

### Hintergrund

Die realistische Nachbildung auditorischer Szenen über Kopfhörer ist eine große Herausforderung, denn eine präzise Vorne-Hinten- und Oben-Unten-Ortung von Schallquellen ist über Kopfhörer besonders schwierig. Dabei optimiert das Hinzufügen von individuellen Merkmalen (die durch die individuelle Form von Ohren, Kopf und Rumpf verursachten klangfärbenden Reflexionen; Außenohrübertragungsfunktionen (head-related transfer functions, HRTFs; Møller et al., 1995)) die Richtungswahrnehmung von Schallereignissen signifikant (Wenzel et al., 1993; Møller et al., 1996; Middlebrooks, 1999). Das Ausmaß in welchem diese hörspezifischen Merkmale in einer Virtualisierung gebraucht werden, um eine realitätsnahe binaurale Darbietung zu ermöglichen, ist Thema dieses Beitrages. Gerade durch die jüngsten technischen Fortschritte im Bereich der Virtual- und Augmented-Reality-Headsets kombiniert mit einem präzisen Head-Tracking ist diese Thematik im Bereich der Grundlagenforschung als auch in der technischen Umsetzung höchst aktuell.

### Ziel

Um die Notwendigkeit und den Einfluss der Individualisierung von HRTFs in der binauralen virtuellen Akustik zu untersuchen, steht hier die 3D-Audio-Wahrnehmung von virtuellen audiovisuellen Klangszene auf Virtual-Reality-Headsets im Vordergrund. Ziel ist es herauszufinden, ob und inwieweit die Wahrnehmungsqualität von Schallsignalen in virtueller Realität durch den Einsatz von individuellen HRTFs verbessert werden kann. Dabei steht in diesem Beitrag und speziell in der virtuellen Realität zur Darbietung von virtuellen Welten eine Fragestellung besonders im Fokus:

**„Wie lassen sich bewegte Schallquellen in einer virtuellen Umgebung möglichst realitätsnahe nachbilden?“**

### Methode

Mittels eines Fragebogens bestehend aus fünf Wahrnehmungsqualitäten basierend auf dem „Qualitätsinventar zur Schallfeldvirtualisierung“ (Spatial Audio Quality Inventory, SAQI; Lindau et al., 2014) und der Studie von Simon et al. (2016) bewerten 39 Probanden individuelle und nicht-individuelle HRTFs in einer audiovisuellen virtuellen Szene. In dieser wird über die VR-Brille Oculus Rift CV1 eine akustisch zu ortende fliegende dynamische Schallquelle (Drohne) in einer Winterlandschaftsumgebung präsentiert. Die Umschaltung der HRTFs erfolgt hierbei via Touch-Controllern und einem eigens entwickelten Plug-In "SOFA Native Spatializer Plugin for Unity" (Jenny et al., 2018). Dabei berücksichtigen die Filteralgorithmen die Hörer-Interaktion in Echtzeit.

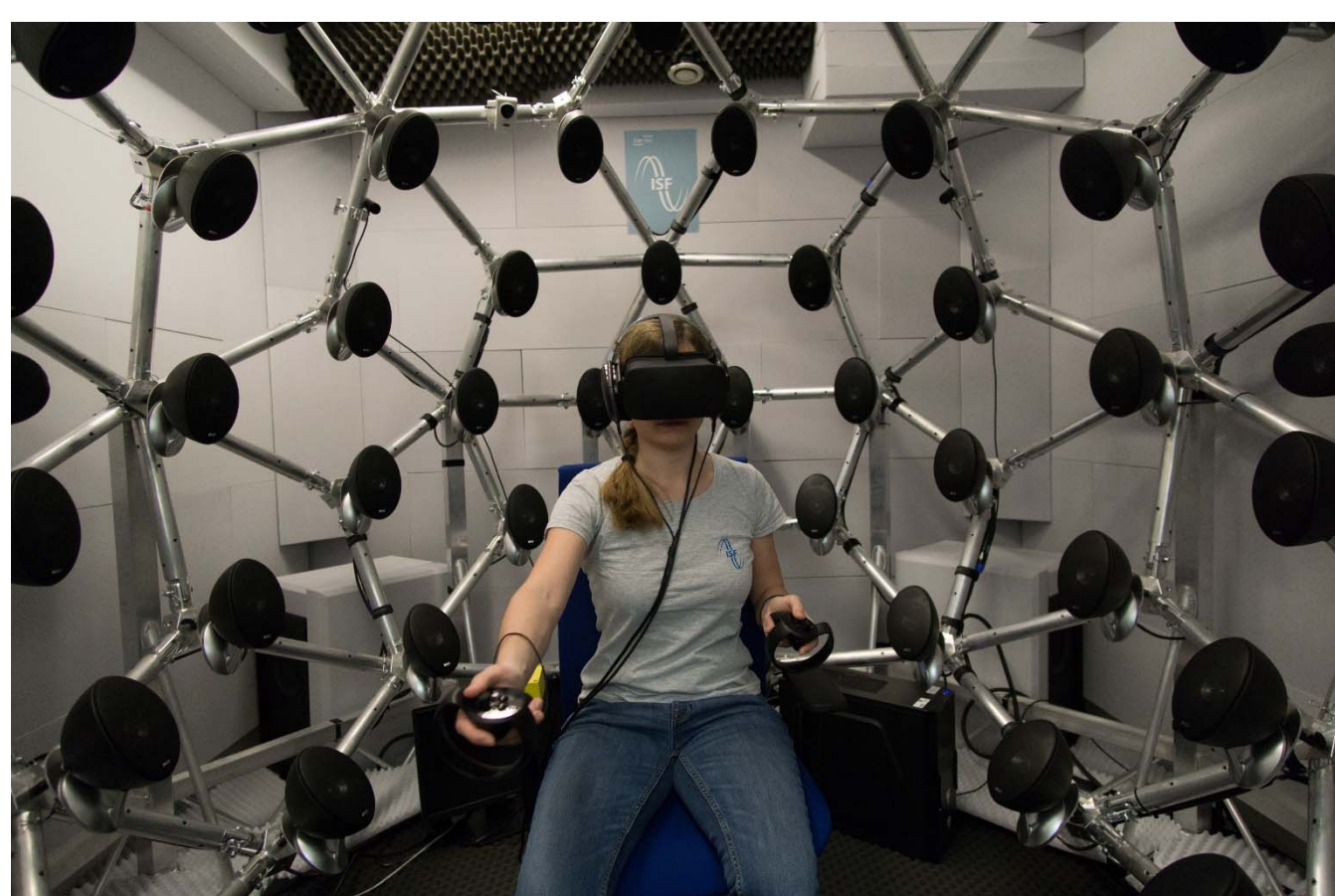


Abb. 1: Versuchsumgebung mit Oculus Rift Virtual Reality Headset.

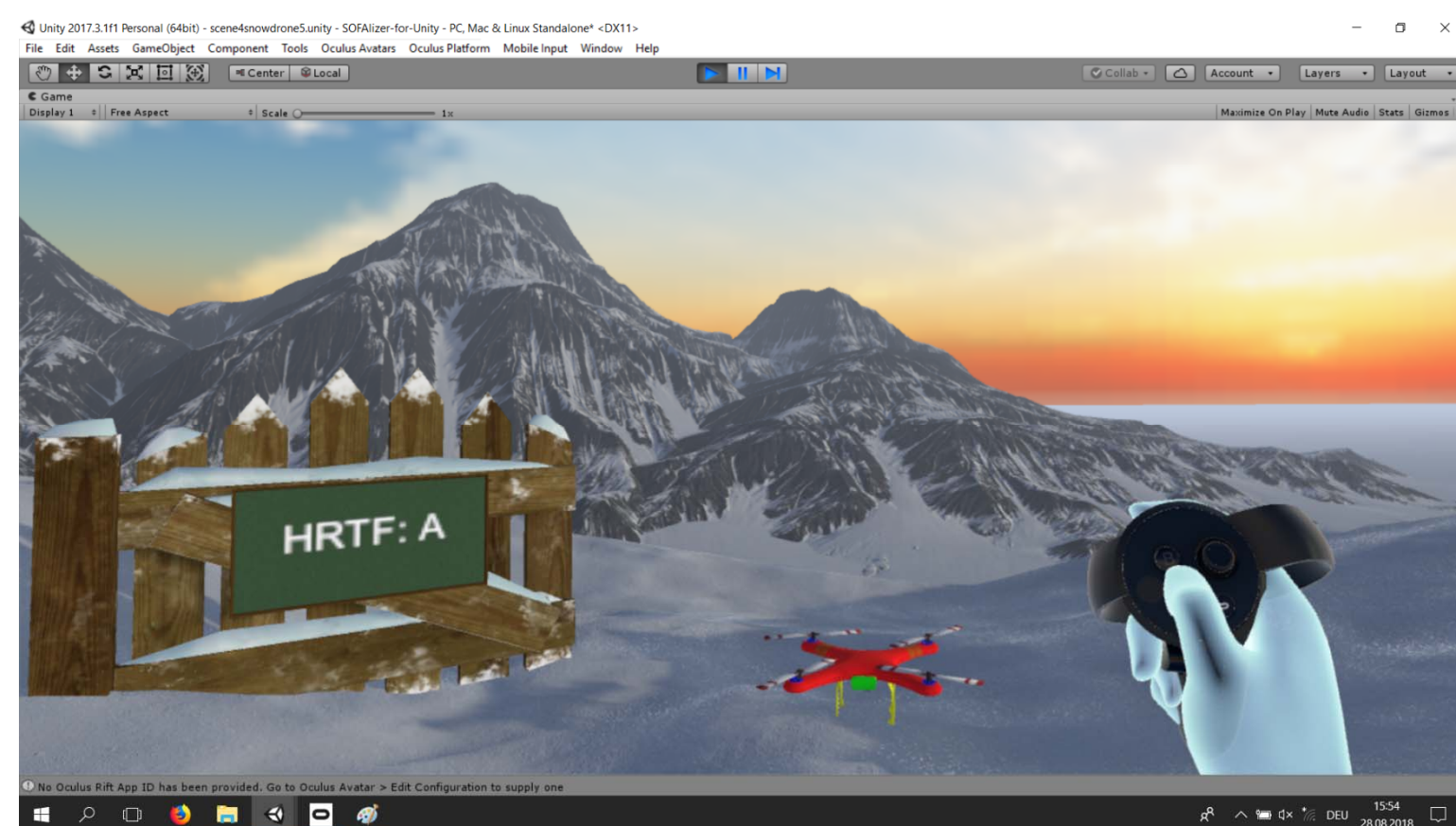


Abb. 2: Winterlandschaftsszene mit Drohne, Tafel für die Anzeige der HRTFs und Touch-Controllern zur Umschaltung.

Der Hörversuch beinhaltet zwei Tests: im ersten Test bewertet der Proband die eigene vs. eine generelle HRTF (MIT KEMAR Kunstkopf; Gardner & Martin, 1995) und im zweiten Test die eigene vs. zwei weitere nicht-individuelle HRTFs aus der HRTF-Datenbank des Instituts für Schallforschung der ÖAW (Majdak et al., 2007). Als Basis zur Auswahl der nicht-individuellen HRTFs dient das Lokalisationsmodell in Sagittalebene von Baumgartner (2014), welches Lokalisationsfehler prädiziert. Für das Experiment werden probanden-spezifisch zwei nicht-individuelle HRTFs mit einer minimalen und maximalen Lokalisationsfehler-Abweichung ausgewählt.

### Fragebogen

Aus einer Auswahl von Attributen aus dem „Qualitätsinventar zur Schallfeldvirtualisierung“ (SAQI) und der Studie von Simon et al. (2016) werden folgende Wahrnehmungsqualitäten zur Bewertung der audiovisuellen Szene abgeleitet:

Wahrnehmungsqualität	Präzisierende Ergänzung	Skalenpole
Lokalisierbarkeit	Bei geringer Lokalisierbarkeit sind räumliche Ausdehnung und Ort einer Schallquelle schlecht abschätzbar bzw. erscheinen diffus. Bei hoher Lokalisierbarkeit erscheint eine Schallquelle dagegen klar umgrenzt. Geringe Lokalisierbarkeit/große Lokalisierbarkeit gehen oft mit großer bzw. geringer wahrgenommener Ausdehnung einer Schallquelle einher. Beispiele: Schallquellen in stark diffusen Schallfeldern sind schlecht lokalisierbar (Lindau et al., 2014).	schwieriger lokalisierbar - einfacher lokalisierbar
Vorne-Hinten-Lage	Beschreibt die wahrgenommene Lage der Schallquelle vor bzw. hinter dem Hörer (Lindau et al., 2014).	vertauscht / nicht vertauscht
Externalisierungsgrad	Beschreibt die Deutlichkeit, mit der eine Schallquelle - unabhängig von ihrer Distanz - innerhalb oder außerhalb des Kopfes wahrgenommen wird. Fachlich oft auch zwischen Phänomenen Im-Kopf-Lokalisation und Außer-Kopf-Lokalisation eingegrenzt. Beispiele: Schlecht/nicht externalisiert = wahrgenommener Schallquellenort bei diotischer Schallpräsentation per Kopfhörer; Gut/stark externalisiert = wahrgenommener Schallquellenort beim Hören einer natürlichen Schallquelle in nachhallbehafteter Umgebung unter Zulassen von Bewegungen des Hörers (Lindau et al., 2014).	internalisierter - externalisierter
Klangfarbe hell-dunkel	Klangeindruck, der durch das Verhältnis hoher zu tiefer Frequenzanteile bestimmt wird (Lindau et al., 2014).	dunkler - heller
Realitätsnähe	Die auditive Quelle scheint von einer realistischen Position zu stammen, die um den Hörer liegt (Simon et al., 2016).	nicht realitätsnah - realitätsnah

### Ergebnisse

Die Versuchsergebnisse zeigen mit Wilcoxon-Vorzeichen-Rang-Test (Test 1) und ANOVA (Test 2) in allen Wahrnehmungsqualitäten Signifikanzen, außer in der Vorne-hinten-Lage zwischen eigener und minimal-abweichendes-nicht-individueller HRTF. Somit werden Klänge unter individuellen HRTFs im Vergleich zu nicht-individuellen HRTFs als einfacher lokalisierbar, externalisierter sowie in der Klangfarbe natürlicher und realitätsnäher bewertet.

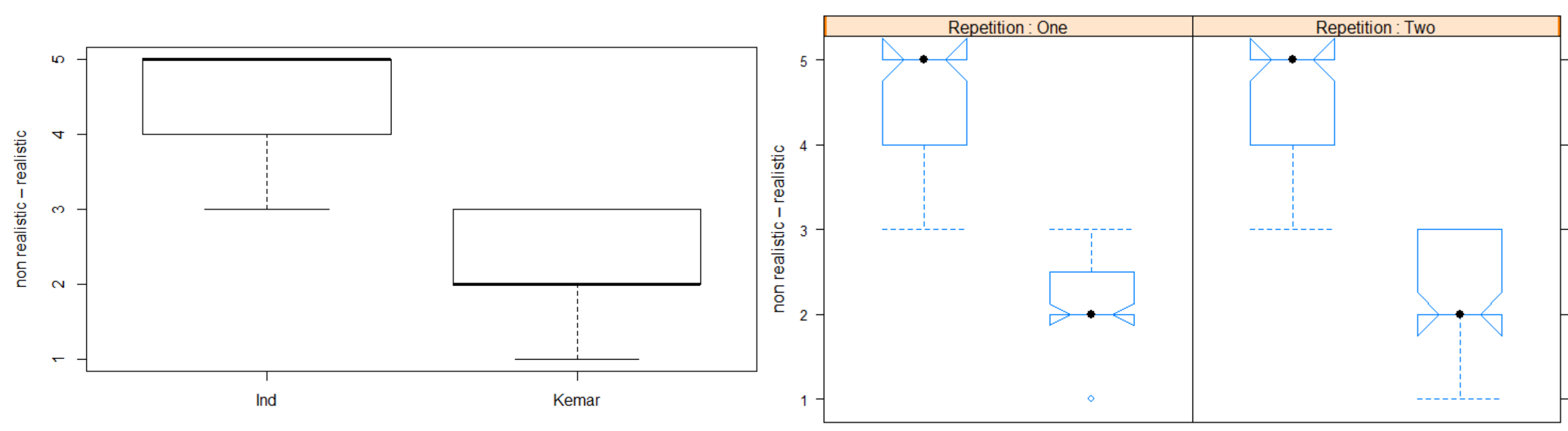


Abb. 3: Test 1: links: Blockdiagramm zur Bewertung des individuellen HRTFs und des KEMAR HRTFs in der Realitätsnähe, rechts: Ergebnis bei Wiederholung der Bewertung zeigt keine Signifikanz.

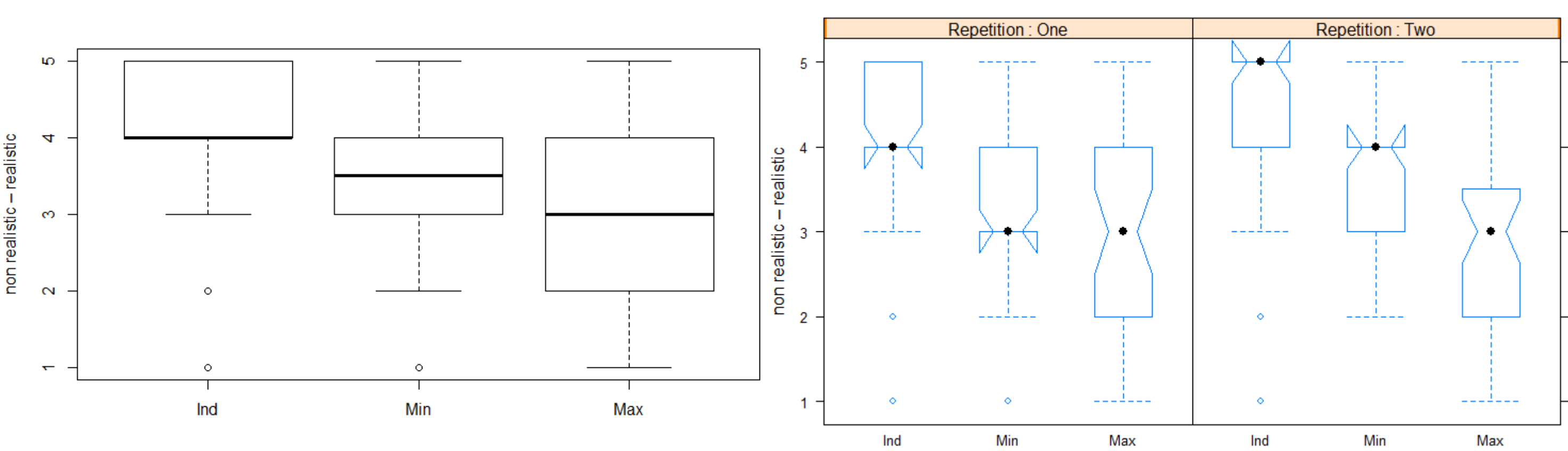


Abb. 4: Test 2: links: Blockdiagramm zur Bewertung des individuellen, des minimal- und maximalabweichenden HRTFs in der Realitätsnähe, rechts: Ergebnis bei Wiederholung der Bewertung zeigt wieder keine Signifikanz.

### Zusammenfassung

Eine möglichst realitätsnahe Nachbildung von Schallquellen erreicht man durch den Einsatz von individuellen HRTFs, die im Bereich folgender Wahrnehmungsqualitäten zu einer Verbesserung führen: Lokalisierbarkeit, Externalisierungsgrad, Klangfarbe und Realitätsnähe.

### Literatur

Baumgartner, R., Majdak, P., Laback, B.: Modeling sound-source localization in sagittal planes for human listeners. In: Journal of the Acoustical Society of America 136 (2014), S. 791–802.  
 Gardner, W. G., Martin, K. D.: HRTF measurements of a KEMAR. In: Journal of the Acoustical Society of America 97/6 (1995), S. 3907–3908.  
 Jenny, C., Majdak, P., Reuter, C.: SOFA Native Spatializer Plugin for Unity - Exchangeable HRTFs in Virtual Reality. In: Proceedings of the 144th Convention of the Audio Engineering Society, Milan, Italy (2018), Convention e-brief 406.  
 Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkmann, F., Weinzierl, S.: A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI). In: Acta Acustica united with Acustica 100/5 (2014), S. 984–994.  
 Majdak, P., Balazs, P., Laback, B.: Multiple exponential sweep method for fast measurement of head-related transfer functions. In: Journal of the Audio Engineering Society 55 (2007), S. 623–637.  
 Middlebrooks, J. C.: Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. In: Journal of the Acoustical Society of America 106/3 (1999), S. 1493–1510.  
 Møller, H., Sørensen, M. F., Hammershøj, D., Jensen, C. B.: Head-related transfer functions of human subjects. In: Journal of the Audio Engineering Society 43/5 (1995), S. 300–321.  
 Møller, H., Sørensen, M. F., Jensen, C. B., Hammershøj, D.: Binaural Technique: Do We Need Individual Recordings?. In: Journal of the Audio Engineering Society 44/6 (1996), S. 451–469.  
 Simon, L. S. R., Zacharov, N., Katz, B. F. G.: Perceptual attributes for the comparison of head-related transfer functions. In: Journal of the Acoustical Society of America 140/5 (2016), S. 3623–3632.  
 Wenzel, E. M., Arrunda, M., Kistler, D. J., Wightman, F. L.: Localization using nonindividualized head-related transfer functions. In: Journal of the Acoustical Society of America 94/1 (1993), S. 111–123.