Noise Reduction for Hearing Aids:
Enabling Communication in Adverse Conditions

Rainer Martin
October 22, 2013
Communication in Adverse Acoustic Conditions

Hearing Loss Prevalence Data

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

Hearing Loss and its Consequences I

- 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

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

Sources: Blackberry, Nokia, Siemens, 2010
Outline

1. Introduction
2. Spectral Analysis and Synthesis
3. Single Channel Noise Reduction
4. Multi-channel Speech Enhancement
5. Summary
Spectral Analysis and Synthesis

Requirements of noise reduction:

- High energy compaction of target signal
  - High spectral resolution of harmonics for voiced speech
    ⇒ good separation of speech and noise
  - High temporal resolution for transient sounds
    ⇒ 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

\[ x(n) \times (n) \times (n) \]

\[ \tilde{x}(n) \]

**analysis**

**synthesis**
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

Low Latency Spectral Analysis / Synthesis with Adaptive Resolution

- High spectral resolution required:
  - High temporal resolution required:
Adaptive Window Switching

Analysis

\[ x(n) \]

\[ n \]

Synthesis

\[ \hat{x}(n) \]

[Diagram showing analysis and synthesis with windows and their corresponding positions]
Single-channel Noise Reduction

In the DFT domain we have:

- **Noisy speech:** \( Y_\mu (l) = S_\mu (l) + N_\mu (l) \)
  - frequency index \( \mu \)
  - time index \( l \)
- **Estimated speech coefficient:** \( \hat{S}_\mu (l) = f(Y_\mu (l)) \)
Noise Reduction: Basic Tasks

$P_n(\mu)$: noise power  \quad $\gamma_\mu(l)$: a posteriori SNR  \quad $\xi_\mu(l)$: a priori SNR
Principle of Single Channel Noise Reduction

Graph showing the spectrum of a noisy signal, car noise, and an enhanced spectrum, with frequency in Hz on the x-axis and power in dB on the y-axis.
Postprocessing in the Cepstrum Domain for the Reduction of Musical Noise

Definition of the real-valued cepstrum:

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

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

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


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

Signal ($f_s = 8$ kHz) vs. DFT bins

Cepstrum vs. cepstral bins

Power vs. DFT bins
Temporal Cepstrum Smoothing

- 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

**Clean Signal**

- Frequency $k$
- Time $l$
- $20 \log_{10} (A_{\mu l})$ [dB]

**Noisy Signal**

- Frequency $k$
- Time $l$
- $20 \log_{10} (A_{\mu l})$ [dB]
NR Example: Speech in Babble Noise

noisy signal

enhanced signal
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]
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.
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

noisy, 5 dB

reconstructed

Introduction Analysis/Synthesis Single Channel NR Multi-Channel Summary Rainer Martin
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]:

\[
\begin{align*}
y_1(k) \\
y_2(k) \\
\vdots \\
y_{NM}(k)
\end{align*}
\]

Fixed beamformer \( H \)

Blocking matrix \( B \)

Multi-channel adaptive noise canceller \( W \)

\( \hat{s}_F(k) \)

\( \hat{s}(k) \)
Extraction of source signal $q$ using PEG:

$$\hat{S}_\mu^{(q)}(l) = (H_\mu^{(q)}(l))^H Y_\mu(l) - (W_\mu^{(q)}(l))^H B_\mu^{(q)}(l) Y_\mu(l)$$
Spectrograms of Speaker 1 and 2

speaker 1

speaker 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:

\[
\hat{\theta}_\mu(l) = \arg\max_\theta J_\mu(l, \theta)
\]
Azimuth Histogram for Single Frame

- Estimation of source posterior distributions via the *expectation-maximization* algorithm
SegIWSIR and SegIWSINR vs. Input SNR

\[ \Delta \text{SegIWSIR} \]: Segmental Intelligibility Weighted Signal to Interference Ratio improvement

\[ \Delta \text{SegIWSINR} \]: 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?
Blind separation of more speech than sensors with less distortion by combining sparseness and ica.

H. Buchner, R. Aichner, and W. Kellermann.

C. Breithaupt, T. Gerkmann, and R. Martin.

C. Breithaupt, T. Gerkmann, and R. Martin.
A Novel A Priori SNR Estimation Approach Based on Selective Cepstro-Temporal Smoothing.

The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking.
References II

C. Breithaupt and R. Martin.

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.

D. Dalga and S. Doclo.

Methodology for quantifying perceptual effects from noise suppression systems.
Robust Localization in Reverberant Rooms.

S. Doclo, A. Spriet, J. Wouters, and M. Moonen.
Speech Distortion Weighted Multi-channel Wiener Filtering Techniques for Noise Reduction.

G.W. Elko and A.-T Nguyen Pong.
A Steerable and Variable First-Order Differential Microphone Array.

Y. Ephraim and H. Van Trees.
A Signal Subspace Approach for Speech Enhancement.

S. Gannot, D. Burshtein, and E. Weinstein.
Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech.
References IV

T. Gerkmann.
Cepstral weighting for speech dereverberation without musical noise.

L. J. Griffiths and C. W. Jim.
An Alternative Approach to Linearly Constrained Adaptive Beamforming.

D.W. Griffin and J.S. Lim.
Signal Estimation from Modified Short-Time Fourier Transform.

V. Hohmann.
Frequency Analysis and Synthesis using a Gammatone Filterbank.

Speech Segregation Based on Pitch Tracking and Amplitude Modulation.

E.W. Healy, S.E. Yoho, Y. Wang, and D.L. Wang.
An Algorithm to Improve Speech Recognition in Noise for Hearing-impaired Listeners.
G. Kim, Y. Lu, Y. Hu, and P.C. Loizou.
An Algorithm that Improves Speech Intelligibility in Noise for Normal-hearing Listeners.

Inventory-based audio-visual speech enhancement.

Multicenter evaluation of signal enhancement algorithms for hearing aids.

H.W. Löllmann and P. Vary.
Generalized Filter-Bank Equalizer for Noise Reduction with Reduced Signal Delay.

H.W. Löllmann and P. Vary.
Low Delay Filter-Banks for Speech and Audio Processing.
N. Madhu, C. Breithaupt, and R. Martin.
Temporal Smoothing of Spectral Masks in the Cepstral Domain for Speech Separation.

D. Mauler and R. Martin.

D. Mauler and R. Martin.
Improved Reproduction of Stops in Noise Reduction Systems with Adaptive Windows and Non-Stationarity Detection.
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.

N. Madhu and R. Martin.
A Versatile Framework for Speaker Separation Using a Model-Based Speaker Localization Approach.
Corpus-Based Speech Enhancement with Uncertainty Modeling and Cepstral Smoothing. 

A. Nagathil and R. Martin. 
Optimal Signal Reconstruction from a Constant-Q Spectrum. 

P. Vary. 
An Adaptive Filterbank Equalizer for Speech Enhancement. 

Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids. 

X. Xiao and R.M. Nickel. 
Speech Enhancement with Inventory Style Speech Resynthesis. 