Binaural Extension and Performance of Single-Channel Spectral Subtraction Dereverberation Algorithms

Alexandros Tsilfidis, Eleftheria Georganti and John Mourjopoulos

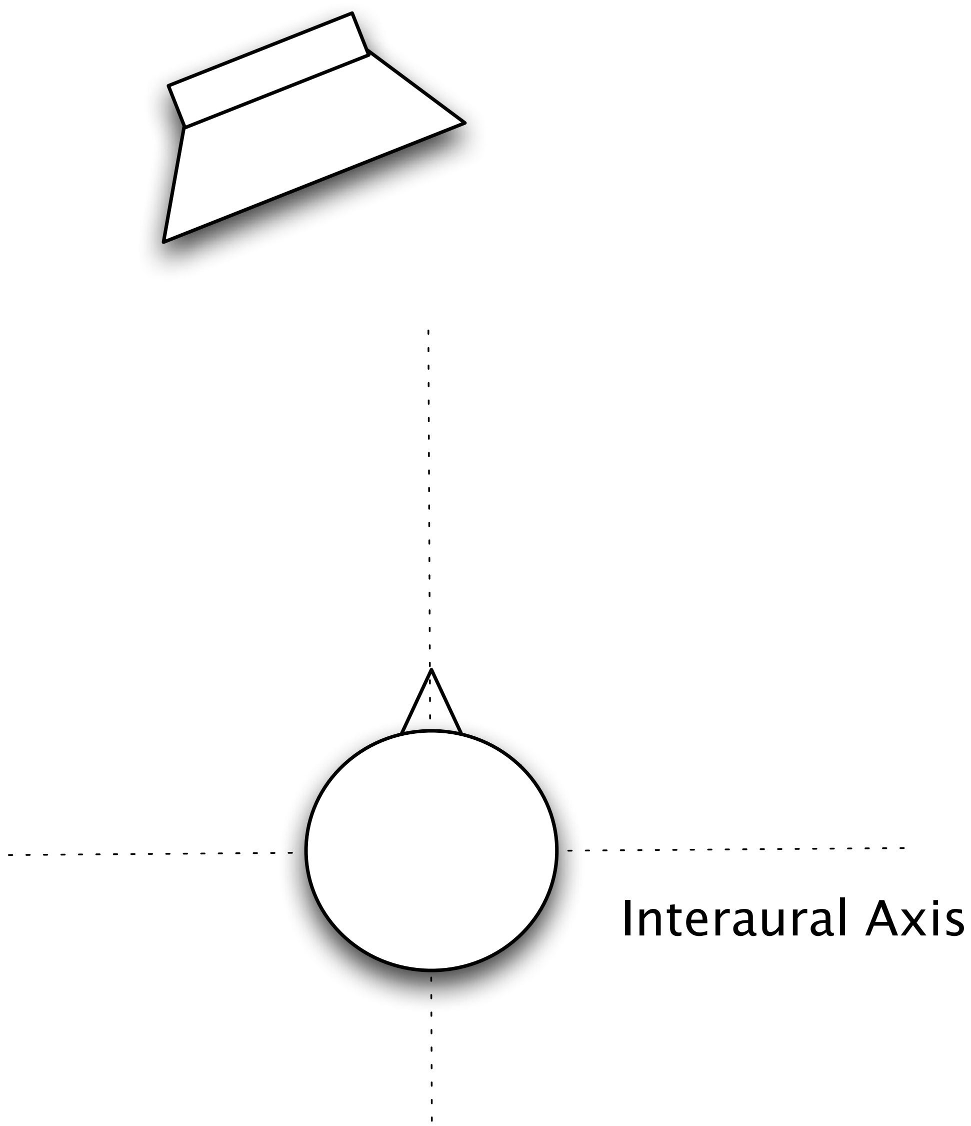
Audio and Acoustic Technology Group
Wire Communications Laboratory
Department of electrical and Computer Engineering
University of Patras
Greece

Tsilfidis et al.
University of Patras
Greece

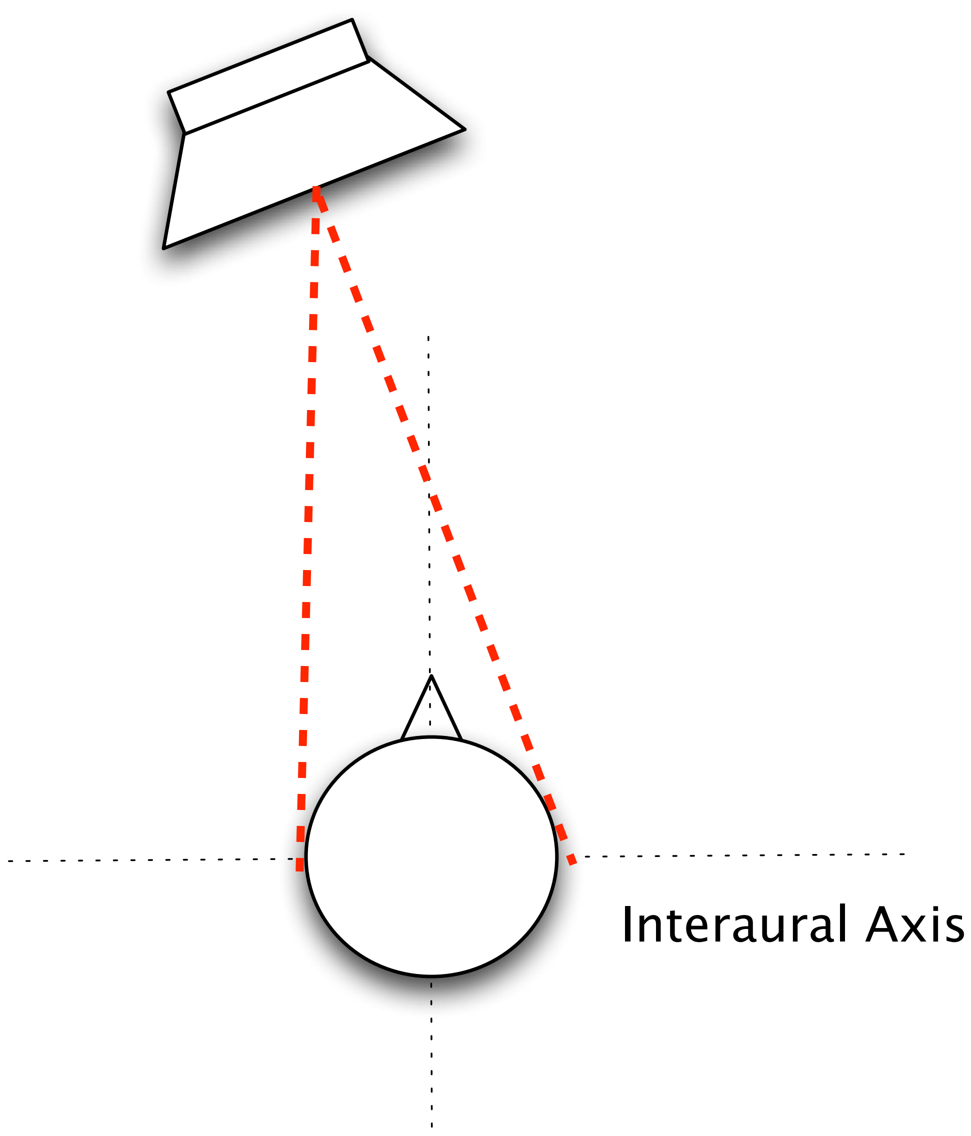
WHY?

Tsilfidis et al.
University of Patras
Greece

WHY?



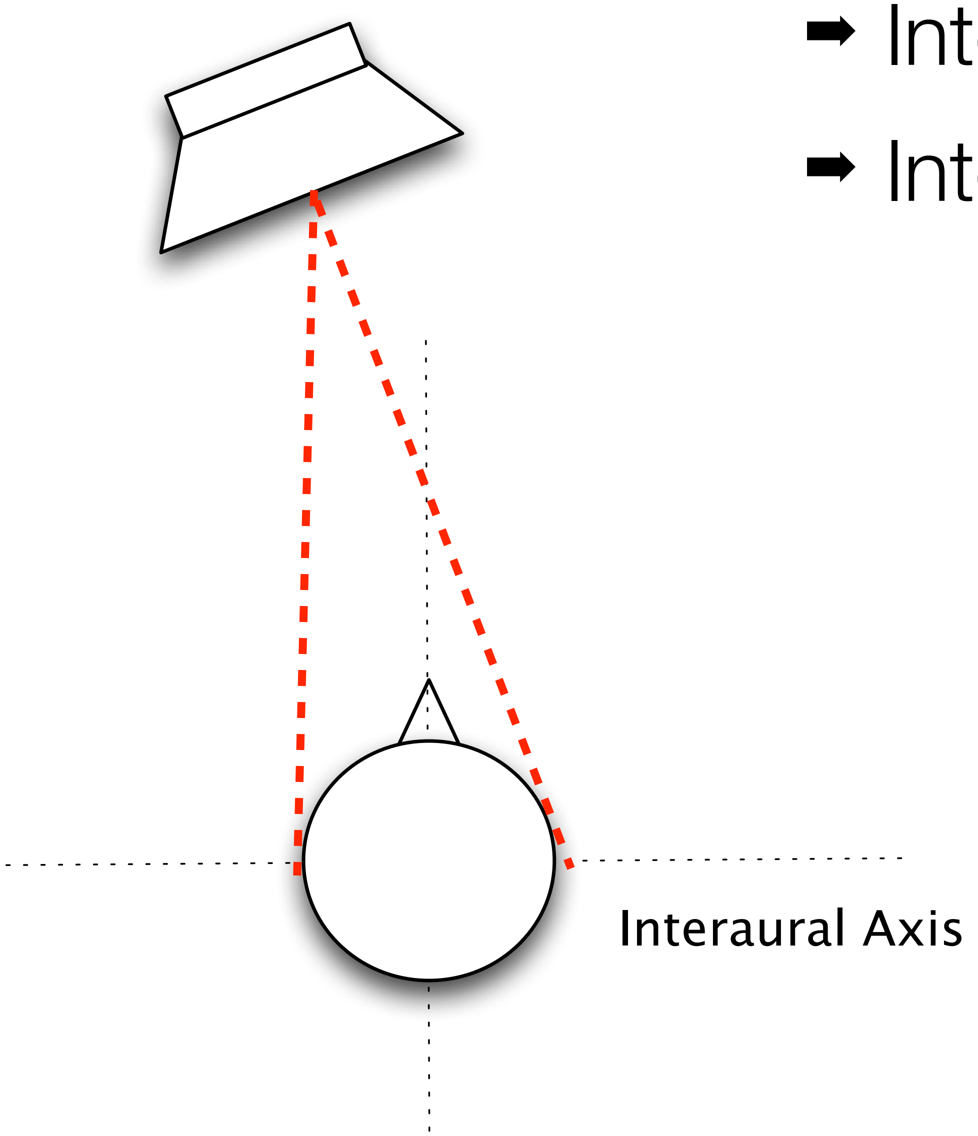
WHY?



Tsilfidis et al.
University of Patras
Greece

WHY?

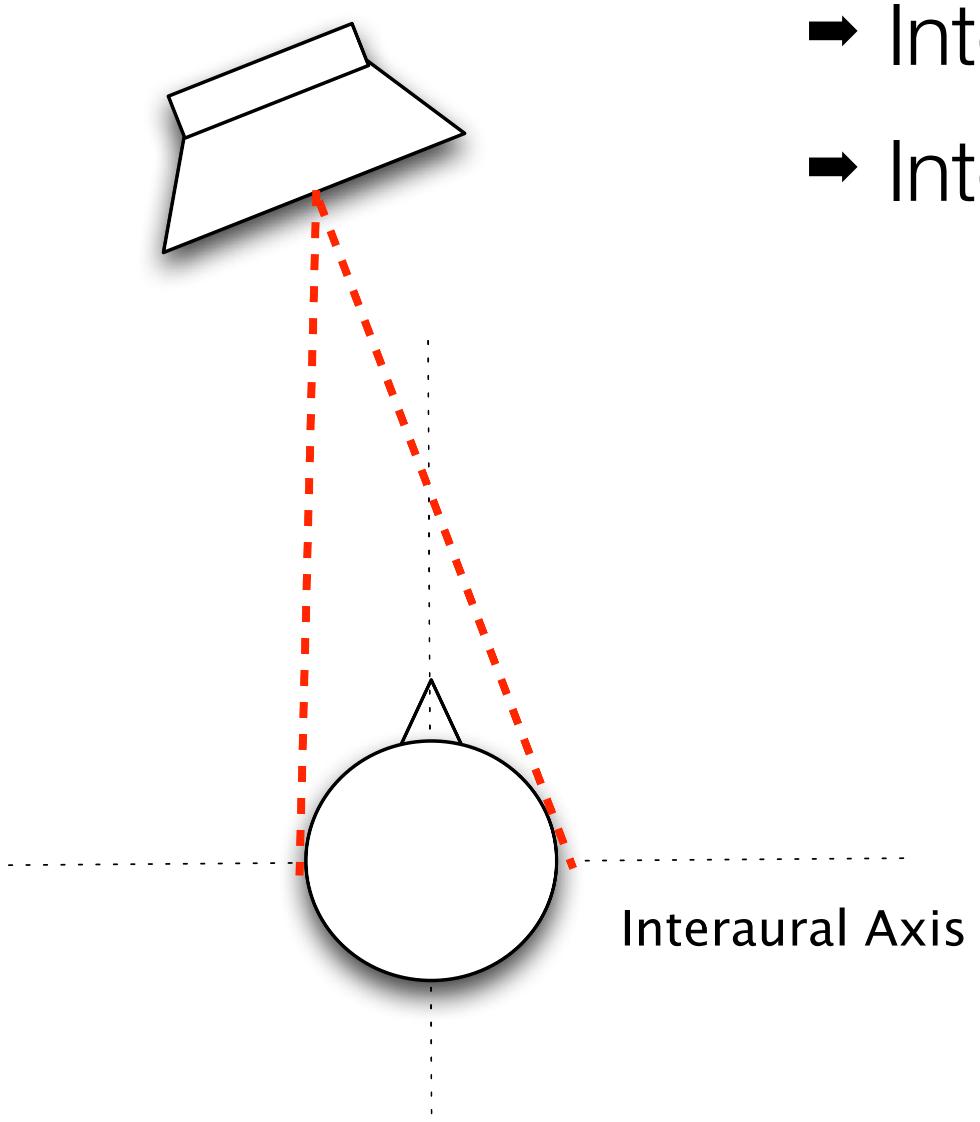
Binaural Cues



- ➡ Interaural Time Difference (ILD)
- ➡ Interaural Level Difference (ITD)

Tsilfidis et al.
University of Patras
Greece

WHY?



Binaural Cues

- ➡ Interaural Time Difference (ILD)
- ➡ Interaural Level Difference (ITD)

Applied in:

- Hearing Aids
- Binaural Telephony
- Hands-Free Devices
- Immersive Audio Applications

HOW?

HOW?

Improved single-channel spectral subtraction dereverberation

A. Tsilfidis and J. Mourjopoulos. "Signal-dependent constraints for perceptually motivated suppression of late reverberation". Signal Processing, 90: 959965, 2010.

A. Tsilfidis and J. Mourjopoulos. "Perceptually-motivated selective suppression of late reverberation". In Digital Signal Processing Conference (DSP), Santorini, Greece, 2009

Dereverberation based on perceptual reverberation modeling

A. Tsilfidis and J. Mourjopoulos. "Blind single-channel suppression of late reverberation based on perceptual reverberation modeling". Journal of the Acoustical Society of America, 129(3): 1439-1451, 2011.

A. Tsilfidis and J. Mourjopoulos. "Blind single-channel dereverberation for music post-processing". In Proc. of the 130th Convention of the Audio Engineering Society, London, UK, 2011.

Semi Blind Dereverberation based on a handclap recording

A. Tsilfidis, E. Kokkinis, and J. Mourjopoulos. "Suppression of late reverberation at multiple speaker positions utilizing a single impulse response measurement". In Forum Acusticum, Aalborg, Denmark, 2011.

A. Tsilfidis, E. Georganti, E. Kokkinis, and J. Mourjopoulos. "Speech dereverberation based on a recorded handclap". In Digital Signal Processing Conference (DSP), Corfu, Greece, 2011.

HOW?

Extend single-channel dereverberation techniques

Improved single-channel spectral subtraction dereverberation

A. Tsilfidis and J. Mourjopoulos. “Signal-dependent constraints for perceptually motivated suppression of late reverberation”. Signal Processing, 90: 959965, 2010.

A. Tsilfidis and J. Mourjopoulos. “Perceptually-motivated selective suppression of late reverberation”. In Digital Signal Processing Conference (DSP), Santorini, Greece, 2009

Dereverberation based on perceptual reverberation modeling

A. Tsilfidis and J. Mourjopoulos. “Blind single-channel suppression of late reverberation based on perceptual reverberation modeling”. Journal of the Acoustical Society of America, 129(3): 1439-1451, 2011.

A. Tsilfidis and J. Mourjopoulos. “Blind single-channel dereverberation for music post-processing”. In Proc. of the 130th Convention of the Audio Engineering Society, London, UK, 2011.

Semi Blind Dereverberation based on a handclap recording

A. Tsilfidis, E. Kokkinis, and J. Mourjopoulos. “Suppression of late reverberation at multiple speaker positions utilizing a single impulse response measurement”. In Forum Acusticum, Aalborg, Denmark, 2011.

A. Tsilfidis, E. Georganti, E. Kokkinis, and J. Mourjopoulos. “Speech dereverberation based on a recorded handclap”. In Digital Signal Processing Conference (DSP), Corfu, Greece, 2011.

HOW?

Extend single-channel dereverberation techniques

Improved single-channel spectral subtraction dereverberation

A. Tsilfidis and J. Mourjopoulos. "Signal-dependent constraints for perceptually motivated suppression of late reverberation". Signal Processing, 90: 959965, 2010.

A. Tsilfidis and J. Mourjopoulos. "Perceptually-motivated selective suppression of late reverberation". In Digital Signal Processing Conference (DSP), Santorini, Greece, 2009

Dereverberation based on perceptual reverberation modeling

A. Tsilfidis and J. Mourjopoulos. "Blind single-channel suppression of late reverberation based on perceptual reverberation modeling". Journal of the Acoustical Society of America, 129(3): 1439-1451, 2011.

A. Tsilfidis and J. Mourjopoulos. "Blind single-channel dereverberation for music post-processing". In Proc. of the 130th Convention of the Audio Engineering Society, London, UK, 2011.

Semi Blind Dereverberation based on a handclap recording

A. Tsilfidis, E. Kokkinis, and J. Mourjopoulos. "Suppression of late reverberation at multiple speaker positions utilizing a single impulse response measurement". In Forum Acusticum, Aalborg, Denmark, 2011.

A. Tsilfidis, E. Georganti, E. Kokkinis, and J. Mourjopoulos. "Speech dereverberation based on a recorded handclap". In Digital Signal Processing Conference (DSP), Corfu, Greece, 2011.

Objectives

Objectives

✓ Propose a simple framework for the extension of single channel spectral subtraction dereverberation algorithms

Objectives

- ✓ Propose a simple framework for the extension of single channel spectral subtraction dereverberation algorithms
- ✓ Implement an efficient way to prevent overestimation errors

Objectives

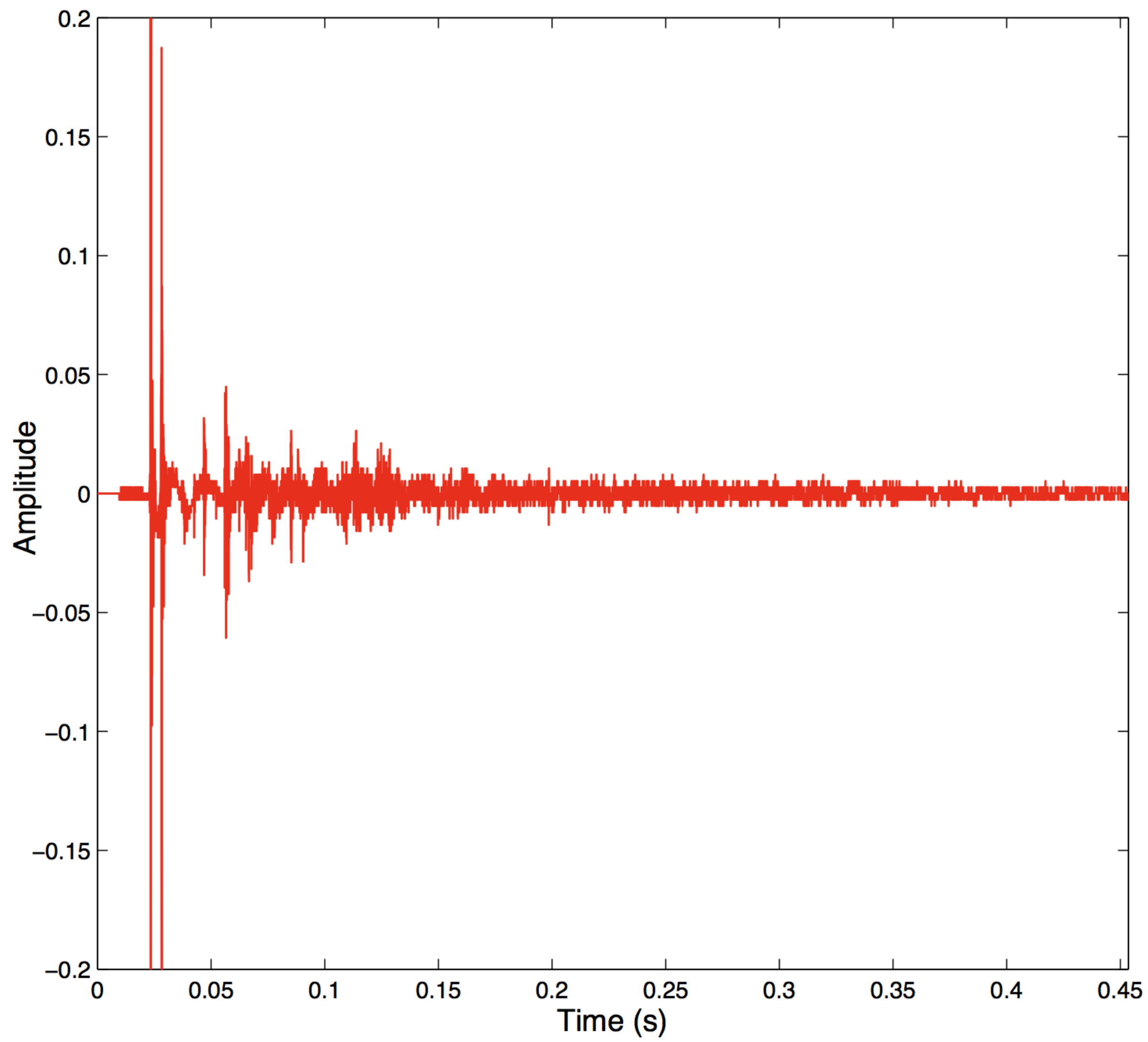
- ✓ Propose a simple framework for the extension of single channel spectral subtraction dereverberation algorithms
- ✓ Implement an efficient way to prevent overestimation errors
- ✓ Evaluate the proposed framework in state-of-the-art spectral subtraction dereverberation techniques

Spectral Subtraction

Spectral Subtraction

✓ originally proposed for denoising applications

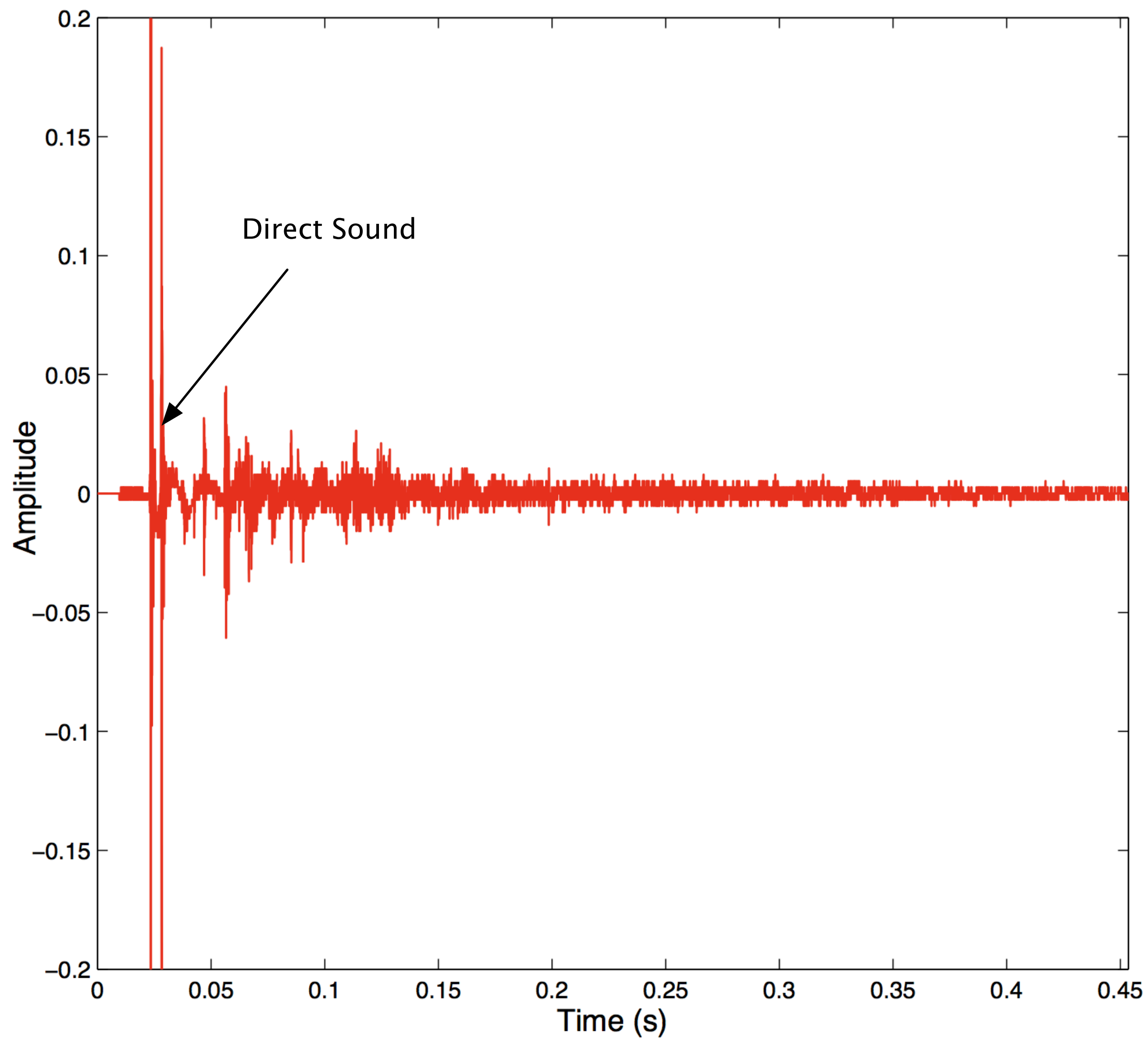
Spectral Subtraction



for denoising applications

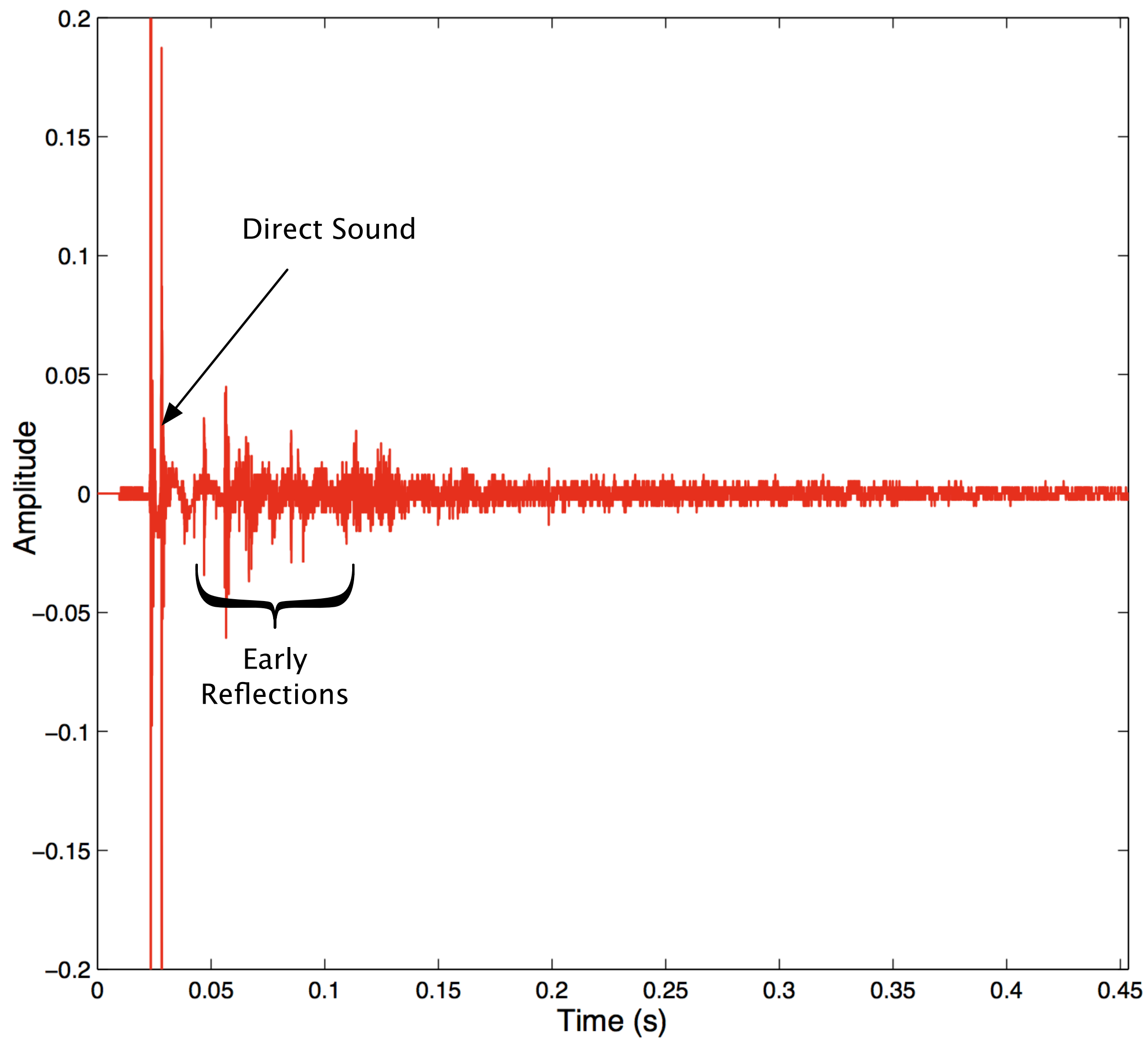
Tsilfidis et al.
University of Patras
Greece

Spectral Subtraction



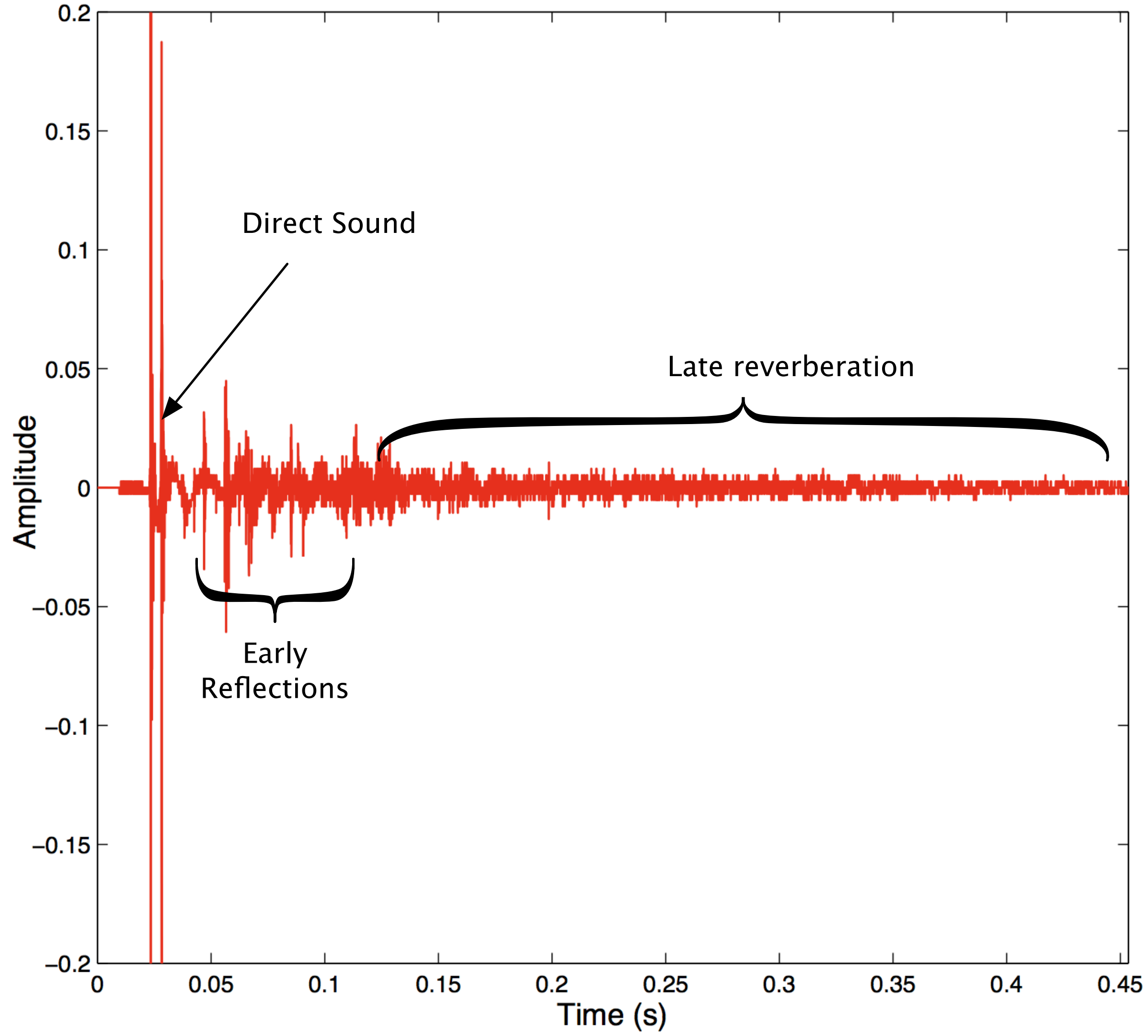
for denoising applications

Spectral Subtraction



for denoising applications

Spectral Subtraction



for denoising applications

Spectral Subtraction

✓ originally proposed for denoising applications

Spectral Subtraction

$$\underbrace{S_e(\omega, j)}_{} = Y(\omega, j) - R(\omega, j)$$

Frequency bin (ω) Time frame (j)

✓ originally proposed for denoising applications

Spectral Subtraction

Anechoic
Estimation

$$\boxed{S_e(\omega, j)} = Y(\omega, j) - R(\omega, j)$$

Frequency bin Time frame

✓ originally proposed for denoising applications

Spectral Subtraction

Anechoic Estimation

Reverberant Signal

$$\underbrace{S_e(\omega, j)}_{} = \boxed{Y(\omega, j)} - R(\omega, j)$$

Frequency bin

Time frame

✓ originally proposed for denoising applications

Tsilfidis et al.
University of Patras
Greece

Spectral Subtraction

Anechoic Estimation — Reverberant Signal — Late Reverberation Estimation

$$\underbrace{S_e(\omega, j)}_{\text{Anechoic Estimation}} = \underbrace{Y(\omega, j)}_{\text{Reverberant Signal}} - \underbrace{\boxed{R(\omega, j)}}_{\text{Late Reverberation Estimation}}$$

ω: Frequency bin

j: Time frame

✓ originally proposed for denoising applications

Spectral Subtraction

Late Reverberation
Estimation

$$R(\omega, j)$$

Spectral Subtraction

Late Reverberation Estimation

$$R(\omega, j) \rightarrow \text{Signal Statistics}$$

- M Wu and D Wang. A two-stage algorithm for one-microphone reverberant speech enhancement. IEEE Transactions on Audio, Speech and Language Processing, 14:774–784, 2006. **(WW)**
- K Furuya and A Kataoka. Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. IEEE Transactions on Audio, Speech and Language Processing, 15:1571–1579, 2007. **(FK)**

Spectral Subtraction

Late Reverberation Estimation

$$R(\omega, j)$$

Signal Statistics

Reverberation Characteristics

- K Lebart, J Boucher, and P. Denbigh. A new method based on spectral subtraction for speech dereverberation. Acta Acustica united with Acustica, 87:359–366, 2001. **(LB)**

- M Wu and D Wang. A two-stage algorithm for one-microphone reverberant speech enhancement. IEEE Transactions on Audio, Speech and Language Processing, 14:774–784, 2006. **(WW)**
- K Furuya and A Kataoka. Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. IEEE Transactions on Audio, Speech and Language Processing, 15:1571–1579, 2007. **(FK)**

Spectral Subtraction

Late Reverberation Estimation

$R(\omega, j)$

Signal Statistics

Reverberation Characteristics

- M Wu and D Wang. A two-stage algorithm for one-microphone reverberant speech enhancement. IEEE Transactions on Audio, Speech and Language Processing, 14:774–784, 2006. **(WW)**
- K Furuya and A Kataoka. Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. IEEE Transactions on Audio, Speech and Language Processing, 15:1571–1579, 2007. **(FK)**

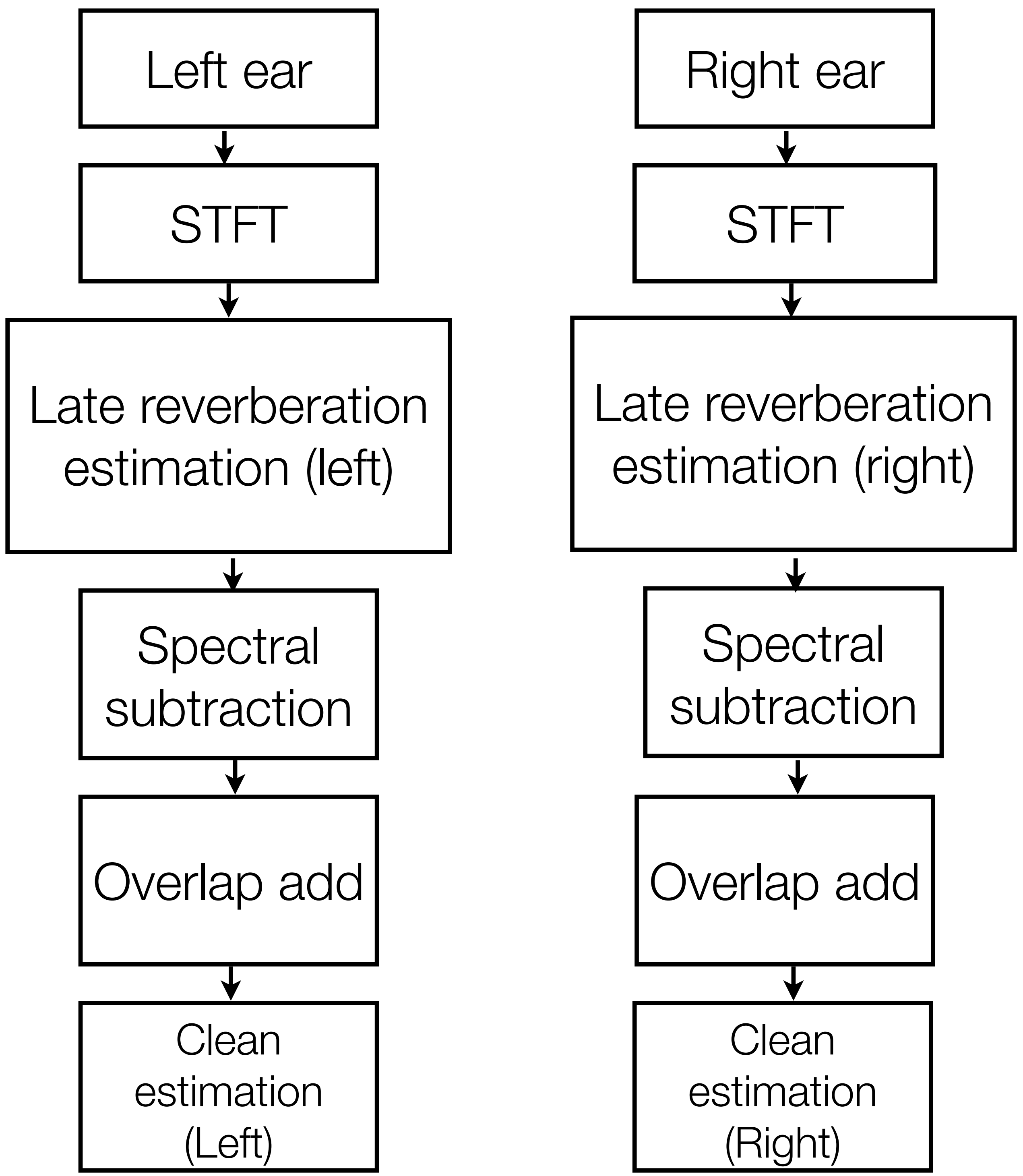
- K Lebart, J Boucher, and P. Denbigh. A new method based on spectral subtraction for speech dereverberation. Acta Acustica united with Acustica, 87:359–366, 2001. **(LB)**

$$S_e(\omega, j) = Y(\omega, j) - R_{late}(\omega, j)$$

$$S_e(\omega, j) = G(\omega, j)Y(\omega, j)$$

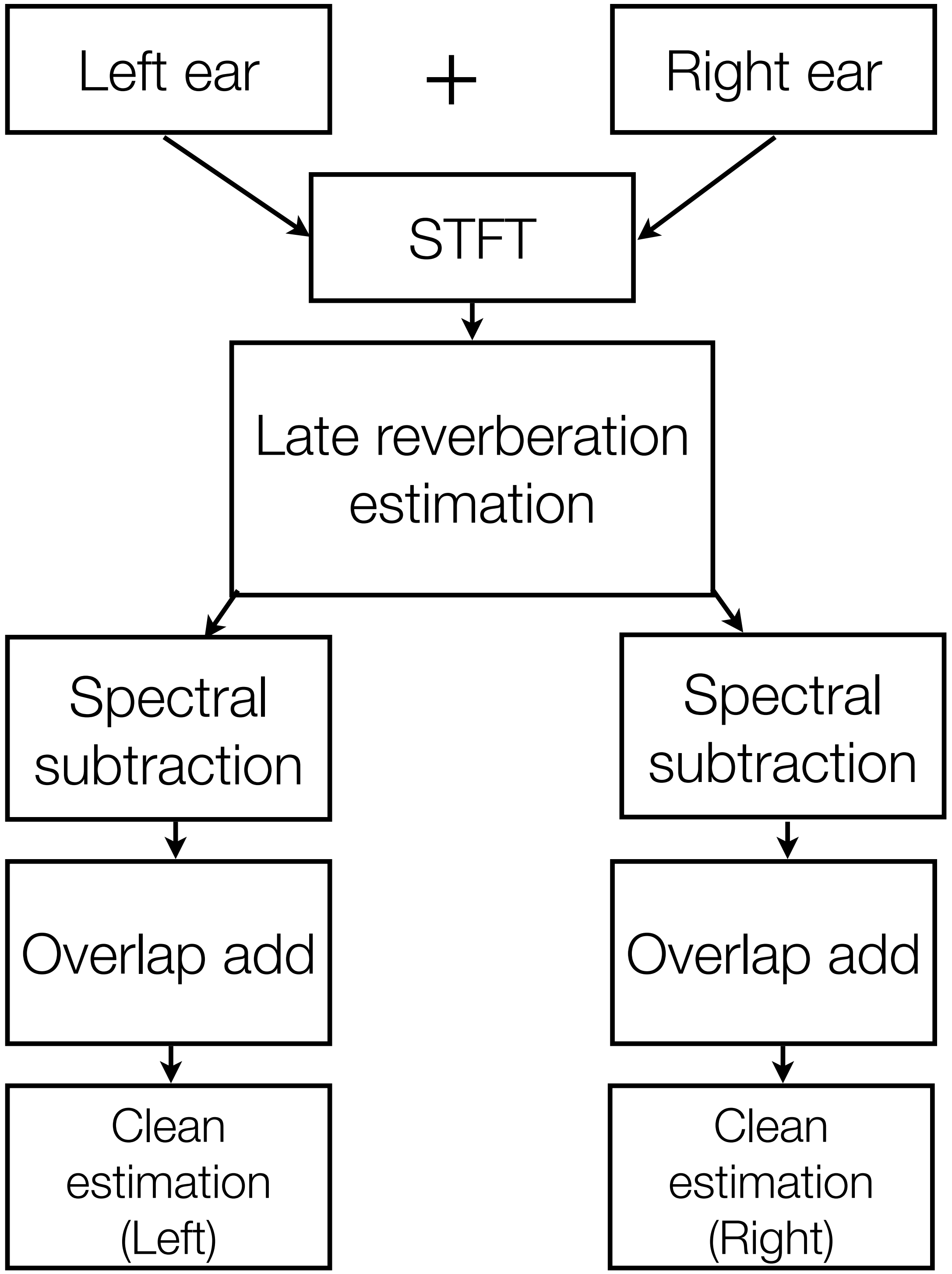
Gain Function: $$G(\omega, j) = \frac{Y(\omega, j) - R(\omega, j)}{Y(\omega, j)}$$

Tsilfidis et al.
University of Patras
Greece



Spectral Subtraction

Bilateral signal processing <u>destroys</u> binaural cues !



Spectral Subtraction

Bilateral signal processing destroys binaural cues !

↓

Spectral Subtraction extension based on a Delay and Sum Beamformer (DSB)

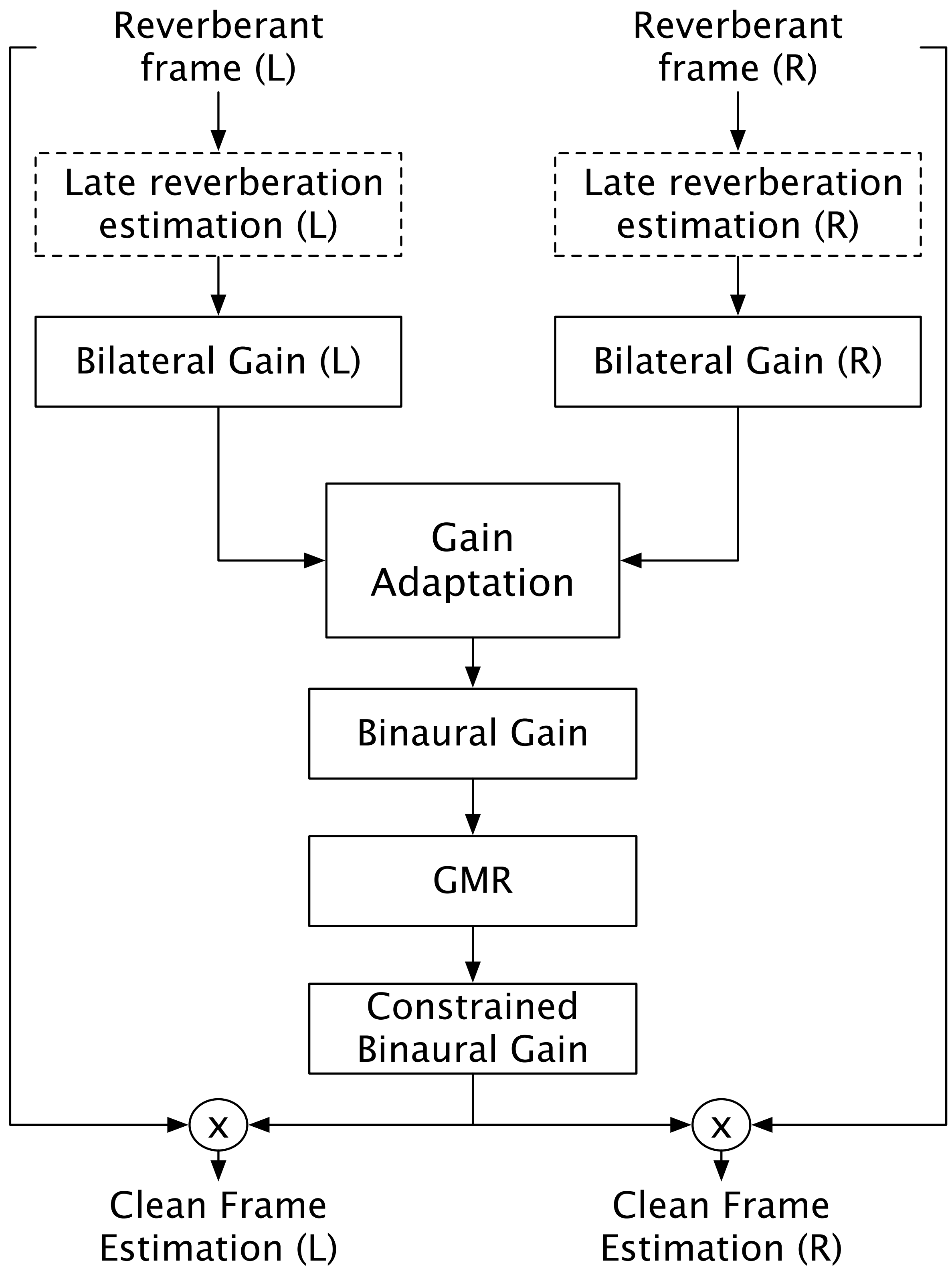
M Jeub, M Schafer, T Esch, and P Vary. *Model-Based Dereverberation Preserving Binaural Cues.* IEEE Transactions on Audio, Speech, and Language Processing, 18:1732–1745, 2010

Here:

- ✓ Relative delay **<** typical analysis windows (depends on the width of the human head)
- ✓ Binaural extension of single-channel spectral subtraction dereverberation based on Bilateral Gain Adaptation

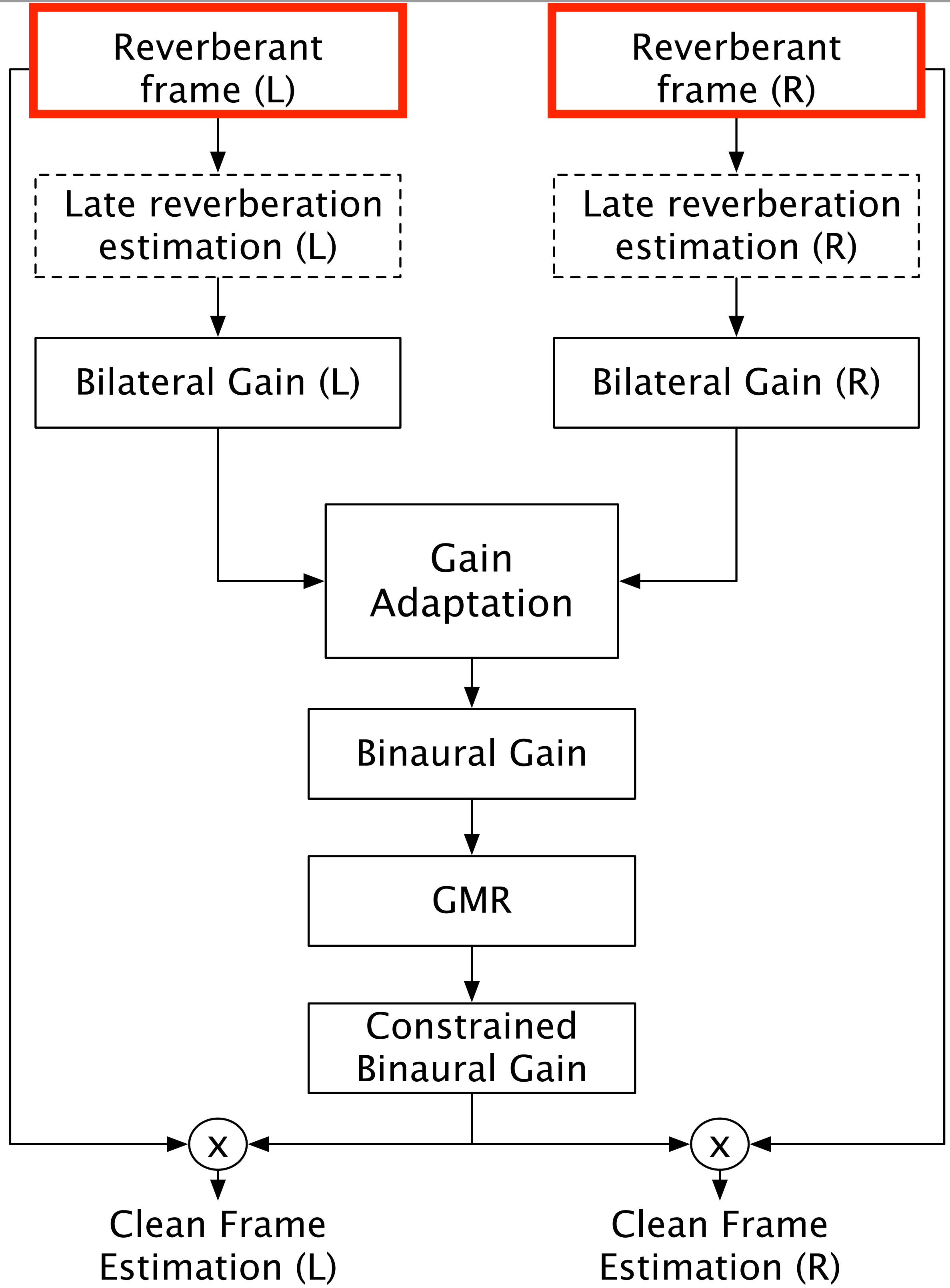
Tsilfidis et al.
University of Patras
Greece

Signal Flow



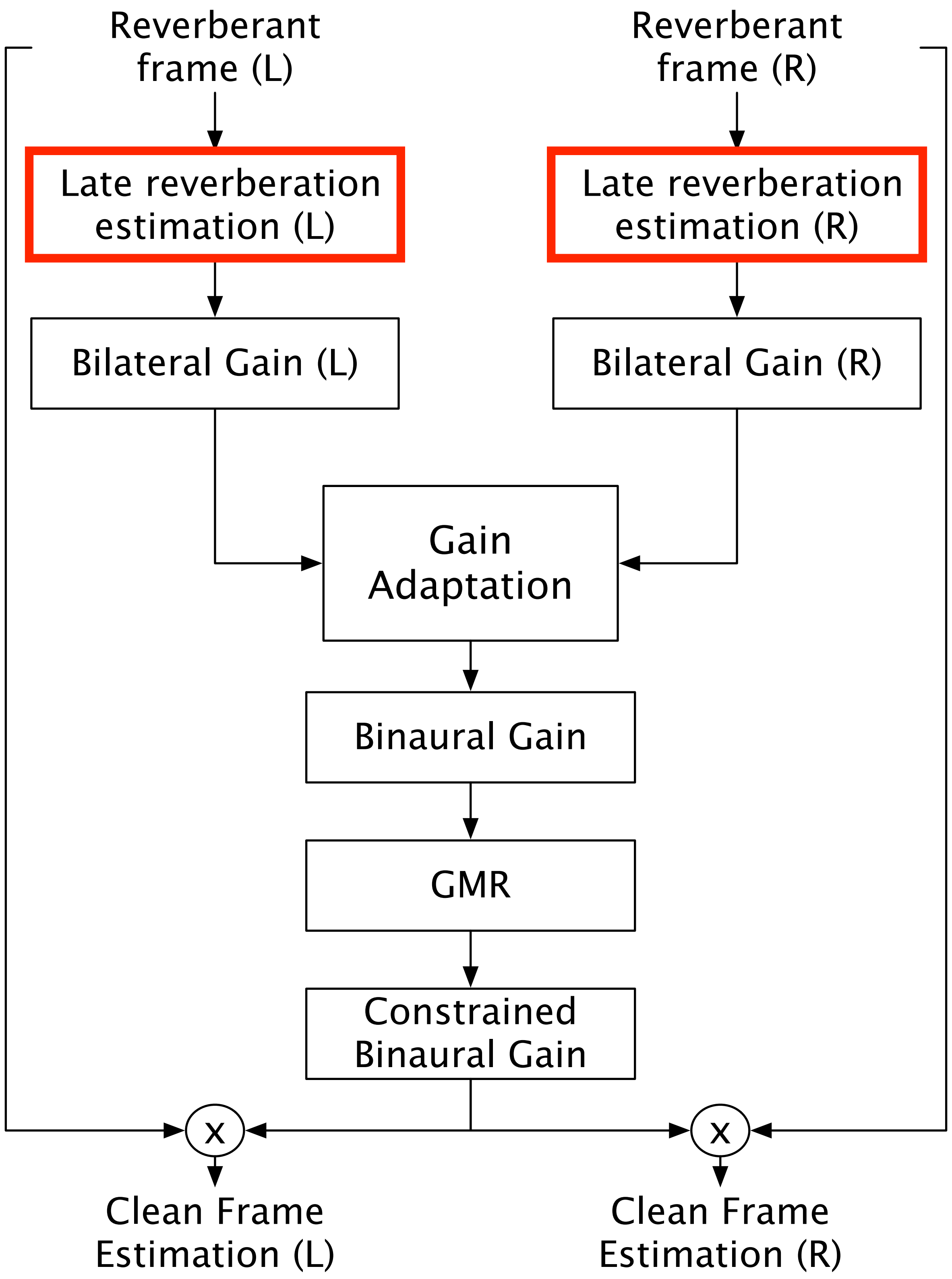
Tsilfidis et al.
University of Patras
Greece

Signal Flow



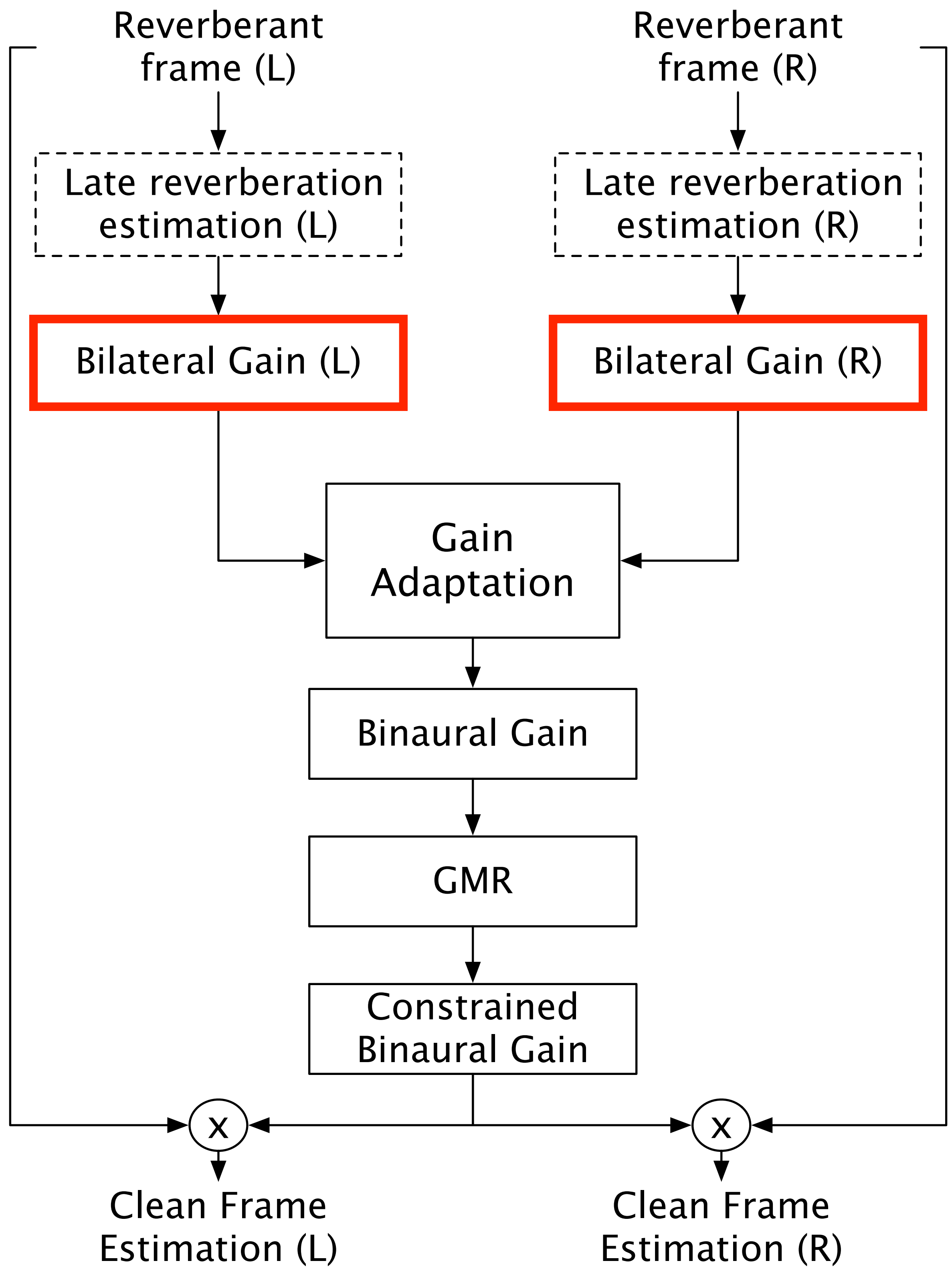
Tsilfidis et al.
University of Patras
Greece

Signal Flow



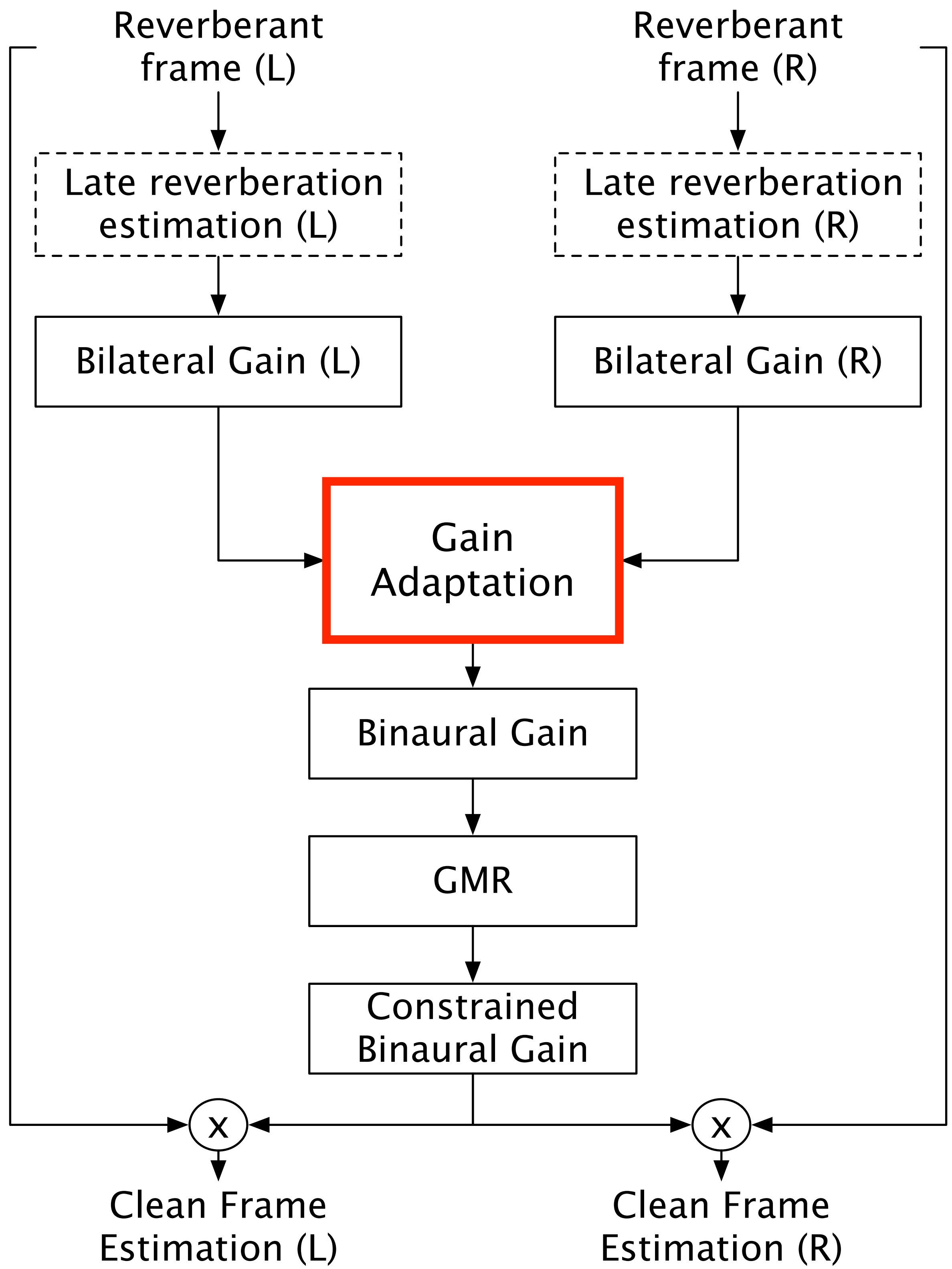
Tsilfidis et al.
University of Patras
Greece

Signal Flow



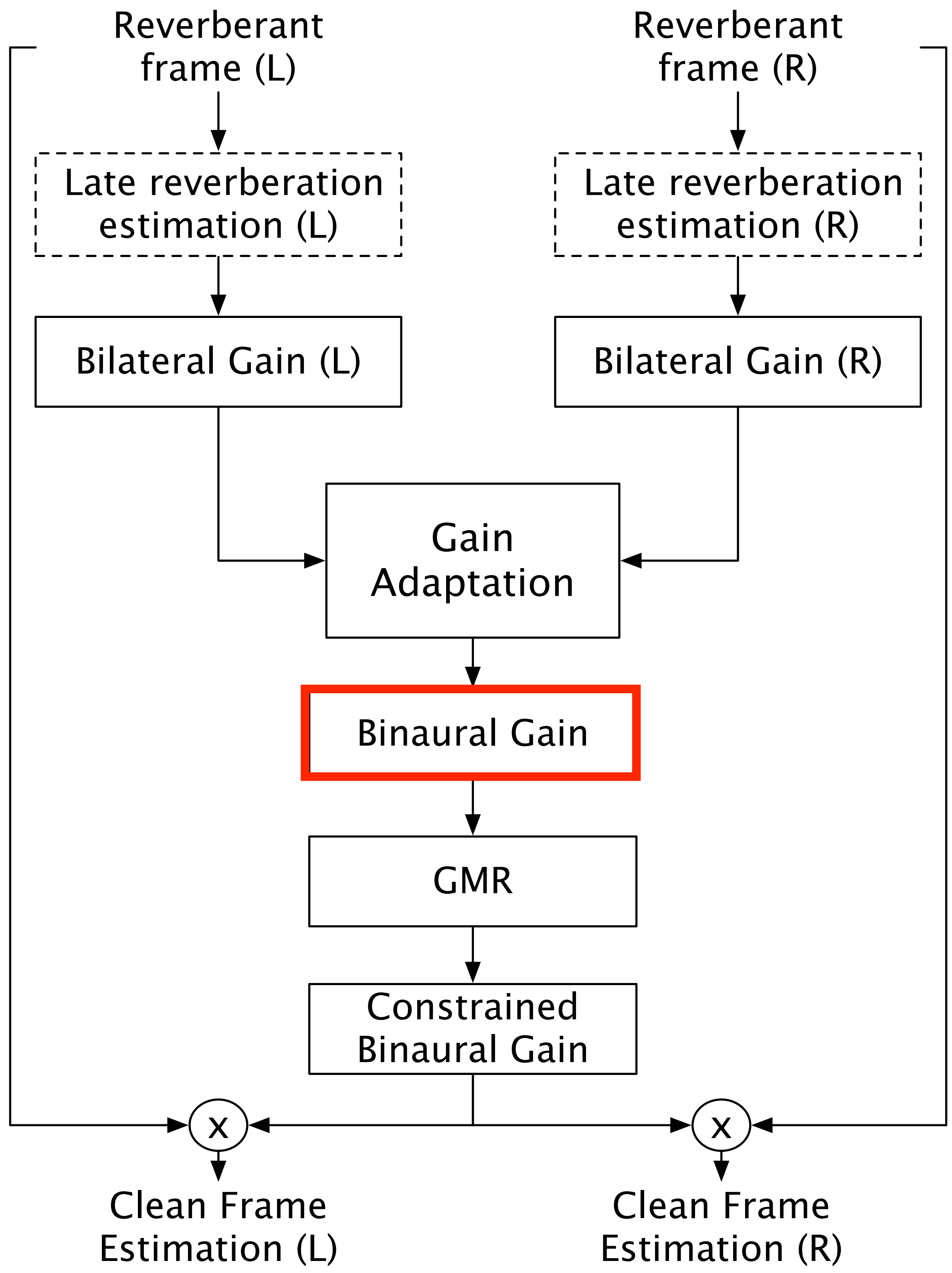
Tsilfidis et al.
University of Patras
Greece

Signal Flow



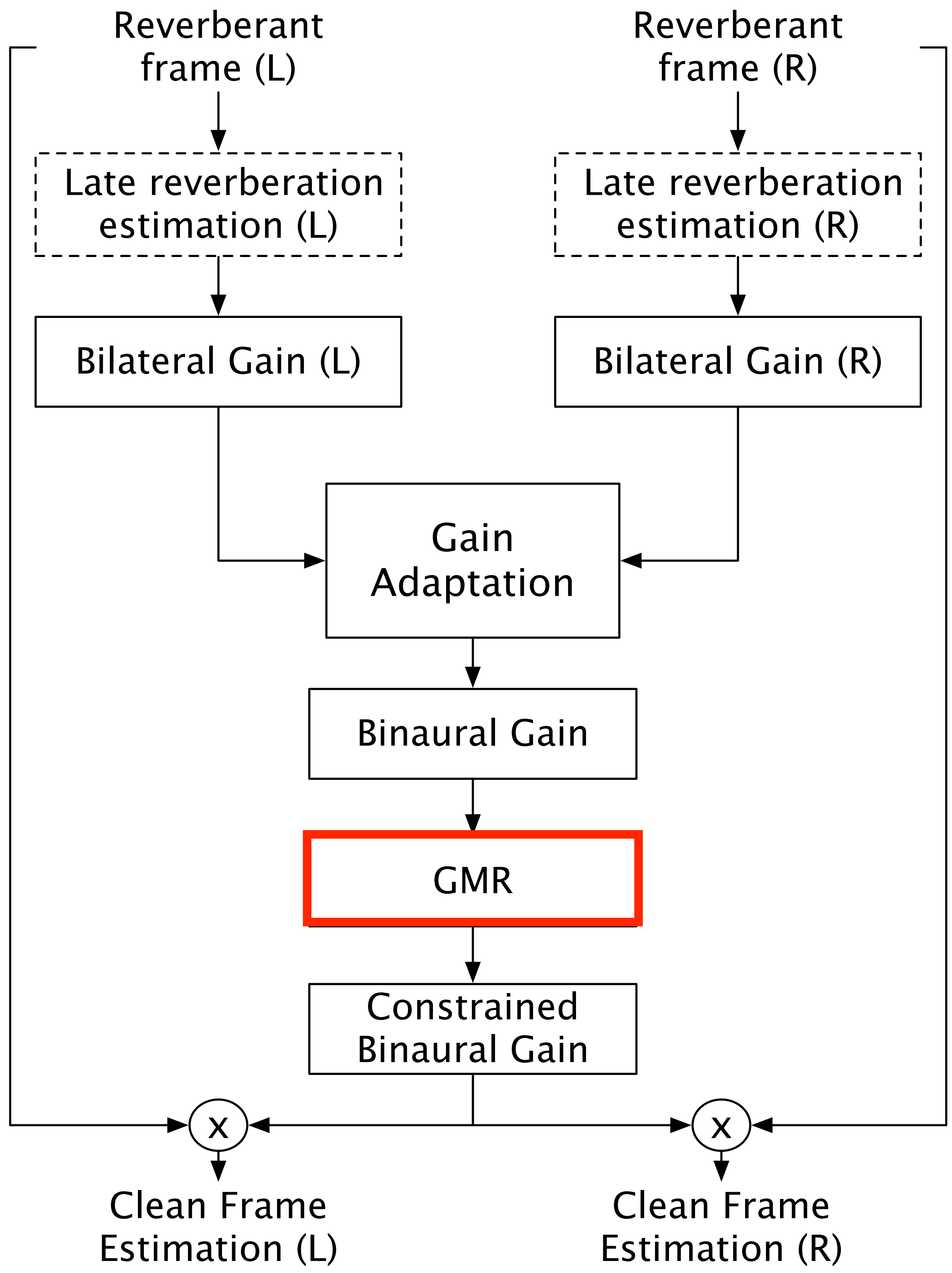
Tsilfidis et al.
University of Patras
Greece

Signal Flow



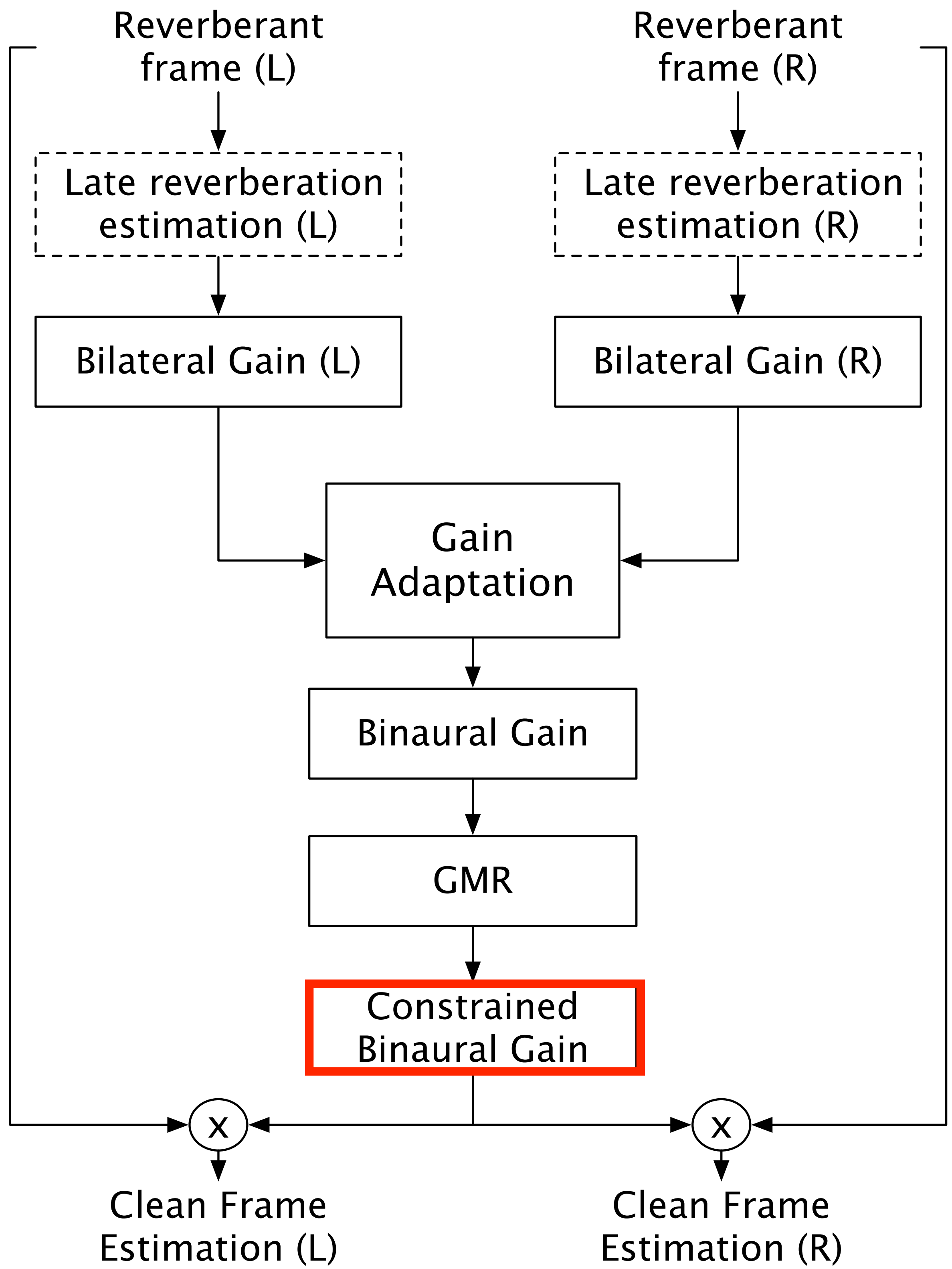
Tsilfidis et al.
University of Patras
Greece

Signal Flow



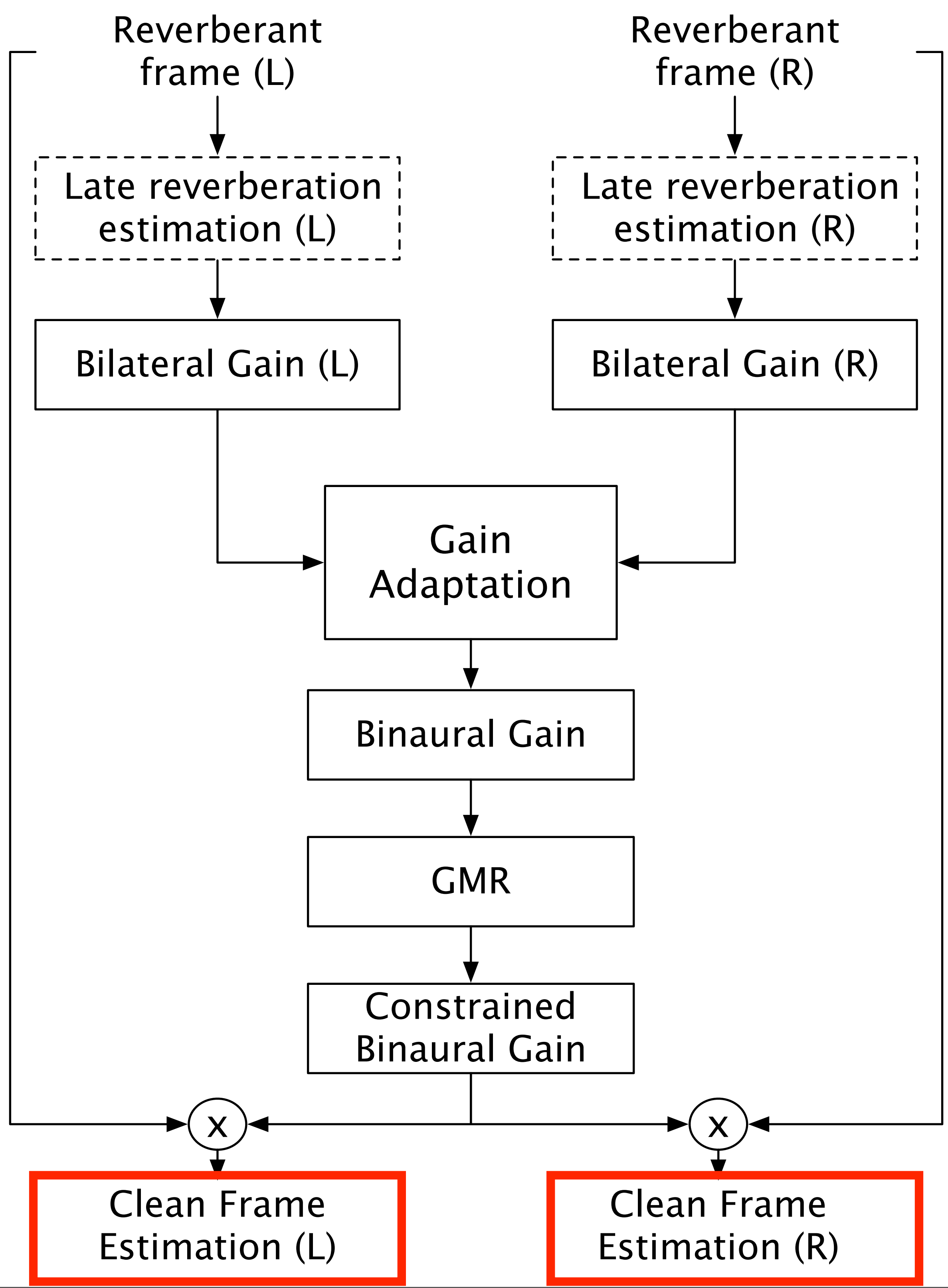
Tsilfidis et al.
University of Patras
Greece

Signal Flow



Tsilfidis et al.
University of Patras
Greece

Signal Flow



Tsilfidis et al.
University of Patras
Greece

Gain Adaptation

maxGain:		
avgGain:		
minGain:		

Tsilfidis et al.
University of Patras
Greece

Gain Adaptation

maxGain:	$G(\omega, j) = max(G_l(\omega, j), G_r(\omega, j))$	moderate suppression, fewer processing artifacts
avgGain:		
minGain:		

Tsilfidis et al.
University of Patras
Greece

Gain Adaptation

maxGain:	$G(\omega, j) = max(G_l(\omega, j), G_r(\omega, j))$	moderate suppression, fewer processing artifacts
avgGain:	$G(\omega, j) = avg(G_l(\omega, j), G_r(\omega, j))$	compromise between reverberation reduction and artifacts
minGain:		

Tsilfidis et al.
University of Patras
Greece

Gain Adaptation

maxGain:	$G(\omega, j) = max(G_l(\omega, j), G_r(\omega, j))$	moderate suppression, fewer processing artifacts
avgGain:	$G(\omega, j) = avg(G_l(\omega, j), G_r(\omega, j))$	compromise between reverberation reduction and artifacts
minGain:	$G(\omega, j) = min(G_l(\omega, j), G_r(\omega, j))$	significant suppression, possible artifacts

Tsilfidis et al.
University of Patras
Greece

Gain Adaptation

maxGain:	$G(\omega, j) = max(G_l(\omega, j), G_r(\omega, j))$	moderate suppression, fewer processing artifacts
avgGain:	$G(\omega, j) = avg(G_l(\omega, j), G_r(\omega, j))$	compromise between reverberation reduction and artifacts
minGain:	$G(\omega, j) = min(G_l(\omega, j), G_r(\omega, j))$	significant suppression, possible artifacts

selection of the gain adaptation scheme according to the application scenario

Tsilfidis et al.
University of Patras
Greece

Gain Magnitude Regularization
(GMR)

Gain Magnitude Regularization (GMR)

$$G'(\omega, j) = \begin{cases} \dfrac{G(\omega, j) - \theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \\ & \text{and } G(\omega, j) < \theta \\ G(\omega, j) & \text{otherwise} \end{cases}$$

Gain Magnitude Regularization (GMR)

$$G'(\omega, j) = \begin{cases} \dfrac{G(\omega, j) - \theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \text{ and } G(\omega, j) < \theta \\ G(\omega, j) & \text{otherwise} \end{cases}$$

Low SRR detector

$$\boxed{\zeta = \frac{\sum_{\omega=1}^{\Omega} G(\omega, j)|Y(\omega, j)|^2}{\sum_{\omega=1}^{\Omega} |Y(\omega, j)|^2}}$$

M Jeub, M Schafer, T Esch, and P Vary. *Model-Based Dereverberation Preserving Binaural Cues.* IEEE Transactions on Audio, Speech, and Language Processing, 18:1732–1745, 2010

Tsilfidis et al.
University of Patras
Greece

Gain Magnitude Regularization (GMR)

$$G'(\omega, j) = \begin{cases} \dfrac{G(\omega, j) - \theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \\ & \text{and } G(\omega, j) < \theta \\ G(\omega, j) & \text{otherwise} \end{cases}$$

Regularization application threshold

Low SRR detector

$$\boxed{\zeta = \frac{\sum_{\omega=1}^{\Omega} G(\omega, j) |Y(\omega, j)|^2}{\sum_{\omega=1}^{\Omega} |Y(\omega, j)|^2}}$$

M Jeub, M Schafer, T Esch, and P Vary. *Model-Based Dereverberation Preserving Binaural Cues.* IEEE Transactions on Audio, Speech, and Language Processing, 18:1732–1745, 2010

Gain Magnitude Regularization (GMR)

$$G'(\omega, j) = \begin{cases} \dfrac{G(\omega, j) - \theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \\ & \text{and } G(\omega, j) < \theta \\ G(\omega, j) & \text{otherwise} \end{cases}$$

Regularization ratio

Regularization application threshold

Low SRR detector

$$\boxed{\zeta = \frac{\sum_{\omega=1}^{\Omega} G(\omega, j)|Y(\omega, j)|^2}{\sum_{\omega=1}^{\Omega} |Y(\omega, j)|^2}}$$

M Jeub, M Schafer, T Esch, and P Vary. *Model-Based Dereverberation Preserving Binaural Cues.* IEEE Transactions on Audio, Speech, and Language Processing, 18:1732–1745, 2010

Tsilfidis et al.
University of Patras
Greece

Gain Magnitude Regularization (GMR)

$$G'(\omega, j) = \begin{cases} \frac{G(\omega, j) - \theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \\ & \text{and } G(\omega, j) < \theta \\ G(\omega, j) & \text{otherwise} \end{cases}$$

Regularization ratio

Regularization application threshold

Low SRR detector

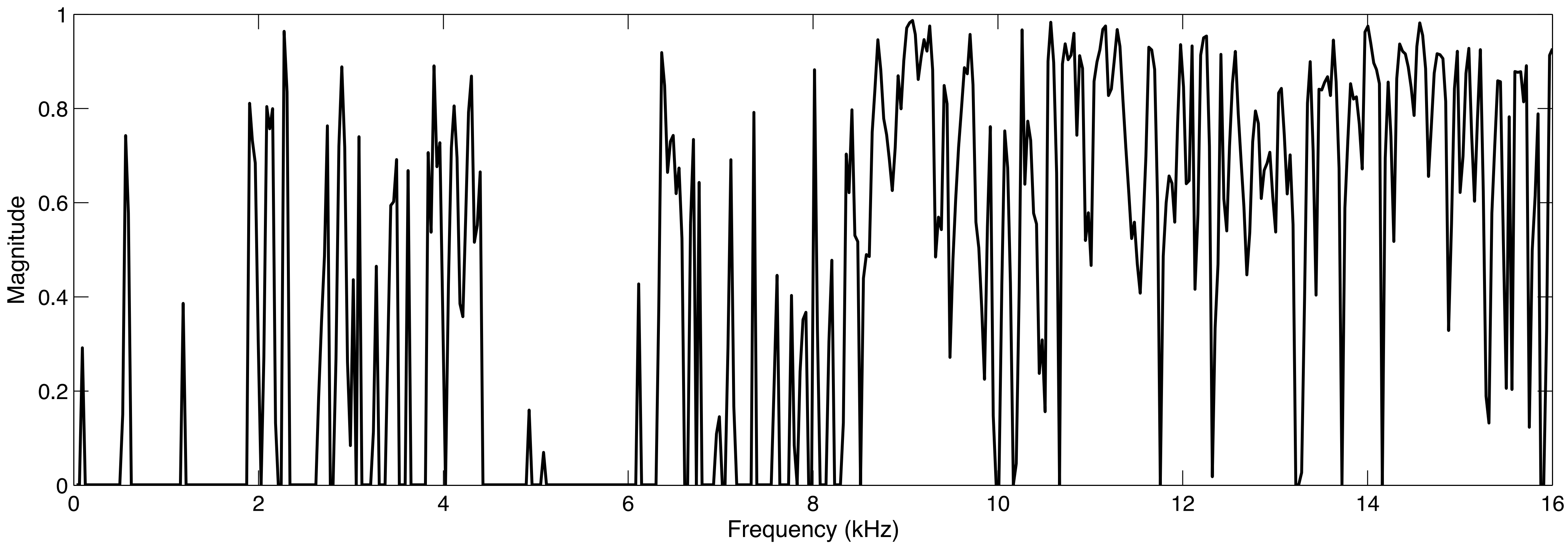
$$\zeta = \frac{\sum_{\omega=1}^{\Omega} G(\omega, j) |Y(\omega, j)|^2}{\sum_{\omega=1}^{\Omega} |Y(\omega, j)|^2}$$

- Compensate for overestimation.
- Prevents musical noise.

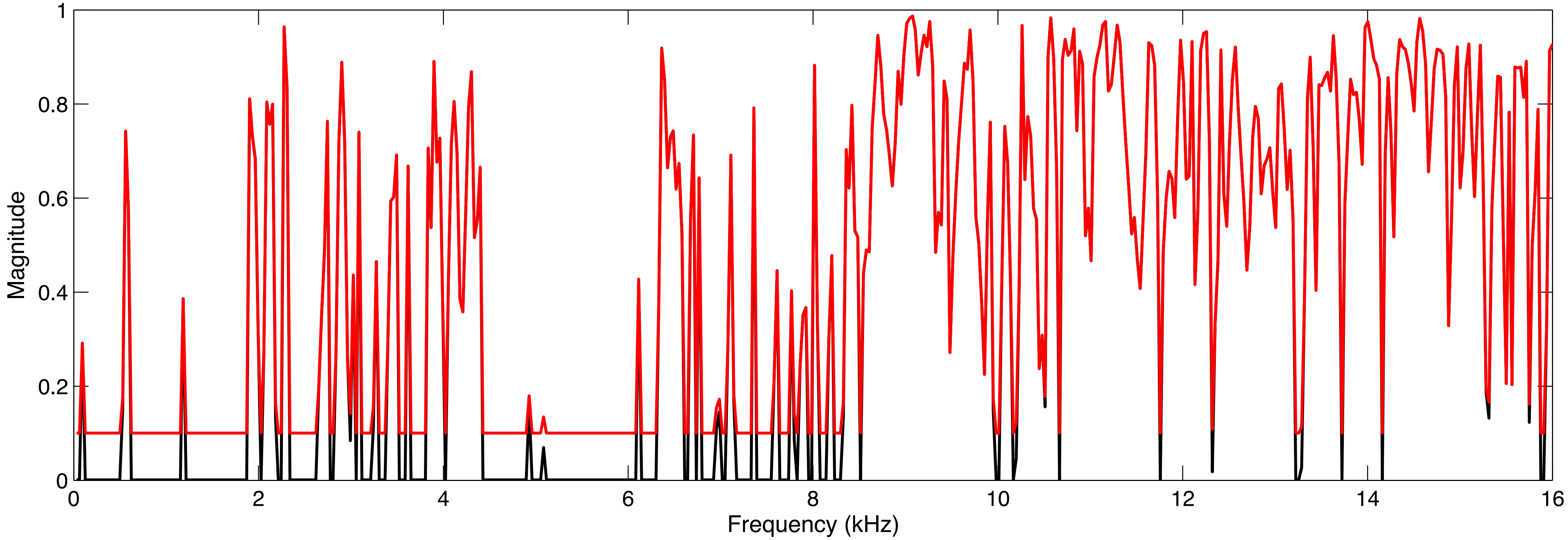
M Jeub, M Schafer, T Esch, and P Vary. *Model-Based Dereverberation Preserving Binaural Cues.* IEEE Transactions on Audio, Speech, and Language Processing, 18:1732–1745, 2010

Tsilfidis et al.
University of Patras
Greece

Binaural Dereverberation



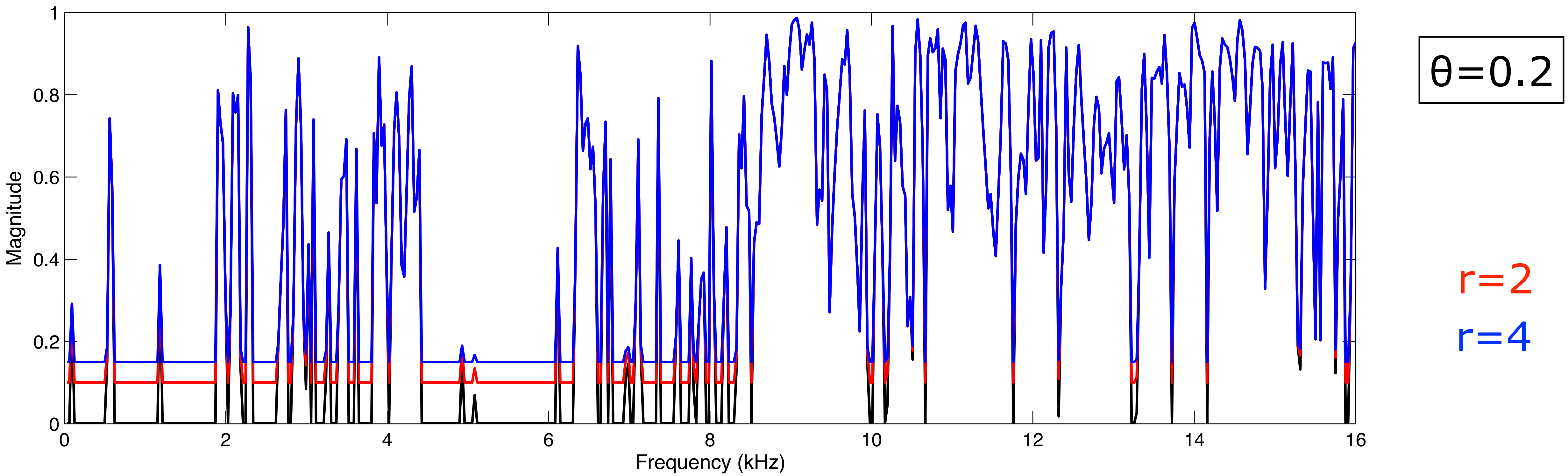
Binaural Dereverberation



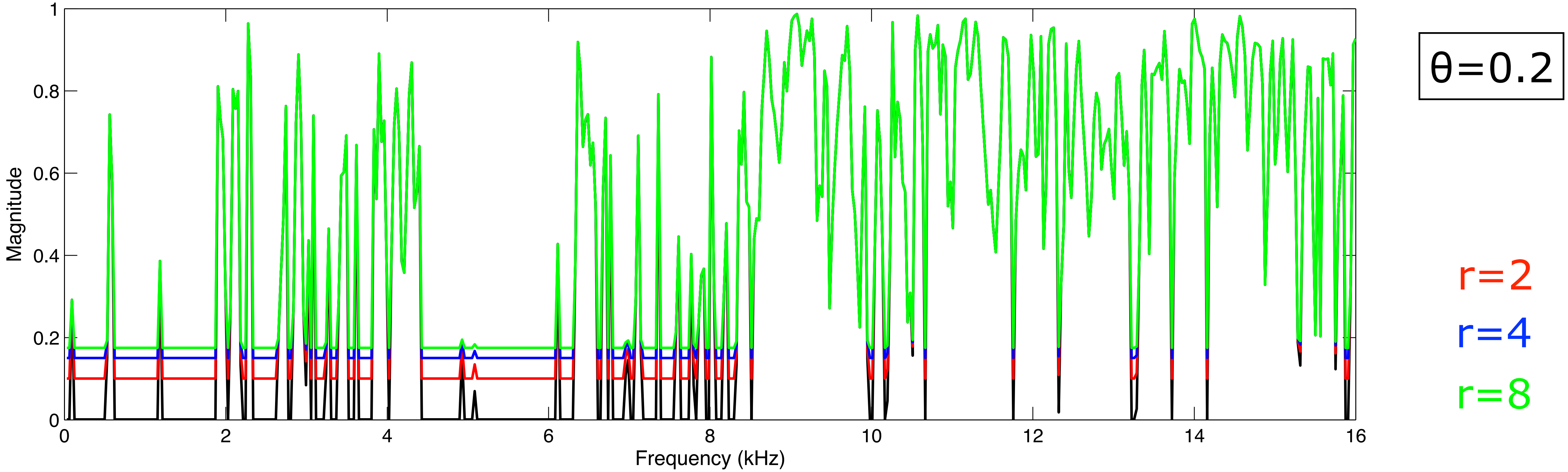
θ=0.2

r=2

Binaural Dereverberation

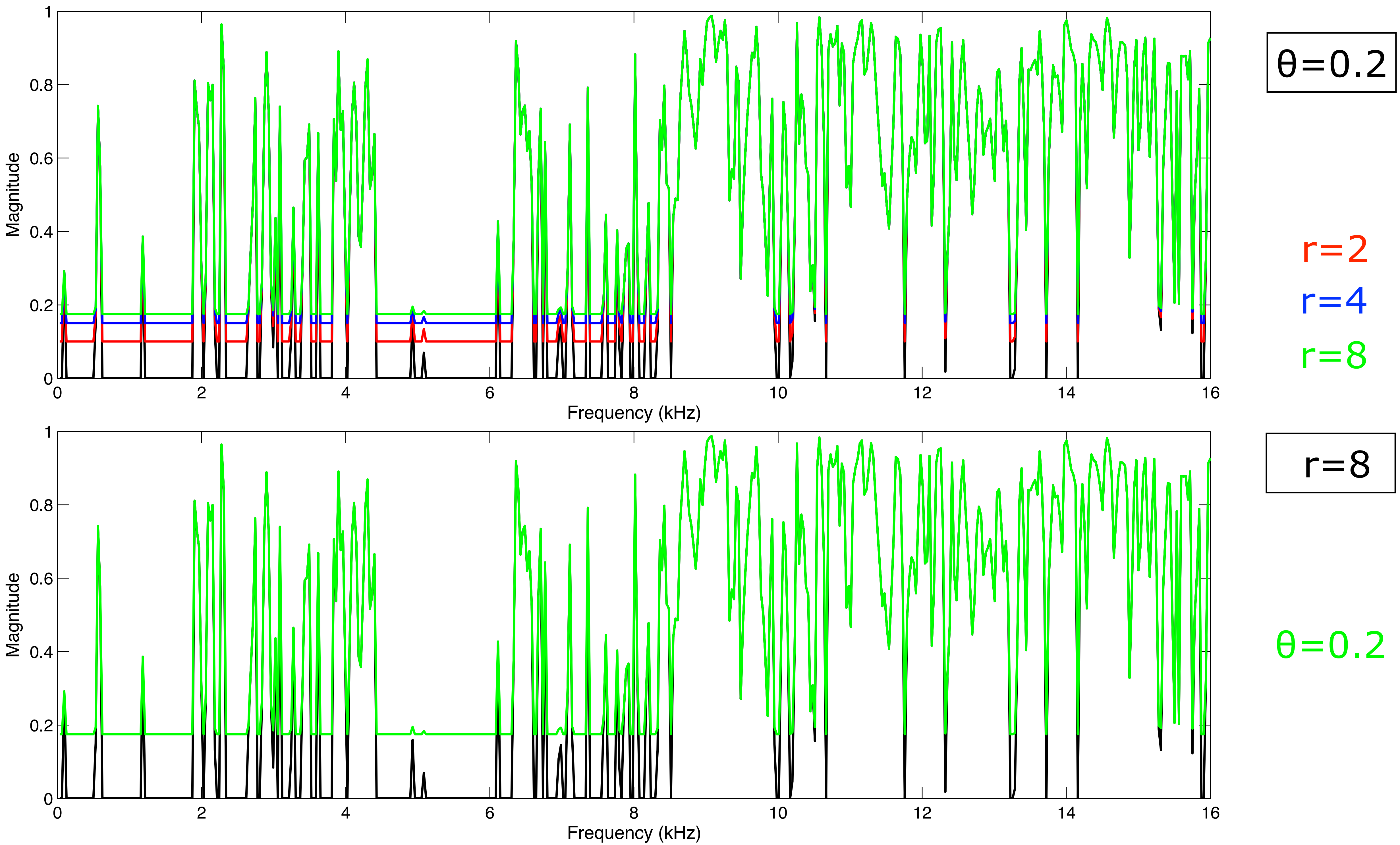


Binaural Dereverberation

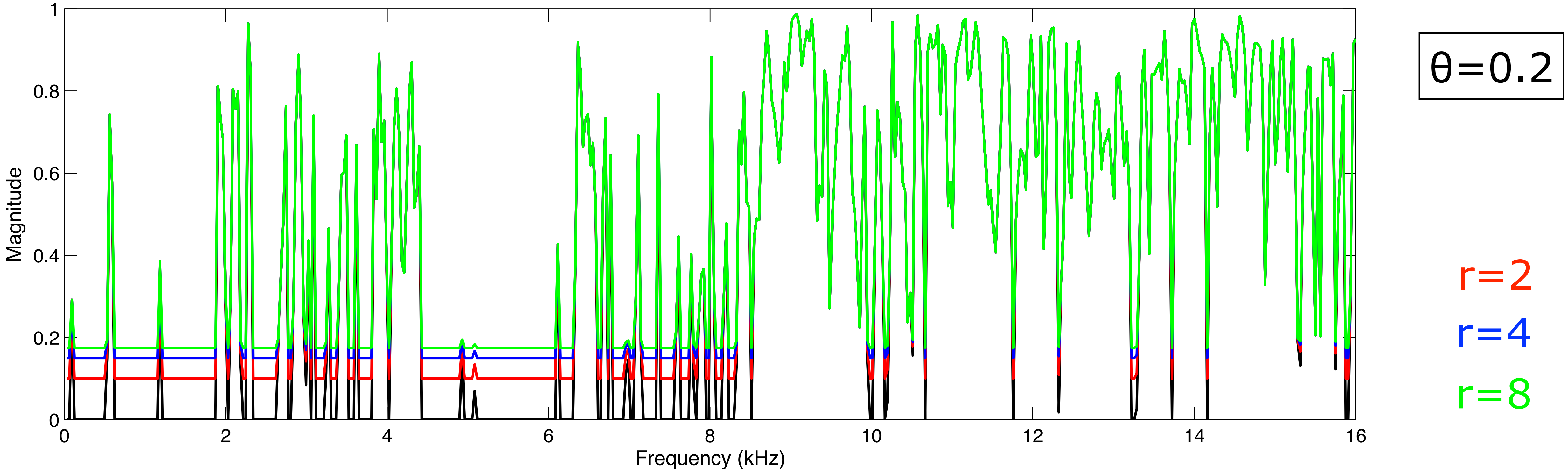


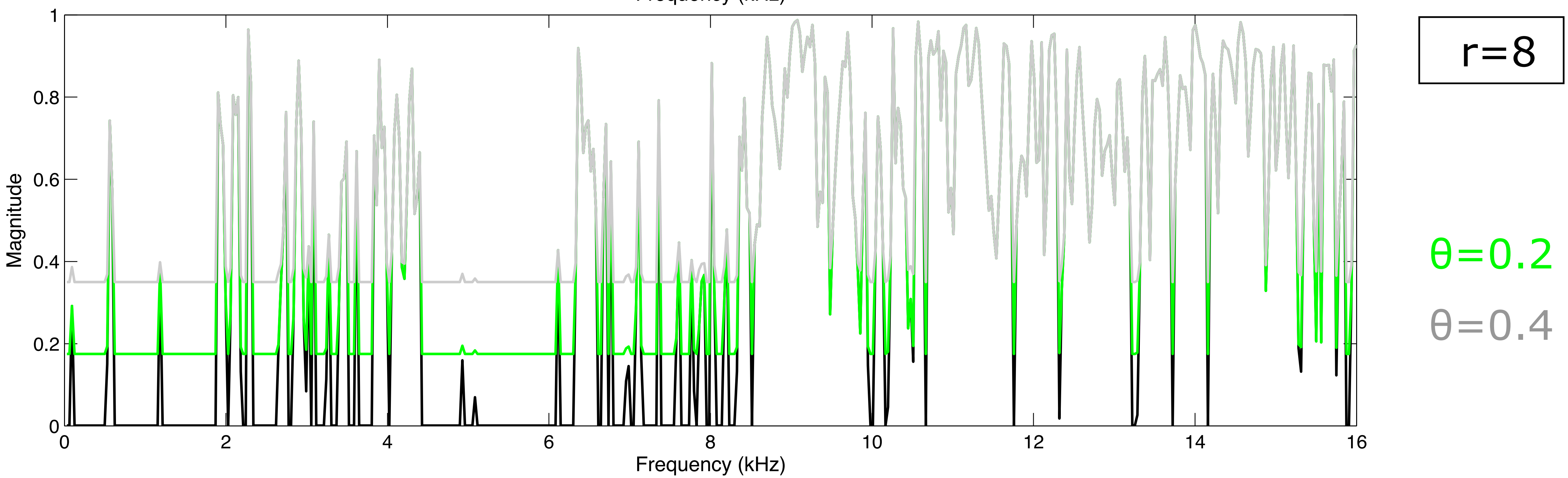
Binaural Dereverberation

Tsilfidis et al.
University of Patras
Greece

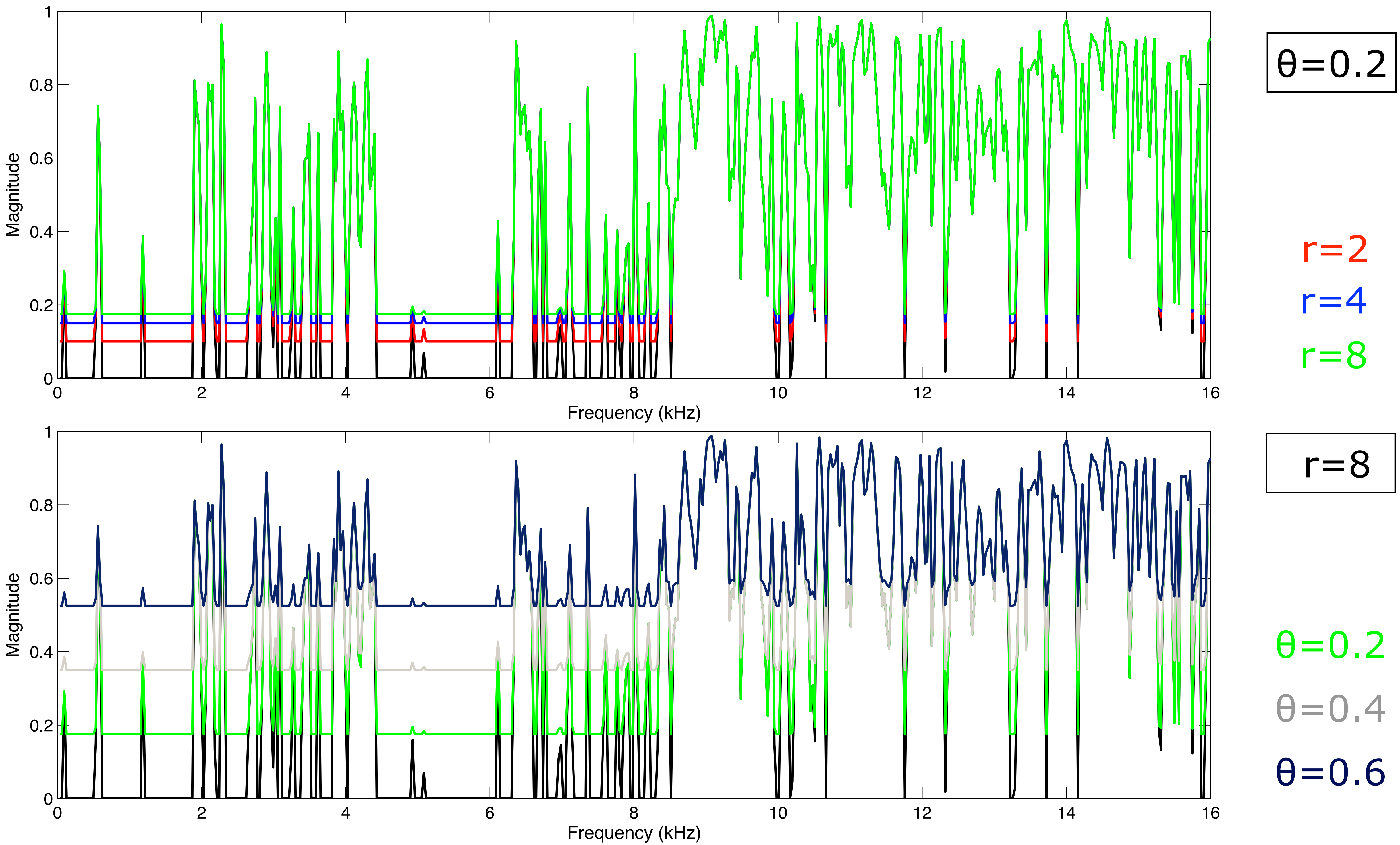


Binaural Dereverberation





Binaural Dereverberation



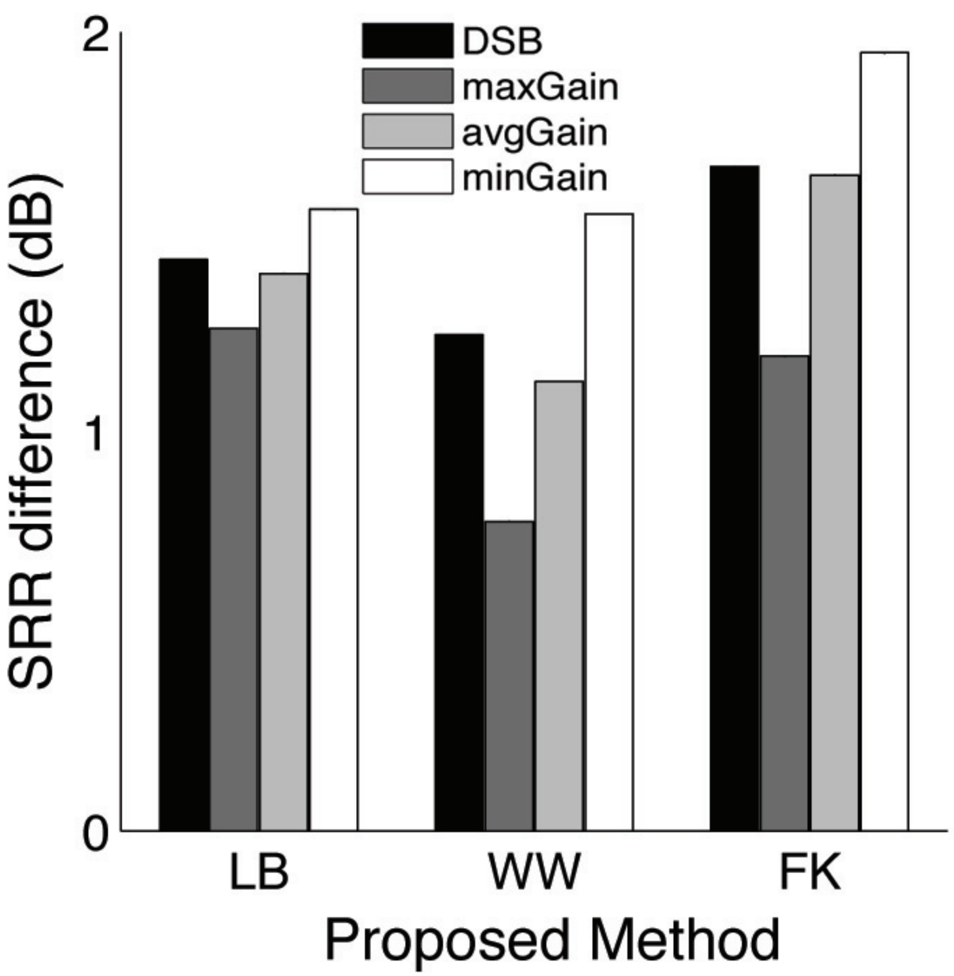
Tsilfidis et al.
University of Patras
Greece

Results

- Stairway Hall

RT = 0.69 s, Distance= 3m, Angles = 0, 30, 60, 90° (Aachen Database)

Sampling Frequency: 16 kHz



	Stairway Hall			
Method	**DSB**	**maxGain**	**avgGain**	**minGain**
LB	0.153	0.142	0.147	**0.158**
WW	0.206	0.160	0.208	**0.258**
FK	0.160	0.180	**0.186**	-0.029

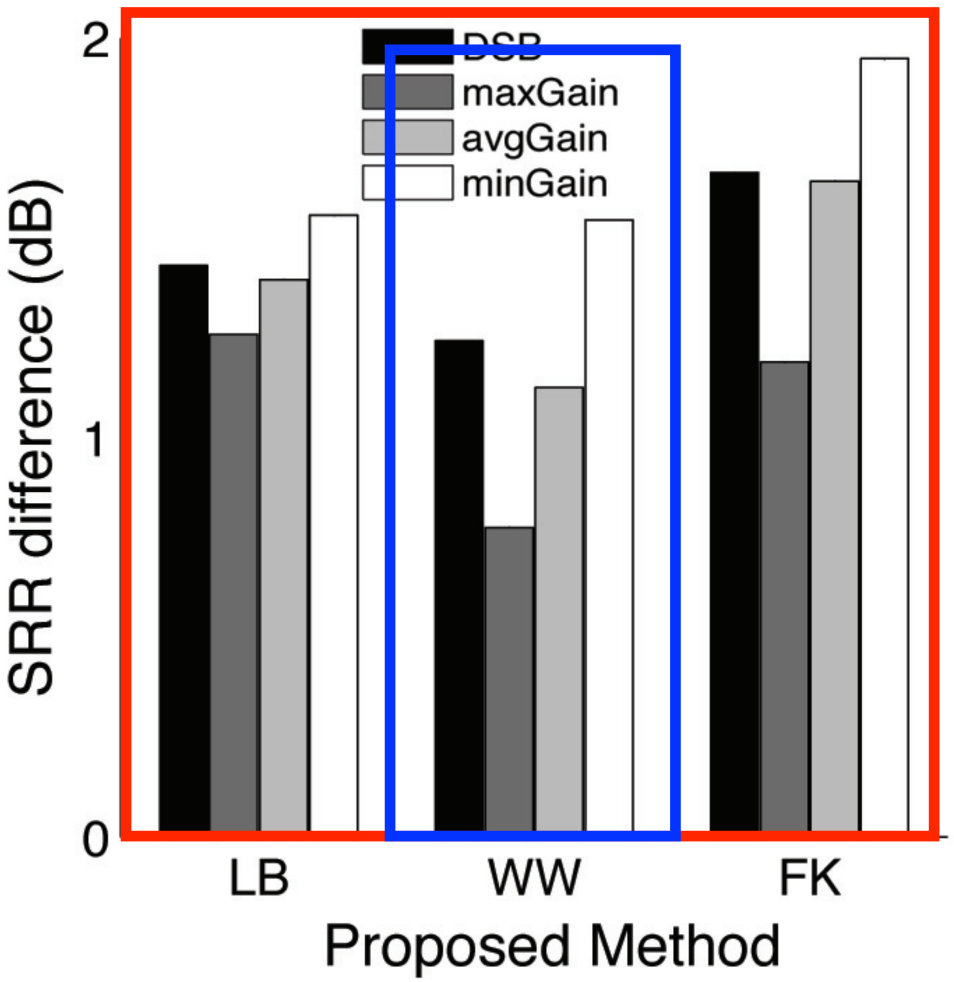
PESQ difference

Tsilfidis et al.
University of Patras
Greece

Results

- Stairway Hall
  RT = 0.69 s, Distance= 3m, Angles = 0, 30, 60, 90° (Aachen Database)

Sampling Frequency: 16 kHz



	Stairway Hall			
Method	**DSB**	**maxGain**	**avgGain**	**minGain**
LB	0.153	0.142	0.147	**0.158**
WW	0.206	0.160	0.208	**0.258**
FK	0.160	0.180	**0.186**	-0.029

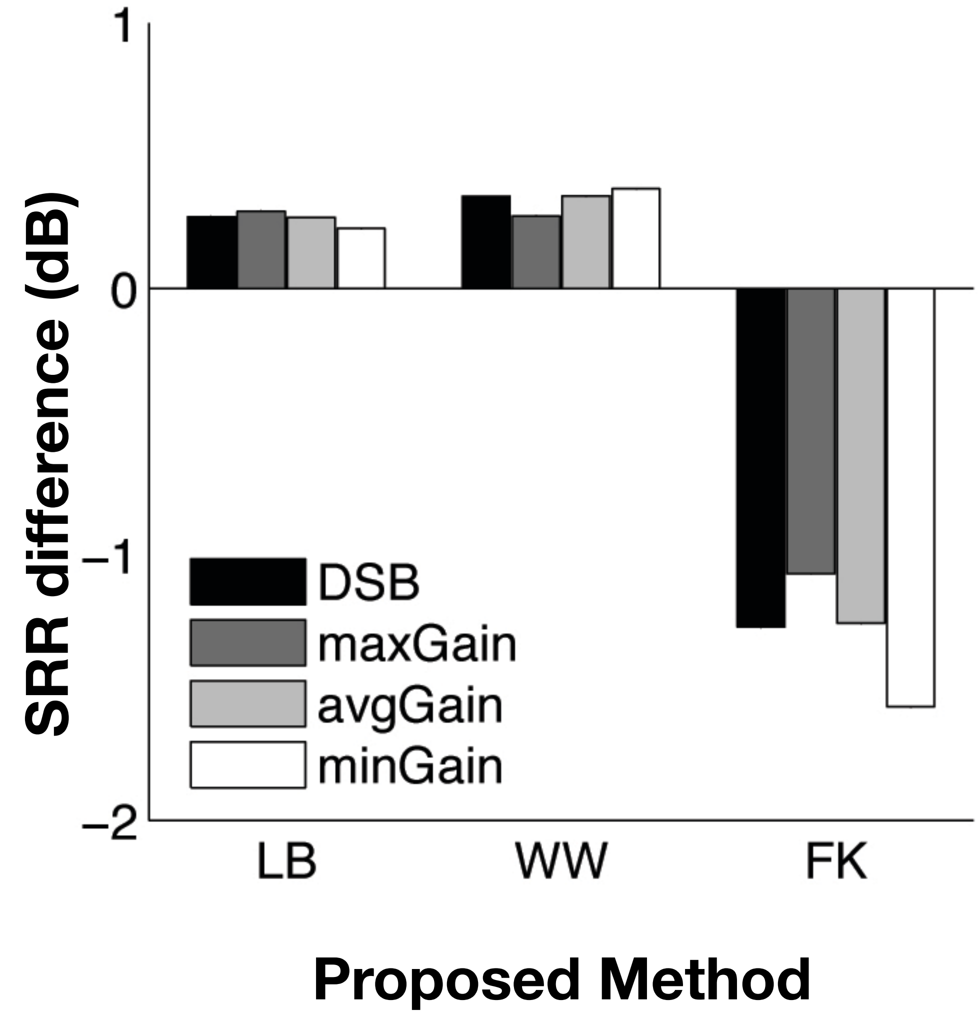
PESQ difference

Tsilfidis et al.
University of Patras
Greece

Results

- Cafeteria
  RT = 1.29 s, Distance= 1.18-1.62m, Angles = -30, 0, 90° (Oldenburg Database)

Sampling Frequency: 16 kHz



	Cafeteria			
Method	**DSB**	**maxGain**	**avgGain**	**minGain**
LB	0.133	**0.136**	0.135	0.133
WW	0.205	0.208	**0.216**	0.198
FK	-0.235	**-0.141**	-0.228	-0.428

PESQ difference

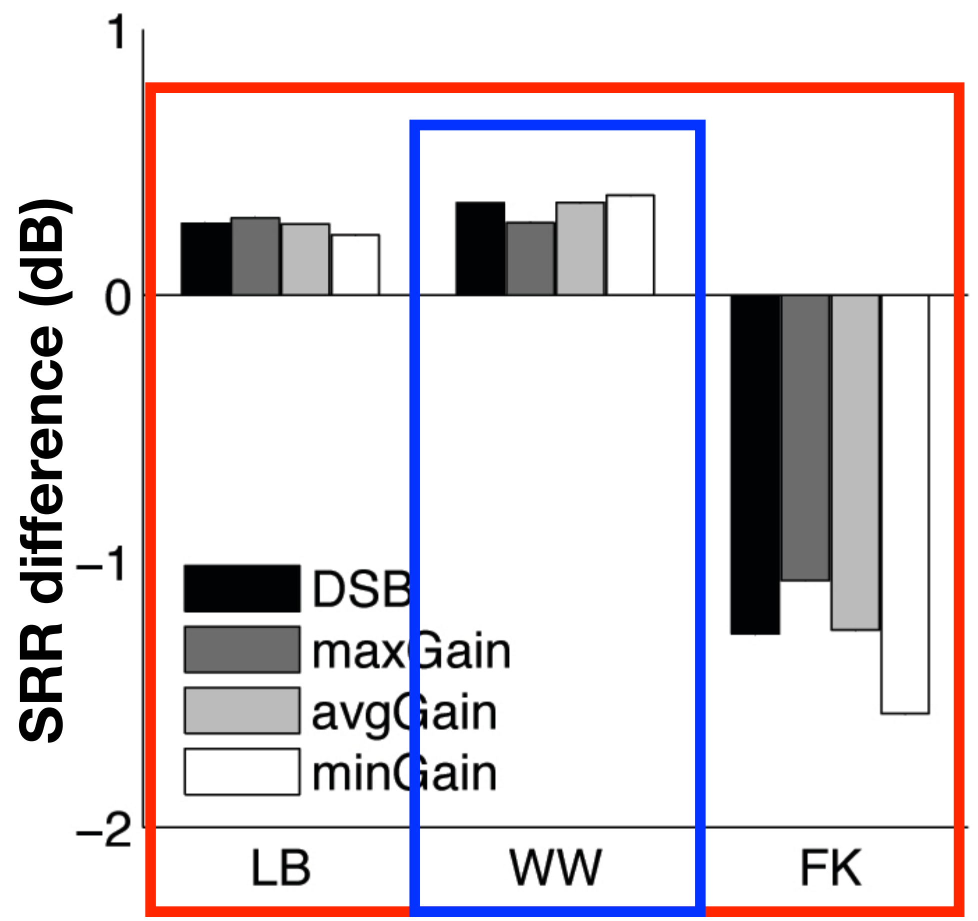
Tsilfidis et al.
University of Patras
Greece

Results

- **Cafeteria**

  RT = 1.29 s, Distance= 1.18-1.62m, Angles = -30, 0, 90° (Oldenburg Database)

Sampling Frequency: 16 kHz



Proposed Method

	Cafeteria			
Method	**DSB**	**maxGain**	**avgGain**	**minGain**
LB	0.133	**0.136**	0.135	0.133
WW	0.205	0.208	**0.216**	0.198
FK	-0.235	**-0.141**	-0.228	-0.428

PESQ difference

Results

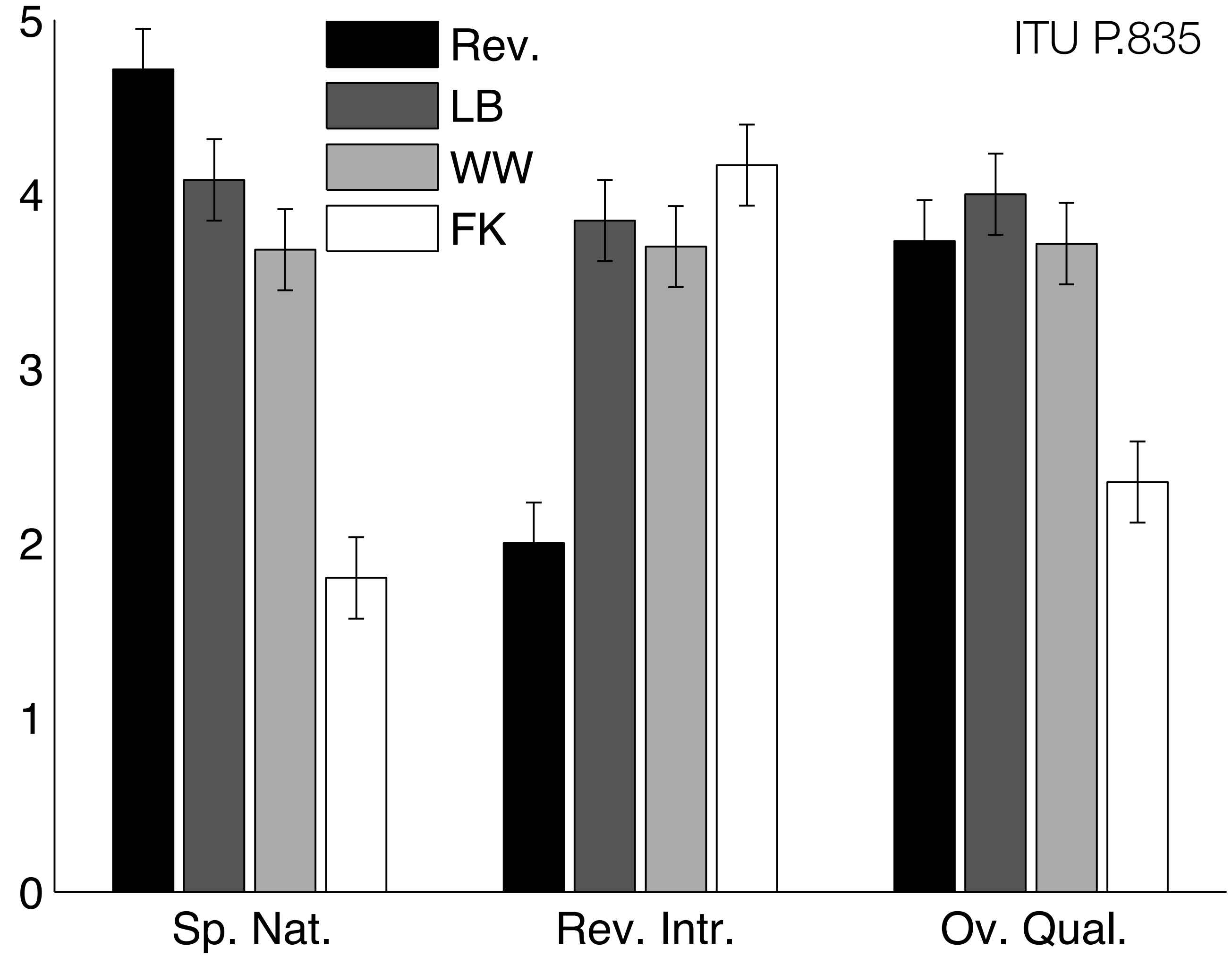
- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°

Sampling Frequency: 44.1 kHz

- 17 test subjects

LB, WW → avgGain

FK → minGain



A. Tsilfidis, E. Georganti, and J. Mourjopoulos. "A binaural framework for spectral subtraction dereverberation". In Forum Acusticum, Aalborg, Denmark, 2011

Tsilfidis et al.
University of Patras
Greece

Results

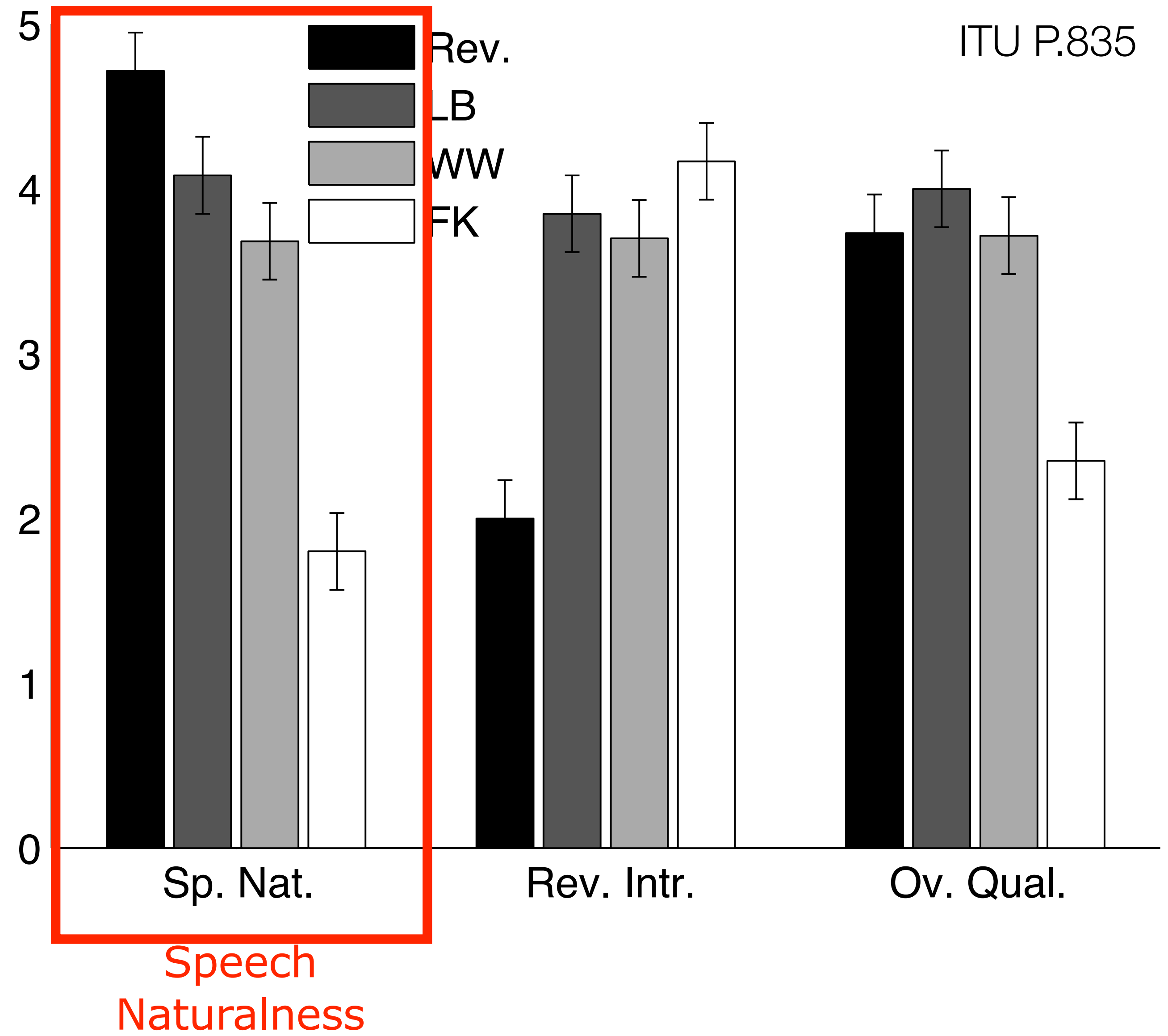
- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°

Sampling Frequency: 44.1 kHz

- 17 test subjects

LB, WW → avgGain

FK → minGain



A. Tsilfidis, E. Georganti, and J. Mourjopoulos. "A binaural framework for spectral subtraction dereverberation". In Forum Acusticum, Aalborg, Denmark, 2011

Results

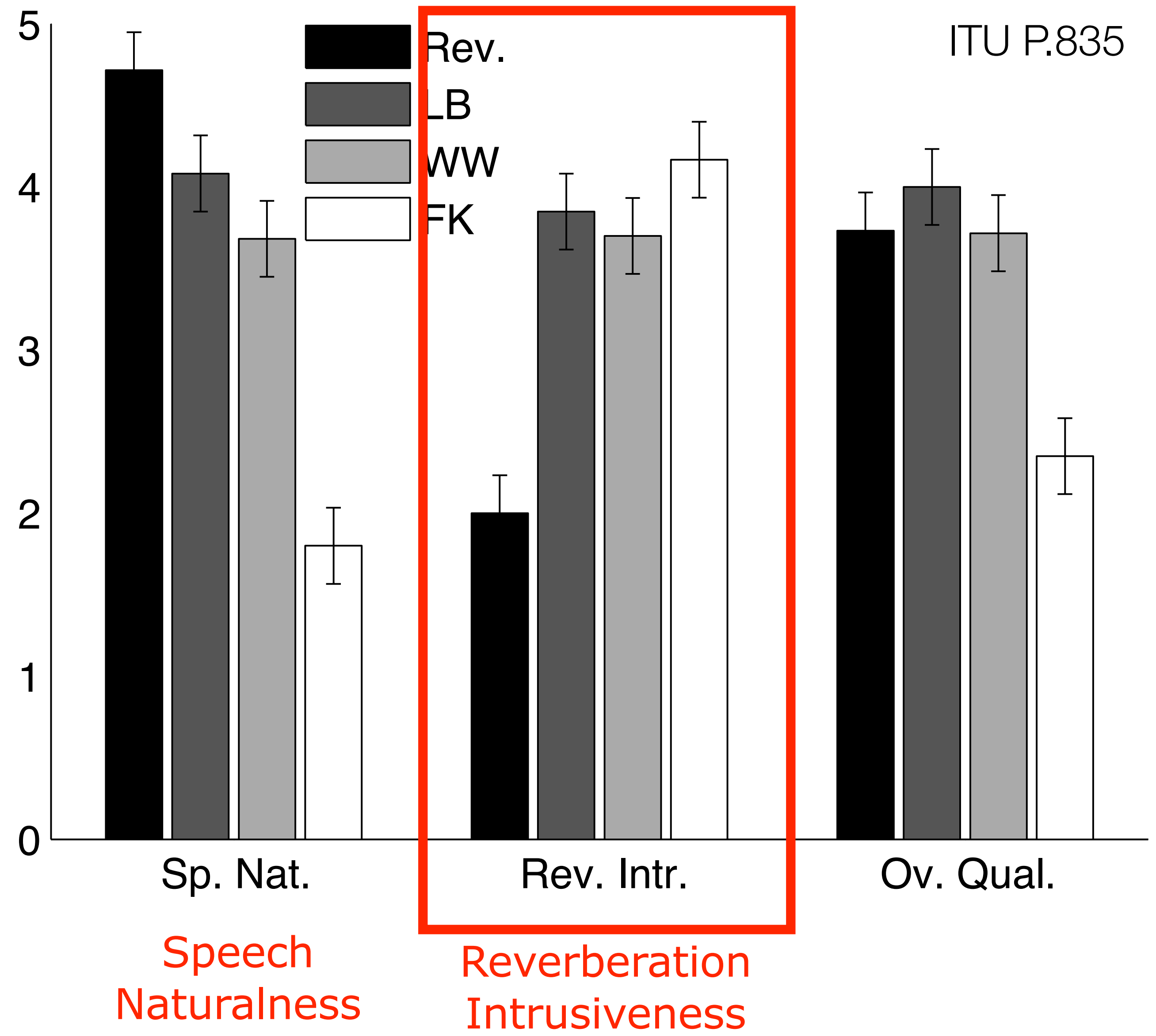
- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°

Sampling Frequency: 44.1 kHz

- 17 test subjects

LB, WW → avgGain

FK → minGain



A. Tsilfidis, E. Georganti, and J. Mourjopoulos. "A binaural framework for spectral subtraction dereverberation". In Forum Acusticum, Aalborg, Denmark, 2011

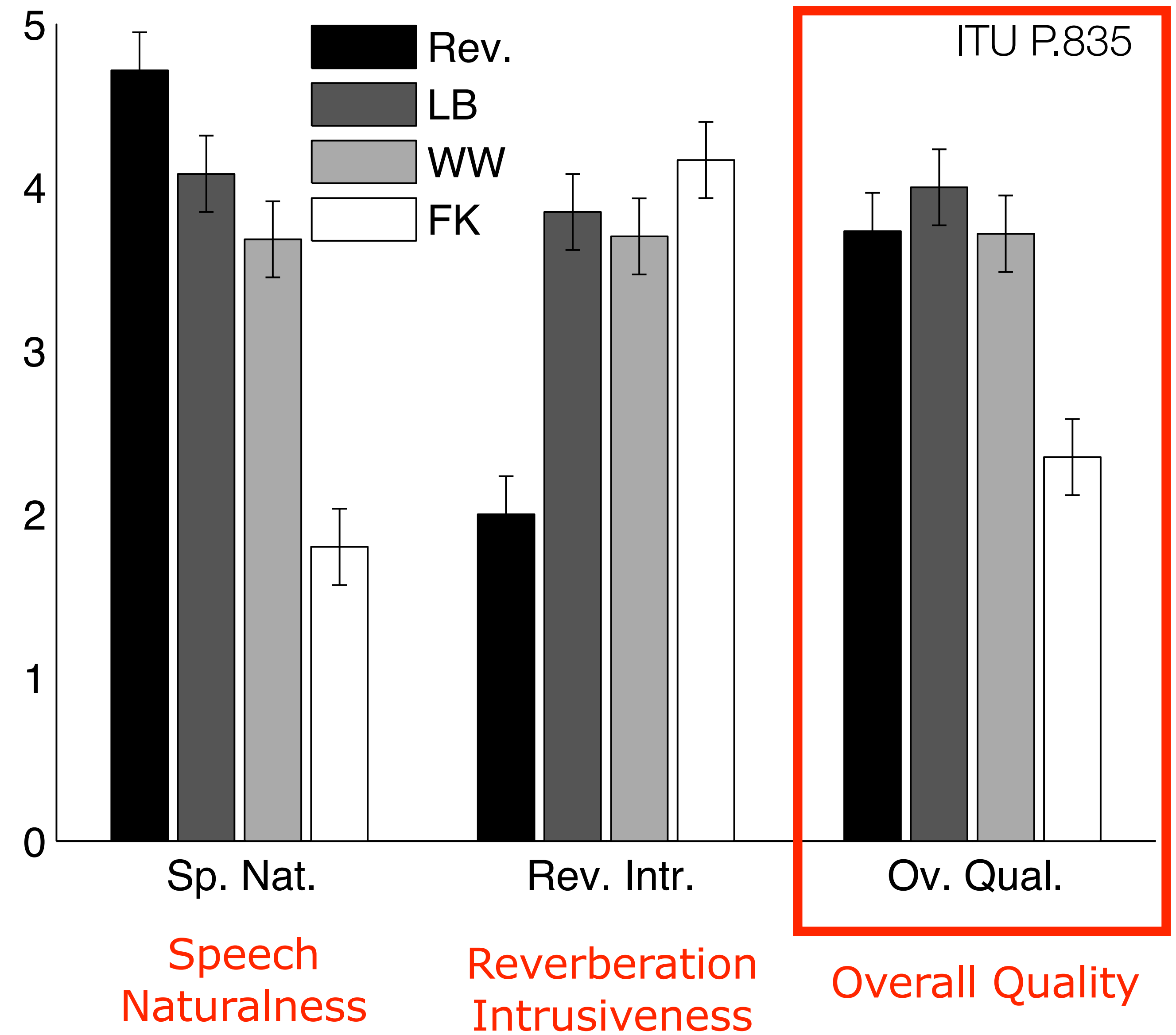
Tsilfidis et al.
University of Patras
Greece

Results

- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°
Sampling Frequency: 44.1 kHz
- 17 test subjects

LB, WW → avgGain

FK → minGain



A. Tsilfidis, E. Georganti, and J. Mourjopoulos. "A binaural framework for spectral subtraction dereverberation". In Forum Acusticum, Aalborg, Denmark, 2011

Tsilfidis et al.
University of Patras
Greece

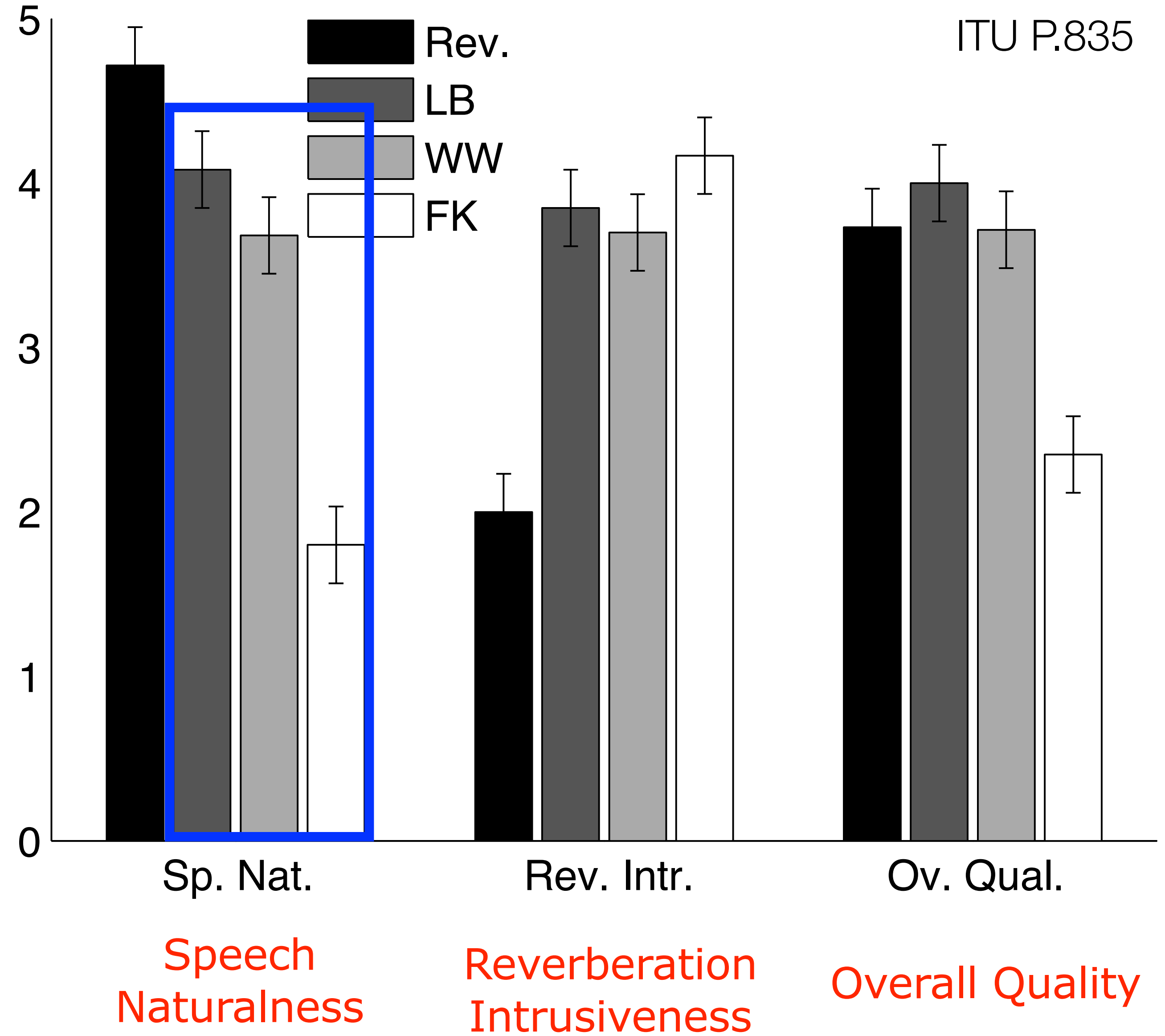
Results

- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°
Sampling Frequency: 44.1 kHz
- 17 test subjects

LB, WW → avgGain

FK → minGain

- Less natural speech



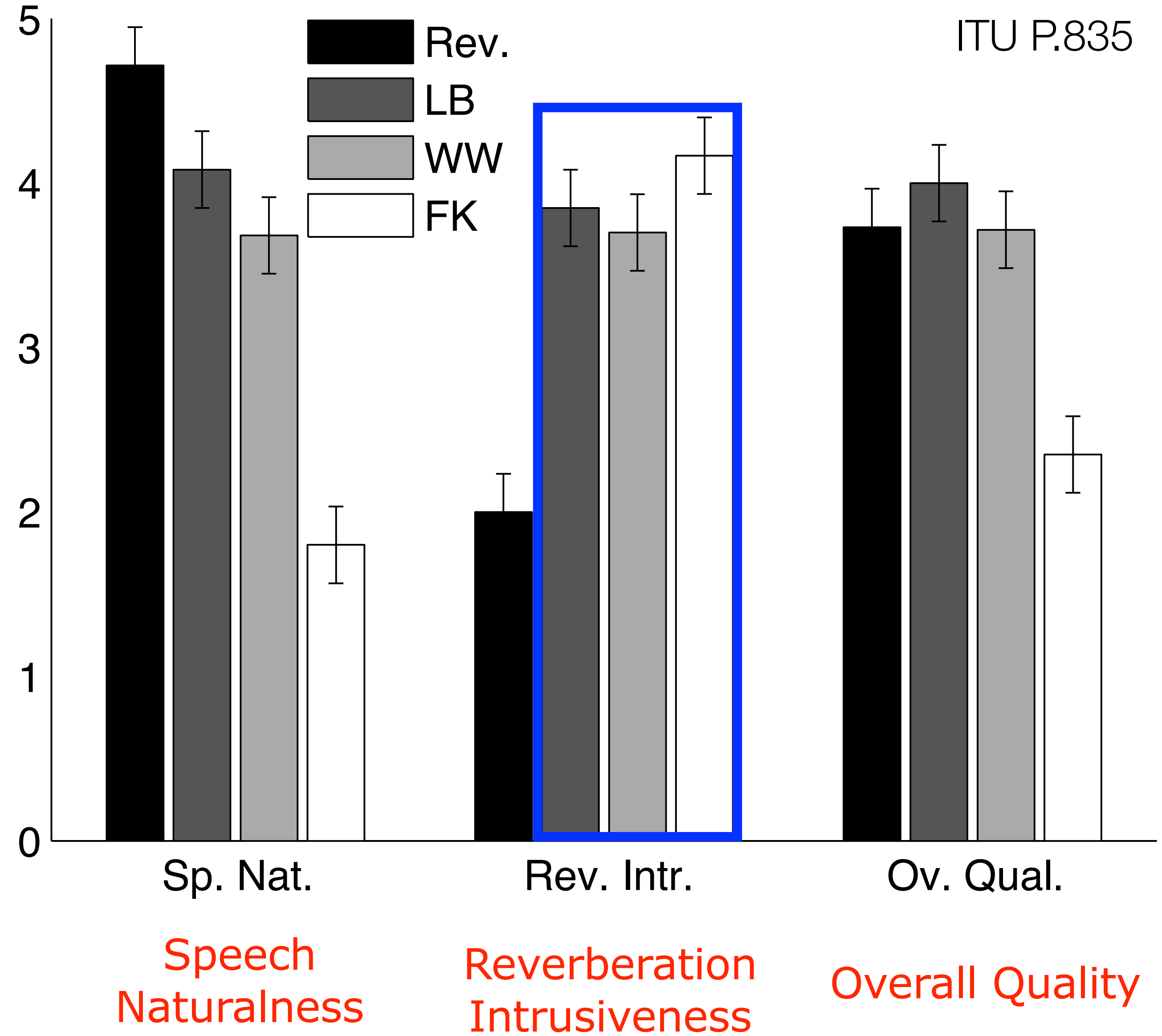
A. Tsilfidis, E. Georganti, and J. Mourjopoulos. "A binaural framework for spectral subtraction dereverberation". In Forum Acusticum, Aalborg, Denmark, 2011

Results

- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°
Sampling Frequency: 44.1 kHz
- 17 test subjects

LB, WW → avgGain

FK → minGain



- Less natural speech
- Significant Reverberation reduction

A. Tsilfidis, E. Georganti, and J. Mourjopoulos. "A binaural framework for spectral subtraction dereverberation". In Forum Acusticum, Aalborg, Denmark, 2011

Results

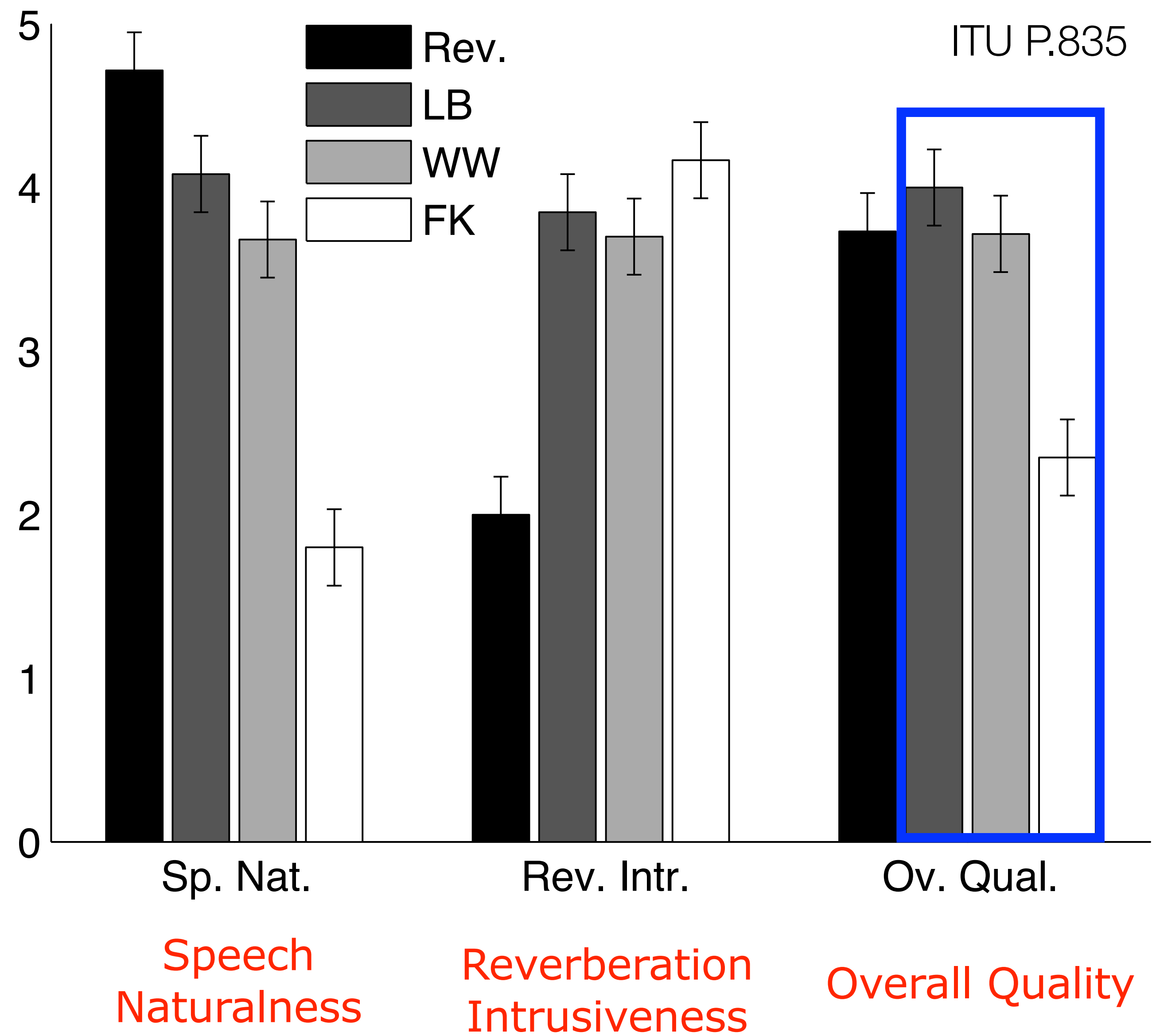
- Stairway Hall
  RT = 0.69 s, Distance= 3m,
  Angles= 0, 45, 90°

Sampling Frequency: 44.1 kHz

- 17 test subjects

LB, WW → avgGain

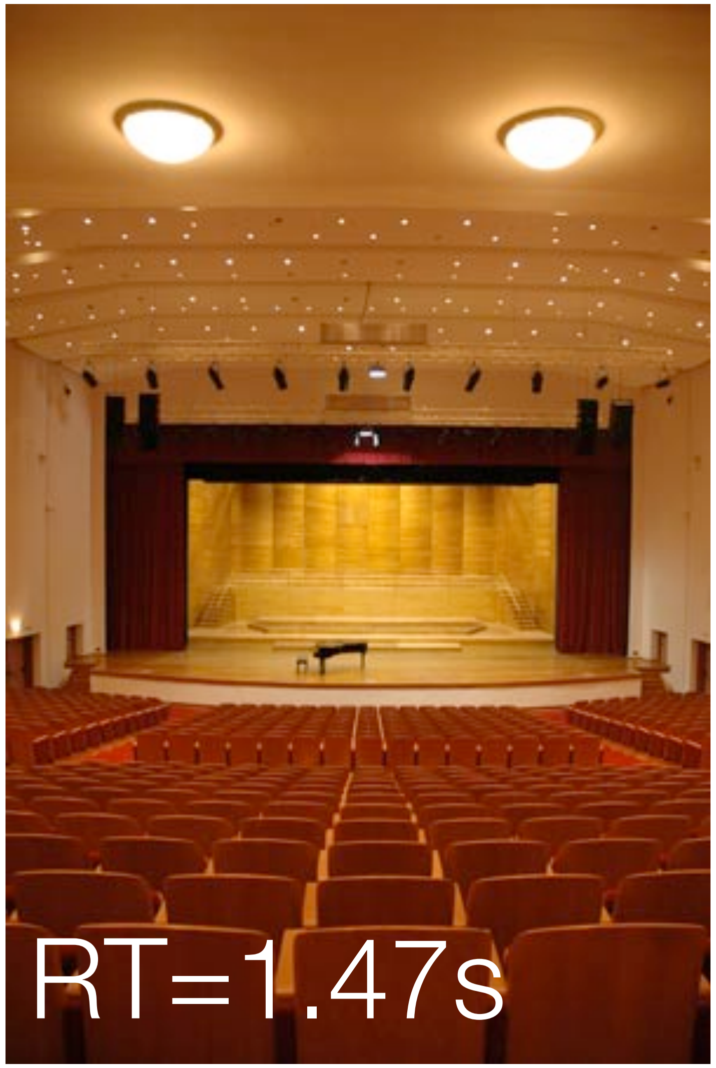
FK → minGain



- Less natural speech
- Significant Reverberation reduction
- LB & WW preserve signal quality

A. Tsilfidis, E. Georganti, and J. Mourjopoulos. “A binaural framework for spectral subtraction dereverberation”. In Forum Acusticum, Aalborg, Denmark, 2011

Audio Demos



Available at:

http://www.wcl.ece.upatras.gr/audiogroup/tools/derev.html

Conclusions

Conclusions

- Framework for binaural spectral subtraction dereverberation based on bilateral gain adaptation
- Gain Magnitude Regularization for the reduction of overestimation errors

Conclusions

- Framework for binaural spectral subtraction dereverberation based on bilateral gain adaptation
- Gain Magnitude Regularization for the reduction of overestimation errors
    - ✓ Selection of the adaptation scheme and the GMR parameters according to the application scenario
    - ✓ Significant reverberation reduction and preservation of overall speech quality and binaural cues

Conclusions

- Framework for binaural spectral subtraction dereverberation based on bilateral gain adaptation
- Gain Magnitude Regularization for the reduction of overestimation errors
  - ✓ Selection of the adaptation scheme and the GMR parameters according to the application scenario
  - ✓ Significant reverberation reduction and preservation of overall speech quality and binaural cues

However...

Loss of speech naturalness → Need for native binaural methods that take into account the binaural properties of the auditory system

Future work

Thank you for attention!