Binaural speech enhancement and cue preservation algorithms

Prof. Dr. Simon Doclo
University of Oldenburg, Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4all

ELOBES Workshop – January 12, 2019
Introduction

- Hearing impaired suffer from a loss of speech understanding in adverse acoustic environments with competing speakers, background noise and reverberation

*Apply acoustic signal pre-processing techniques* in order to improve speech quality and intelligibility
Hearing impaired suffer from a loss of speech understanding in adverse acoustic environments with competing speakers, background noise and reverberation.

**Multiple microphones available** → spatial + spectral processing
This presentation:
- **Binaural noise reduction algorithms** based on minimum variance distortionless response (MVDR) beamformer and multi-channel Wiener filter (MWF)
- Integration with **external microphone**

Main objectives of algorithms:
- Improve speech intelligibility and avoid signal distortions
- Preserve spatial awareness and directional hearing (binaural cues)
Binaural cues

- **Interaural Time/Phase Difference (ITD/IPD)**
- **Interaural Level Difference (ILD)**
- **Interaural Coherence (IC)**

- ITD: $f < 1500 \text{ Hz}$, ILD: $f > 2000 \text{ Hz}$
- IC: describes spatial characteristics, e.g. perceived width, of diffuse noise, and determines when ITD/ILD cues are *reliable*.

- Binaural cues, in addition to spectro-temporal cues, play an important role in **auditory scene analysis** (source localization/segregation) and **speech intelligibility**.

[Bronkhorst, Plomp, 1988] [Wightman, Kistler, 1992] [Blauer, 1997] [Hawley et al., 2004] [Faller, Merimaa, 2004] [Dietz et al., 2011]
Binaural noise reduction
Binaural hearing aid configuration:

- Two hearing aids with in total $M$ microphones
- All microphone signals $Y$ are assumed to be available at both hearing aids (perfect wireless link)

Apply a filter $W_0$ and $W_1$ at the left and the right hearing aid, generating binaural output signals $Z_0$ and $Z_1$

$$Z_0(\omega) = W_0^H(\omega)Y(\omega), \quad Z_1(\omega) = W_1^H(\omega)Y(\omega)$$
Binaural noise reduction: Two main paradigms

Spectral post-filtering (based on multi-microphone noise reduction)

Binaural spatial filtering techniques

- Binaural cue preservation
- Possible single-channel artifacts

- Larger noise reduction performance
- Merge spatial and spectral post-filtering
- Binaural cue preservation not guaranteed
Minimum-Variance-Distortionless-Response (MVDR) beamformer

**Goal:** minimize output noise power without distorting speech component in reference microphone signals

\[
\begin{align*}
\min_{W_0} & \quad W_0^H R_v W_0 \quad \text{subject to} \quad W_0^H A = A_0 \\
\min_{W_1} & \quad W_1^H R_v W_1 \quad \text{subject to} \quad W_1^H A = A_1
\end{align*}
\]

+ Noise reduction
+ Distortionless constraint

**Requires** estimate/model of noise coherence matrix (e.g. diffuse) and estimate/model of relative transfer function (RTF) of target speech source

Multi-channel Wiener Filter (MWF)

**Goal:** estimate speech component in reference microphone signals + trade off noise reduction and speech distortion

\[
J_{\text{MWF}}(W) = \mathcal{E} \left\{ \left\| \begin{bmatrix} X_0 - W_0^H X \\ X_1 - W_1^H X \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} W_0^H V \\ W_1^H V \end{bmatrix} \right\|^2 \right\}
\]

+ Speech distortion
+ Noise reduction

**Requires** estimate of speech and noise covariance matrices, e.g. based on VAD

Can be decomposed as binaural MVDR beamformer and spectral postfilter

---

Good noise reduction performance, what about binaural cues?

Binaural MVDR and MWF
Binaural cues (diffuse noise)

Note: MSC = Magnitude Squared Coherence
Binaural MVDR and MWF
Binaural cues (diffuse noise)

Binaural cues for residual noise and interference in binaural MVDR/MWF are not preserved
Binaural noise reduction
Extensions for diffuse noise
Binaural MWF: Extensions for diffuse noise

Binaural MWF

- SNR improvement
- Binaural cues of speech source
- Binaural cues of noise

Interaural coherence preservation (MWF-IC)

Partial noise estimation (MWF-N)

\[ J_{MWF-IC}(W) = J_{MWF}(W) \frac{W_0^H R_v W_1}{\sqrt{W_0^H R_v W_0 W_1^H R_v W_1}} - IC_{des} \]

\[ J_{MWF-N}(W) = \mathcal{E} \left\{ \left[ \frac{X_0 - W_0^H X}{X_1 - W_1^H X} \right]^2 + \mu \left( \frac{\eta V_0 - W_0^H V}{\eta V_1 - W_1^H V} \right)^2 \right\} \]

- No closed-form solution, iterative optimization procedures required
- Closed-form solution

Trade-off between SNR improvement and binaural cue preservation, depending on parameters (\( \eta \) and \( \lambda \))

Binaural MWF-N: Partial noise estimation

- **Closed-form filter expression** → mixing of binaural MWF output signals and reference microphone signals

- \( \eta = 0 \) : binaural MWF (optimal noise reduction, but no cue preservation)
- \( \eta = 1 \) : reference microphone signals (perfect cue preservation, but no noise reduction)

Binaural MWF-N: Trade-off parameter $\eta$

- **Fixed broadband values** ($\eta = 0.1 \ldots 0.3$)
- **Frequency-dependent values** based on IC discrimination ability of human auditory system for diffuse noise

- IC discrimination ability depends on magnitude of reference IC
- **Boundaries on Magnitude Squared Coherence** ($\text{MSC} = |\text{IC}|^2$):
  - For $f < 500$ Hz ("large" IC): frequency-dependent MSC boundaries (blue)
  - For $f > 500$ Hz ("small" IC): fixed MSC boundary, e.g. 0.36 (red) or 0.04 (green)

Binaural MWF-N: Trade-off parameter $\eta$

- **Fixed broadband values** ($\eta = 0.1 \ldots 0.3$)
- **Frequency-dependent values** based on IC discrimination ability of human auditory system for diffuse noise
  - **Trade-off parameter** $\eta$ achieving desired MSC:
    - **MWF-N**: exhaustive search for optimal trade-off parameter
    - **MVDR-N** (i.e. special case of MWF-N):
      - Closed-form expression for optimal trade-off parameter
      - No spectral filtering as in MWF-N
    - **MVDR-N + spectral postfilter**
      - Not equivalent to MWF-N, but combining spatial and spectral filtering with closed-form expression for both filter and trade-off parameter

Binaural MWF-N: Instrumental evaluation

Office ($T_{60} \approx 700\text{ms}$), M=4 (binaural RIRs), recorded ambient noise, target speaker at -45°, 0 dB input iSNR (left hearing aid)

MVDR: anechoic RTF of target speaker (DOA known), diffuse spatial coherence matrix (from anechoic ATFs) / MWF = MVDR + postfilter (SPP-based)

[Marquardt and Doclo, *IEEE/ACM TASLP*, 2018]
Binaural MWF-N: Trade-off parameter $\eta$

- **Fixed broadband values** ($\eta = 0.1 \ldots 0.3$)
- **Frequency-dependent values** based on IC discrimination ability of human auditory system for diffuse noise
- **Frequency-dependent function** of MSC between noisy reference microphone signals and output signals of BMVDR beamformer

High MSC/SNR: more important to keep maximum noise reduction (BMVDR)

Low MSC/SNR: more important to preserve binaural cues (scaled input signals)

Subjective Evaluation: Test setup

- Binaural hearing aid recordings (M=4 mics) in cafeteria ($T_{60} \approx 1250 \text{ ms}$)
  - Target speaker at -35°
  - Realistic cafeteria ambient noise
- **Algorithms**: binaural MVDR and binaural MVDR-N with different trade-off parameters:
  - MVDR-IC
  - MVDR-MSC1: $\eta_{\text{max}}=0.7$, $\text{MSC}_{\text{min}}=0$
  - MVDR-MSC2: $\eta_{\text{max}}=1.0$, $\text{MSC}_{\text{min}}=0.1$
- **Subjective listening experiments**:
  - 11 normal-hearing subjects
  - SRT using Oldenburg Sentence Test (OLSA)
  - Spatial quality (diffuseness) using MUSHRA

---

Does binaural unmasking compensate for SNR decrease of cue preservation algorithms (MVDR-N)?

[Klug, Marquardt, Doclo, *Proc. ITG Speech Communication*, 2018]
Subjective Evaluation: Spatial quality (MUSHRA)

- Evaluate spatial difference between reference and output signal
- MVDR-N (IC/MSC) outperforms BMVDR
  - Trade-off parameters: MSC-based better than IC-based
  - Using MSC2 hardly any difference to input!

Binaural cue preservation for diffuse noise significantly improves spatial quality

[Klug, Marquardt, Doclo, Proc. ITG Speech Communication, 2018]
All algorithms show a highly significant speech reception threshold (SRT) improvement.

No significant SRT difference between BMVDR and MVDR-N (IC/MSC).

Subjective Evaluation: Speech intelligibility (SRT)

Binaural cue preservation for diffuse noise does not affect speech intelligibility.

Klug, Marquardt, Doclo, *Proc. ITG Speech Communication*, 2018
Binaural noise reduction
Extensions for interfering sources
Binaural MVDR: Extensions for interfering source

Relative transfer function (BMVDR-RTF)

\[
\min_{W_0, W_1} \left\{ W_0^H R_v W_0 + W_1^H R_v W_1 \right\} \\
\text{s.t. } W_0^H A = A_0, W_1^H A = A_1, \frac{W_0^H B}{W_1^H B} = \frac{B_0}{B_1}. 
\]

Interference rejection (BMVDR-IR)

\[
\min_{W_0} \left\{ W_0^H R_v W_0 \right\} \text{ s.t. } W_0^H A = A_0, W_0^H B = \eta B_0 \\
\min_{W_1} \left\{ W_1^H R_v W_1 \right\} \text{ s.t. } W_1^H A = A_1, W_1^H B = \eta B_1. 
\]

- SNR improvement
- Binaural cues of speech source
- Binaural cues of interferer
- Binaural cues of speech source and interfering source preserved
- Also binaural MWF-based versions (incl. spectral filtering) can be derived
- MSC of background not exactly preserved, RTF estimation difficult

Binaural noise reduction
Integration with external microphone(s)
Exploit the availability of one or more external microphones (acoustic sensor network) with hearing aids


Integrating external microphone(s) with hearing aid microphones may lead to:

- Improved noise reduction and binaural cue preservation performance
- Low-complexity method to estimate relative transfer function (RTF) of target speaker

\[ w_L = \frac{R_n^{-1}a_L}{a_L^HR_n^{-1}a_L}, \quad w_R = \frac{R_n^{-1}a_R}{a_R^HR_n^{-1}a_R} \]
Binaural noise reduction: External microphone(s)

- Including external microphone in **binaural MVDR-N beamformer** leads to:
  - Larger output **SNR** for same trade-off parameter $\eta$
  - Same output SNR with larger trade-off parameter $\eta \rightarrow$ **better cue preservation**

Starkey database with real-world recordings ($T_{60} \approx 620$ms), $M=4$, target speaker $S_1$, multi-talker babble noise, 0 dB input $i$SNR (right hearing aid)

MVDR: perfectly estimated noise correlation matrix, RTF of target speaker estimated using covariance whitening method

---

Binaural noise reduction: External microphone(s)

- **Estimate RTF of target speaker** to steer binaural MVDR beamformer
- **Spatial coherence (SC) method**: assume that *external microphone* is spatially separated from *HA microphones*, such that noise components in external microphone and HA microphones can be assumed to be uncorrelated

\[
\begin{align*}
\tilde{a}_L^{SCE} &= \frac{\tilde{R}_y e_E}{e_L^T \tilde{R}_y e_E}, \quad \tilde{a}_R^{SCE} = \frac{\tilde{R}_y e_E}{e_R^T \tilde{R}_y e_E} \\
\bar{w}_L^{SCE} &= \begin{bmatrix} \alpha \cdot [I_{2M}, \ 0_{2M \times 1}] \bar{w}_L \\
\alpha (1 + \beta) \cdot e_E^T \bar{w}_L \end{bmatrix}
\end{align*}
\]

Binaural noise reduction: External microphone(s)

- MVDR with external microphone (SCE) leads to **better SNR** compared to MVDR using only HA microphones (SC, FIX) and external microphone (EM)
- MVDR using estimated RTFs (SCE, SC) **preserves binaural cues of target speaker** compared to fixed MVDR (FIX) and external microphone (EM)

Oldenburg Varechoic Lab ($T_{60} \approx 350\text{ms}$), $M=4 + 1$ external mic (1.5m/0.5m), moving speaker, pseudo-diffuse babble noise, $iSNR=0\text{dB}$ (right HA)

STFT: 32 ms, 50% overlap, sqrt-Hann; SPP in external microphone; smoothing: 100 ms (speech), 1 s (noise)
Subjective evaluation of binaural speech enhancement algorithms with **HA/CI users** ongoing

- **Complex and time-varying scenarios**: incorporate computational acoustic scene analysis (CASA) into control path of developed algorithms

- Integration of **multiple external microphones** (acoustic sensor network)
Binaural noise reduction algorithms: 2 main paradigms
- Spectral post-filtering
- "True" binaural spatial filtering

Extensions of binaural MVDR/MWF for diffuse noise and interfering speaker, preserving binaural cues of residual noise/interference

Evaluation of binaural MVDR extensions (MVDR-N) for diffuse noise:
- Binaural cue preservation improves spatial quality
- Binaural cue preservation does not/hardly affect speech intelligibility

Extensions with external microphone possible
Acknowledgments

Funding:

- Cluster of Excellence Hearing4all (DFG)
- Research Unit Individualized Hearing Acoustics (DFG)
- Joint Lower-Saxony Israel Project “Acoustic scene aware speech enhancement for binaural hearing aids” (Partner: Bar-Ilan University, Israel)
Recent publications


http://www.sigproc.uni-oldenburg.de → Publications