Integrating Beamforming with Binaural Sound Reproduction using a Spherical Microphone Array

By Michael Jeffet, Noam R. Shabtai and Boaz Rafaely

Department of Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer-Sheva 84105, Israel.

January 8, 2015

Outline

Introduction

- Video conferencing or telepresence applications.

Introduction

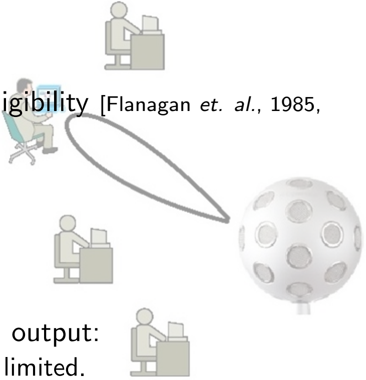
- Video conferencing or telepresence applications.
- Disturbing signals reduce speech intelligibility:
  - Other talkers.
  - Background noise.
  - Reverberation.

Introduction

- Video conferencing or telepresence applications.
- Disturbing signals reduce speech intelligibility:
  - Other talkers.
  - Background noise.
  - Reverberation.
- Microphone arrays increase intelligibility [Flanagan *et. al.*, 1985, Benesty, 2001]:
  - Beamforming.
  - Noise reduction.
  - Echo cancellation.
  - Dereverberation.
  - Blind source separation.

Introduction

- Video conferencing or telepresence applications.
- Disturbing signals reduce speech intelligibility:
  - Other talkers.
  - Background noise.
  - Reverberation.
- Microphone arrays increase intelligibility [Flanagan *et. al.*, 1985, Benesty, 2001]:
  - Beamforming.
  - Noise reduction.
  - Echo cancellation.
  - Dereverberation.
  - Blind source separation.
- Usually employing single-channel output:
  - Inherent spatial information is limited.
  - More compatible with machine receiver applications.

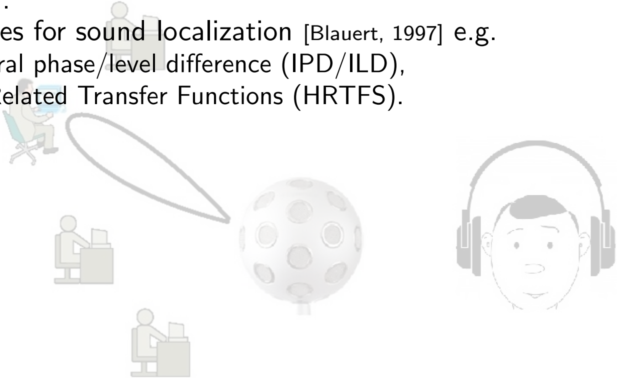


Introduction

- Binaural output with spatial information for cocktail party effect and spatial release from masking (SRM) [Jones and Litovsky, 2011].

Introduction

- Binaural output with spatial information for cocktail party effect and spatial release from masking (SRM) [Jones and Litovsky, 2011].
- Binaural cues for sound localization [Blauert, 1997] e.g.
  - interaural phase/level difference (IPD/ILD),
  - Head Related Transfer Functions (HRTFS).



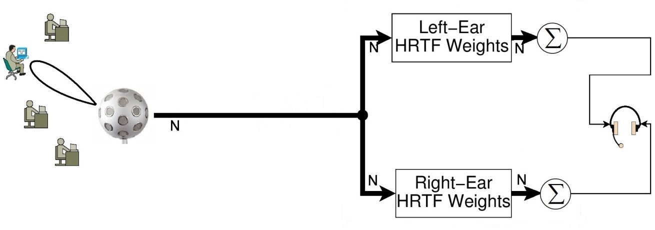
Introduction

- Binaural output with spatial information for cocktail party effect and spatial release from masking (SRM) [Jones and Litovsky, 2011].
- Binaural cues for sound localization [Blauert, 1997] e.g.
  - interaural phase/level difference (IPD/ILD),
  - Head Related Transfer Functions (HRTFS).
- Beamformers which preserve binaural cues:
  - Binaural Wiener filtering [Klasen *et. al.*, 2007, Doclo *et. al.*, 2009].
  - Partial communication between sensors in hearing aids [Bertrand and Moonen, 2009, Bertrand and Moonen, 2010].
  - Model-based dereverberation for binaural cues preservation [Jeub *et. al.*, 2010].
  - Binaural extension of spectral-subtraction dereverberation [Tsildis *et. al.*, 2011].

Introduction

- Binaural output with spatial information for cocktail party effect and spatial release from masking (SRM) [Jones and Litovsky, 2011].
- Binaural cues for sound localization [Blauert, 1997] e.g.
  - interaural phase/level difference (IPD/ILD),
  - Head Related Transfer Functions (HRTFS).
- Beamformers which preserve binaural cues:
  - Binaural Wiener filtering [Klasen *et. al.*, 2007, Doclo *et. al.*, 2009].
  - Partial communication between sensors in hearing aids [Bertrand and Moonen, 2009, Bertrand and Moonen, 2010].
  - Model-based dereverberation for binaural cues preservation [Jeub *et. al.*, 2010].
  - Binaural extension of spectral-subtraction dereverberation [Tsildis *et. al.*, 2011].
- The preserved spatial information is limited to the target source.

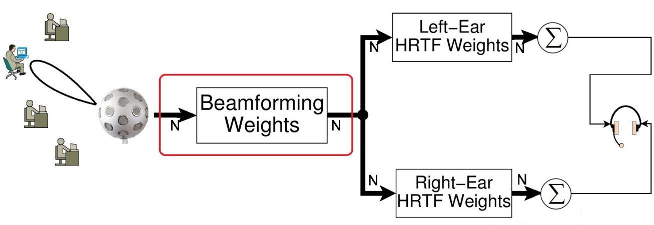
Binaural Reproduction- Plane Wave Domain



- Every direction of arrival is given its own binaural cue using the HRTFs. [Rafaely and Avni, 2010],

$$y_{l/r}(k) = \int_{\Omega \in S^2} H_{l/r}(k, \Omega) a(k, \Omega) d\Omega$$

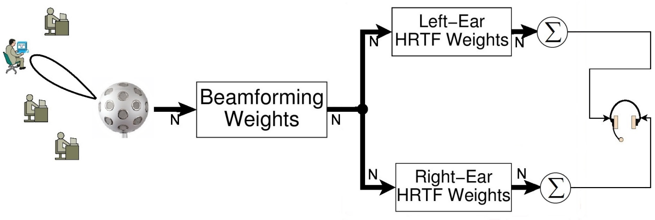
Generalized Spherical Beamforming



- The product is multiplied by a beamforming weight function to suppress interference [Shabtai and Rafaely, 2014],

$$y_{l/r}(k) = \int_{\Omega \in S^2} w^*(k,\Omega) H_{l/r}(k,\Omega) a(k,\Omega) d\Omega$$

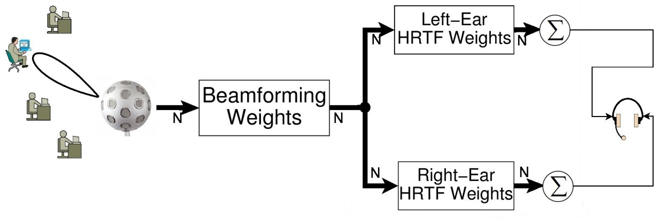
Generalized Spherical Beamforming



- The GSB incorporates beamforming and binaural reproduction [Shabtai and Rafaely, 2014],

$$y_{l/r}(k) = \int_{\Omega \in S^2} w^*(k,\Omega) H_{l/r}(k,\Omega) a(k,\Omega) d\Omega$$

Generalized Spherical Beamforming



- The GSB incorporates beamforming and binaural reproduction [Shabtai and Rafaely, 2014],

$$y_{l/r}(k) = \int_{\Omega \in S^2} \underbrace{w^*(k,\Omega)H_{l/r}(k,\Omega)}_{\widetilde{w}^*_{l/r}(k,\Omega)} a(k,\Omega)d\Omega$$

- $\widetilde{w}_{l/r}(k,\Omega) = w(k,\Omega)H^*_{l/r}(k,\Omega)$- Generalized weight function.

Truncation Problem

- The GSB in the Spherical Harmonics domain,

$$y(k) = \int w^*(k,\Omega)H(k,\Omega)a(k,\Omega)d\Omega = \sum_{n_a=0}^{N_a} \sum_{m_a=-n_a}^{n_a} \widetilde{w}^*_{n_a m_a}(k) a_{n_a m_a}(k)$$

Truncation Problem

- The GSB in the Spherical Harmonics domain,

$$y(k) = \int w^*(k,\Omega)H(k,\Omega)a(k,\Omega)d\Omega = \sum_{n_a=0}^{N_a} \sum_{m_a=-n_a}^{n_a} \widetilde{w}^*_{n_a m_a}(k) a_{n_a m_a}(k)$$

- $a(k,\Omega)$ is order limited by N_a due to the sampling of the sphere.

Truncation Problem

- The GSB in the Spherical Harmonics domain,

$$y(k) = \int w^*(k,\Omega)H(k,\Omega)a(k,\Omega)d\Omega = \sum_{n_a=0}^{N_a} \sum_{m_a=-n_a}^{n_a} \widetilde{w}^*_{n_a m_a}(k) a_{n_a m_a}(k)$$

- $a(k,\Omega)$ is order limited by N_a due to the sampling of the sphere.
- $\widetilde{w}(k,\Omega) = w(k,\Omega)H^*(k,\Omega)$, is truncated by N_a.

Truncation Problem

- The GSB in the Spherical Harmonics domain,

$$y(k) = \int w^*(k,\Omega)H(k,\Omega)a(k,\Omega)d\Omega = \sum_{n_a=0}^{N_a}\sum_{m_a=-n_a}^{n_a} \widetilde{w}^*_{n_a m_a}(k)a_{n_a m_a}(k)$$

- $a(k,\Omega)$ is order limited by N_a due to the sampling of the sphere.
- $\widetilde{w}(k,\Omega) = w(k,\Omega)H^*(k,\Omega)$, is truncated by N_a.

$$w(k,\Omega) = \sum_{n_w=0}^{N_w}\sum_{m_w=-n_w}^{n_w} w_{n_w m_w}(k)Y_{n_w}^{m_w}(\Omega),$$

$$H(k,\Omega) = \sum_{n_h=0}^{N_h}\sum_{m_h=-n_h}^{n_h} H_{n_h m_h}(k)Y_{n_h}^{m_h}(\Omega)$$

- The order of $\widetilde{w}(k,\Omega)$ is $(N_w + N_h)$.

Truncation Problem

- The GSB in the Spherical Harmonics domain,

$$y(k) = \int w^*(k,\Omega)H(k,\Omega)a(k,\Omega)d\Omega = \sum_{n_a=0}^{N_a}\sum_{m_a=-n_a}^{n_a}\widetilde{w}^*_{n_a m_a}(k)a_{n_a m_a}(k)$$

- $a(k,\Omega)$ is order limited by N_a due to the sampling of the sphere.
- $\widetilde{w}(k,\Omega) = w(k,\Omega)H^*(k,\Omega)$, is truncated by N_a.

$$w(k,\Omega) = \sum_{n_w=0}^{N_w}\sum_{m_w=-n_w}^{n_w} w_{n_w m_w}(k)Y_{n_w}^{m_w}(\Omega),$$

$$H(k,\Omega) = \sum_{n_h=0}^{N_h}\sum_{m_h=-n_h}^{n_h} H_{n_h m_h}(k)Y_{n_h}^{m_h}(\Omega)$$

- The order of $\widetilde{w}(k,\Omega)$ is (N_w+N_h).
- High orders of $w(k,\Omega)$ and $H(k,\Omega)$ are reflected in low orders of $\widetilde{w}(k,\Omega)$

No truncation of $\widetilde{w}(k, \Omega)$

$$(N_w + N_h) \leq N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k, \Omega) = w^*(k, \Omega) H_{l/r}(k, \Omega).$$

No truncation of $\widetilde{w}(k, \Omega)$

$$(N_w + N_h) \leq N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k, \Omega) = w^*(k, \Omega) H_{l/r}(k, \Omega).$$

- N_h affect the spatial complexity of $H_{l/r}(k, \Omega)$.

No truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) \leq N_a \stackrel{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) = w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- N_h affect the spatial complexity of $H_{l/r}(k,\Omega)$.
- Given N_h, the optimal directivity of the GSB is given for $N_w = N_a - N_h$.

No truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) \leq N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) = w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- N_h affect the spatial complexity of $H_{l/r}(k,\Omega)$.
- Given N_h, the optimal directivity of the GSB is given for $N_w = N_a - N_h$.

Truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) > N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) \neq w^*(k,\Omega)H_{l/r}(k,\Omega).$$

No truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) \leq N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) = w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- N_h affect the spatial complexity of $H_{l/r}(k,\Omega)$.
- Given N_h, the optimal directivity of the GSB is given for $N_w = N_a - N_h$.

Truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) > N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) \neq w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- There is a truncation error at the GSB output.

No truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) \leq N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) = w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- N_h affect the spatial complexity of $H_{l/r}(k,\Omega)$.
- Given N_h, the optimal directivity of the GSB is given for $N_w = N_a - N_h$.

Truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) > N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) \neq w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- There is a truncation error at the GSB output.
- How is the directivity of the GSB affected?

No truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) \leq N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) = w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- N_h affect the spatial complexity of $H_{l/r}(k,\Omega)$.
- Given N_h, the optimal directivity of the GSB is given for $N_w = N_a - N_h$.

Truncation of $\widetilde{w}(k,\Omega)$

$$(N_w + N_h) > N_a \overset{beam-pattern}{\Longrightarrow} y_{l/r}(k,\Omega) \neq w^*(k,\Omega)H_{l/r}(k,\Omega).$$

- There is a truncation error at the GSB output.
- How is the directivity of the GSB affected?
- Given N_w, does an increase in N_h improves the approximation of $H_{l/r}(k,\Omega)$?

Simulations- Methodology

- **Goal**- To objectively evaluate the truncation affect on the performance of the GSB.

Simulations- Methodology

- **Goal**- To objectively evaluate the truncation affect on the performance of the GSB.
- **Measures**-
  - **Directivity Factor (DF)**, spatial selectivity property [Gerzon, 1973].

$$DF_{l/r} = \frac{\left|y_{l/r}(\Omega_l)\right|^2}{\frac{1}{4\pi}\int_{\Omega_o \in S^2}\left|y_{l/r}(\Omega)\right|^2 d\Omega}$$

Simulations- Methodology

- **Goal**- To objectively evaluate the truncation affect on the performance of the GSB.
- **Measures**-
  - **Directivity Factor (DF)**, spatial selectivity property [Gerzon, 1973].

$$DF_{l/r} = \frac{\left|y_{l/r}(\Omega_l)\right|^2}{\frac{1}{4\pi}\int_{\Omega_o \in S^2} \left|y_{l/r}(\Omega)\right|^2 d\Omega}$$

  - **Binaural Error(BE)**, binaural reproduction effectiveness estimation.

$$\frac{y_l(\Omega)}{y_r(\Omega)} \overset{(N_w+N_h)\leq N_a}{=} \frac{H_l(\Omega)}{H_r(\Omega)}$$

BLE- Normalized ILD error.
BPE- Normalized IPD error.

Simulations- Methodology

- **PWD**- Limited to $N_a = 5$.

Simulations- Methodology

- **PWD**- Limited to $N_a = 5$.
- **Simulated Orders**- $0 \leq N_w \leq 15$, $0 \leq N_h \leq 15$.

Simulations- Methodology

- **PWD**- Limited to $N_a = 5$.
- **Simulated Orders**- $0 \leq N_w \leq 15$, $0 \leq N_h \leq 15$.
- **Beamformer**- Maximum directivity, $DF = 20\log(N_w + 1)$[dB].

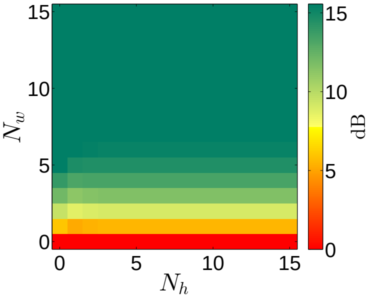
Simulations- Methodology

- **PWD**- Limited to $N_a = 5$.
- **Simulated Orders**- $0 \leq N_w \leq 15$, $0 \leq N_h \leq 15$.
- **Beamformer**- Maximum directivity, $DF = 20\log(N_w + 1)$[dB].
- **Neumann KU100 HRTF**- 2354 measured directions sampled with Lebedev's sampling scheme, (N_h up to 40).
  [Bernschütz, 2013]

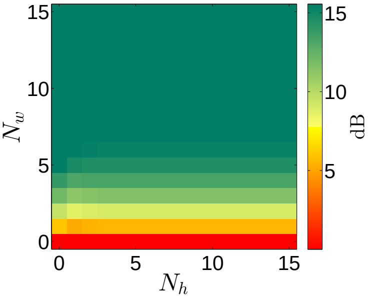
Simulations- Methodology

- **PWD**- Limited to $N_a = 5$.
- **Simulated Orders**- $0 \leq N_w \leq 15$, $0 \leq N_h \leq 15$.
- **Beamformer**- Maximum directivity, $DF = 20\log(N_w + 1)$[dB].
- **Neumann KU100 HRTF**- 2354 measured directions sampled with Lebedev's sampling scheme, (N_h up to 40).
  [Bernschütz, 2013]
- **Smoothing**- The results were averaged over the look direction, Ω_l, to compensate for the HRTF variance.

Simulations- DF results



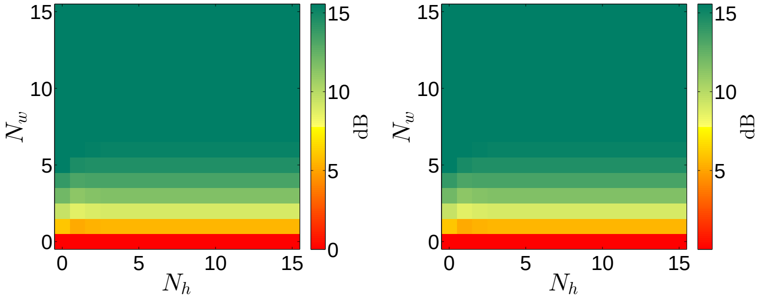
(a) Left ear



(b) Right ear

Figure: Average DF for $N_a = 5$ at frequency of 1000Hz.

Simulations- DF results

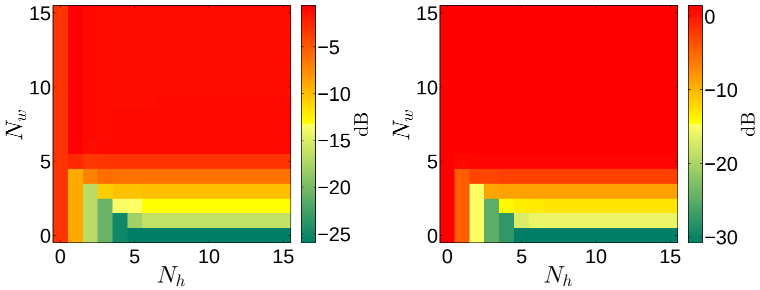


(a) Left ear (b) Right ear

Figure: Average DF for $N_a = 5$ at frequency of 1000Hz.

- The DF is mainly effected by N_w.

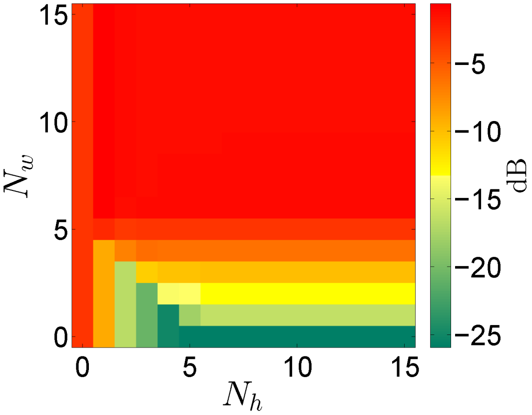
Simulations- BE results



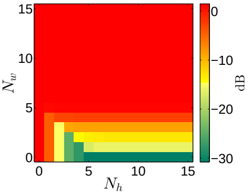
(a) BLE (b) BPE

Figure: Average BE for $N_a = 5$ at frequency of 1000Hz.

Simulations- BE results



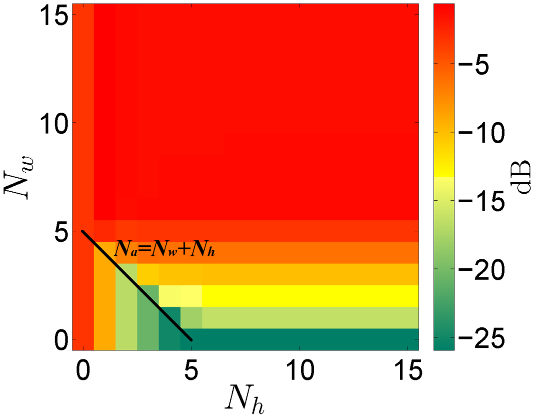
(a) BLE



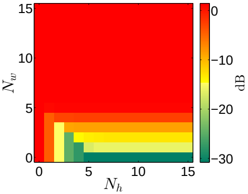
(b) BPE

Figure: Average BE for $N_a = 5$ at frequency of 1000Hz.

Simulations- BE results



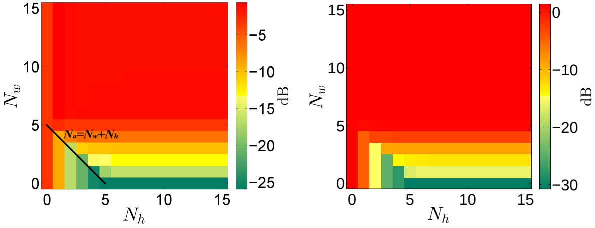
(a) BLE



(b) BPE

Figure: Average BE for $N_a = 5$ at frequency of 1000Hz.

Simulations- BE results



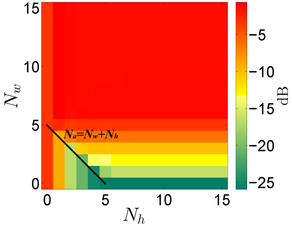
(a) BLE (b) BPE

Figure: Average BE for $N_a = 5$ at frequency of 1000Hz.

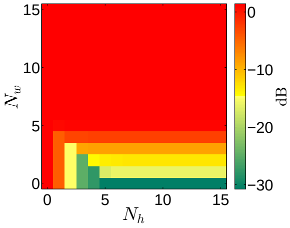
Results

- Best configurations on the line $N_a = N_h + N_w$.

Simulations- BE results



(a) BLE



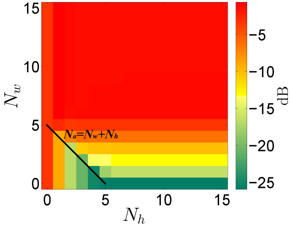
(b) BPE

Figure: Average BE for $N_a = 5$ at frequency of 1000Hz.

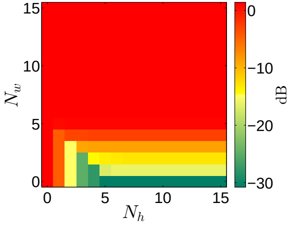
Results

- Best configurations on the line $N_a = N_h + N_w$.
- Spatial selectivity vs. binaural reproduction trade-off.

Simulations- BE results



(a) BLE



(b) BPE

Figure: Average BE for $N_a = 5$ at frequency of 1000Hz.

Results

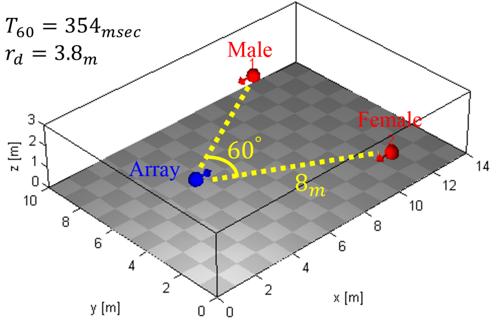
- Best configurations on the line $N_a = N_h + N_w$.
- Spatial selectivity vs. binaural reproduction trade-off.
- Is this relation perceptually noticeable?

Listening Test- Methodology

- **Goal**- Subjective effect of N_w and N_h on the GSB.

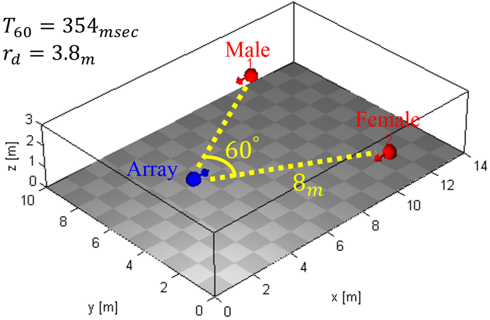
Listening Test- Methodology

- **Goal**- Subjective effect of N_w and N_h on the GSB.



Listening Test- Methodology

- **Goal**- Subjective effect of N_w and N_h on the GSB.



- **Criteria answered**:
  - **Spatial perception**- Realism of the sound scene.
  - **Interference suppression**- Clearly perceive the male speaker without the female speaker interfering.

Listening Test- Methodology

- **Test**- MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA).

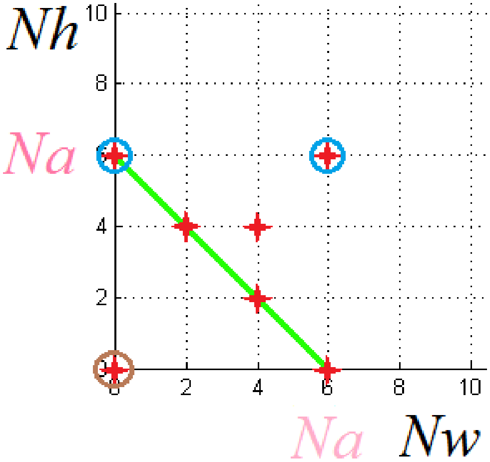


Figure: Test stimuli setup points

Listening Test- Methodology

- **Test**- MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA).
- **Subjects**- Eighteen normal hearing subjects, 5 female, 13 males, ages 24-37.

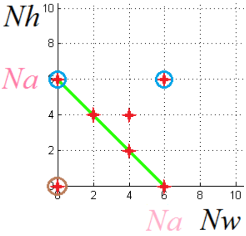


Figure: Test stimuli setup points

Listening Test- Methodology

- **Test**- MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA).
- **Subjects**- Eighteen normal hearing subjects, 5 female, 13 males, ages 24-37.
- **Head tracking**- Horizontal.

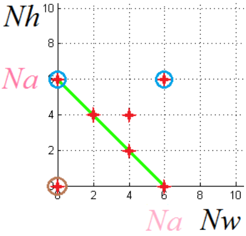


Figure: Test stimuli setup points

Listening Test- Methodology

- **Test**- MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA).
- **Subjects**- Eighteen normal hearing subjects, 5 female, 13 males, ages 24-37.
- **Head tracking**- Horizontal.
- **Look direction**- Male speaker.

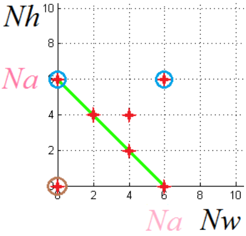


Figure: Test stimuli setup points

Listening Test- Methodology

- **Test**- MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA).
- **Subjects**- Eighteen normal hearing subjects, 5 female, 13 males, ages 24-37.
- **Head tracking**- Horizontal.
- **Look direction**- Male speaker.
- **Post beamforming compensations**- Headphones and timbre compensation filters.

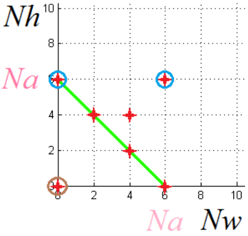


Figure: Test stimuli setup points

Listening Test- Methodology

- **Test**- MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA).
- **Subjects**- Eighteen normal hearing subjects, 5 female, 13 males, ages 24-37.
- **Head tracking**- Horizontal.
- **Look direction**- Male speaker.
- **Post beamforming compensations**- Headphones and timbre compensation filters.
- **Playback**- Sound Scape Renderer (SSR) auralization engine.

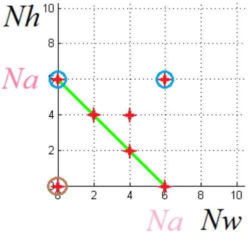
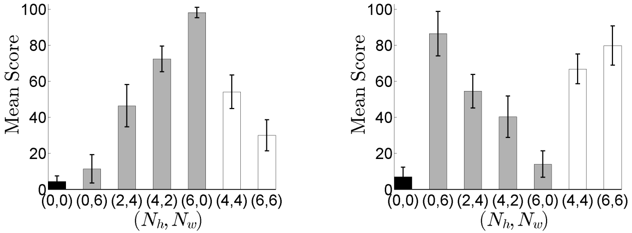


Figure: Test stimuli setup points

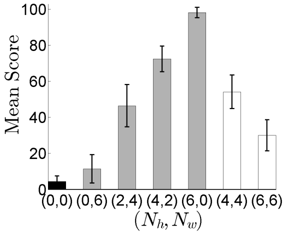
Listening Test- Results



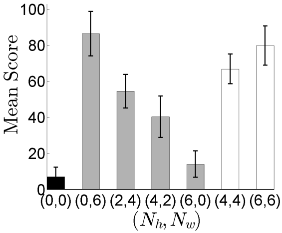
(a) Spatial perception (b) Interference suppression

Figure: Listening test results for $N_a = 6$

Listening Test- Results



(a) Spatial perception



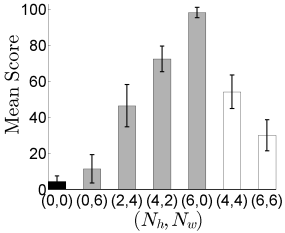
(b) Interference suppression

Figure: Listening test results for $N_a = 6$

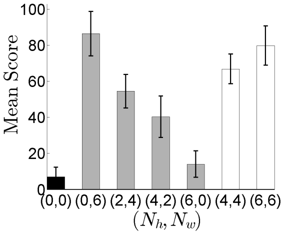
Results

- On-the-line trade off, spatial selectivity vs. spatial perception.

Listening Test- Results



(a) Spatial perception



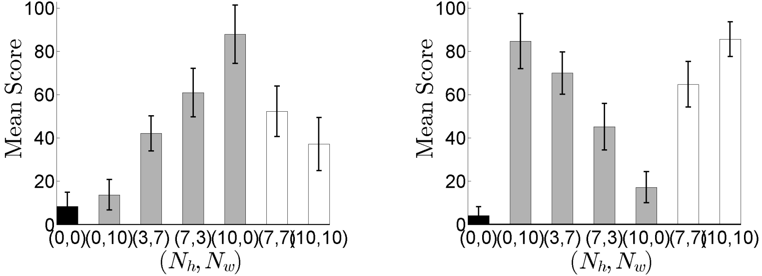
(b) Interference suppression

Figure: Listening test results for $N_a = 6$

Results

- On-the-line trade off, spatial selectivity vs. spatial perception.
- Potential future research for out-of-line stimuli.

Listening Test- Results



(a) Spatial perception

(b) Interference suppression

Figure: Listening test results for $N_a = 10$

Conclusions

- Limited $N_a \rightarrow$ Spatial selectivity and spatial perception trade-off.

Conclusions

- Limited $N_a \rightarrow$ Spatial selectivity and spatial perception trade-off.
- Expected operation on the line $N_a = N_h + N_w$.

Conclusions

- Limited $N_a \rightarrow$ Spatial selectivity and spatial perception trade-off.
- Expected operation on the line $N_a = N_h + N_w$.
- Encapsulation of beamforming, binaural reproduction, or a mixed mode of operation.

Conclusions

- Limited $N_a \rightarrow$ Spatial selectivity and spatial perception trade-off.
- Expected operation on the line $N_a = N_h + N_w$.
- Encapsulation of beamforming, binaural reproduction, or a mixed mode of operation.
- A Tunable GSB is proposed in order to benefit from the advantage of the two methods integration in a varying environment.

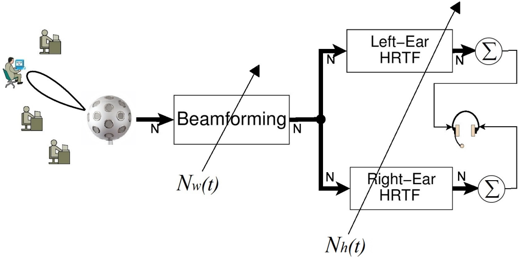


Figure: Tunable GSB

Thank you

GSB Evaluation Measures- Binaural Error

- The GSB's output maintain the following relation,

$$\frac{y_l^{\Omega_o}}{y_r^{\Omega_o}} \overset{Na \to \infty}{=} \frac{H_l(\Omega_o)}{H_r(\Omega_o)}$$

GSB Evaluation Measures- Binaural Error

- The GSB's output maintain the following relation,

$$\frac{y_l^{\Omega_o}}{y_r^{\Omega_o}} \overset{Na\to\infty}{=} \frac{H_l(\Omega_o)}{H_r(\Omega_o)}$$

- To account for the ILD and IPD binaural cues, the binaural level error (BLE) and the binaural phase error (BPE) should be examined separately,

$$BLE = \sqrt{\int_{\Omega_o \in S^2} \left(\frac{\left| y_l^{\Omega_o} H_r(\Omega_o) \right| - \left| y_r^{\Omega_o} H_l(\Omega_o) \right|}{\frac{1}{2}\left(\left| y_l^{\Omega_o} H_r(\Omega_o) \right| + \left| y_r^{\Omega_o} H_l(\Omega_o) \right| \right)} \right)^2 d\Omega_o}$$

$$BPE = \sqrt{\int_{\Omega_o \in S^2} \left(\angle \left(y_l^{\Omega_o} H_r(\Omega_o) \right) - \angle \left(y_r^{\Omega_o} H_l(\Omega_o) \right) \right)^2 d\Omega_o}$$