia, Siemens, 2010

- Users expect effortless communication in complex acoustic environments
  - many spatially distributed sources
  - non-stationary, non-Gaussian signals
  - ambient noise and reverberation
  - time-varying signal paths
  - very long impulse responses
- This requires optimization of both intelligibility and quality.
- Hardware restrictions
  - very small size of device
  - very low latency < 10 ms
  - very low power < 1 mW

Outline

Spectral Analysis and Synthesis

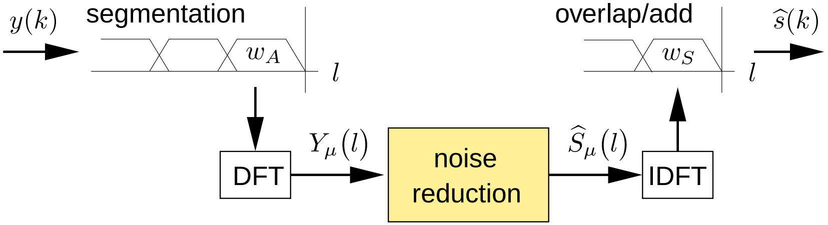
Requirements of noise reduction:

- High energy compaction of target signal
  - High spectral resolution of harmonics for voiced speech
    $\Rightarrow$ good separation of speech and noise
  - High temporal resolution for transient sounds
    $\Rightarrow$ accurate reproduction of transient speech sounds
- High stop-band attenuation
- Perfect reconstruction
- Low algorithmic delay
- High computational efficiency

Spectral Analysis / Synthesis

- DFT and uniform filter banks, e.g. [Griffin and Lim 1984]
  - high-resolution
  - perfect reconstruction
  - highly efficient
- Non-uniform filter banks, e.g. [Hohmann 2002]
  - resolution according to perceptual model
  - near-perfect reconstruction
- Low-delay filter-bank equalizer,
  e.g. [Löllmann and Vary 2005], [Vary 2006], [Löllmann and Vary 2008]
- Eigenvalue / eigenvector decomposition,
  e.g. [Ephraim and van Trees 1995]
  - signal adaptive / optimal
  - computationally expensive

Overlap-Add Analysis and Synthesis

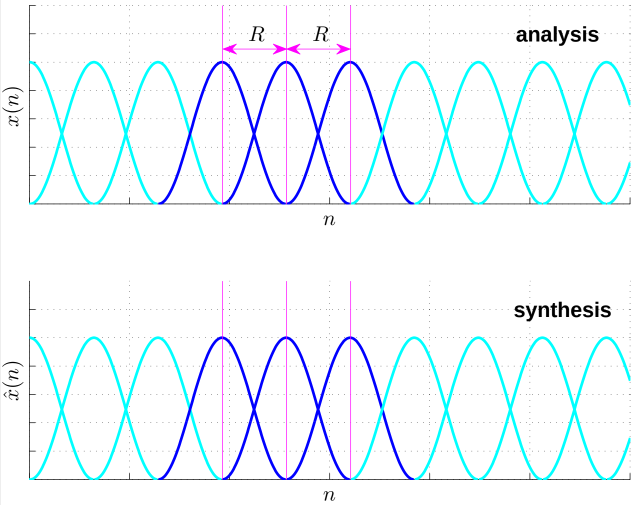


To achieve *perfect reconstruction* the product of these window functions must satisfy the *constant-overlap-add* constraint

$$\sum_{k=-\infty}^{\infty} w_A(n-R)w_S(n-R) = 1$$

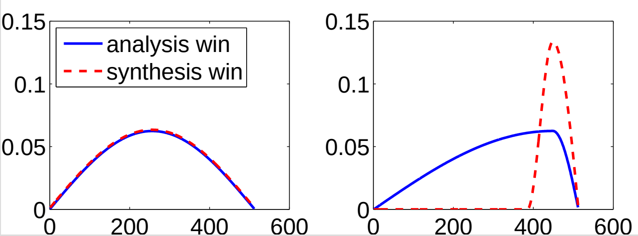
where R is the block shift of the windows.

Overlap-Add with Symmetric Windows



Low Latency Spectral Analysis / Synthesis

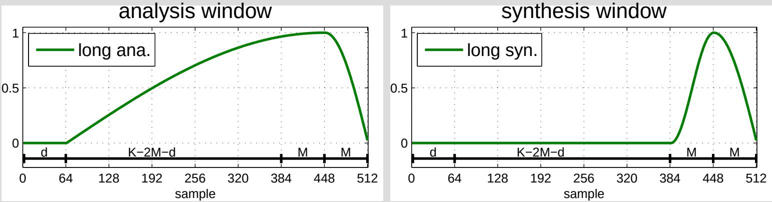
- Latency is identical to the length of the synthesis window
- Use non-symmetric analysis window and short window for synthesis
- Family of non-symmetric windows
  - right-hand side of all analysis and all synthesis windows is identical
  - left-hand side is variable
  - use different windows for different speech sounds



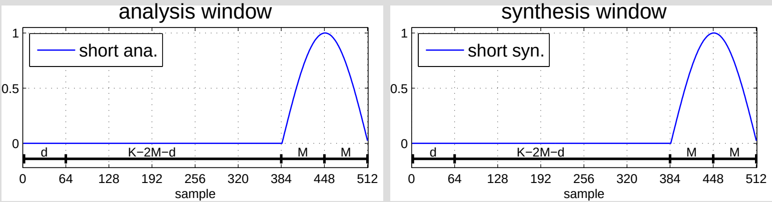
DFT: [Mauler and Martin 2007, 2009, 2010], CQT: [Nagathil and Martin, 2012]

Low Latency Spectral Analysis / Synthesis with Adaptive Resolution

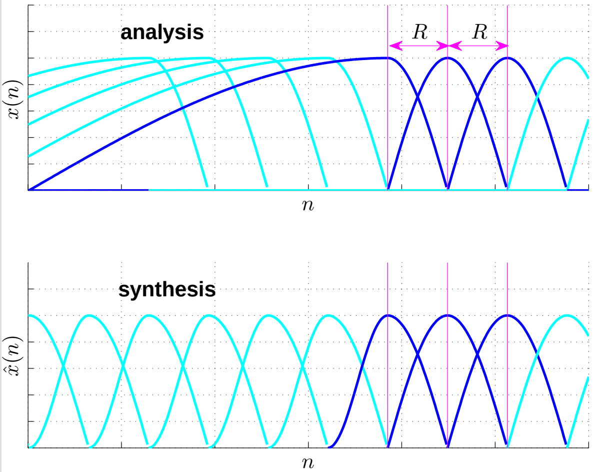
- High spectral resolution required:



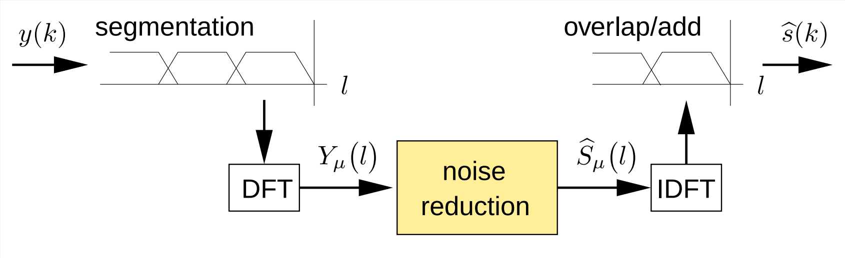
- High temporal resolution required:



Adaptive Window Switching



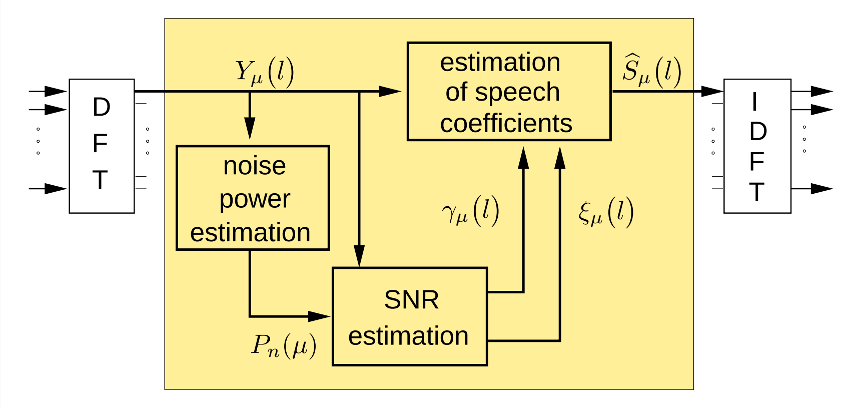
Single-channel Noise Reduction



In the DFT domain we have:

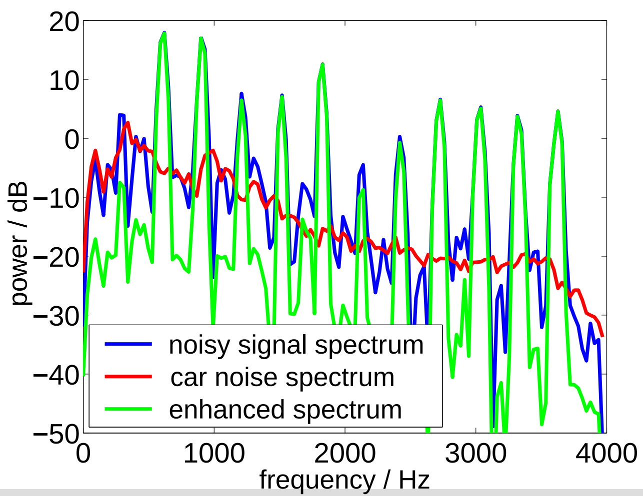
- Noisy speech: $Y_\mu(l) = S_\mu(l) + N_\mu(l)$
  - frequency index μ
  - time index l
- Estimated speech coefficient: $\widehat{S}_\mu(l) = f(Y_\mu(l))$

Noise Reduction: Basic Tasks



$P_n(\mu)$: noise power $\gamma_\mu(l)$: *a posteriori* SNR $\xi_\mu(l)$: *a priori* SNR

Principle of Single Channel Noise Reduction



Postprocessing in the Cepstrum Domain for the Reduction of Musical Noise

Definition of the **real-valued cepstrum**:

$$c_y(q) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln\left(\left|Y(e^{j\Omega})\right|\right) e^{j\Omega q} \mathrm{d}\Omega$$

where $Y(e^{j\Omega})$ is the spectrum of time domain signal $y(i)$.

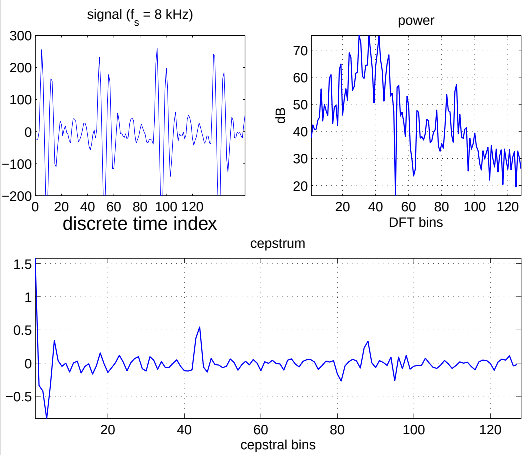
Some (strange) terminology: cepstrum, quefrency, rahmonic, ...

[B.P. Bogert, M.J.R. Healy and J.W. Tukey, 1963]

The cepstrum is very well suited to group speech components:

- coarse spectral features (envelope),
- harmonic structure, and
- fine structure of spectrum.

Cepstrum of a Voiced Speech Sound



Temporal Cepstrum Smoothing

Principle [Breithaupt, Gerkmann, Martin, IEEE Signal Proc. Lett. 2007] :

- separation of coarse and fine spectral features
- relatively strong smoothing of spectral fine structure
- relatively little smoothing of coarse spectral structures.

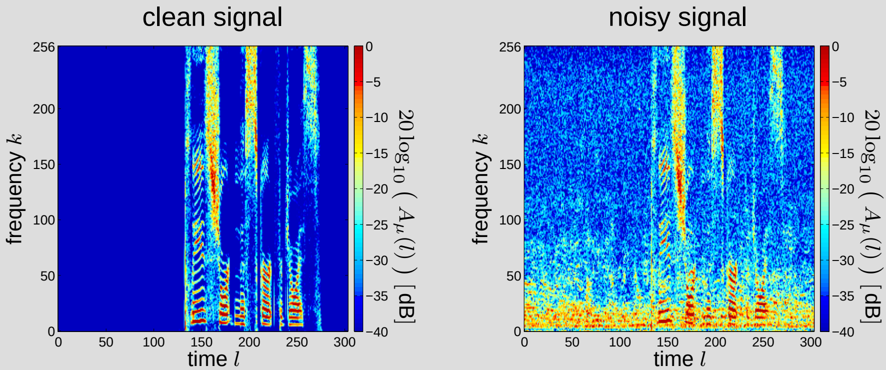
Advantages with respect to other smoothing methods:

- reduction of variance of residual noise
- negligible impact on speech signal
- preservation of harmonic spectral structure of voiced speech.

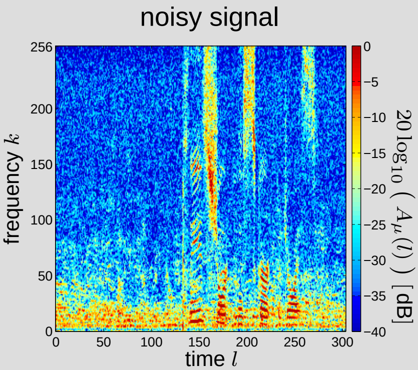
Applications:

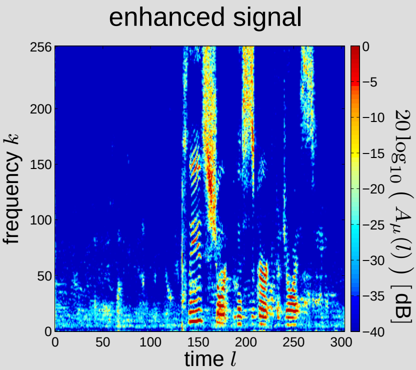
- single channel noise red. [Breithaupt, Gerkmann, Martin 2008]
- blind source separation [Madhu, Breithaupt and Martin 2008]
- automatic speech recognition [Breithaupt and Martin 2006, 2008]
- binaural dereverberation [Gerkmann 2011]

NR Example: Speech in Babble Noise



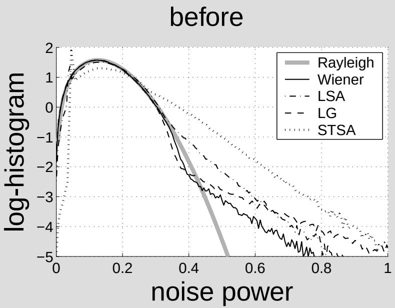
NR Example: Speech in Babble Noise





Analysis of Spectral Outliers

Log-histogram of residual noise before and after noise reduction for various estimators and white Gaussian noise:

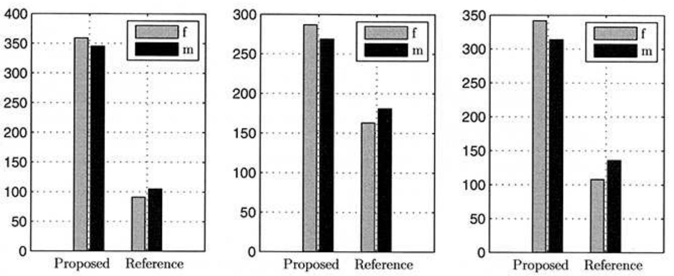


- Heavy tails result in unnatural fluctuations!
- Smoothing in the cepstro-temporal domain results in a significant reduction

Evaluation: Preference Test

- Proposed algorithm with adaptive window switching, temporal cepstrum smoothing, amplification of transient sounds
- Reference algorithm with standard components.
- Male and female speakers, 3 noise types, 3 SNRs (5, 10, 15 dB)
- 27 normally-hearing listeners

speech quality noise quality overall quality



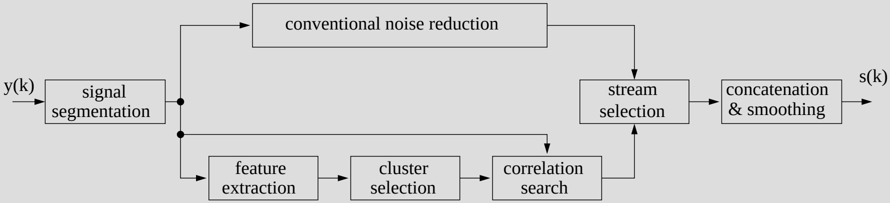
[Mauler 2010]

Performance of Single Channel NR

- Many studies, e.g. [Dahlquist et al., 2005] and [Luts et al. 2010], have shown that generic single channel methods
  - show SNR improvements in of 2-12 dB,
  - improve subjective quality,
  - reduce listener fatigue,
  - **but do not** improve intelligibility
- Improvements of about 1-2 dB are reported for CI users.
- Improvements of intelligibility are reported for the **ideal and estimated binary masks** [Hu and Wang, 2001, [Kim et al. 2008], [Healy et al., 2013].
- Alternative approach: synthesis using corpus of clean signal segments.

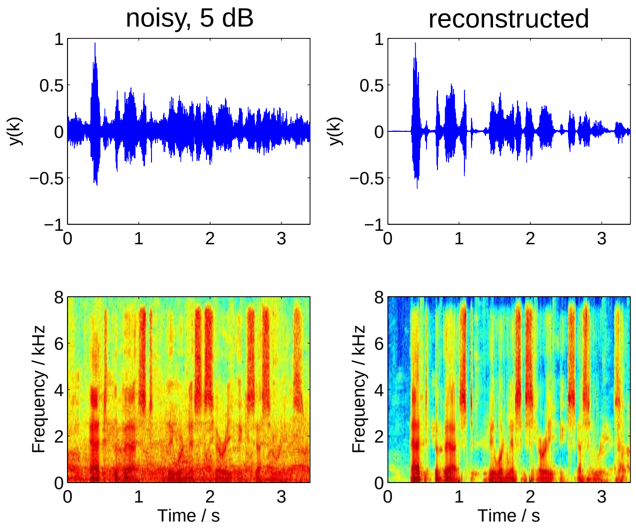
Corpus-based Speech Enhancement

- Conventional noise reduction systems suffer from
  - insufficient noise attenuation and/or
  - target signal distortions
- Idea: Resynthesize speech from clean speech segments
  [Xiao and Nickel, 2010]



[Nickel et al., 2013], also with audio-visual front-end processing: [Kolossa et al., 2012]

Example: Speech + Babble noise

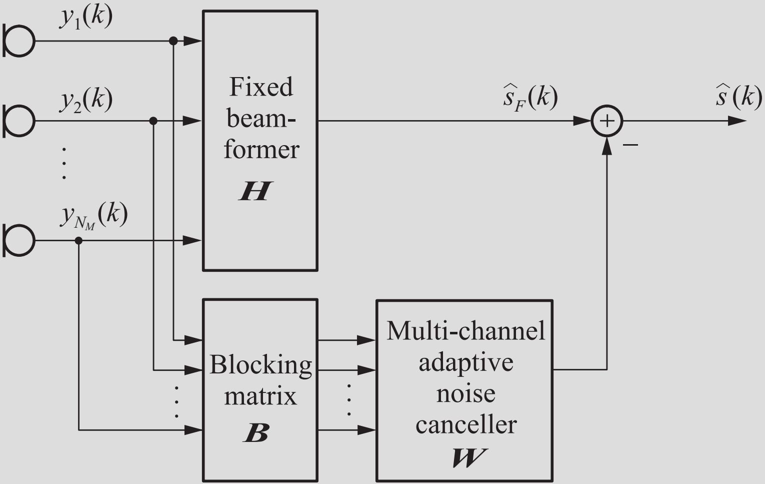


From Single- to Multi-channel Processing

- Performance of single-channel algorithms is limited because only temporal and spectral information can be used.
- Multichannel systems allow to exploit spatial information such as location of sources and spatial sound field statistics.
- Sensors can be distributed in space for improved signal pick-up.
- Many successful multichannel approaches:
  - (Adaptive) differential microphones, e.g. [Elko and Pong, 1997]
  - MVDR beamforming, e.g. [Cox et al. 1986]
  - Generalized sidelobe canceler, e.g. [Griffiths and Lim, 1982] , [Gannot et al., 2001]
  - Blind source separation, e.g. [Araki, Makino et al. 2003, ...], [Buchner, Aichner, Kellermann 2003, ...]
  - Speech distortion weighted multi-channel Wiener filter, e.g. [Doclo et al. 2005], [van den Bogaert et al. 2009]

Adaptive Beamformer

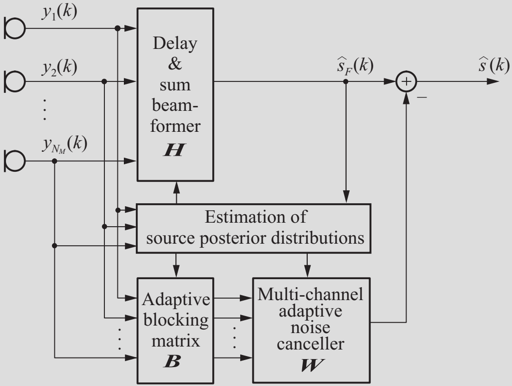
Basic *Generalized Sidelobe Canceler* (GSC) [Griffiths and Jim, 1982]:



Parsimonious Excitation-based Generalized Sidelobe Canceller (PEG)

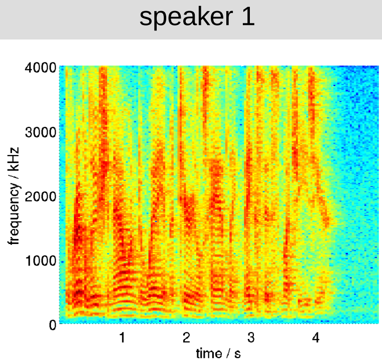
Extraction of source signal q using PEG:

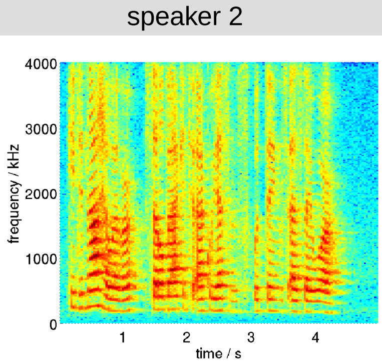
$$\hat{S}_{\mu}^{\langle q \rangle}(l) = (\boldsymbol{H}_{\mu}^{\langle q \rangle}(l))^{H} \boldsymbol{Y}_{\mu}(l) - (\boldsymbol{W}_{\mu}^{\langle q \rangle}(l))^{H} \boldsymbol{B}_{\mu}^{\langle q \rangle}(l) \boldsymbol{Y}_{\mu}(l)$$



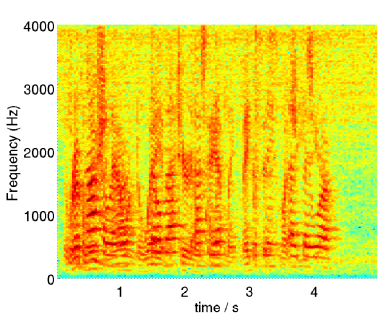
[Madhu and Martin, 2011]

Spectrograms of Speaker 1 and 2





Mixed Signal with Ambient Noise

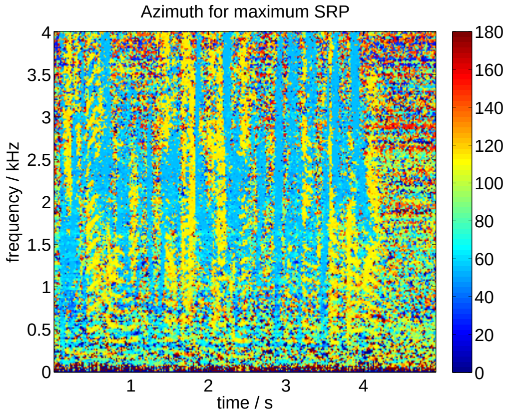


- Signal separation in two steps:
  - Source localization via steered response power (SRP-PHAT) [DiBiase et al. 2001]
  - Target signal extraction using parsimonious excitation-based GSC [Madhu and Martin 2011]

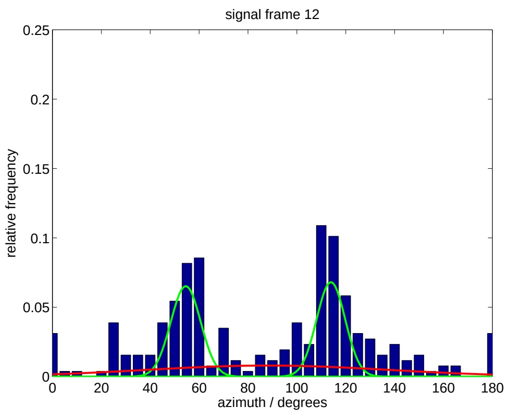
Optimal Azimuth per TF-Bin

- Optimize SRP-PHAT cost function in each time-frequency (TF) bin:

$$\widehat{\theta}_\mu(l) = \underset{\theta}{\operatorname{argmax}} J_\mu(l,\theta)$$

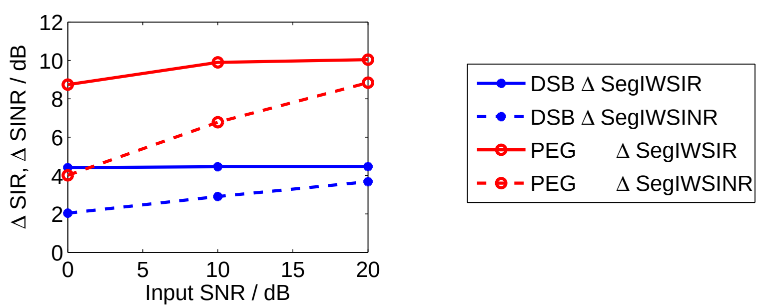


Azimuth Histogram for Single Frame



- Estimation of source posterior distributions via the *expectation-maximization* algorithm

SegIWSIR and SegIWSINR vs. Input SNR



ΔSegIWSIR: Segmental Intelligibility Weighted Signal to Interference Ratio improvement

ΔSegIWSINR: Segmental Intelligibility Weighted Signal to Interference plus Noise Ratio improvement

Summary

- Modern hearing systems are highly complex signal processing devices
- Signal enhancement is at the core of speech processing tasks in hearing aids
  - single- and multi-channel noise reduction
  - microphone array processing and source separation
- The challenge continues ...
  - enable effortless speech communication for normal-hearing people and people with a hearing loss,
  - find low complexity / low power / low latency implementations.
- New solutions and opportunities arise from
  - including more *a priori* knowledge about speech and hearing
  - the availability of sensor networks and
  - inclusion of top-down cognitive processes.

Contributors and Acknowledgments

Dr.-Ing. Colin Breithaupt, Prof. Dr.-Ing. Timo Gerkmann
Dr.-Ing. Dirk Mauler, Dr.-Ing. Nilesh Madhu
Dipl.-Ing. Anil Nagathil
Prof. Robert Nickel, Ph.D., Prof. Dr.-Ing. Dorothea Kolossa

These works were sponsored by grants from DFG, the EU FP7 project HEARCOM, and the EU FP7 Marie Curie project InventHI.

The Future of Hearing Aids?

Source: www.bioaid.org.uk, accessed on Sept 20, 2013

References I

S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada.
Blind separation of more speech than sensors with less distortion by combining sparseness and ica.
In *Proc. Intl. Workshop Acoustic Echo and Noise Control (IWAENC)*, pages 271–274, 2003.

H. Buchner, R. Aichner, and W. Kellermann.
A Generalization of a Class of Blind Source Separation Algorithms for Convolutive Mixtures.
In *Proc. Int. Symp. Independent Component Analysis*, 2003.

C. Breithaupt, T. Gerkmann, and R. Martin.
Cepstral Smoothing of Spectral Filter Gains for Speech Enhancement without Musical Noise.
IEEE Signal Proc. Letters, 14(12):1036–1039, 2007.

C. Breithaupt, T. Gerkmann, and R. Martin.
A Novel A Priori SNR Estimation Approach Based on Selective Cepstro-Temporal Smoothing.
In *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pages 4897–4900, 2008.

B.P. Bogert, M.J.R. Healy, and J.W. Tukey.
The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking.
In *Proc. of the Symposium on Time Series Analysis*, pages 209–243, 1963.

References II

C. Breithaupt and R. Martin.
Statistical Analysis and Performance of DFT Domain Noise Reduction Filters for Robust Speech Recognition.
In *Proc. 9th International Conference on Spoken Language Processing (ICSLP)*, pages 365–368, 2006.

C. Breithaupt and R. Martin.
DFT-based Speech Enhancement for Robust Automatic Speech Recognition.
In *Proc. ITG-Conference on Voice Communication (Sprachkommunikation)*, Aachen, 2008.

H. Cox, R.M. Zeskind, and T. Kooij.
Practical Supergain.
IEEE Trans. Acoustics, Speech and Signal Processing, 34(3):393–398, June 1986.

D. Dalga and S. Doclo.
Influence of Secondary Path Estimation Errors on the Performance of ANC-Motivated Noise Reduction Algorithms for Hearing Aids.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, 2013.

M. Dahlquist, M.E. Lutman, S. Wood, and A. Leijon.
Methodology for quantifying perceptual effects from noise suppression systems.
Int. J. of Audiology, 44:721–732, 2005.

References III

J.H. DiBiase, H.F. Silverman, and M.S. Brandstein.
Robust Localization in Reverberant Rooms.
In M. Brandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, Berlin, 2001.

S. Doclo, A. Spriet, J. Wouters, and M. Moonen.
Speech Distortion Weighted Multi-channel Wiener Filtering Techniques for Noise Reduction.
In J. Benesty, S. Makino, and J. Chen, editors, *Speech Enhancement*, chapter 9, pages 199–228. Springer-Verlag, 2005.

G.W. Elko and A.-T Nguyen Pong.
A Steerable and Variable First-Order Differential Microphone Array.
In *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pages 223–226, 1997.

Y. Ephraim and H. Van Trees.
A Signal Subspace Approach for Speech Enhancement.
IEEE Trans. Speech and Audio Processing, 3(4):251–266, July 1995.

S. Gannot, D. Burshtein, and E. Weinstein.
Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech.
IEEE Trans. Signal Processing, 49(8):1614–1626, 2001.

References IV

T. Gerkmann.
Cepstral weighting for speech dereverberation without musical noise.
In *Proc. European Signal Processing Conference (EUSIPCO)*, pages 2309–2313, 2011.

L. J. Griffiths and C. W. Jim.
An Alternative Approach to Linearly Constrained Adaptive Beamforming.
IEEE Transactions on Antennas and Propagation, 30(1):27–34, 1982.

D.W. Griffin and J.S. Lim.
Signal Estimation from Modified Short-Time Fourier Transform.
IEEE Trans. Acoustics, Speech and Signal Processing, 32(2):236–243, April 1984.

V. Hohmann.
Frequency Analysis and Synthesis using a Gammatone Filterbank.
Acta Acoustica united with Acoustica, 88(3), 2002.

G. Hu and D.L. Wang.
Speech Segregation Based on Pitch Tracking and Amplitude Modulation.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 79–82, 2001.

E.W. Healy, S.E. Yoho, Y. Wang, and D.L. Wang.
An Algorithm to Improve Speech Recognition in Noise for Hearing-impaired Listeners.
J. Acoust. Soc. Am., 134:3029–3038, 2013.

References V

G. Kim, Y. Lu, Y. Hu, and P.C. Loizou.
An Algorithm that Improves Speech Intelligibility in Noise for Normal-hearing Listeners.
J. Acoust. Soc. Am., 126(3):1486 – 1494, 2009.

D. Kolossa, R.M. Nickel, S. Zeiler, and R. Martin.
Inventory-based audio-visual speech enhancement.
In *Proc. Interspeech*, September 2012.

H. Luts, K. Eneman, J. Wouters, M. Schulte, M. Vormann, M. Buechler, N. Dillier, R. Houben, W. Dreschler, M. Froehlich, H. Puder, G. Grimm, V. Hohmann, A. Leijon, A. Lombard, D. Mauler, and A. Spriet.
Multicenter evaluation of signal enhancement algorithms for hearing aids.
Journal of the Acoustical Society of America (JASA), 127(3):1491–1505, March 2010.

H.W. Löllmann and P. Vary.
Generalized Filter-Bank Equalizer for Noise Reduction with Reduced Signal Delay.
In *European Conference on Speech Communication and Technology (INTERSPEECH)*, September 2005.

H.W. Löllmann and P. Vary.
Low Delay Filter-Banks for Speech and Audio Processing.
In E. Hänsler and G. Schmidt, editors, *Speech and Audio Processing in Adverse Environments*, chapter 2, pages 13 – 61. Springer-Verlag, August 2008.

References VI

N. Madhu, C. Breithaupt, and R. Martin.
Temporal Smoothing of Spectral Masks in the Cepstral Domain for Speech Separation.
In *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pages 45 – 48, 2008.

D. Mauler and R. Martin.
A Low Delay, Variable Resolution, Perfect Reconstruction Spectral Analysis-Synthesis System for Speech Enhancement.
In *Proc. Euro. Signal Processing Conf. (EUSIPCO)*, 2007.

D. Mauler and R. Martin.
Improved Reproduction of Stops in Noise Reduction Systems with Adaptive Windows and Non-Stationarity Detection.
Journal on Advances in Signal Processing, 2009.
in special issue on Digital Signal Processing for Hearing Instruments.

D. Mauler and R. Martin.
Optimization of Switchable Windows for Low-Delay Spectral-Analysis-Synthesis.
In *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2010.

N. Madhu and R. Martin.
A Versatile Framework for Speaker Separation Using a Model-Based Speaker Localization Approach.
IEEE Trans. Audio, Speech and Language Processing, 19(7):1900–1912, 2011.

References VII

R.M. Nickel, R. Fernandez Astudillo, D. Kolossa, and R. Martin.
Corpus-Based Speech Enhancement with Uncertainty Modeling and Cepstral Smoothing.
IEEE Trans. Audio, Speech and Language Processing, 21(5):983 – 997, 2013.

A. Nagathil and R. Martin.
Optimal Signal Reconstruction from a Constant-Q Spectrum.
In *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pages 349–352, 2012.

P. Vary.
An Adaptive Filterbank Equalizer for Speech Enhancement.
SIGPROC, 86(6):1206 – 1214, 2006.

T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen.
Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids.
J. Acoust. Soc. Am., 125(1):360 – 371, 2009.

X. Xiao and R.M. Nickel.
Speech Enhancement with Inventory Style Speech Resynthesis.
IEEE Trans. Audio, Speech and Language Processing, 18(6):1243 – 1257, 2010.