Spectral and binaural cue reweighting for sound localization in real and virtual environments

Ondrej Spišák[a], Maike Klingel[b], Peter Lokša[a], René Šebeňa[a], Bernhard Laback[b], Norbert Kopčo[a]

[A] Perception and Cognition Lab, Institute of Computer Science
P. J. Šafárik University in Košice, Slovakia

[B] Institut für Schallforschung, ÖAW

[Work supported by Danube Partnership Grant DS-FR-19-0025]

Introduction

Weighting of binaural cues in sound localization

- is frequency-dependent (Strutt, 1907):
    - for low-frequency (LF) sounds ITD dominates,
    - for high-frequency (HF) sounds ILD dominates.
- has been typically measured as **trading ratios** using headphones (e.g., Colburn and Durlach, 1965), but also attempted using loudspeakers (Leakey and Cherry, 1957),
- can vary dynamically, e.g., in reverberation (Rakerd and Hartman, 2010), or due to attention (Lang & Buchner, 2008).

Reweighting:

- has been achieved for **binaural cues** using audiovisual training under headphones (Klingel et al., 2021; Moore et al, 2020; Kumpik et al., 2019), but was not always successful (Jeffers and McFadden, 1971),
- has been achieved for increasing **monaural** vs. **binaural** cues by plugging one ear (Kumpik et al, 2010),
- occurs commonly for different **spectral components**, e.g., in speech perception (Stilp et al., 2016).

Current study

1. Examine whether **audiovisual training with dynamic cues** can be used to induce **spectral reweighting for horizontal localization** in real reverberant environment, by either increasing the weight of:
   - HF spectral components, or
   - LF spectral components.
2. Check **generalization** to **non-trained mid-frequencies**.
3. Test whether such spectral reweighting **generalizes to binaural reweighting at mid frequencies** in a virtual environment (like in Klingel et al., 2021), i.e., whether:
   - increased HF weight leads to increased ILD weight, and/or,
   - increased LF weight leads to increased ITD weight.

Approach

Methods:

- Behavioral experiment using broadband multi-component noise stimuli in **real environment.**
- Train two groups of subjects using visual guiding signals:
  - to increase the weight of HF (f > 2.8 kHz) components – **HF group** (12 normal-hearing listeners)
  - to increase the weight of LF (f < 2.8kHz) components – **LF group** (12 normal-hearing listeners)

Questions:

- Test:
  - whether the training induces spectral reweighting,
  - whether the spectral reweighting generalizes to stimuli with an untrained frequency component (2.8 kHz),
  - whether the spectral reweighting generalizes to ITD/ILD reweighting at 2.8 kHz (using VR).

Real Environment (re. Virtual Environment):

- no issues with veridicality/accuracy of localization, externalization, easy to generate dynamic cues,
- cannot independently manipulate binaural cues.

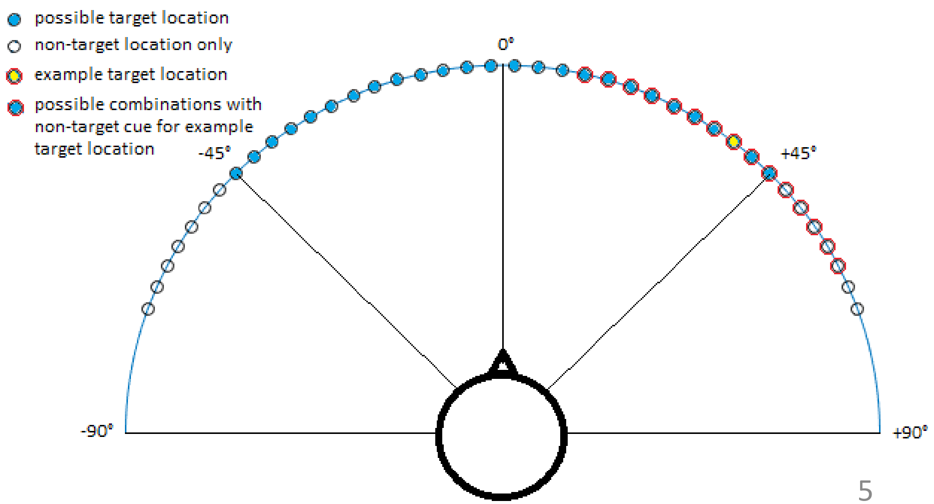
Setup

Real environment (RE) – testing & training

- 11 speakers in semicircle from -56° to 56 (11° spacing),
- 300-ms 0.5-oct noise bursts in channels centered at:
  - LF: .35 or .7 kHz, and
  - HF: 5.6 or 11.2 kHz
- 2 (1 LF & 1 HF) or 4 components (2 LF & 2HF) played from the same or neighboring speakers (up to 2 speakers apart),
- visual stimuli projected on screen above speakers,
- head-turns used to indicate perceived location.

Virtual environment (VE) – testing only

- 1-octave noise bursts (Fc=2.8 kHz) presented with ITD/ILD location inconsistency of up to 25° within a range of ±70°.
- Head-mounted display (*Oculus*) used to track head turns to perceived location.

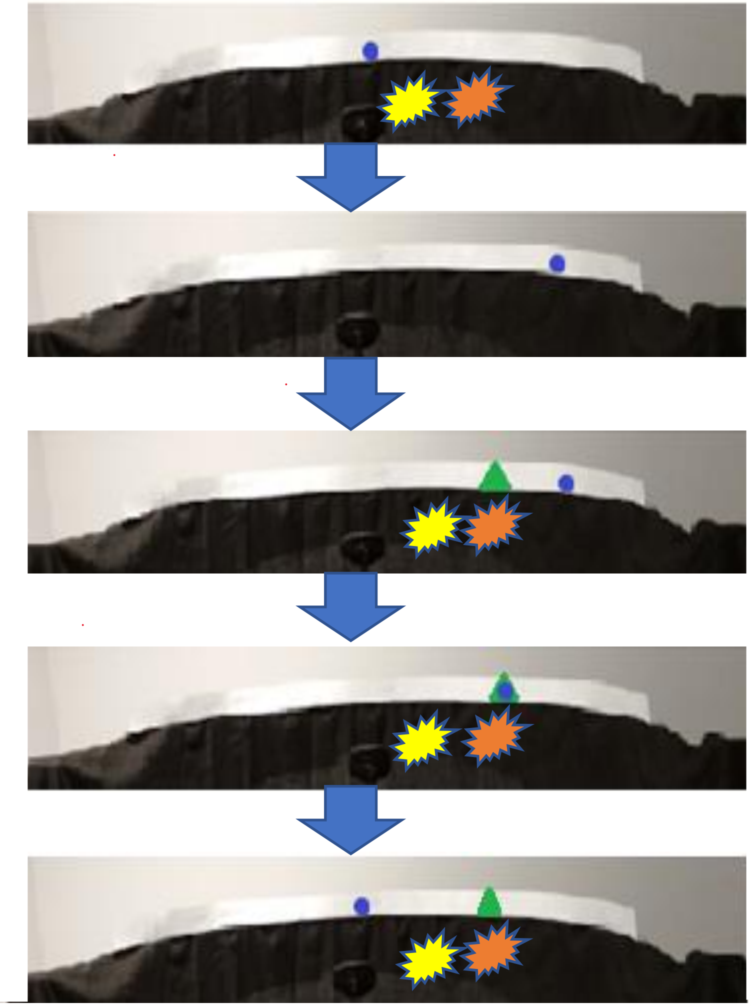


Overall Procedure

Experiment consisted of four 2-3 hr sessions, performed on consecutive days:

Training Procedure

Procedure for 2-component sounds (identical procedure for 4-component sounds):



Initial position & stimulus presentation

- Head-turn to 0°
- Present stimulus once

↓

Response

- Head-turn to perceived target location & press Enter

↓

Visual feedback

- Green triangle shows correct location
- Present stimulus continuously (200 ms gap)

↓

Training with dynamic cues

- Head-turn to visual feedback & press Enter
- Stimulus continues

↓

Return to initial position while stimulus playback continues

- Head-turn to 0° & press Enter
- Stimulus stops

Legend:

- Head orientation
- Visual feedback
- HF sound
- LF sound

Visual feedback aligned with HF component(s) for HF group, with LF component(s) for LF group.

Test procedure identical to steps 1 & 2 of training procedure.

Weight Analysis

Real Environment:

- Regression used to estimate coefficients k_{LF} / k_{HF} relating change in LF/HF component location Δ_{LF} / Δ_{HF} to a change in response R at azimuth α (Q is overall bias):

$$R(\alpha, \Delta_{LF}, \Delta_{HF}) = k_{LF}(\alpha) * \Delta_{LF} + k_{HF}(\alpha) * \Delta_{HF} + Q(\alpha)$$

- Relative weight of HF vs. LF components:

$$w_{HL}(\alpha) = \frac{atan\left(\frac{k_{LF}(\alpha)}{k_{HF}(\alpha)}\right)}{90}$$

such that $w_{HL} = 1$ if only HFs are used, $w_{HL} = 0$ if only LFs are used.

Virtual Environment:

- Similar regression estimating the relative weight of ILD vs. ITD cues w_{LT}:

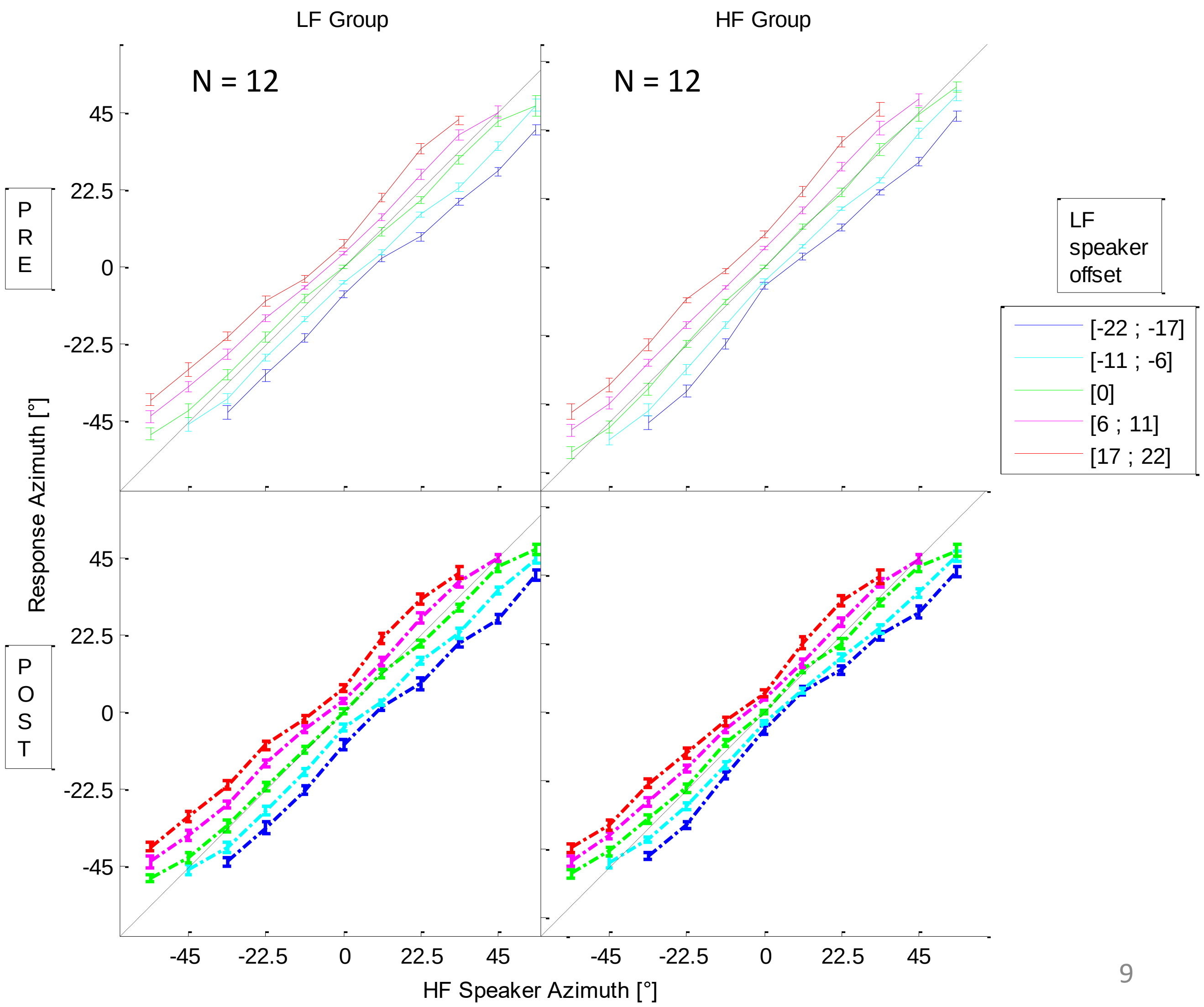
$$R(\alpha, \Delta_{ITD}, \Delta_{ILD}) = k_{ITD}(\alpha) * \Delta_{ITD} + k_{ILD}(\alpha) * \Delta_{ILD} + Q(\alpha)$$

$$w_{LT}(\alpha) = \frac{atan\left(\frac{k_{ILD}(\alpha)}{k_{ITD}(\alpha)}\right)}{90}$$

Raw data in Real Environment

Response azimuth in pretest vs. posttest:
- as a function of position of HF speaker azimuth,
- parametrized by azimuth of LF speaker *re.* HF speaker.

- Groups comparable in pretest
- Posttest:
- in LF group, lines vertically separated → responses closer to position of LF speaker,
- in HF group, lines vertically closer together → responses closer to position of HF speaker.



In both groups, the posttest responses shifted in the direction of visually reinforced frequency component.

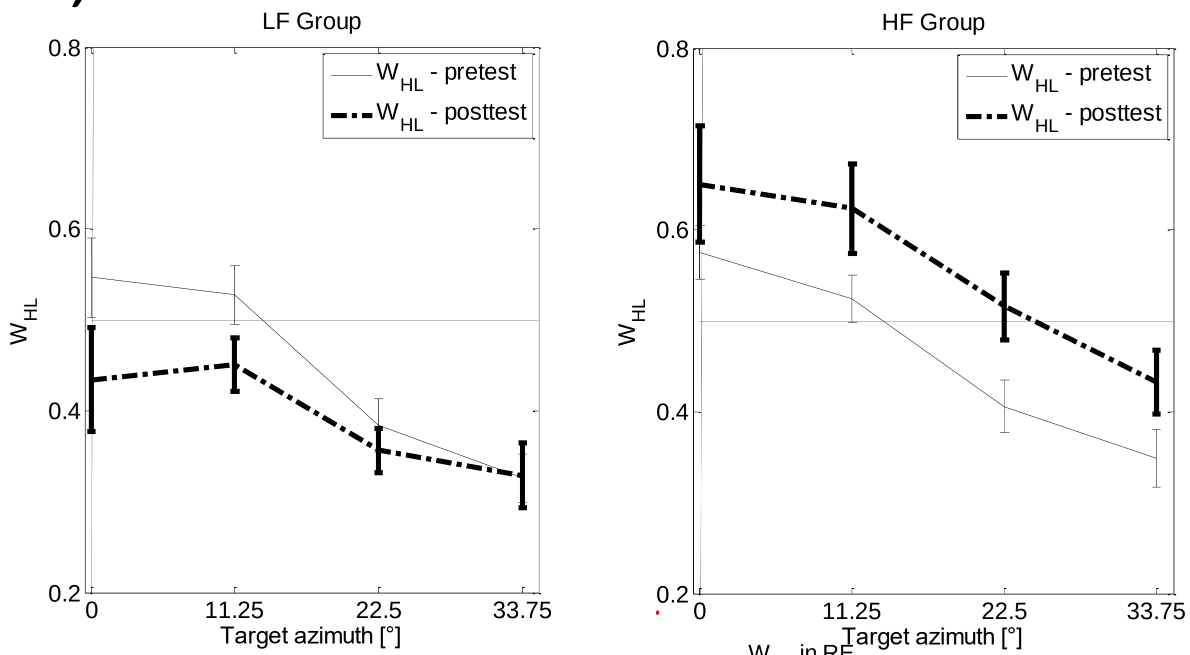
Estimated Weights, Real Environment

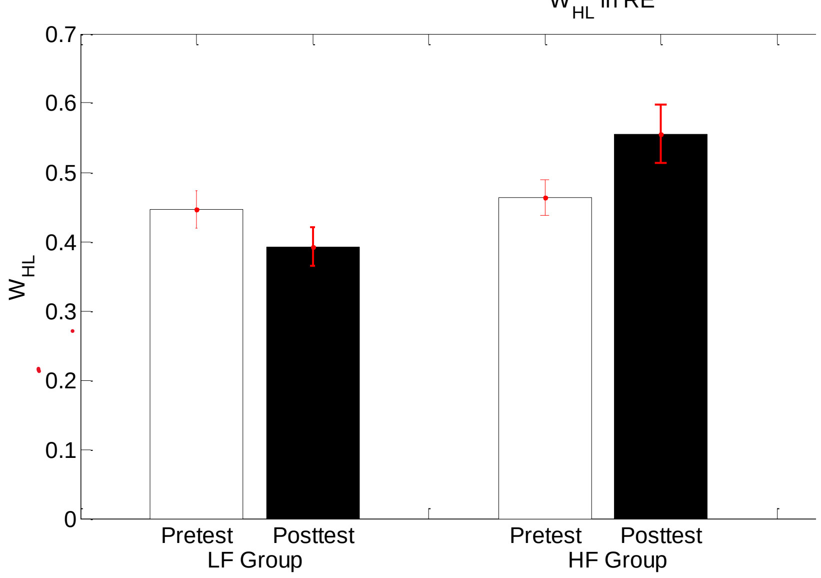
Weights w_{HL} for data collapsed across hemifields:

- significant change from pretest to posttest in both groups (p<0.01),
- HF group increased weight
  - strongly, and at all azimuths,
- LF group decreased weight
  - the decrease was weaker,
  - mainly in the center (n. s.).

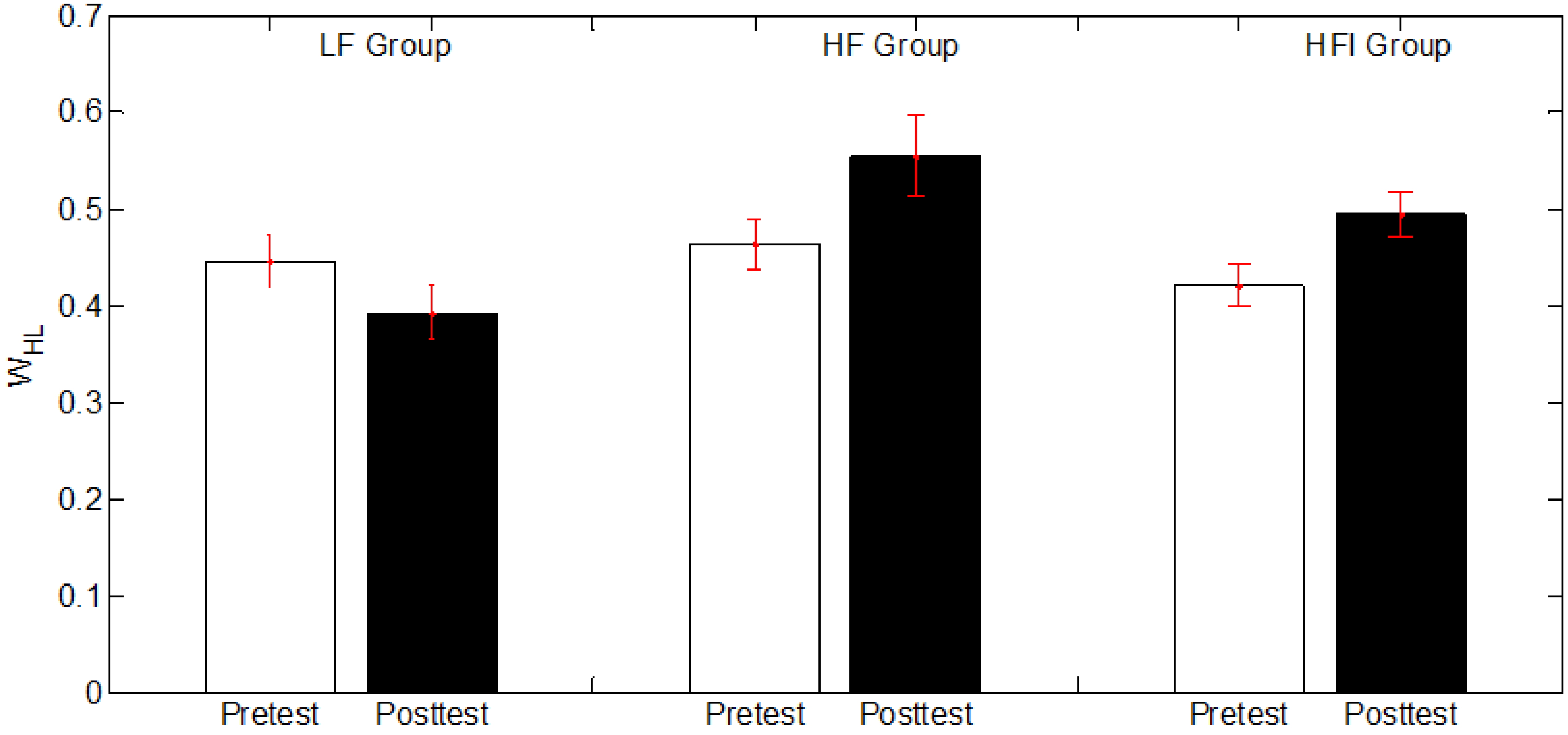
Weights w_{HL} averaged across azimuths:

- **Reweighting stronger and more uniform for the HF group**.





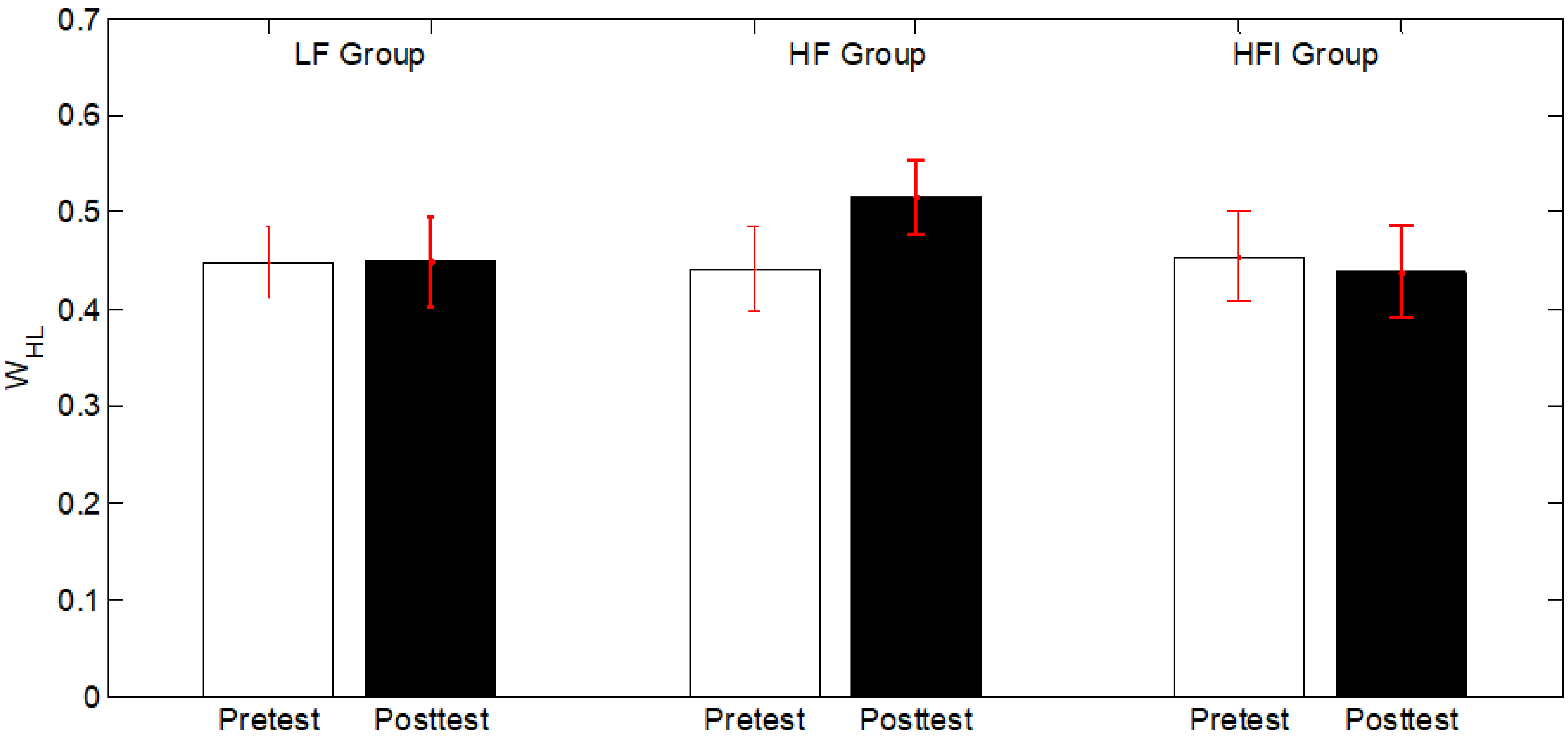
Estimated Weights – “Informed” HF group



- Followup experiment with additional HFI group was performed.
- HFI group had the same training as HF group, but was informed about composition of stimuli and instructed to indicate the position of sound with higher frequency already prior to pretest.
- No significant difference between HF and HFI group →

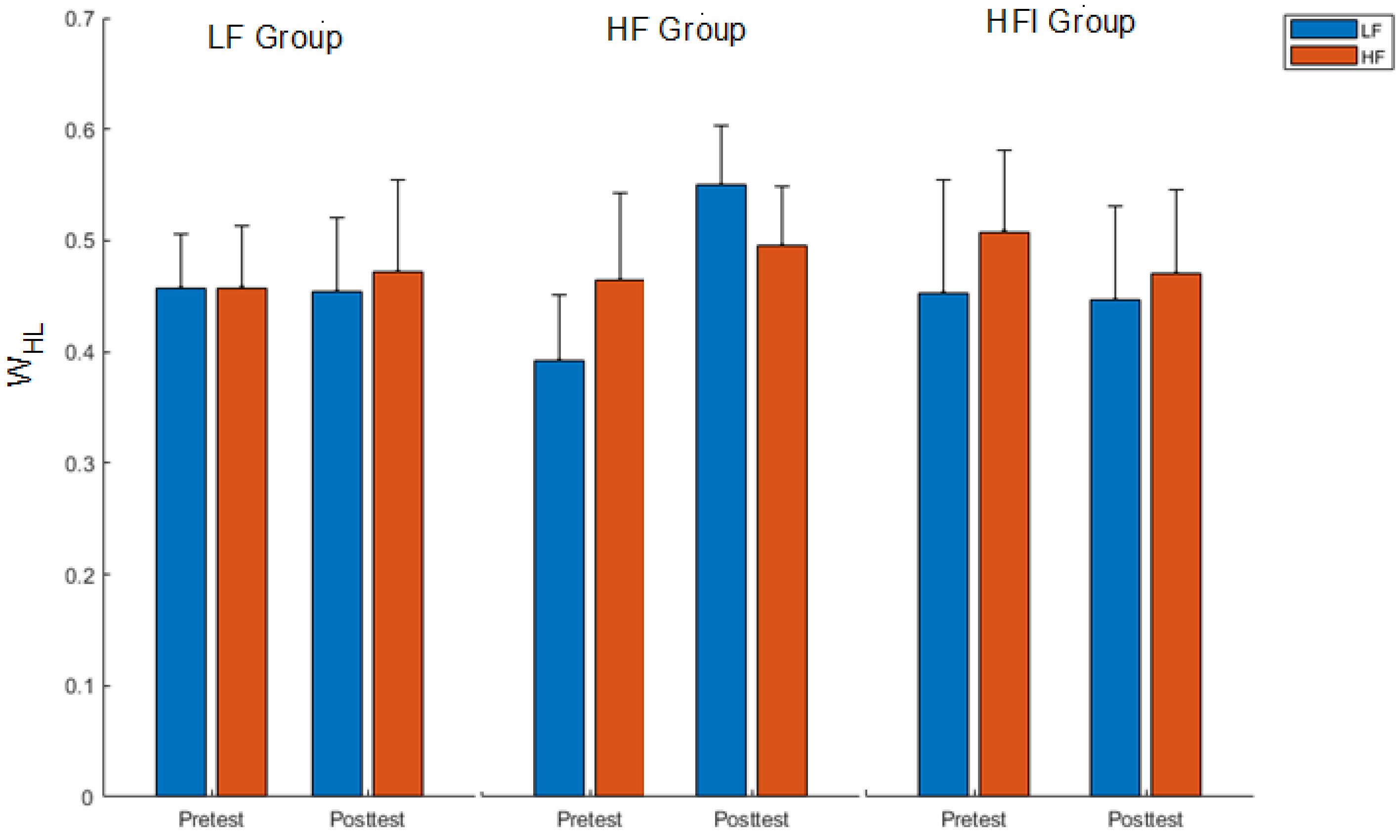
Observed effects not due to change in strategy during training.

Generalization to untrained frequency: 2.8 kHz



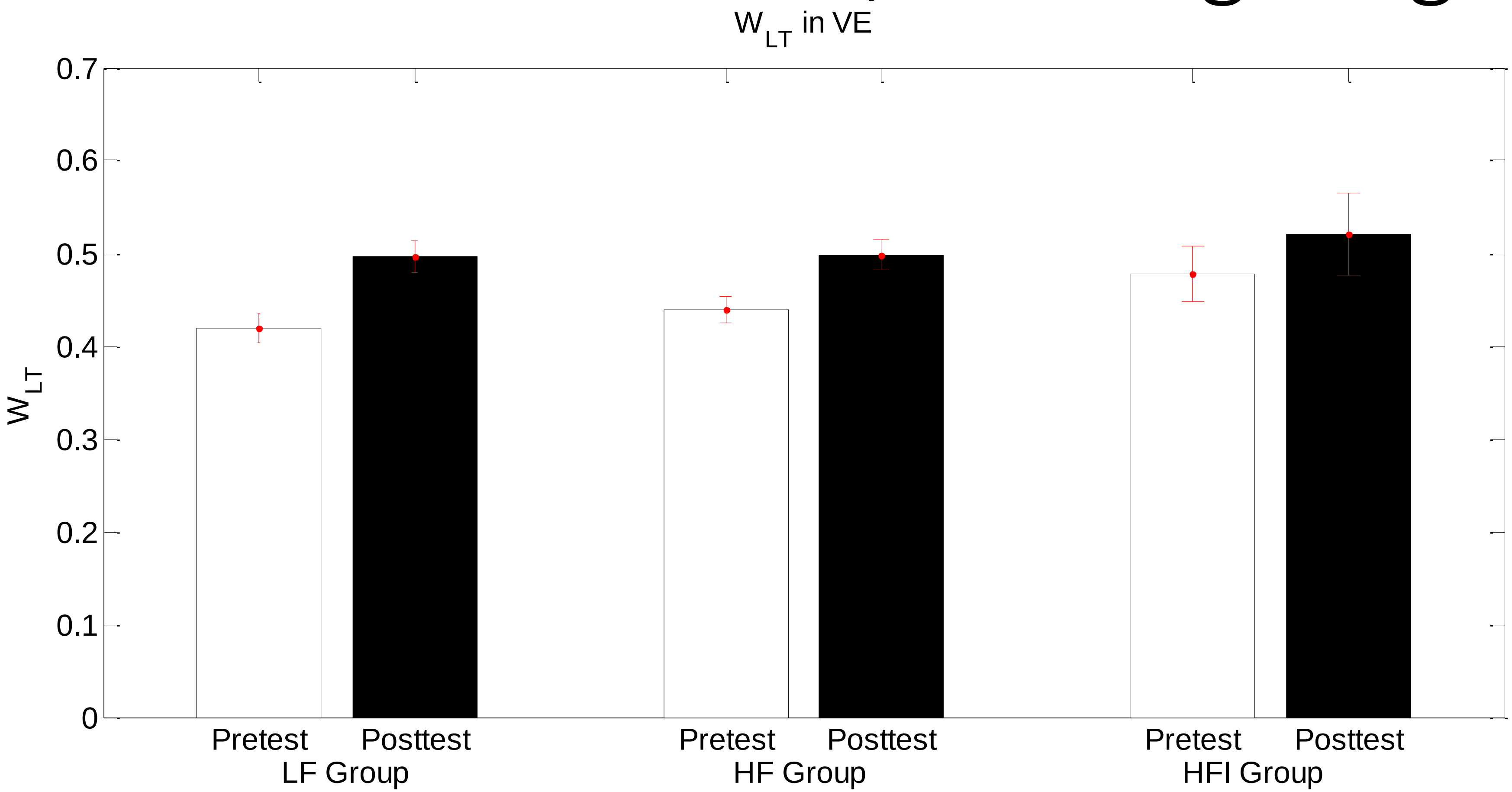
- Significant difference between groups in terms of weight change from pretest to posttest ($p<0.05$).
- Only HF group changed weighting from pretest to posttest significantly ($p<0.01$) → **Generalization of training to new frequencies only partial.**

Generalization to untrained frequency: 2.8 kHz



For HF group:
- most of the increase in weight occurs for the LF stimuli,
- very little for HF stimuli.

Generalization to ITD/ILD weighting in VE



- Significant effect of time for all groups ($p<0.01$).
- No significant difference between groups.
- Increase of ILD weight (increased w_{LT}) at all azimuths for all groups from pretest to posttest.

All groups changed from pretest to posttest, but not in direction consistent with training in case of LF group.

Summary and Discussion I

Reweighting of spectral cues for sound localization**:**

- can be achieved by **visual training in real environment with dynamic cues**,
- both for **increase** in **HF** and **LF component weighting**,
- is asymmetrical, **stronger for HF** training than **LF** training,
- is **independent** of whether subjects are **informed** about the goal of training,
- partially **generalizes** to **untrained** mid **frequencies**, **only** for **HF training** and only when subjects are **not informed** about training goal.

Reweighting asymmetries possibly due to ILD processing, dominant at HF, being more adaptive (Dahmen et al., 2010).

Training people to use the most reliable spectral components can improve localization performance even in normal-hearing listeners, e.g., in environments in which certain components are masked.

Summary and Discussion II

Generalization of **spectral** reweighting to **ITD/ILD reweighting at mid frequencies**:
- **no direct evidence of generalization**, possibly due to testing not at trained frequencies.

Alternative interpretation:
- for **HF (& HFI) training**, increased ILD weighting, as expected,
- for **LF training**, increased ILD weighting, contrary to expectation,
    - partially explainable by no generalization to mid frequencies.

Possible explanations of all groups increasing ILD weight:
- **Procedural training in VE** (no no-training baseline)
  - not likely because no such effect in Klingel et al. (2021).
- **Adaptation to preceding environment** ---------------------------->
  - anechoic VE pretraining preceded VE pretest, but reverberant RE posttest preceded VE posttest,
  - adaptation to reverberation observed for localization (Shinn-Cunningham, 2000) or speech perception (Vlahou et al., 2021).
- **Combination of spectral training for HF/HFI and procedural VE training / adaptation for LF group**
  - similar increase in ILD weight w/o training in Kumpik et al. (2019).