Dereverberation and Reverberation of Audio, Music, and Speech

Patrick A. Naylor - Dep. of Electrical and Electronic Eng., Imperial College London, UK
Enzo De Sena - Dep. of Music and Media, University of Surrey, UK
Toon van Waterschoot - Dep. of Electrical Engineering, KU Leuven, Belgium

Imperial College London

UNIVERSITY OF SURREY

KU LEUVEN

August 28, 2017

Outline of the Tutorial

1. Introduction

2. Fundamentals of Room Acoustics

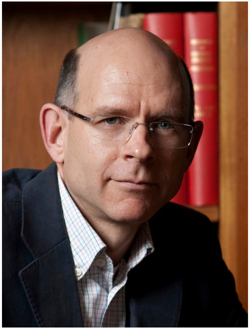
3. Measures of Reverberation

4. Measurement and Estimation of Acoustic Impulse Responses

5. Room Acoustics Modelling and Simulation

6. Dereverberation Processing Methods

Contributors



Patrick A. Naylor

Enzo De Sena

Toon van Waterschoot

Acknowledgements

DREAMS

Dereverberation and
Reverberation of
Audio, Music, and Speech

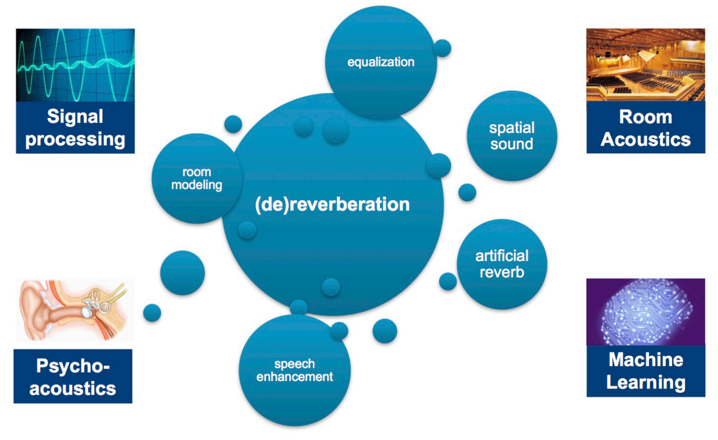
MARIE CURIE ACTIONS

Introduction

Aims

- Set the topic in its scientific context
- Provide an *overview* of the topic and *classification* of its subtopics
- Outline the key technical approaches and give references to relevant algorithms
  - *it's not a class*
- Give insights and perspectives

Overview



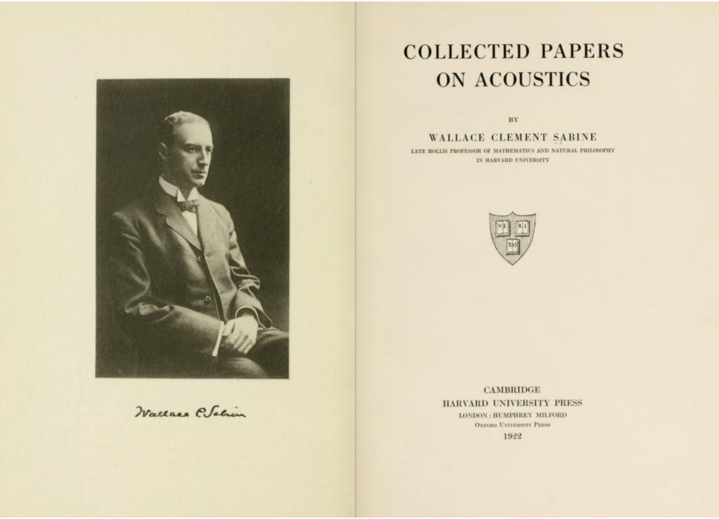
Signal processing
Room Acoustics
Psycho-acoustics
Machine Learning
equalization
spatial sound
room modeling
(de)reverberation
artificial reverb
speech enhancement

History

Earliest known mention of reverberation, "The Republic", written by Plato around 380 BC:
"And what if sound echoed off the prison wall opposite them? When any of the passers-by spoke, don't you think they'd be bound to assume that the sound came from a passing shadow?"

Pioneering scientific work:

- Rayleigh, "*The theory of sound*," 1877
- Sabine, "*Collected papers on acoustics*," 1922
- Bolt, "*Theory of speech masking by reverberation*," 1949
- Schroeder, "*Natural sounding artificial reverberation*," 1961
- Haas, "*The influence of a single echo on the audibility of speech*," 1972
- Allen, "*Image method for efficiently simulating small-room acoustics*," 1979


Wallace C Sabine

COLLECTED PAPERS
ON ACOUSTICS

BY

WALLACE CLEMENT SABINE

LATE HOLLIS PROFESSOR OF MATHEMATICS AND NATURAL PHILOSOPHY
IN HARVARD UNIVERSITY

CAMBRIDGE
HARVARD UNIVERSITY PRESS
LONDON : HUMPHREY MILFORD
OXFORD UNIVERSITY PRESS
1922

More Recent Influences

Hand-held $\rightarrow$ Hands-free (1990's)

More Recent Influences

Hand-held $\rightarrow$ Hands-free (2000's)

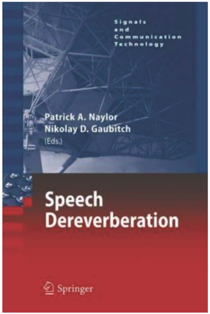
More Recent Influences

Hand-held $\rightarrow$ Hands-free (2010's)

Other Influences

- Speech-in-noise intelligibility is significantly degraded by reverberation
- Many children are schooled in a second language

Immersive Audio



Overview of geometrical room acoustic modeling techniques

Lauri Savioja[a)]
Department of Computer Science, Aalto University, Otaniementie 17, P.O. Box 15500, FI-00076 Aalto, Finland

U. Peter Svensson
Acoustics Research Centre, Department of Electronics and Telecommunications, Norwegian University of Science and Technology, O.S. Bragstads plass 2B, NO-7491 Trondheim, Norway

J. Acoust. Soc. Am. **138** (2), August 2015

REVERB CHALLENGE

DREAMS

Dereverberation and Reverberation of Audio, Music, and Speech

MARIE CURIE ACTIONS

IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 5, JULY 2012

Fifty Years of Artificial Reverberation

Vesa Välimäki, *Senior Member, IEEE*, Julian D. Parker, Lauri Savioja, *Senior Member, IEEE*, Julius O. Smith, *Member, IEEE*, and Jonathan S. Abel, *Member, IEEE*

Tutorial Outline

1. Introduction
2. Fundamentals of Room Acoustics
3. Measures of Reverberation
4. Measurement and Estimation of Acoustic Impulse Responses
5. Room Acoustics Modelling and Simulation
6. Dereverberation Processing Methods

Summary and Questions/Comments

- Context and motivation for reverberation and dereverberation
- Emphasize the importance for telecommunications but also other sectors and influences

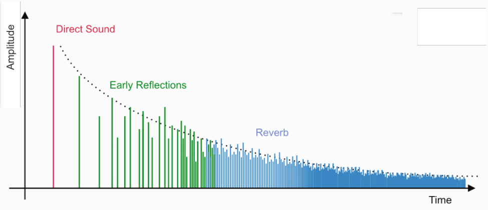
- Any questions so far?

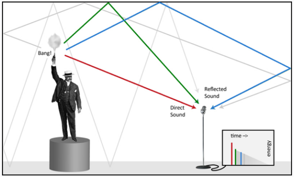
Fundamentals of Room Acoustics

Outline of the Section

- Start with components of acoustic impulse response (AIR), some definitions, and examples of measured AIR
- Fundamentals on physical modelling of sound, wave equation and modal description of reverberation
- Elements of perception of room acoustics

Acoustic Impulse Response (AIR)





Components of AIR $h(t)$ in rooms:

- Direct line-of-sight (LOS)
- Early reflections: relatively sparse first echoes
- Late reverberation: so densely populated with echoes that it is best to characterise the response *statistically*.

Some definitions

Mixing time:

- Transition point between early reflections and late reverberation

Reverberation time (T60):

- Time taken for sound to decay 60 dB from its initial level (more detailed definition in next section)

Measured AIR samples

- St Patrick's Church in Patrington (recording by Foteinou and Murphy) $T_{60} = 1.86$ s

- Bathroom (recording by van Saane) $T_{60} = 0.35$ s

- Inchindown oil storage (recording by Cox) $T_{60} = 75$ s

Wave equation

- Sound propagation governed by the PDE [1]:

$$\triangle p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = s$$

  where $c = 343$ [m/s], p pressure, s source distribution
- Need initial and boundary conditions to find solution
- Example of boundary condition:

$$\frac{\partial p}{\partial t} = -cZ_w \nabla p \cdot \mathbf{n}$$

  where $\mathbf{n}$ normal at boundary, Z_w wall impedance
- Equation admits closed form solution only in few cases

Modal description of reverberation

- Using monochromatic sound source, Helmholtz equation [1]:

$$\triangle \hat{p} + \left(\frac{\omega}{c}\right)^2 \hat{p} = \hat{s}$$

- Using separation of variables, and point source:

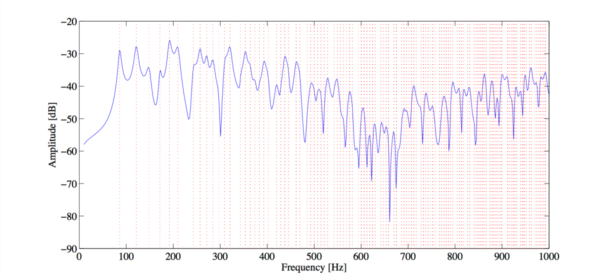
$$\hat{p} = \frac{c^2}{V} \sum_{m=0}^{\infty} \frac{\psi_m(\mathbf{x}')\psi_m(\mathbf{x})}{\omega^2 - \omega_m^2 - 2j\delta_m\omega_m},$$

  where V volume, $\mathbf{x}$ observation point, $\mathbf{x}'$ source position, $\psi_m(\mathbf{x})$ eigenfunctions of problem, ω_m and δ_m real and imaginary part of problem's eigenvalues, respectively

- Equivalent to a parallel of second-order resonant modes

Modal description of reverberation (cont'd)

- Density of modes increases as f^2 [Kuttruff, 2000] [1]
- Similarly to early/late in impulse response, frequency response of a reverberant room can be divided in:
  - Low-frequency sparse distribution of resonant modes
  - Modes packed so densely that they merge to form *random frequency response*

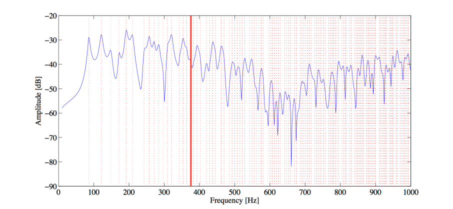


Schroeder frequency

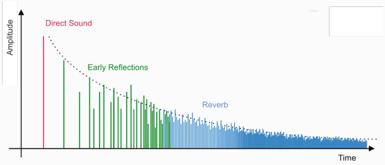
- Transition point between two regions called Schroeder frequency [Schroeder, 1962] [2]: $F_c = 2000\sqrt{\frac{T_{60}}{V}}$

Examples

- Bathroom $V = 10\,\text{m}^3$, $T_{60} = 0.35$ s $\Rightarrow F_c = 374$ Hz
- Concert hall $V = 2700\,\text{m}^3$, $T_{60} = 2$ s $\Rightarrow F_c = 54$ Hz



Perception of Sound in Rooms



Copyright © 2015 Rational Acoustics

- Governed by complex and not fully understood perceptual phenomena [3, 4]

1. Early reflections: affect spaciousness, envelopment, and apparent source width.
2. Late reverberation: precise structure not important, but
   2.1 $T_{60}(\omega)$: affects impression of size
   2.2 Echo density: affects perceived texture of reverberation
   2.3 Mode density: if insufficient can yield metallic sound

Summary

- AIR in a room: LOS, early reflection and late reverb
- Wave equation gives physical model for propagation
- Wave equation requires initial and boundary conditions to find solution, and solution hard to find in closed form
- Solution for point-like sound source yields modal description of reverberation
- Modes well separated at low frequencies
- Room perception governed by complex phenomena
- Accurate rendering of early reflections is important
- We are not sensitive to precise structure of late reverb

Questions/Comments

- Questions

Measures of Reverberation

Why are Reverberation Measures Important?

- Significant aspect of speech quality
  - not specifically included in general measures
- Significant aspect of audio/music quality
  - usually aesthetically judged
- Adaptively control dereverberation processing
  - switch off if not needed
- 'Awareness' of room acoustics
  - exploited in other processing
- Modelling for Automatic Speech Recognition (ASR)
  - multi-condition training
  - control distribution of reverberation in the training set

Parts of this tutorial will emphasize speech applications

Why are Reverberation Measures Important?

- Significant aspect of speech quality
  - not specifically included in general measures
- Significant aspect of audio/music quality
  - usually aesthetically judged
- Adaptively control dereverberation processing
  - switch off if not needed
- 'Awareness' of room acoustics
  - exploited in other processing
- Modelling for Automatic Speech Recognition (ASR)
  - multi-condition training
  - control distribution of reverberation in the training set

Parts of this tutorial will emphasize speech applications

Why are Reverberation Measures Important?

- Significant aspect of speech quality
  - not specifically included in general measures
- Significant aspect of audio/music quality
  - usually aesthetically judged
- Adaptively control dereverberation processing
  - switch off if not needed
- 'Awareness' of room acoustics
  - exploited in other processing
- Modelling for Automatic Speech Recognition (ASR)
  - multi-condition training
  - control distribution of reverberation in the training set

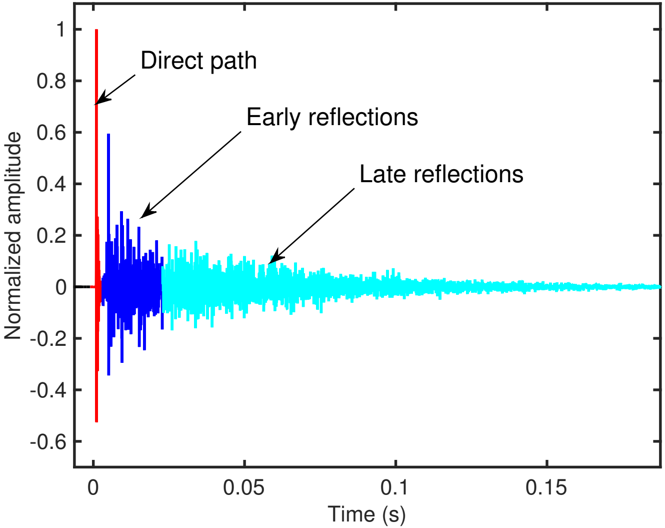
Parts of this tutorial will emphasize speech applications

Why are Reverberation Measures Important?

- Significant aspect of speech quality
  - not specifically included in general measures
- Significant aspect of audio/music quality
  - usually aesthetically judged
- Adaptively control dereverberation processing
  - switch off if not needed
- ‘Awareness’ of room acoustics
  - exploited in other processing
- Modelling for Automatic Speech Recognition (ASR)
  - multi-condition training
  - control distribution of reverberation in the training set

Parts of this tutorial will emphasize speech applications

Acoustic Impulse Response



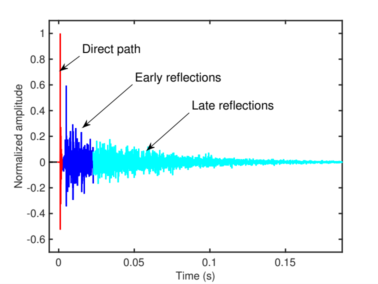
Classes of Measures of Reverberation

- Room acoustics based
  - T60 - reverberation time
- Model based
  - DRR - direct to reverberation ratio
  - C50 - clarity index
- Perceptually based
  - PLR - Perceived Level of Reverberation (opinion-based)
  - Q_m - (instrumental)
- ... others

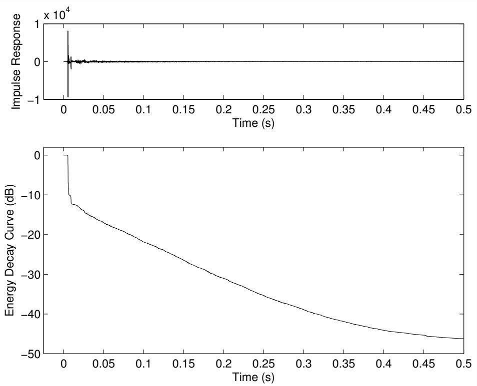
Physical Acoustical Characteristics

- Early reflections
  - echoes
- Late reflections
  - reverberation

$\rightarrow$ ratios
$\rightarrow$ spectral variation (RTS)



Energy Decay Curve



If the AIR of the room, $h(t)$, is known, the energy decay curve (EDC) can be obtained from the Schroeder integral [1]

$$\mathrm{EDC}(t) = \int_t^{\infty} h^2(\tau) d\tau$$

Taxonomy

measurement method	instrumental vs. opinion-based
prior information	intrusive vs. non-intrusive

Reverberation Time

The reverberation time, T_{60}, is defined for a diffuse sound field as the time in seconds required for the EDC to decay by 60 dB

Sabine's formula [5]:

$$T_{60} \propto \frac{V}{\alpha_{Sabine} A} \quad \mathrm{s}$$

- V is the enclosed volume
- $\alpha_{Sabine} A$ is the total sound absorption in the room with surface area A

Examples

Reverberation Time

> 5 s Loss of speech intelligibility is substantial

1.5 to 3.5 s **Auditoria** - reverb is good for music, significant loss of speech intelligibility
- New York, Carnegie Hall: 1.7 s

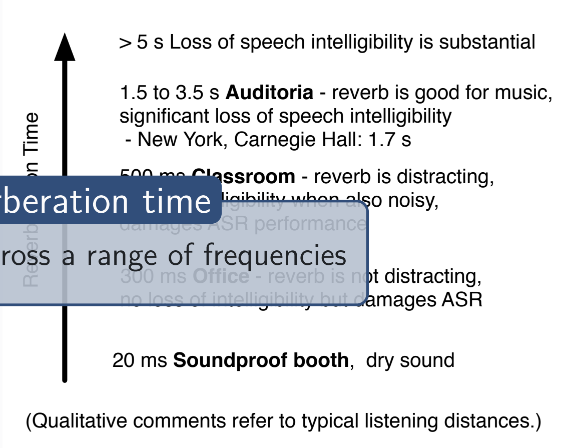
500 ms **Classroom** - reverb is distracting, loss of intelligibility when also noisy, damages ASR performance

300 ms **Office** - reverb is not distracting, no loss of intelligibility but damages ASR

20 ms **Soundproof booth**, dry sound

(Qualitative comments refer to typical listening distances.)

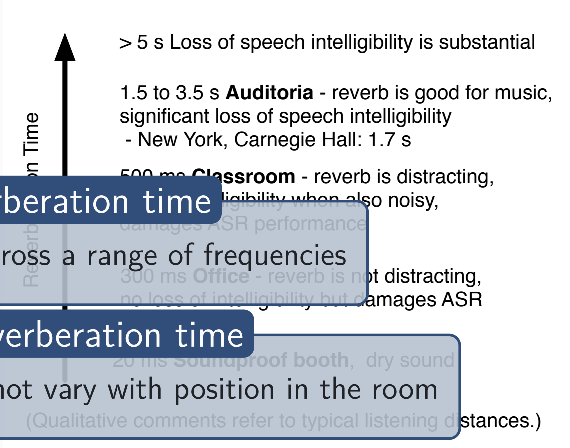
Examples



Reverberation time

varies across a range of frequencies

Examples



Reverberation time

varies across a range of frequencies

Reverberation time

does not vary with position in the room

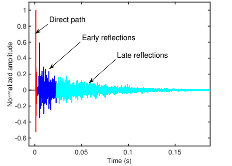
Critical Distance

Distance D_c from the source at which the sound energy density due to the direct-path component, E_d, equals the sound energy density due to the reverberant component, E_r [1]

$$D_c \approx 0.11 \sqrt{\frac{QV}{\pi T_{60}}}$$

with source directivity factor Q and room volume V

Direct to Reverberation Ratio



Direct to Reverberation Ratio (DRR) is given by

$$\mathrm{DRR} = 10\log_{10}\left(\frac{\sum_{n=0}^{n_\mathrm{d}} h^2(n)}{\sum_{n=n_\mathrm{d}+1}^{\infty} h^2(n)}\right) \quad \text{dB},$$

- samples of the AIR, h_n, indexed $n = 0, \ \ldots, \ n_\mathrm{d}$ represent direct-path propagation
- samples of the AIR indexed $n > n_\mathrm{d}$ represent late reverberation

Clarity Index

Clarity Index is given by

$$C = 10\log_{10}\left(\frac{\sum_{n=0}^{n_\mathrm{e}} h^2(n)}{\sum_{n=n_\mathrm{e}+1}^{\infty} h^2(n)}\right) \quad \text{dB}$$

- where $n_\mathrm{e} f_\mathrm{s}$ can be chosen as 50 ms
  - denoted by C_{50}
- motivated by the human auditory system
  - interprets multipath signal components as a single signal if the components' times of arrival differ by less than around 50 ms

Speech-to-Reverberation Modulation Energy Ratio

Falk *et al*, 2010: "A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech" [6]

- SRMR operates in the modulation domain
  - spectrum of temporal envelopes in gammatone bands

$$\mathrm{SRMR} = \frac{\sum_{k=1}^{4} \bar{\mathcal{E}}_k}{\sum_{k=5}^{K*} \bar{\mathcal{E}}_k}$$

where $\bar{\mathcal{E}}_k$ is the average modulation energy in band k

- Approach
  - anechoic speech contains modulation frequencies ranging 2-20 Hz
  - reverberation causes smearing of modulation energy into higher modulation frequencies

Examples

- Building entrance with ambient noise
- Source distance 2.5 m
- T_{60} 0.85 s
- Reverberant signal to noise ratio -10 dB to 30 dB
- Source data CHiME 3 (WSJ)
- Kaldi ASR

Word Error Rate Results (%) vs RSNR

RSNR [dB]	reverb	DSB	DSB+SS	MVDR	MVDR +SS	GWPE	GWPE +DSB	GWPE +DSB +SS	GWPE+ MVDR	GWPE+ MVDR +SS
-10	99.98	99.79	99.35	99.91	99.61	100.0	99.89	99.57	99.96	99.61
0	99.18	97.12	94.47	98.09	95.72	99.36	95.95	92.49	96.88	93.95
10	95.16	86.46	79.47	91.48	83.30	87.00	76.24	68.04	79.69	71.70
20	92.56	79.19	70.56	86.64	77.08	66.73	56.79	51.54	61.27	55.33
30	92.25	78.05	69.14	86.55	76.93	58.77	51.20	48.44	55.09	51.04

RSNR - reverberant signal-to-noise ratio

Effects on speech intelligibility

STOI Results vs RSNR

RSNR [dB]	reverb	DSB	DSB+SS	MVDR	MVDR+SS	GWPE	GWPE+DSB	GWPE+DSB+SS	GWPE+MVDR	GWPE+MVDR+SS
-10	0.43	0.44	0.42	0.44	0.42	0.41	0.44	0.41	0.43	0.41
0	0.59	0.62	0.63	0.62	0.62	0.59	0.63	0.63	0.62	0.63
10	0.72	0.75	0.77	0.74	0.76	0.75	0.78	0.79	0.77	0.78
20	0.77	0.81	0.82	0.8	0.81	0.83	0.85	0.86	0.84	0.85
30	0.79	0.83	0.84	0.81	0.82	0.85	0.87	0.87	0.86	0.86

Combination of noise and reverberation is strongly problematic

Perceived Level of Reverberation

How does reverberation affect the human perceptual system?

- Perceived Level of Reverberation (PLR) measure
- adaptively control hearing aid signal processing

σ_R is the Room Spectral Variance - a measure of colouration due to convolution with the RIR

Perceived Level of Reverberation

How does reverberation affect the human perceptual system?

- Perceived Level of Reverberation (PLR) measure
- adaptively control hearing aid signal processing

Previous instrumental estimators of PLR (*intrusive*):

- Allen [7] $$P \propto -T_{60}\ \sigma_R^2$$
- de Lima [8] $$Q = -\frac{T_{60}\ \sigma_R^2}{(DRR)^\gamma}$$
- Vallado [9] $$Q_m = -\frac{(T_{60})^\alpha\ (\sigma_R^2)^\beta}{(DRR)^\gamma}$$

σ_R is the Room Spectral Variance - a measure of colouration due to convolution with the RIR

Non-intrusive PLR estimation

Eaton *et al*, 2017, "Estimation of the perceived level of reverberation using non-intrusive single-channel variance of decay rates"[10]

- **Feature**: Negative Side Variance of decay rates
  - NSV
- **Data**: Use an existing intrusive measure Q_m
  - As much data for training as desired!

Data-driven approaches like this are effective given sufficient labelled training data

Why Q_m ?

- Q_m measure has been extensively validated
- Q_m is therefore used to label reverberant speech to be used for training the proposed algorithm
  - Q_m is intrusive $\Rightarrow$ OK for training
  - Q_m is instrumental $\Rightarrow$ plenty of examples to learn from

Energy decay rate

- At each freq in STFT domain, the amplitude envelope of each sound/phoneme has a decay phase
  - anechoic: decays are intrinsic decay envelopes of speech
  - reverb: decay rates are slower and extended in time

Energy decay rate

- At each freq in STFT domain, the amplitude envelope of each sound/phoneme has a decay phase
  - anechoic: decays are intrinsic decay envelopes of speech
  - reverb: decay rates are slower and extended in time
- Energy decay rate λ_h of AIR can be found by linear regression[11] in each frequency bin over short windows

Energy decay rate

- At each freq in STFT domain, the amplitude envelope of each sound/phoneme has a decay phase
  - anechoic: decays are intrinsic decay envelopes of speech
  - reverb: decay rates are slower and extended in time
- Energy decay rate λ_h of AIR can be found by linear regression[11] in each frequency bin over short windows
- In reverberation, decay rate results from decays from anechoic speech, x, convolved with the AIR h

$$\lambda_{x*h}(t, f)$$

Negative Side Variance of Decay Rates

Hypothesis:
"The variance in decay rates in both the diffuse and near fields is related to PLR"

- somewhat similar to a measure of modulation
- we only consider negative gradients for 'decays'
- written as $\sigma^2(\lambda_{x*h-})$

$$\mathrm{PLR} \propto Q_m = g_W(\sigma^2(\lambda_{x*h-}))$$

Negative Side Variance of Decay Rates

Hypothesis:
"The variance in decay rates in both the diffuse and near fields is related to PLR"

- somewhat similar to a measure of modulation
- we only consider negative gradients for 'decays'
- written as $\sigma^2(\lambda_{x*h-})$

$$\mathrm{PLR} \propto Q_m = g_W(\sigma^2(\lambda_{x*h-}))$$

Mapping Function

Learn from training data

Negative Side Variance of Decay Rates

Hypothesis:
"The variance in decay rates in both the diffuse and near fields is related to PLR"

- somewhat similar to a measure of modulation
- we only consider negative gradients for 'decays'
- written as $\sigma^2(\lambda_{x*h-})$

$$\mathrm{PLR} \propto Q_m = g_W(\sigma^2(\lambda_{x*h-}))$$

Mapping Function

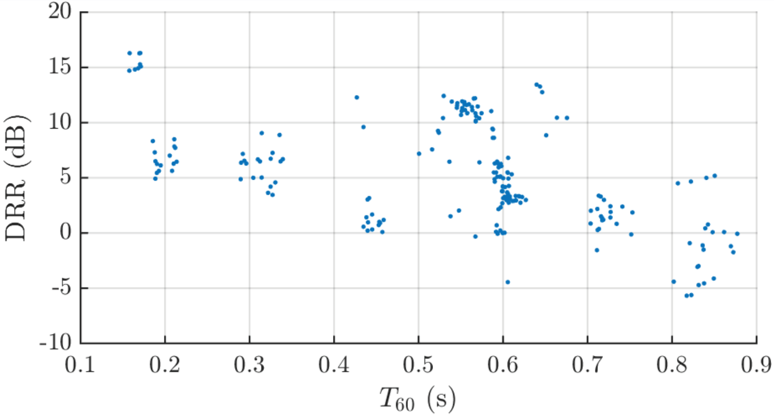
Learn from training data

Decay Rates

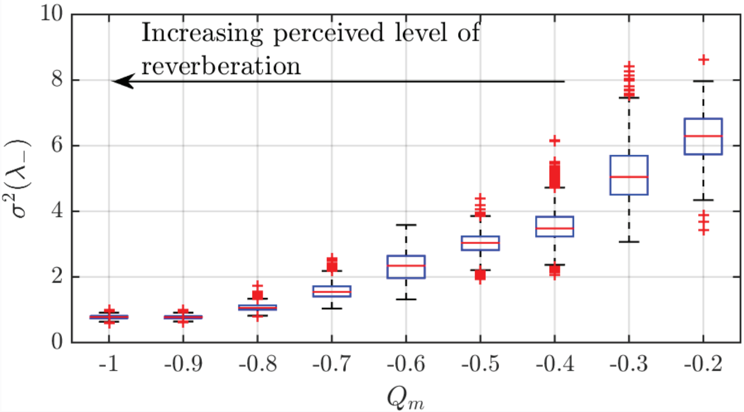
Compute from the signals

Training

Each point represents one AIR in the training corpus of 203



NSV vs Q_m in the training data



- 20,300 reverberant speech files from 203 AIRs in the training corpus grouped by ranges of Q_m
- learn a 3rd order polynomial by least squares regression

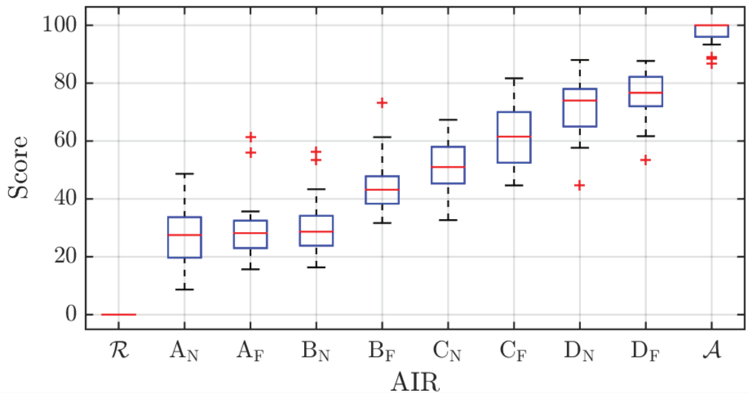
Listening Test

Aim to determine the true PLR of the evaluation corpus

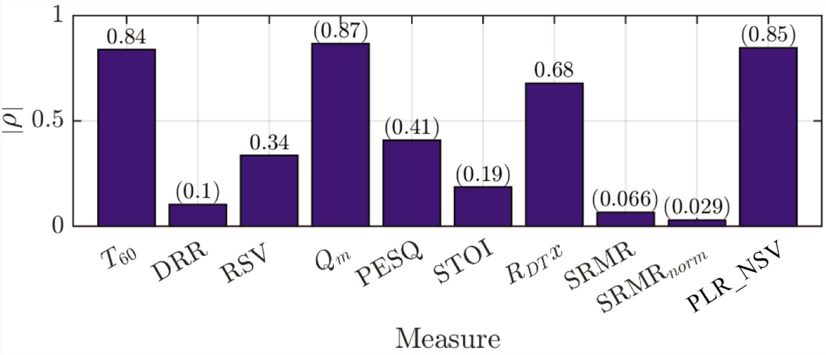
- 28 listeners with self-assessed normal hearing
- 3 speech utterances for each of 10 AIRs
- MUSHRAR [12]
  - scoring range is inverted - more reverberant signal receives a higher score
  - hidden reference, R, and anchor, A, signals use low and high reverberation

See www.acousp.org for a list of acoustic signal processing resources, specifically impulse response datasets

Listening Test Results



Comparative PLR Estimation Results



Absolute Pearson correlation coefficients for each measure evaluated against the listening test results

Computational Complexity

Algorithm	RRTF
PESQ [7]	0.096
STOI [8]	0.1
$R_{\mathrm{DT}}\mathrm{x}$ [12]	0.57
SRMR [14]	1
$\mathrm{SRMR}_{\mathrm{norm}}$ [15]	0.98
PLR-NSV	0.031

- Relative Real-time Factor normalized to SRMR

Summary and Questions

- Motivation for reverberation measures
- Intrusive vs non-intrusive
- Examples
- Measures based on room acoustics and AIR models
- New method for perceived level of reverberation based on efficient data-driven approach

- Questions

Measurement and Estimation of Acoustic Impulse Responses

AIR measurement: Introduction

- Direct AIR measurement requires **reproduction of impulse signal** (band-limited to freq. range of interest)
- Physically challenging and undesirable for several reasons

©Trevor Cox, Creative Commons License

©Pat Brown, Synergetic Audio Concepts

Indirect AIR measurement = system identification

- Reproduce stimulus and measure room response
- Estimate AIR from {stimulus, response} pair

AIR measurement: Challenges

©DREAMS ITN

©Magnus Schäfer, RWTH Aachen

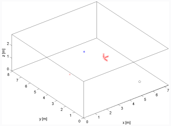
Challenges when measuring/estimating AIRs

- Reproducibility
- Loudspeaker artefacts
- Noise robustness
- Identifiability / persistence of excitation

AIR measurement: Reproducibility

- Reproduction of stimuli in different measurement trials requires **loudspeaker playback of prerecorded stimuli**
- Accurate measurement of **loudspeaker and microphone positions** crucial when measuring AIRs for spatial audio applications (e.g. beamforming, rendering, localization)
- **Documenting room and measurement setup** (room geometry, boundary materials, temperature, measurement hardware, stimuli) strongly recommended
- Covering hardware and other objects with absorption foam advisable to **avoid spurious reflections**

©Jesper Kjær Nielsen, Aalborg University



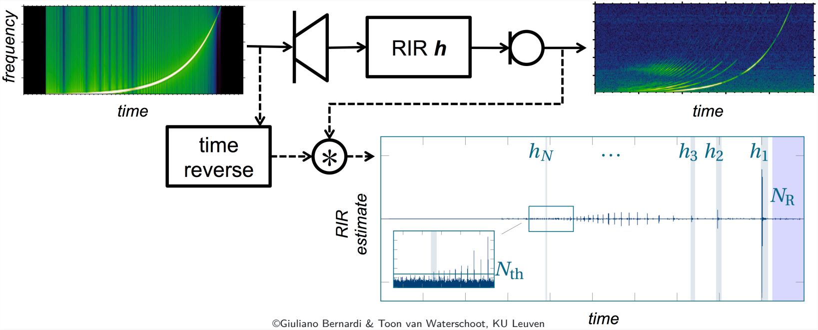
©Jesper Kjær Nielsen, Aalborg University

©Neofytos Kaplanis, Bang & Olufsen

©Jesper Kjær Nielsen, Aalborg University

AIR measurement: Loudspeaker artefacts

- Linear loudspeaker distortion: usually not compensated
- Nonlinear even-harmonic distortion: eliminable with **Exponential Sine Sweep (ESS)** method [13, 14]



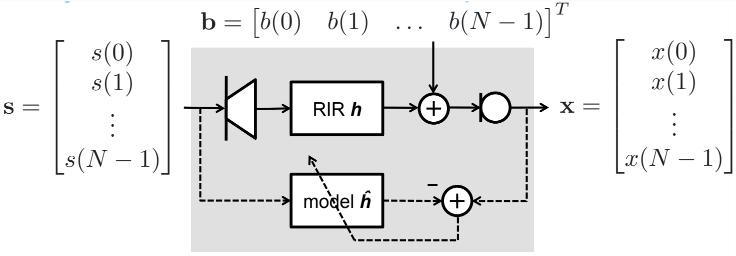
- Nonlinear odd-harmonic and impulsive (e.g. rub & buzz) distortion: requires careful **level calibration** [15]

AIR measurement: Noise robustness

- Organize measurements during silent periods (overnight)
- Increase **length of measurement stimuli**:
  - Doubling amount of measurement data yields 3 dB increase of measurement SNR
  - Result only holds for broadband random noise, not for impulsive noise or nonlinear artefacts
- **Synchronous averaging of responses** to short stimuli[16]:
  - SNR increase as above, with possibility to discard responses affected by impulsive noise
  - Applicable to many types of stimuli, e.g. ESS, maximum-length sequences (MLS), inverse repeated sequence (IRS)

AIR measurement: Identifiability

- System identification setup for AIR estimation



©Toon van Waterschoot, KU Leuven

Signal model: $\mathbf{x} = \mathbf{S}\mathbf{h} + \mathbf{b}$

$$\begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} = \begin{bmatrix} s(0) & s(-1) & \dots & s(-L+1) \\ s(1) & s(0) & \dots & s(-L+2) \\ \vdots & \vdots & \ddots & \vdots \\ s(N-1) & s(N-2) & \dots & s(N-L) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{L-1} \end{bmatrix} + \begin{bmatrix} b(0) \\ b(1) \\ \vdots \\ b(N-1) \end{bmatrix}$$

AIR measurement: Identifiability

AIR identifiable only when stimulus is persistently exciting:
rank$(S) \geq L$

- Densely sampled AIRs are excessively long: $O(L) = 10^4$
- Lack of identifiability occurs with speech/audio stimuli
- **Regularization** is crucial for (online) AIR estimation

AIR measurement: Identifiability

- Tikhonov regularization (e.g. least squares AIR estimate)

$$\min_{\hat{\mathbf{h}}} \|\mathbf{x} - \mathbf{S}\hat{\mathbf{h}}\|^2 + \lambda\|\hat{\mathbf{h}}\|^2 \Rightarrow \hat{\mathbf{h}} = (\mathbf{S}^T\mathbf{S} + \lambda\mathbf{I})^{-1}\mathbf{S}^T\mathbf{x}$$

- Levenberg-Marquardt regularization (e.g. normalized least mean squares adaptive AIR estimate)

$$\min_{\hat{\mathbf{h}}(n)} \| \underbrace{x(n) - \mathbf{s}^T(n)\hat{\mathbf{h}}(n)}_{e(n)} \|^2 + \lambda\|\hat{\mathbf{h}}(n) - \hat{\mathbf{h}}(n-1)\|^2$$

$$\Rightarrow \hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n) + \frac{\mathbf{s}(n)e(n)}{\mathbf{s}^T(n)\mathbf{s}(n) + \lambda}$$

- Optimal and generalized regularization achievable in Bayesian framework [17]

AIR measurement databases

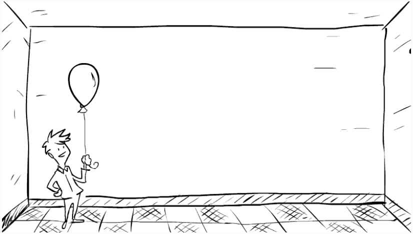
Selection of open-source AIR measurement databases for different room types and measurement setups (www.acousp.org):

- GTAC Database (2017): 1 room, 34560 AIRs [18]
- SUBRIR Database (2017): 1 room, 48 very LF AIRs [15]
- ACE Challenge Corpus (2016): 7 rooms, 700 AIRs, incl. noise measurements [19]
- SMARD Database (2014): 1 room, ca. 1000 AIRs, incl. reverberant signals [20]
- C4DM Database (2010): 3 rooms, 468 AIRs, incl. B-format recordings [21]
- Aachen IR Database (2009-2010): 10 rooms, 58 AIRs, binaural/dual-mic [22]
- Oldenburg IR Database (2009): 5 rooms, 2846 AIRs, binaural, incl. noise/HRTF measurements [23]
- MARDY Database (2006): 1 room, 72 AIRs [24]
- Open AIR Library (2010-2017): platform for sharing AIRs [25]

AIR interpolation

AIR is point-to-point response

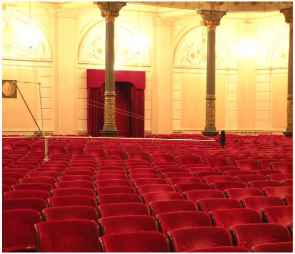
- How to measure spatial variations of room acoustics?
- **AIR interpolation** yields AIR estimates for virtual source/observer positions
- Interpolation often relies on wave-based model

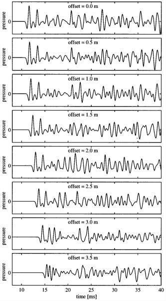


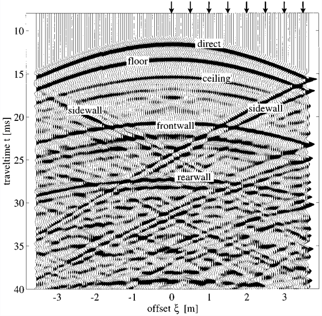
©Niccoló Antonello, KU Leuven

Wave Field Analysis

- AIR measurements from dense linear array (microphone spacing = 5 cm) reveal wave patterns [26]
- Patterns can be interpolated and extrapolated using plane-wave assumption

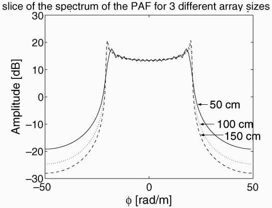


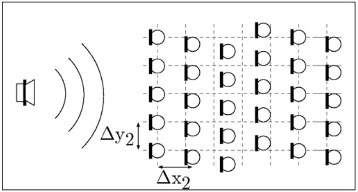




Plenacoustic Sampling

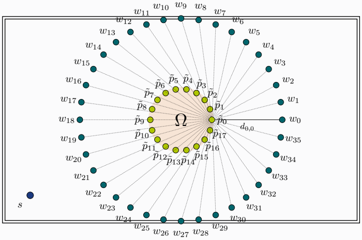
- **How densely should sound field be sampled?**
- Plenacoustic function (PAF) describing spatiotemporal sound field is approximately spatially bandlimited [27]
  - *Example:* Slice of PAF spectrum for 1-D spatial sampling
- **Plenacoustic sampling theory:** Optimal spatiotemporal sampling schemes and reconstruction bounds [27]
  - *Example:* Hexagonal 2-D spatial sampling scheme
  - *Example:* Reconstructing sound field up to 1.3 kHz with 100 dB SNR requires microphone spacing = 12.35 cm



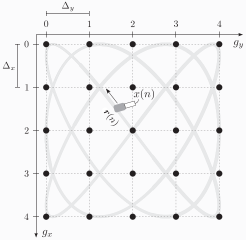


Sparse AIR interpolation in compressed sensing framework

- Temporal sparsity of early reflections [28]
- Sparsity of low-frequency room modes in plane-wave decomposition domain [29]
- Spatio-temporal sparsity of equivalent sources [30]
- Random sampling with moving microphone [31]



©Niccoló Antonello, KU Leuven



©Fabrice Katzberg, University of Lübeck

Summary and Questions

- AIR measurement challenges: reproducibility, loudspeaker artefacts, noise robustness, identifiability
- AIR measurement databases overview
- AIR interpolation: wave field analysis, plenacoustic sampling, sparse interpolation

- Questions?

Blind System Identification (BSI)

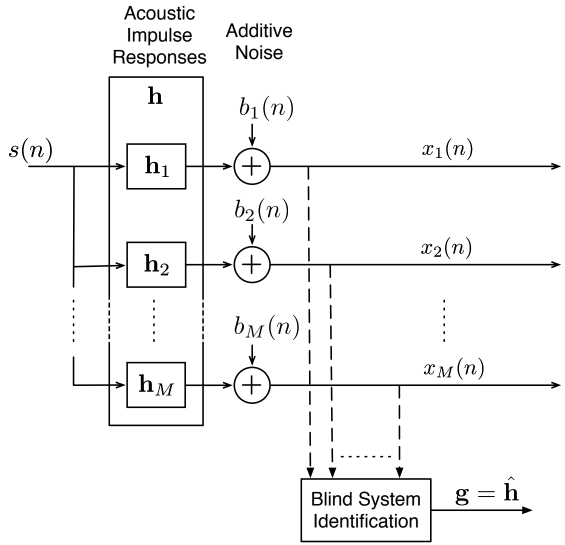
1 sound source, 2 or more microphones $\rightarrow$ find the AIRs

- Blind - just use the microphone signals

Conventional BSI methods rely on the multichannel cross-relation (CR) property

- only valid for fully-modelled systems

Acoustic SIMO System



Signal Model

In a reverberant environment with a single sound source and an M-element microphone array at time n and channel i

$$\mathbf{x}_i(n) = \mathbf{H}_i \cdot \mathbf{s}(n) + \mathbf{b}_i(n) \quad i = 1, 2, ..., M$$

Signal Model

In a reverberant environment with a single sound source and an M-element microphone array at time n and channel i

$$\mathbf{x}_i(n) = \mathbf{H}_i \cdot \mathbf{s}(n) + \mathbf{b}_i(n) \quad i = 1, 2, ..., M$$

- $\mathbf{x}_i(n)$: the $K \times 1$ signal vector of the i-th microphone

Signal Model

In a reverberant environment with a single sound source and an M-element microphone array at time n and channel i

$$\mathbf{x}_i(n) = \mathbf{H}_i \cdot \mathbf{s}(n) + \mathbf{b}_i(n) \quad i = 1, 2, ..., M$$

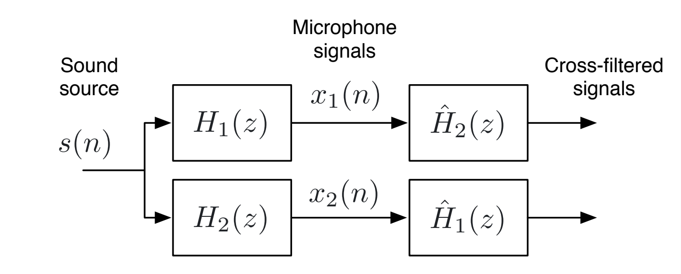
- $\mathbf{x}_i(n)$: the $K \times 1$ signal vector of the i-th microphone

$$\mathbf{H}_i = \begin{bmatrix} h_{i,0} & h_{i,1} & \cdots & h_{i,L-1} & 0 & \cdots & 0 \\ 0 & h_{i,0} & \cdots & h_{i,L-2} & h_{i,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{i,0} & h_{i,1} & \cdots & h_{i,L-1} \end{bmatrix}$$

- constructed from $\mathbf{h}_i = [h_{i,0} \; h_{i,1} \; \ldots \; h_{i,L-1}]^{\mathrm{T}}$
- $\mathbf{s}(n)$: the $K + L - 1 \times 1$ source signal vector
- $\mathbf{b}_i(n)$: the $K \times 1$ noise signal vector

Cross-filtered signals

2-channel example:



Cross-filtered signals are equal if $\hat{H}_{1,2}(z) = \alpha H_{1,2}(z)$ or zero

Fully-modelled BSI

temporarily suppress time dependence n

Cross-relation (CR) property

$$x_1 = s * h_1 \qquad\qquad x_2 = s * h_2$$

$$x_1 * h_2 = s * h_1 * h_2 \qquad\qquad x_2 * h_1 = s * h_2 * h_1$$

temporarily suppress time dependence n

Cross-relation (CR) property

$$x_1 = s * h_1 \qquad\qquad x_2 = s * h_2$$

$$x_1 * h_2 = s * h_1 * h_2 \qquad\qquad x_2 * h_1 = s * h_2 * h_1$$

$$\mathbf{h}_i^{\mathrm{T}} \mathbf{x}_j = \mathbf{h}_j^{\mathrm{T}} \mathbf{x}_i$$
$$i, j = 1{,}2, \ldots, M, i \neq j$$

Fully-modelled BSI

temporarily suppress time dependence n

Cross-relation (CR) property

$$x_1 = s * h_1 \qquad\qquad x_2 = s * h_2$$

$$x_1 * h_2 = s * h_1 * h_2 \qquad\qquad x_2 * h_1 = s * h_2 * h_1$$

$$\mathbf{h}_i^{\mathrm{T}} \mathbf{x}_j = \mathbf{h}_j^{\mathrm{T}} \mathbf{x}_i$$

$$i, j = 1{,}2, ..., M, i \neq j$$

Let $\mathbf{g}$ be an estimate of $\mathbf{h}$, cross-filtered signals $\tilde{x}_{ij} = \mathbf{g}_i^{\mathrm{T}} \mathbf{x}_j$

CR error

$$\varepsilon_{ij} = \tilde{x}_{ij} - \tilde{x}_{ji}$$

Sum squared error

$$\chi(\mathbf{g}) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \mathbb{E}\{[\varepsilon_{ij}]^2\}$$

AIRs estimated by $\mathbf{g}$ that minimize $\chi(\mathbf{g})$ with unit-norm

Identifiability Conditions

The AIRs can be identified up to an unknown scale factor (unknown filter in the case of subband processing) provided that:

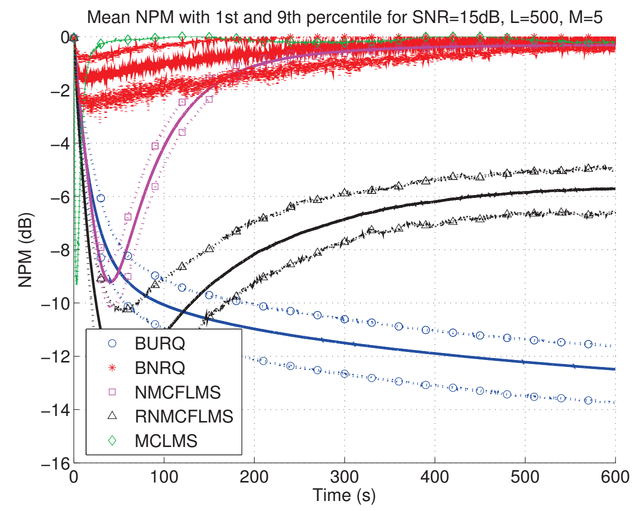
- C1 - The polynomials formed from $\mathbf{h}_i = [h_{i,0}\ h_{i,1}\ \ldots]$ are co-prime so that the channels' transfer functions $H_i(z)$ contain no common zeros
- C2 - The autocorrelation matrix of the source signal, $\mathbf{R}_{ss} = E\{\mathbf{s}(n)\mathbf{s}^T(n)\}$, is full rank

- Undermodelling is always a problem in realistic cases

Non-unique Solutions

- Cross-relation approach is not robust to additive noise
- Iterative solutions use an instantaneous estimate $\mathbf{R}_n$ of the cross-relation matrix $\mathbf{R}$
  - the gradient descent direction consistently steers the adaption towards the null space of $\mathbf{R}_n$, which contains $\mathbf{h}$ only in the absence of noise
- Recent research on block-based Rayleigh quotient cost functions with and without normalization has shown improved noise robustness [32]

NPM (mis-)Convergence Curves



Under-modelled BSI

For BSI with $\mathbf{h}$ under-modelled by $\mathbf{g}$

$$\tilde{x}_{ij}(n) = \tilde{x}_{ij,e}(n) + \tilde{x}_{ij,l}(n),$$

with $\tilde{x}_{ij,e}$ and $\tilde{x}_{ij,l}$ early and late cross-filtered signals and again temporarily suppressing n

Under-modelled BSI

For BSI with $\mathbf{h}$ under-modelled by $\mathbf{g}$

$$\tilde{x}_{ij}(n) = \tilde{x}_{ij,e}(n) + \tilde{x}_{ij,l}(n),$$

with $\tilde{x}_{ij,e}$ and $\tilde{x}_{ij,l}$ early and late cross-filtered signals and again temporarily suppressing n

If $\mathbf{g}_i = \alpha \mathbf{h}_{i,e}$, then

- early part is correctly modelled: $\tilde{x}_{ij,e} = \tilde{x}_{ji,e}$
- late part acts as unwanted noise: $\tilde{x}_{ij,l} \neq \tilde{x}_{ji,l}$

Under-modelled BSI

For BSI with $\mathbf{h}$ under-modelled by $\mathbf{g}$

$$\tilde{x}_{ij}(n) = \tilde{x}_{ij,e}(n) + \tilde{x}_{ij,l}(n),$$

with $\tilde{x}_{ij,e}$ and $\tilde{x}_{ij,l}$ early and late cross-filtered signals and again temporarily suppressing n

If $\mathbf{g}_i = \alpha \mathbf{h}_{i,e}$, then

- early part is correctly modelled: $\tilde{x}_{ij,e} = \tilde{x}_{ji,e}$
- late part acts as unwanted noise: $\tilde{x}_{ij,l} \neq \tilde{x}_{ji,l}$

Under-modelling

CR property holds *only* for the early cross-filtered signals

Under-modelling Error Analysis

Under the reverberant signal model[1]

$$\mathbb{E}\{[\varepsilon_{ij}]^2\} = \mathbb{E}\{[\tilde{x}_{ij,e} - \tilde{x}_{ji,e}]^2\} + \underbrace{\mathbb{E}\{[\tilde{x}_{ij,l} - \tilde{x}_{ji,l}]^2\}}_{bias\ term}$$

- bias term is dominated by $\mathbb{E}\{[\tilde{x}_{ij,l}]^2\} + \mathbb{E}\{[\tilde{x}_{ji,l}]^2\}$
  $\Rightarrow$ always large

[1] $\mathbf{x}^n_{i,e}$ and $\mathbf{x}^n_{j,l}$ are uncorrelated

Under-modelling Error Analysis

Under the reverberant signal model[1]

$$\mathbb{E}\{[\varepsilon_{ij}]^2\} = \mathbb{E}\{[\tilde{x}_{ij,e} - \tilde{x}_{ji,e}]^2\} + \underline{\mathbb{E}\{[\tilde{x}_{ij,l} - \tilde{x}_{ji,l}]^2\}}_{bias\ term}$$

- bias term is dominated by $\mathbb{E}\{[\tilde{x}_{ij,l}]^2\} + \mathbb{E}\{[\tilde{x}_{ji,l}]^2\}$
  $\Rightarrow$ always large
- as estimated AIRs approach the true values
  $\mathbb{E}\{[\tilde{x}_{ij,e} - \tilde{x}_{ji,e}]^2\} \rightarrow 0$, bias term will dominate
  $\Rightarrow$ no convergence!

[1] $\mathbf{x}^n_{i,e}$ and $\mathbf{x}^n_{j,l}$ are uncorrelated

Cross-correlation of Cross-filtered Signals

Consider the cross-correlation between cross-filtered signals:

$$\gamma_{ij} = \mathbb{E}\{\tilde{x}_{ij} \cdot \tilde{x}_{ji}\}$$

Under-modelled case

$$\begin{aligned}\gamma_{ij} &= \mathbb{E}\{[\tilde{x}_{ij,e} + \tilde{x}_{ij,l}][\tilde{x}_{ji,e} + \tilde{x}_{ji,l}]\} \\ &= \mathbb{E}\{[\tilde{x}_{ij,e}\tilde{x}_{ji,e}]\} + \underline{\mathbb{E}\{[\tilde{x}_{ij,l}\tilde{x}_{ji,l}]\}}_{bias\ term}\end{aligned}$$

Cross-correlation of Cross-filtered Signals

Consider the cross-correlation between cross-filtered signals:

$$\gamma_{ij} = \mathbb{E}\{\tilde{x}_{ij} \cdot \tilde{x}_{ji}\}$$

Under-modelled case

$$\begin{aligned}\gamma_{ij} &= \mathbb{E}\{[\tilde{x}_{ij,e} + \tilde{x}_{ij,l}][\tilde{x}_{ji,e} + \tilde{x}_{ji,l}]\} \\ &= \mathbb{E}\{[\tilde{x}_{ij,e}\tilde{x}_{ji,e}]\} + \underline{\mathbb{E}\{[\tilde{x}_{ij,l}\tilde{x}_{ji,l}]\}}_{bias\ term}\end{aligned}$$

- bias term is smaller than for CR
    - cross-correlation of noise-like late reverberation

Cross-correlation of Cross-filtered Signals

Consider the cross-correlation between cross-filtered signals:

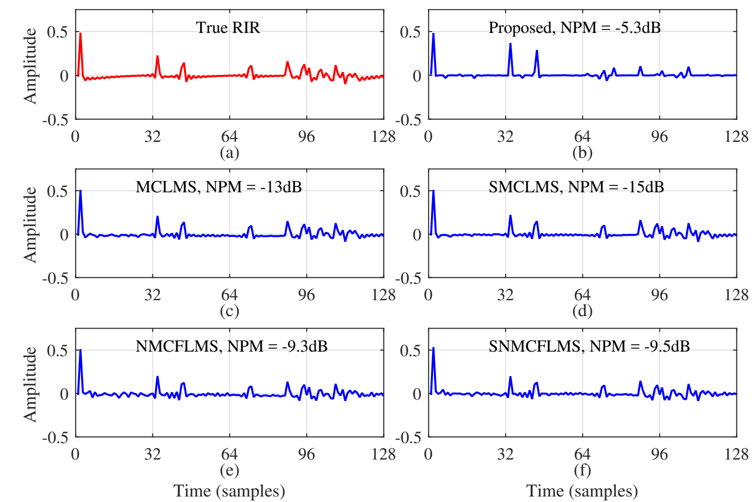
$$\gamma_{ij} = \mathbb{E}\{\tilde{x}_{ij} \cdot \tilde{x}_{ji}\}$$

Under-modelled case

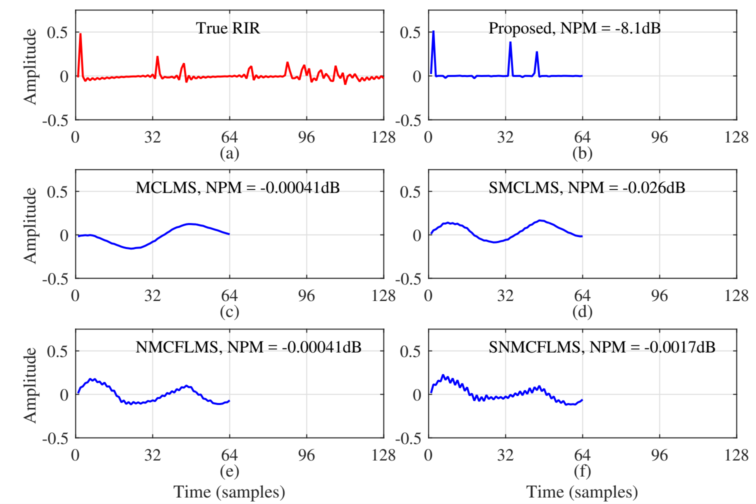
$$\begin{aligned}\gamma_{ij} &= \mathbb{E}\{[\tilde{x}_{ij,e} + \tilde{x}_{ij,l}][\tilde{x}_{ji,e} + \tilde{x}_{ji,l}]\} \\ &= \mathbb{E}\{[\tilde{x}_{ij,e}\tilde{x}_{ji,e}]\} + \underline{\mathbb{E}\{[\tilde{x}_{ij,l}\tilde{x}_{ji,l}]\}}_{bias\ term}\end{aligned}$$

- bias term is smaller than for CR
  - cross-correlation of noise-like late reverberation
- when estimated AIRs approach the true values, $\mathbb{E}\{[\tilde{x}_{ij,e}\tilde{x}_{ji,e}]\}$ increases and bias becomes *less* dominant

Comparison Simulation Results - Fully-modelled



Comparison Simulation Results - Under-modelled



Summary

- Determining the acoustic channels between source and sensors gives crucial information regarding the reverberation process
- Identifying the acoustic channels (system) blindly is challenging because the channels are long and time-varying
- Current research on improved procedures continues to bring benefits
- With sufficient accuracy of BSI, dereverberation can be performed by channel equalization

Questions/Comments

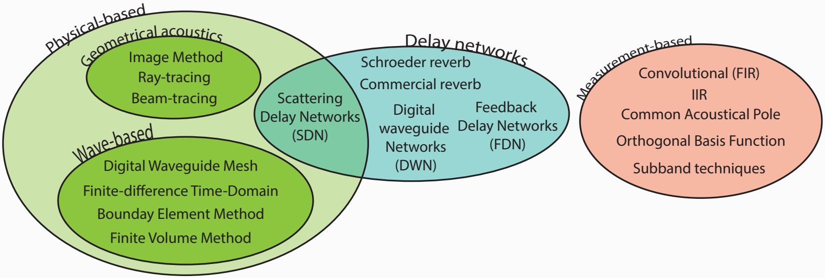
- Questions

Room Acoustics Modelling and Simulation

Outline of the Section

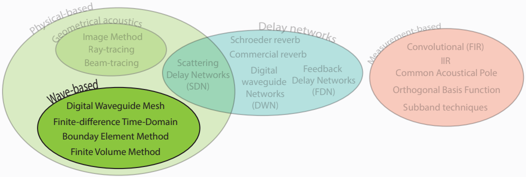
- Overview of available classes of room acoustic models and some recent work in each class
- Focus on popular geometric-based models:
  - Image method
  - Ray tracing
- Focus on perception-based models:
  - Feedback Delay Networks
  - Scattering Delay Networks

Overview



- Excellent overview of past 50 years and more of artificial reverberation by Välimäki *et al.* [33, 34]

Overview



Wave-based models

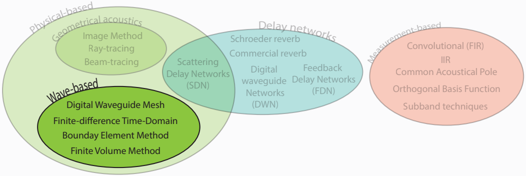
Discretise wave equation in time/frequency and space/boundary/volume

- E.g. FDTD [35] approx. derivatives with finite differences:

$$\frac{\partial^2 p}{\partial t^2} \approx \frac{p_{l,m,i}^{n+1} - 2p_{l,m,i}^{n} + p_{l,m,i}^{n-1}}{T^2} \qquad \frac{\partial^2 p}{\partial x^2} \approx \frac{p_{l-1,m,i}^{n} - 2p_{l,m,i}^{n} + p_{l+1,m,i}^{n}}{X^2}$$

- Convert wave equation into set of linear equations

Overview

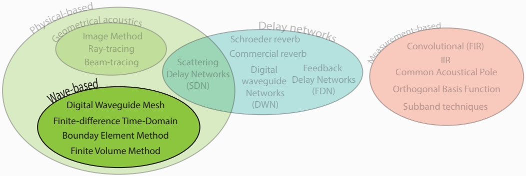


Wave-based models

Discretise wave equation in time/frequency and space/boundary/volume

- Recent work aimed at characterising boundary condition [36, 37, 38], source excitation [39], GPU parallelisation [40, 41]

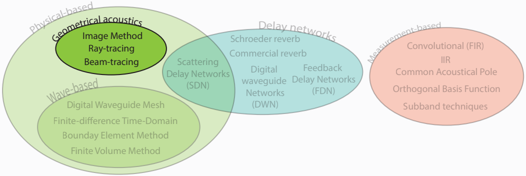
Overview



Wave-based models

Discretise wave equation in time/frequency and space/boundary/volume

- High physical accuracy...
- ...but extremely high computational complexity

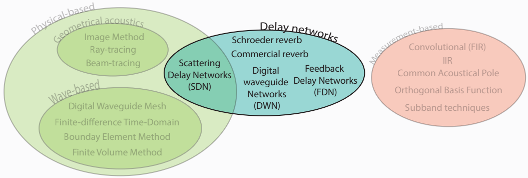


Geometrical acoustics models

Models approximating sound propagation using rays

- Lower computational complexity...
- ...but lower accuracy

Overview

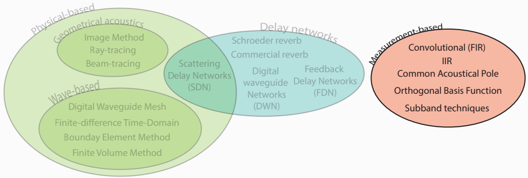


Delay Networks

Methods that do not physically model sound propagation in rooms, but aim to create pleasing reverberant sound

- Very low computational complexity (historically first type of artificial reverberators)...
- ..but no physical accuracy and no explicit modelling

Overview

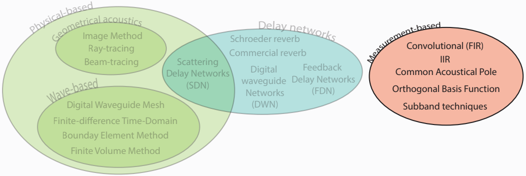


Measurement-based methods

Form parametric representation of room acoustics from real measurements

- E.g. finite impulse response model:
  - Simplest model
  - Very large number of parameters (if $F_s = 40$ kHz, $T_{60} = 2$ s, then $80,000$ parameters)

Overview

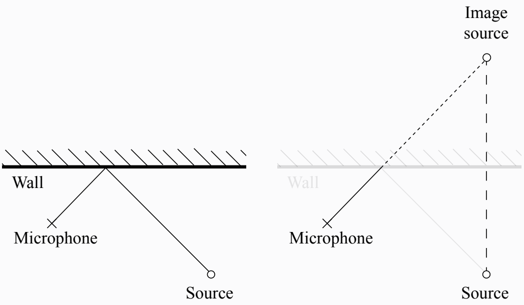


Measurement-based methods

Form parametric representation of room acoustics from real measurements

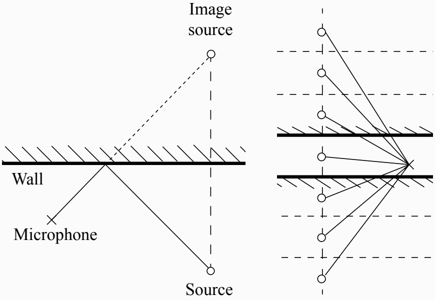
- More compact parametric representation possible with other models, e.g. pole-zero models [42], common acoustical poles [43], orthogonal basis function models [44]

Image method (IM) for single reflector



- Wave propagation in half space is equivalent for:
  1. source and wall
  2. source and image source **(no wall)**
- Exact for rigid wall ($\nabla p \cdot \mathbf{n} = 0$)
- Approximation for non-rigid wall

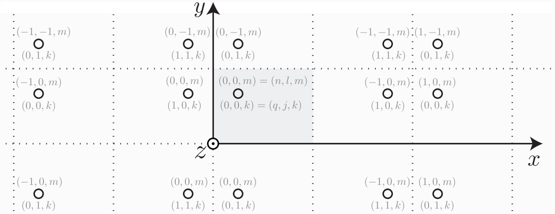
Image method (IM) for multiple surfaces



- **Question:** what about multiple walls?
- Remove wall, mirror source *and* opposite wall
- Repeat until no wall in the region of interest

Image method (IM) for rectangular room

- Rigid wall: exact solution of wave equation [45]
- Non-rigid wall: geometric approximation
- Due to regularity of lattice: simple software implementation

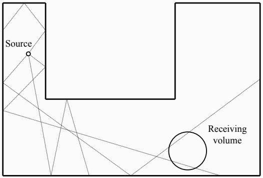


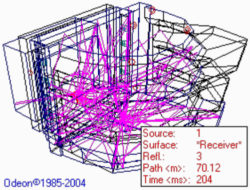

```matlab
function h=rim(mi, so, ro, be, Np, Nr, Tw, Fc)
% mi (microphone), so (source) and ro (room) are
% three-dimensional column vectors.
% Np: num of samples of the RIR.
% Nr: num of random samples (Nr=0 for original IM).
% Tw: samples of low-pass filter, Fc: cut-off freq.
% All quantities above are in sample periods.
% be: matrix of refl. coeff. [x1,y1,z1;x2,y2,z2]
% Example 1: Fig.2a, h=rim([2;1.5;1]/343*4E4,...
% [1;2;2]/343*4E4, [4;4;4]/343*4E4,...
% 0.93.*ones(2,3), 4E4, 0, 40, 0.9);
% Example 2: Fig.11, h=rim([2;1.5;1]/343*4E4,...
% [1;2;2]/343*4E4, [4;4;4]/343*4E4,...
% 0.93.*ones(2,3), 4E4,0.08/343*4E4, 40, 0.9);

h=zeros(Np,1); ps=perm([0,1],[0,1],[0,1]);
Rps=repmat(so,[1,8])+(2.*ps-1).*repmat(mi,[1,8]);
or=floor(Np./(ro.*2))+1;
rs=perm(-or(1):or(1),-or(2):or(2),-or(3):or(3));
for i=1:size(rs'); r=rs(:, i);
 for j=1:8; p=ps(:,j); Rp=Rps(:,j);
  d=norm(2*ro.*r+Rp)+1+Nr*(2*rand-1);
  if round(d)>Np || round(d)<1; continue; end;
  am=be(1,:)'.^abs(r+p).*be(2,:)'.^abs(r);
  if Tw==0; n=round(d); else
    n=(max(ceil(d-Tw/2),1):min(floor(d+Tw/2),Np))';
  end;
  s=(1+cos(2*pi*(n-d)/Tw)).*sinc(Fc*(n-d))/2;
  s(isnan(s))=1; h(n)=h(n)+s*prod(am)/(4*pi*(d-1));
end; end;
function res=perm(varargin);
[res{1:nargin}]=ndgrid(varargin{1:nargin});
res=reshape(cat(nargin+1,res{:}),[],nargin)';
```

- Spatial periodicity of image sources can be exploited for fast rendering in multiple positions [46]

Ray tracing [47]

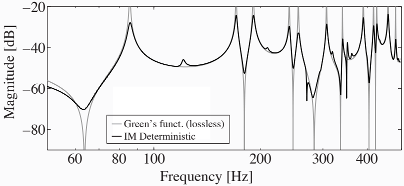
- Source emits rays in all directions
- Specular reflections
- Diffraction and scattering also possible
- Build AIR by recording time and amplitude at receiver
- Choice of receiver size and number of rays is critical





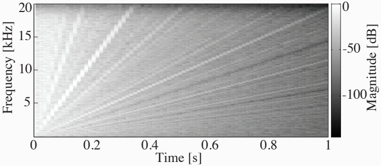
Comparison Ray tracing and Image Method

- Ray tracing
  - Complexity can be controlled by number of rays
  - Can model edge diffraction, scattering
  - No guarantee of low-order reflections
- Image Method
  - Elegance, solution of wave equation for rigid walls
  - Guaranteed all reflections up to certain order present
  - Good rendering of low frequency modes [48]
  - High computational complexity for long AIRs



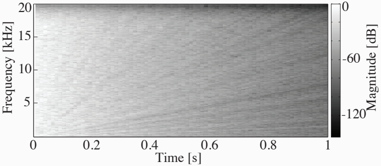
Sweeping echoes in perfectly rectangular rooms

- Perfectly rectangular rooms cause sweeping echoes
- Due to orderly alignment of images along 3 axes [48]
- Regular simulation setups yield stronger sweeping echoes



Room dimension: $4 \times 4 \times 4$
Mic. position $[1, 2, 2]$

Source position $[2, 1.5, 1]$



Room dimension: $4.1 \times 4.2 \times 4.3$
Mic. position $[1.4, 2.5, 2.6]$

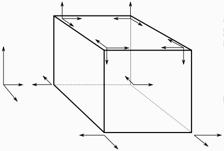
Source position $[2.7, 1.8, 1.9]$

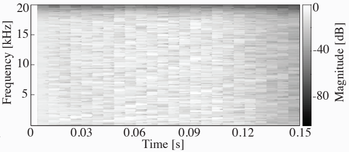
Sweeping echoes in real rooms

- Sweeping echoes can actually be perceived in very regular, empty rooms (e.g. squash court) [49]

But why not in typical rooms?

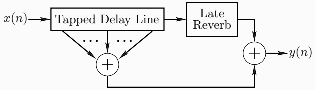
- Room imperfections, objects
- Simulated room with out-of-square imperfections
- Regularity of lattice already broken with 1% imperfections





Perception-based models

- Overview paper by Hacıhabiboğlu *et al.* [50]
- Often separate modules for early and late reverb
- Early reflections using IM and should be spatialised correctly

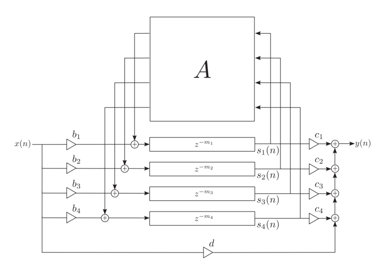


(J. O. Smith, https://ccrma.stanford.edu/~jos)

Desired qualities for late reverb:

- Smooth decay: high echo density
- Smooth frequency response: high mode density
- Moorer's ideal reverb: exponentially decaying white noise

Feedback delay network



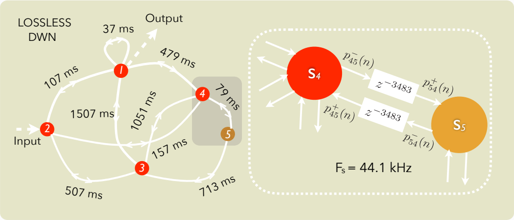
(Schlecht and Habets, 2017)[51]

- Generalization of Schroeder reverberator (Stautner and Puckette, 1982) [52]
- **Design:** start with lossless prototype ($T_{60} = \infty$) to obtain noise-like reverb and add losses to obtain desired reverberation time in each band

Advancements in FDNs

- Jot and Chaigne (1991) [53]:
  - Practical procedure to design delays and FDN matrix to obtain desired echo density and frequency-dependent reverberation time
- Rocchesso and Smith (2002) [54]:
  - Equivalence with DWN
  - Circulant feedback matrix with increased efficiency
- Schlecht and Habets (2015, 2017) [51, 55, 56]:
  - Time-varying FDNs: reduce artifacts and obtain more lively reverberation tail
  - Unilosslessness: new definition of lossless FDN matrix
  - Closed-form and approximated formulas for echo density
  - Procedure to design delays for desired mixing time

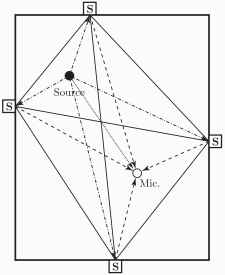
Digital waveguide networks (DWN)



- Network of bi-directional delay lines connected at scattering junctions (Smith, 1985) [57]
- Can be interpreted as network of acoustic tubes
- **Question:** How to set parameters (delay line lengths, network connections, scattering matrix..)?

Scattering delay network (SDN)

- Design DWN based on characteristics of a physical room



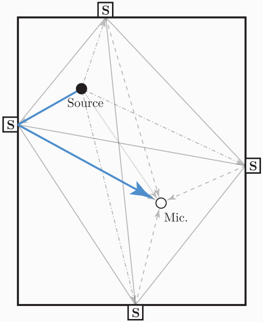
- Position nodes at first-order reflection points
- Fully connected DWN network
- Mono-directional lines for source-junction and junction-mic (De Sena *et al.*, 2015) [58]

Two interpretations:

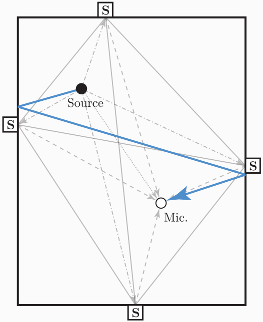
- Physical network of acoustic tubes
- Approximation of image method

SDN: approximation of image method

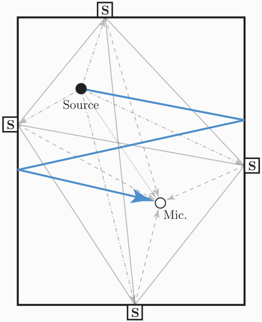
- Correct rendering of LOS and first-order reflections in time, amplitude and direction
- Approximation of second and higher-order reflections, less important perceptually



I-order reflection



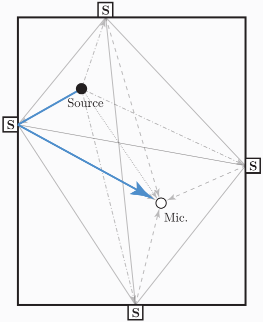
II-order reflection



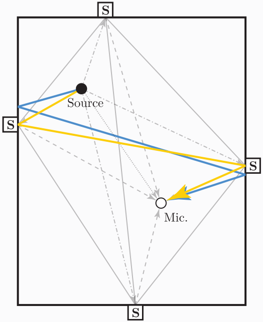
Another II-order reflection

SDN: approximation of image method

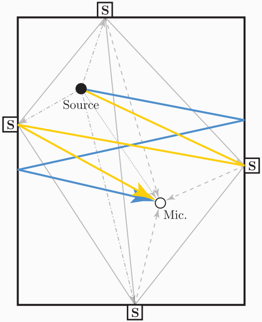
- Correct rendering of LOS and first-order reflections in time, amplitude and direction
- Approximation of second and higher-order reflections, less important perceptually



I-order reflection



II-order reflection



Another II-order reflection

SDN performance

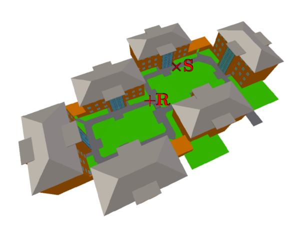
- Significantly faster than convolution alone
- All parameters of model derived from physical properties
- Perceptually more important information given precedence

Advantages w.r.t. delay networks:

- No need for hands-on parameters tuning
- Physical interpretation $\Rightarrow$ spatialisation possible
- More elegant solution than separate early/late modules

Recent advancements in SDN

- Stevens *et al.* (2017) [59]:
  - Extension to exact second-order reflections
  - Implementation of direction-dependent scattering (e.g. modelling of trees)
  - Modelling of outdoor scenes (sky absorbing nodes)
- Schlecht and Habets (2017) [51]:
  - Showed scattering matrix is unilossless



(Stevens et al., 2017) [51]

Summary

- Wide variety of room acoustic models and simulators
- Wave-based models: most accurate available but computationally expensive
- Geometric-models: ray-like assumption, lower complexity but also lower accuracy
- Perception-based models: very fast, attempt to reconstruct only perceptually relevant features of reverberation
- Measurement-based models: parametric representation of room acoustics based on measured AIRs
- Significant advancements have been made in all classes
- Interesting research direction is to find connections between classes (e.g. SDN) and to combine advantages of different classes

Questions/Comments

- Questions

Dereverberation Processing Methods

Overview of dereverberation methods

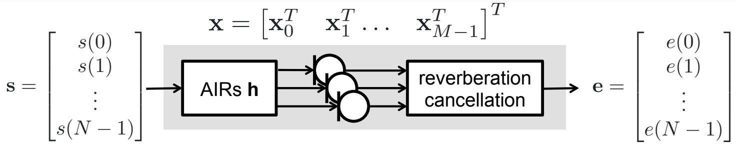
Traditionally, dereverberation processing methods are classified into two categories [60]:

- **Reverberation Cancellation Methods**
  - Modelling reverberation as convolutive interference
  - Recovering speech source signal by (multi-channel) deconvolution
  - Relying on AIR or speech source model
- **Reverberation Suppression Methods**
  - Modelling reverberation as additive interference
  - Recovering speech source signal by spectral/spatial enhancement
  - Relying on spectral/spatial reverberation measure

Recently, some **hybrid methods** have been proposed as well.

Reverberation cancellation

In reverberation cancellation, reverberation is modelled as **convolutive interference**



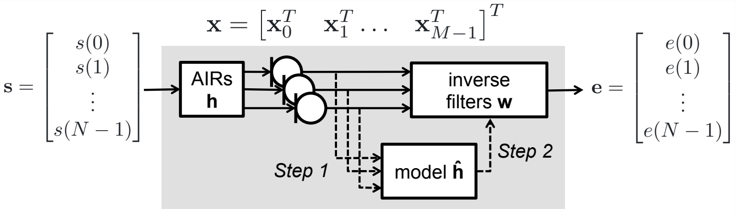
Main methods

- Blind system identification and inversion
- Multi-channel linear prediction

Blind system identification and inversion

Two-step procedure

- Step 1: Blind system identification (BSI, see Section 4)
- Step 2: Multi-channel inverse filter design



©Toon van Waterschoot, KU Leuven

Multi-channel inverse filter design

- Least-squares FIR inverse filter design problem:

$$\min_{\mathbf{w}} \|\mathbf{d} - \hat{\mathbf{H}}\mathbf{w}\|^2$$

  - Target response[61] $\mathbf{d} = [\ \overbrace{0 \quad \dots \quad 0}^{\text{equalization delay}} \quad 1 \;\; 0 \quad \dots \quad 0]^T$
  - SIMO convolution matrix $\hat{\mathbf{H}} = \left[\hat{\mathbf{H}}_0 \quad \dots \quad \hat{\mathbf{H}}_{M-1}\right]$

- **Multiple-input/output Inverse Theorem (MINT)[62]:** Minimum-norm solution $\mathbf{w} = \hat{\mathbf{H}}^T(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)^{-1}\mathbf{d}$ exists if both
  - Estimated AIRs $\hat{\mathbf{h}}_i$ do not share common zeros
  - FIR inverse filter length $L_w = \lceil \frac{L-1}{M-1} \rceil$

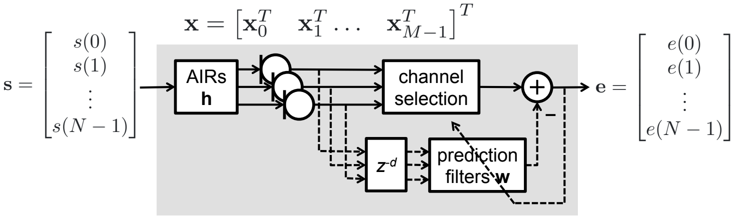
- In case of perfect BSI ($\hat{\mathbf{H}} \equiv \mathbf{H}$), perfect dereverberation is achieved

Multi-channel inverse filter design

- **Challenges** related to MINT-based inverse filter design:
  - Online BSI required in dynamic scenarios
  - Inversion highly sensitive to estimation errors
  - Accurate channel order estimate required
- **Recent advances** in multi-channel inverse filter design:
  - **Channel shortening/reshaping:** maximize early reflections energy while minimizing late reverberation (Rayleigh quotient criterion) [63, 64]
  - **Partial MINT:** maintain early reflections in target $\mathbf{d}$ [65]
  - **Relaxed multi-channel LS:** remove equations for early reflections (weighted LS criterion) [66]
  - **Regularization:** (generalized) Tikhonov regularization in LS [67, 68] or Rayleigh quotient criterion [69]
  - **Sparse regularization:** promote spectral [68] or time-frequency sparsity [70] of equalized speech signal

Multi-channel linear prediction (MCLP)

- Prediction of clean speech signal from multiple microphone signals
- No AIR estimates required
- Operates in time or time-frequency domain



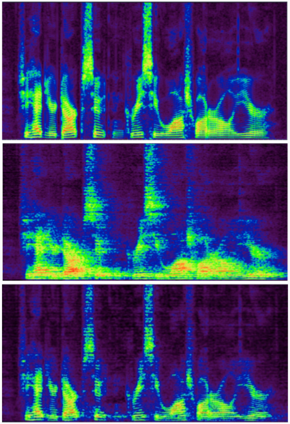
Multi-channel linear prediction

MCLP signal model: $$x_0(n) = \underbrace{\sum_{i=1}^{M-1} \sum_{l=d}^{L_w} w_{i,l} x_i(n-l)}_{\text{reverberation}} + \underbrace{e(n)}_{\substack{\text{enhanced}\\ \text{signal}}}$$

- For **white source signals**, MCLP achieves perfect dereverberation under MINT conditions
- For **speech source signals**, "excessive whitening" of source is alleviated by
  - Increasing prediction delay d [71]
  - Prewhitening microphone signals with inverse source signal model [72]
  - Probabilistic modeling of speech source signal [73, 74]
- **Adaptive MCLP algorithms** based on RLS [75, 76, 77] and Kalman filters [78, 79, 80] have recently been proposed

Multi-channel linear prediction

Simulation example: MCLP with sparse time-frequency prior for speech source signal (Jukić *et al.*, 2015 [74])



©Ante Jukić, University of Oldenburg

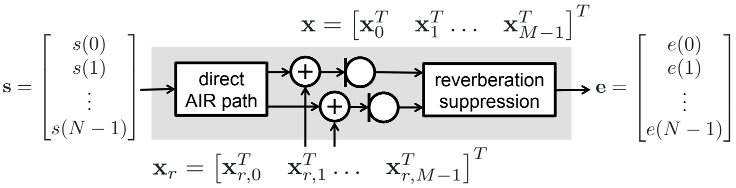
Clean speech

Reverberant speech ($T_{60} = 750$ ms)

Enhanced speech ($M = 4$)

Reverberation suppression

In reverberation suppression, reverberation is modelled as **additive interference**



©Toon van Waterschoot, KU Leuven

Main methods

- Single- & multi-channel spectral enhancement
- Data-independent & data-dependent beamforming

Spectral enhancement

Spectral signal model (short-time power spectral density):

$$\phi_x(p,k) = \phi_{x_d}(p,k) + \phi_{x_r}(p,k)$$

Spectral enhancement

- Step 1: Estimate microphone signal, direct-path signal, and/or reverberant signal PSDs
- Step 2: Apply spectral gain function $g(p,k)$, e.g.

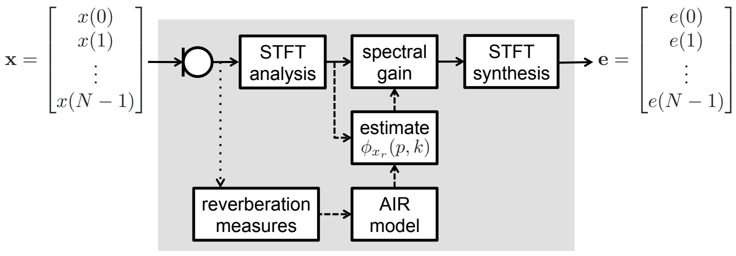
$$g(p,k) = \frac{\hat{\phi}_{x_d}(p,k)}{\hat{\phi}_x(p,k)}, \quad g(p,k) = 1 - \frac{\hat{\phi}_{x_r}(p,k)}{\hat{\phi}_x(p,k)}$$

Key assumption: direct-path and reverberant signals statistically uncorrelated

Single-channel spectral enhancement

Single-channel PSD estimation

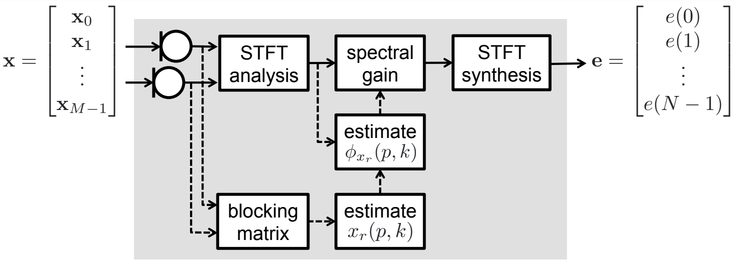
- Use **statistical AIR model** (e.g. Polack's model) to estimate (late) reverberant signal PSD [81]
- Statistical AIR model requires prior information on room acoustics, e.g. by means of **reverberation measures** (C50, T60, etc., see Section 3)



Multi-channel spectral enhancement

Multi-channel PSD estimation: Spatial blocking approach

- Estimate reverberant signal from microphone signals by **spatially blocking direct-path signal** [82]
- Blocking matrix design relies on estimate of source signal direction of arrival (DOA)

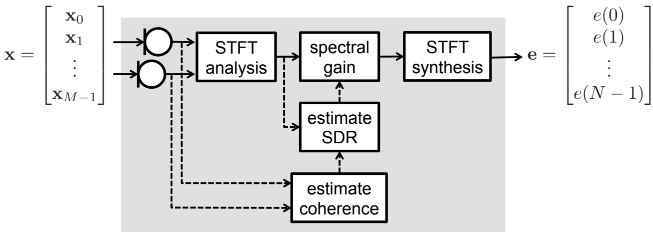


©Toon van Waterschoot, KU Leuven

Multi-channel spectral enhancement

Multi-channel PSD estimation: Spatial coherence approach

- **Key assumption:** direct-path signal is coherent, reverberant signal is diffuse
- Estimate spatial coherence from microphone signals to derive **signal-to-diffuse ratio** (SDR, equivalent to DRR) and design SDR-based spectral gain function [83, 84]
- SDR estimation relies on estimate of source signal DOA or use of directional microphones



Data-independent beamforming

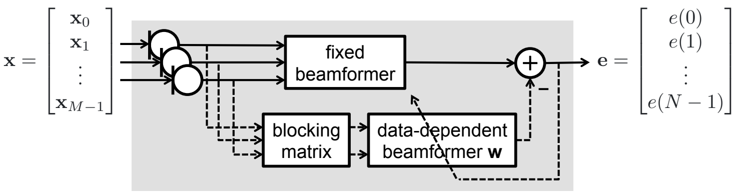
Superdirective beamforming

- **Key assumption:** direct-path signal is coherent, reverberant signal is diffuse
- Beamformer that minimizes diffuse interference (i.e. maximizes directivity) is superdirective beamformer [85]
- **Superdirective beamformer** design requires source signal DOA estimate and diffuse reverberation covariance matrix
- Performance depends strongly on number of microphones, array geometry, source signal DOA, and frequency

Data-dependent beamforming

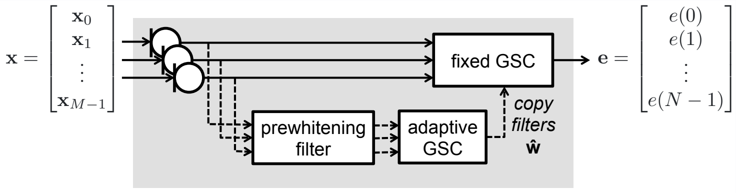
Generalized Sidelobe Canceller (GSC)

- Adaptive implementation of MVDR beamformer
- **No diffuseness assumption** on reverberant signal
- Fixed beamformer (FB) and blocking matrix (BM) design relies on source signal DOA estimate
- Superdirective FB design: data-dependent beamforming **always outperforms** data-independent beamforming



Data-dependent beamforming

- Wiener solution for GSC is **biased** due to source signal coloration [86]
- Impact of bias on enhanced signal very similar to "excessive whitening" in MCLP
- **Prewhitening** proposed for MCLP [72] can also be used in GSC [86]

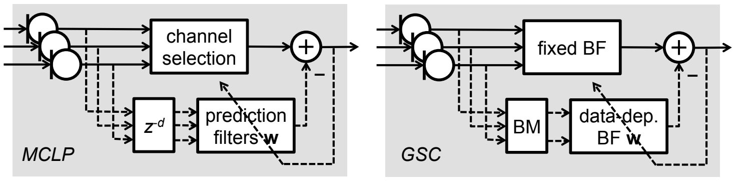


©Toon van Waterschoot, KU Leuven

- Note: Multi-channel Wiener Filter has recently also been employed for dereverberation processing [87]

Hybrid methods

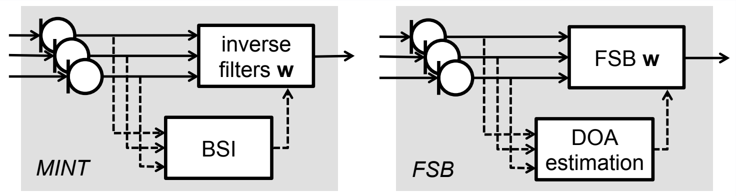
- While based on different signal models, **structural equivalence** exists between MCLP and GSC
- Both methods suffer from bias and excessive whitening due to source signal coloration
- With ideal blocking matrix design (relying on AIR early reflections estimate), **GSC performs equivalent to MCLP**, providing perfect dereverberation for white source signals under MINT conditions [88]
- GSC additionally provides coherent noise cancellation [88]



Hybrid methods

- Similar equivalence can be observed between MINT-based inverse filter design (requiring BSI) and filter-and-sum beamformer (FSB) design (requiring source signal DOA estimate)
- **MINTFormer [89]: hybrid and tunable method** trading off MINT-based dereverberation performance with FSB robustness by weighting MINT and FSB criteria

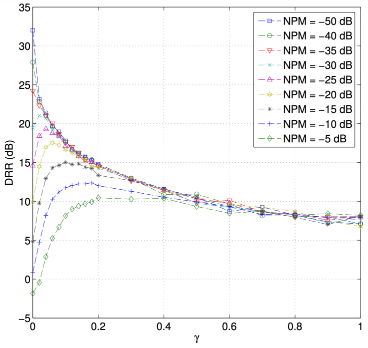
$$\min_{\mathbf{w}} \; \gamma J_{\mathrm{FSB}}(\mathbf{w}) + (1-\gamma) J_{\mathrm{MINT}}(\mathbf{w})$$



Hybrid methods

Simulation example [89]: Output DRR vs. MINTFormer tuning parameter for varying AIR estimation quality (measured by AIR normalized projection misalignment (NPM))

$$\min_{\mathbf{w}} \ \gamma J_{\mathrm{FSB}}(\mathbf{w}) + (1-\gamma) J_{\mathrm{MINT}}(\mathbf{w})$$



Conclusion

Conclusion

- Quantifying **level of reverberation** is highly useful. Several “blind” methods are available with good accuracy in some cases. Recent research includes also prediction of human perceived level of reverberation.
- **AIR measurement** follows well-established procedure and various open-source databases are available.
- Substantial progress in **AIR estimation** (blind and non-blind) has been made while some challenges still remain particularly for real acoustic scenarios.
- **Regularization** turns out to be crucial when addressing AIR estimation and deconvolution problems.
- Broad variety of **room acoustic models** have been proposed in past 50+ years, that can be classified into physical-based models, delay networks, and measurement-based models.

Conclusion

- For **reverberation synthesis** applications, models can broadly be ordered on scale from high accuracy and complexity (physical models) to low accuracy and complexity (delay networks).
- Significant advancements have been made in different model classes but more work remains to be done to find links between models and to combine advantages of different classes.
- Recent work in **dereverberation processing** has brought increased robustness and provides outlook towards adaptivity and scalability.
- Solving dereverberation problem using system identification can be considered as partially solving acoustic scene analysis problem.

Thanks

Niccoló Antonello
Giuliano Bernardi
Zoran Cvetković
Thomas Dietzen
Mathieu Hu
Alastair Moore
Julius O. Smith
Giacomo Vairetti
Wei Xue

Questions/Comments

- Questions

References

[1] H. Kuttruff. *Room Acoustics*. Taylor & Francis, 2000.

[2] M. Schroeder and H. Kuttruff. "On frequency response curves in rooms. Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima". In: *J. Acoust. Soc. Am.* 34.1 (1962), pp. 76–80.

[3] L. L. Beranek. *Acoustics*. Acoustical Society of America, 1993.

[4] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot. "Perception of reverberation in small rooms: A literature study". In: *Proc. AES 55th Int. Conf. Spatial Audio*. Helsinki, Finland, Aug. 2014.

[5] W. C. Sabine. *Collected Papers on Acoustics*. Peninsula Publishing, 1993 (Originally 1921).

[6] T. H. Falk, C. Zheng, and W.-Y. Chan. "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech". In: *IEEE Trans. Audio, Speech, Lang. Process.* 18.7 (Sept. 2010), pp. 1766–1774.

[7] J. B. Allen. "Effects of Small Room Reverberation on Subjective Preference". In: *J. Acoust. Soc. Am.* 71.S5 (1982).

[8] A. A. de Lima, T. de M. Prego, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker, and M. Fozunbal. "On the quality-assessment of reverberated speech". In: *Speech Communication* 54 (2012), pp. 394–401.

[9] J. M. F. del Vallado, A. A. de Lima, T. de M. Prego, and S. L. Netto. "Feature analysis for the reverberation perception in speech signals". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (2013), pp. 8169–8173.

[10] J. Eaton, H. A. Javed, and P. A. Naylor. "Estimation of the perceived level of reverberation using non-intrusive single-channel variance of decay rates". In: *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*. San Francisco, California, USA, 2017.

References

[11] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor. "Blind estimation of reverberation time based on the distribution of signal decay rates". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Las Vegas, USA, Apr. 2008, pp. 329–332.

[12] H. A. Javed, B. Cauchi, S. Doclo, P. A. Naylor, and S. Goetze. "Measuring, modelling and predicting perceived reverberation". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2017, pp. 381–385.

[13] A. Farina. "Simultaneous measurement of impulse response and distortion with a swept-sine technique". In: *Proc. Audio Eng. Soc. (AES) Convention*. Paris, France, 2000.

[14] S. Müller and P. Massarani. "Transfer-function measurement with sweeps". In: *J. Audio Eng. Soc. (AES)* 49.6 (June 2001), pp. 443–471.

[15] G. Vairetti, N.Kaplanis, S. Bech, E. De Sena, M. Moonen, and T. van Waterschoot. "The Subwoofer Room Impulse Response database (SUBRIR)". In: *J. Audio Eng. Soc. (AES)* 65.5 (May 2017), pp. 389–401.

[16] A. Torras-Rosell and F. Jacobsen. "Measuring long impulse responses with pseudorandom sequences and sweep signals". In: *Proc. 39th Int. Congr. Noise Control Eng. (INTERNOISE 2010)*. Lisbon, Portugal, 2010.

[17] T. van Waterschoot, G. Rombouts, and M. Moonen. "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement". In: *Signal Processing* 88.3 (Mar. 2008), pp. 594–611.

[18] *GTAC Database*. Audio and communications signal processing group (GTAC), UPV, Spain. `www.gtac.upv.es/room.php`. 2017.

[19] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor. "Estimation of room acoustic parameters: The ACE challenge". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 24.10 (Oct. 2016), pp. 1681–1693.

References

[20] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen. "The single- and multichannel audio recordings database (SMARD)". In: *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*. Antibes, France, Sept. 2014, pp. 40–44.

[21] R. Stewart and M. Sandler. "Database of omnidirectional and B-format impulse responses". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Dallas, TX, USA, Mar. 2010.

[22] M. Jeub, M. Schäfer, and P. Vary. "A binaural room impulse response database for the evaluation of dereverberation algorithms". In: *Proc. IEEE Intl. Conf. Digital Signal Processing (DSP)*. Santorini, Greece, July 2009.

[23] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier. "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses". In: *EURASIP J. on Advances in Signal Processing* 2009 (2009), pp. 1–10.

[24] J. Y. Wen, N. D. Gaubitch, E. A. Habets, T. Myatt, and P. A Naylor. "Evaluation of speech dereverberation algorithms using the MARDY database". In: *Proc. Intl. on Workshop Acoust. Echo and Noise Control (IWAENC)*. Paris, France, Sept. 2006.

[25] D. Murphy and S. Shelley. *Open AIR Library*. Audiolab, Department of Electronics, University of York, UK. `www.openairlib.net`. 2010.

[26] A. J. Berkhout, D. de Vries, and J. J. Sonke. "Array technology for acoustic wave field analysis in enclosures". In: *J. Acoust. Soc. Am.* (Nov. 1997), pp. 2757–2770.

[27] T. Ajdler, L. Sbaiz, and M. Vetterli. "The plenacoustic function and its sampling". In: *IEEE Trans. Signal Process.* 54.10 (Oct. 2006), pp. 3790–3804.

References

[28] R. Mignot, L. Daudet, and F. Ollivier. "Room reverberation reconstruction: Interpolation of the early part using compressed sensing". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 21.11 (Nov. 2013), pp. 2301–2312.

[29] R. Mignot, G. Chardon, and L. Daudet. "Low frequency interpolation of room impulse responses using compressed sensing". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 22.1 (Jan. 2014), pp. 205–216.

[30] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot. "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* (2017, in press).

[31] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins. "Sound-field measurement with moving microphones". In: *J. Acoust. Soc. Am.* (May 2017), pp. 3220–3235.

[32] M. Hu, S. Doclo, D. Sharma, M. Brookes, and P. A. Naylor. "Noise robust blind system identification algorithms based on a Rayleigh quotient cost function". In: *Proc. European Signal Processing Conf. (EUSIPCO)*. Nice, France, Aug. 2015.

[33] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel. "Fifty years of artificial reverberation". In: *IEEE Trans. Audio, Speech, Lang. Process.* 20.5 (2012), pp. 1421–1448.

[34] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel. "More than 50 years of artificial reverberation". In: *Proc. Audio Eng. Soc. (AES) Conf.: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*. Leuven, Belgium, Feb. 2016.

[35] D. Botteldooren. "Finite-difference time-domain simulation of low-frequency room acoustic problems". In: *J. Acoust. Soc. Am.* 98.6 (1995), pp. 3302–3308.

[36] K. Kowalczyk and M. van Walstijn. "Room acoustics simulation using 3-D compact explicit FDTD schemes". In: *IEEE Trans. Audio, Speech, Lang. Process.* 19.1 (2011), pp. 34–46.

References

[37] S. Bilbao. "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation". In: *IEEE Trans. Audio, Speech, Lang. Process.* 21.7 (2013), pp. 1524–1533.

[38] S. Bilbao, B. Hamilton, J. Botts, and L. Savioja. "Finite volume time domain room acoustics simulation under general impedance boundary conditions". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 24.1 (2016), pp. 161–173.

[39] D. Murphy, A. Southern, and L. Savioja. "Source excitation strategies for obtaining impulse responses in finite difference time domain room acoustics simulation". In: *Applied Acoustics* 82 (2014), pp. 6–14.

[40] R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, and D. Manocha. "An efficient GPU-based time domain solver for the acoustic wave equation". In: *Applied Acoustics* 73.2 (2012), pp. 83–94.

[41] C. J. Webb and S. Bilbao. "Computing room acoustics with CUDA-3D FDTD schemes with boundary losses and viscosity". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP).* Prague, Czech Republic, May 2011, pp. 317–320.

[42] H. Brandenstein and R. Unbehauen. "Least-squares approximation of FIR by IIR digital filters". In: *IEEE Trans. Signal Process.* 46.1 (1998), pp. 21–30.

[43] Y. Haneda, S. Makino, and Y. Kaneda. "Common acoustical pole and zero modeling of room transfer functions". In: *IEEE Trans. Speech Audio Process.* 2.2 (Apr. 1994), pp. 320–328.

[44] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot. "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 25.7 (2017), pp. 1547–1561.

[45] J. B. Allen and D. A. Berkley. "Image method for efficiently simulating small-room acoustics". In: *J. Acoust. Soc. Am.* 65.4 (1979), pp. 943–950.

References

[46] J. Martinez. “Low complexity computer simulation of multichannel room impulse responses”. PhD thesis. TU Delft, 2013.

[47] A. Krokstad, S. Strøm, and S. Sørsdal. “Calculating the acoustical room response by the use of a ray tracing technique”. In: *J. of Sound and Vibration* 8.1 (1968), pp. 118–125.

[48] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot. “On the modeling of rectangular geometries in room acoustic simulations”. In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 23.3 (Apr. 2015), pp. 774–786.

[49] K. Kiyohara, K. Furuya, and Y. Kaneda. “Sweeping echoes perceived in a regularly shaped reverberation room”. In: *J. Acoust. Soc. Am.* 111.2 (2002), pp. 925–930.

[50] H. Hacıhabiboğlu, E. De Sena, Z. Cvetković, J. Johnston, and J.O. Smith. “Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics”. In: *IEEE Signal Process. Mag.* 34.3 (2017), pp. 36–54.

[51] S. J. Schlecht and E. A. P. Habets. “On lossless feedback delay networks”. In: *IEEE Trans. Signal Process.* 65.6 (2017), pp. 1554–1564.

[52] J. Stautner and M. Puckette. “Designing multichannel reverberators”. In: *Computer Music J.* 6.1 (1982), pp. 52–65.

[53] J.-M. Jot and A. Chaigne. “Digital delay networks for designing artificial reverberators”. In: *Proc. Audio Eng. Soc. (AES) Convention*. Paris, France, 1991.

[54] D. Rocchesso and J. O. Smith. “Circulant and elliptic feedback delay networks for artificial reverberation”. In: *IEEE Trans. Speech Audio Process.* 5.1 (Jan. 1997), pp. 51–63.

References

[55] S. J. Schlecht and E. A. P. Habets. "Time-varying feedback matrices in feedback delay networks and their application in artificial reverberation". In: *J. Acoust. Soc. Am.* 138.3 (2015), pp. 1389–1398.

[56] S. J. Schlecht and E. A. P. Habets. "Feedback delay networks: Echo density and mixing time". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 25.2 (2017), pp. 374–383.

[57] J. O. Smith. "A new approach to digital reverberation using closed waveguide networks". In: *Proc. 11th Int. Comput. Music Conf.* Burnaby, Canada, 1985, pp. 47–53.

[58] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith. "Efficient synthesis of room acoustics via scattering delay networks". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 23.9 (Sept. 2015), pp. 1478–1492.

[59] F. Stevens, D. T. Murphy, L. Savioja, and V. Välimäki. "Modeling sparsely reflecting outdoor acoustic scenes using the waveguide web". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 25.8 (Aug. 2017), pp. 1566–1578.

[60] E. A. P. Habets. "Single- and multi-microphone speech dereverberation using spectral enhancement". PhD thesis. The Netherlands: Eindhoven University of Technology, 2007.

[61] T. Hikichi, M. Delcroix, and M. Miyoshi. "Inverse filtering for speech dereverberation less sensitive to noise". In: *Proc. Intl. on Workshop Acoust. Echo and Noise Control (IWAENC).* Paris, France, Sept. 2006.

[62] M. Miyoshi and Y. Kaneda. "Inverse filtering of room acoustics". In: *IEEE Trans. Acoust., Speech, Signal Process.* ASSP-36.2 (Feb. 1988), pp. 145–152.

[63] M. Kallinger and A. Mertins. "Multi-channel room impulse response shaping – a study". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP).* Toulouse, France, May 2006, pp. 101–104.

References

[64] W. Zhang, E. A. P. Habets, and P. A. Naylor. “On the use of channel shortening in multichannel acoustic system equalization”. In: *Proc. Intl. on Workshop Acoust. Echo and Noise Control (IWAENC)*. Tel Aviv, Israel, Sept. 2010.

[65] I. Kodrasi and S. Doclo. “Robust partial multichannel equalization techniques for speech dereverberation”. In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Kyoto, Japan, Mar. 2012, pp. 537–540.

[66] F. Lim, W. Zhang, E. A. P. Habets, and P. A. Naylor. “Robust multichannel dereverberation using relaxed multichannel least squares”. In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 22.9 (Sept. 2014), pp. 1379–1390.

[67] T. Hikichi, M. Delcroix, and M. Miyoshi. “Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations”. In: *EURASIP J. on Advances in Signal Processing* 2007 (2007), pp. 1–12.

[68] T. van Waterschoot, B. Defraene, M. Diehl, and M. Moonen. “Embedded optimization algorithms for multi-microphone dereverberation”. In: *Proc. European Signal Processing Conf. (EUSIPCO)*. Marrakech, Morocco, Sept. 2013.

[69] I. Kodrasi, S. Goetze, and S. Doclo. “Regularization for partial multichannel equalization for speech dereverberation”. In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 21.9 (Sept. 2013), pp. 1879–1890.

[70] I. Kodrasi, A. Jukić, and S. Doclo. “Robust sparsity-promoting acoustic multi-channel equalization for speech dereverberation”. In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai, China, Mar. 2016, pp. 166–170.

[71] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi. “Suppression of late reverberation effect on speech signal using long-term multiple-step linear prediction”. In: *IEEE Trans. Audio, Speech, Lang. Process.* 17.4 (May 2009), pp. 534–545.

References

[72] M. Triki and D. T. M. Slock. "Blind dereverberation of a single source based on multichannel linear prediction". In: *Proc. Intl. on Workshop Acoust. Echo and Noise Control (IWAENC)*. Eindhoven, The Netherlands, Sept. 2005, pp. 173–176.

[73] T. Nakatani, B. H. Juang, T. Yoshioka, K. Kinoshita, M. Delcroix, and M. Miyoshi. "Speech dereverberation based on maximum-likelihood estimation with time-varying Gaussian source model". In: *IEEE Trans. Audio, Speech, Lang. Process.* 16.8 (Nov. 2008), pp. 1512–1527.

[74] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo. "Multi-channel linear prediction-based speech dereverberation with sparse priors". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 23.9 (Sept. 2015), pp. 1509–1520.

[75] T. Yoshioka, H. Tachibana, T. Nakatani, and M. Miyoshi. "Adaptive dereverberation of speech signals with speaker-position change detection". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Taipei, Taiwan, Apr. 2009, pp. 3733–3736.

[76] T. Yoshioka and T. Nakatani. "Dereverberation for reverberation-robust microphone arrays". In: *Proc. European Signal Processing Conf. (EUSIPCO)*. Marrakech, Morocco, Sept. 2013.

[77] A. Jukić, Z. Wang, T. van Waterschoot, T. Gerkmann, and S. Doclo. "Constrained multi-channel linear prediction for adaptive speech dereverberation". In: *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*. Xi'an, China, Sept. 2016.

[78] T. Dietzen, A. Spriet, W. Tirry, S. Doclo, M. Moonen, and T. van Waterschoot. "Partitioned block frequency domain Kalman filter for multi-channel linear prediction based blind speech dereverberation". In: *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*. Xi'an, China, Sept. 2016.

[79] S. Braun and E. A. P. Habets. "Online dereverberation for dynamic scenarios using a Kalman filter with and autoregressive model". In: *IEEE Signal Process. Lett.* 23.12 (Dec. 2016), pp. 1741–1745.

References

[80] T. Dietzen, S. Doclo, A. Spriet, W. Tirry, M. Moonen, and T. van Waterschoot. "Low complexity Kalman filter for multi-channel linear prediction based blind speech dereverberation". In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, NY, USA, Oct. 2017.

[81] E. A. P. Habets. "Multi-channel speech dereverberation based on a statistical model of late reverberation". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Philadelphia, USA, Mar. 2005, pp. 173–176.

[82] E. A. P. Habets and S. Gannot. "Dual-microphone speech dereverberation using a reference signal". In: *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Honolulu, USA, Apr. 2007, pp. 901–904.

[83] O. Thiergart, G. Del Galdo, and E. A. P. Habets. "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation". In: *J. Acoust. Soc. Am.* 132.4 (Oct. 2012), pp. 2337–2346.

[84] A. Schwarz and W. Kellermann. "Coherent-to-diffuse power ratio estimation for dereverberation". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 23.6 (June 2015), pp. 1006–1018.

[85] E. A. P. Habets and J. Benesty. "A two-stage beamforming approach for noise reduction and dereverberation". In: *IEEE Trans. Audio, Speech, Lang. Process.* 21.5 (May 2013), pp. 945–958.

[86] T. Dietzen, N. Huleihel, A. Spriet, W. Tirry, S. Doclo, M. Moonen, and T. van Waterschoot. "Speech dereverberation by data-dependent beamforming with signal pre-whitening". In: *Proc. European Signal Processing Conf. (EUSIPCO)*. Nice, France, Aug. 2015.

[87] A. Kuklasiński and J. Jensen. "Multi-channel Wiener filters in binaural and bilateral hearing aids – speech intelligibility improvement and robustness to DoA errors". In: *J. Audio Eng. Soc. (AES)* 65.1/2 (Jan. 2017), pp. 8–16.

References

[88] T. Dietzen, A. Spriet, W. Tirry, S. Doclo, M. Moonen, and T. van Waterschoot. "On the relation between data-dependent beamforming and multichannel linear prediction for dereverberation". In: *Proc. Audio Eng. Soc. (AES) Conf.: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*. Leuven, Belgium, Feb. 2016.

[89] F. Lim, M. R. P. Thomas, and P. A. Naylor. "MINTFormer: A spatially aware channel equalizer". In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, NY, USA, Oct. 2013.