PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Evaluation of the effect of head-mounted display on individualized head-related transfer functions

Maria CUEVAS-RODRIGUEZ, David Lou ALON, Samuel W. CLAPP, Philip W. ROBINSON, Ravish MEHRA

Facebook Reality Labs, Facebook, 1 Hacker Way, Menlo Park, CA 94025, USA, Email: davidalon@fb.com

ABSTRACT

Head Mounted Displays (HMDs) have become popular for acoustic experiments, such as perceptual comparison of virtual and real sound sources, or Head Related Transfer Function (HRTF) measurement. However, wearing an HMD while listening to real sound sources modifies the sound waves that reaches the listener's ears. In this work, we study the effect of HMDs on individualized HRTFs. We have collected acoustic HRTF measurements, with and without the HMD, from 24 human participants and a manikin head, over 612 directions (including different azimuth and elevation angles). An objective evaluation indicates that, although there are frequency-dependent directional regions where the distortion caused by the HMD is higher, other regions are less affected by the HMD. A MUSHRA test with 15 subjects compares individualized HRTFs and a generic HRTF, measured with and without HMD. The perceptual evaluation reveals a significant effect of the HMD on the measured individualized HRTF, but it also shows that this effect is significantly smaller than the difference between individualized HRTFs and a generic HRTF, as indicated by the objective evaluation. This implies that, although distorted by the presence of the HMD, some of the individualized HRTF characteristics might be preserved when the HMD is worn.

Keywords: Head-related transfer function, 3D-audio

1 INTRODUCTION

The use of Head Mounted Display (HMDs) is becoming increasingly common in virtual and augmented reality applications and experiments that involve auditory stimuli. Some examples are: natural listening to real sound sources while watching video on the HMD; Head-Related Transfer Function (HRTF) measurement processes while wearing an HMD; or immersive audio experiences in mixed reality environments.

The HRTF represents how the listener's head, pinna and torso filter the sound before it reaches his/her ears. The HRTF is an individualized characteristic of the subject and plays an important role in the synthesis of binaural spatial sound [6]. In the case of sound localization of sources in the horizontal plane, the Interaural Time Difference (ITD) provides one of the primary cues that determine the perceived location of the sound [12, 14]. However, for median plane localization, relevant information that allows the user to get elevation cues is derived from the spectral features of the sound [3].

Wearing an HMD will modify the listener's HRTF since it diffracts the acoustic waves that travel from the sound source to the listener's ears. Some studies have been carried out to evaluate how an HMD affects the HRTF. Gupta et al [9] presented an analysis of the effect of different HMDs, including the Oculus Rift™ (Rift), on the HRTF for a KEMAR manikin [5] over the horizontal plane. The objective and subjective analysis revealed noticeable changes in the HRTF spectrum in the 1-16 kHz frequency range. In the case of Rift, the maximum error was observed at $-60°$ azimuth. Another study that analyzes the distortion caused by HMDs on the HRTF was presented by Genoveses et al [8]. This work is focused on headsets used for mixed and augmented reality (more specifically the Microsoft Hololens™ and a Metavision Meta-2™), and presents the effect of these headsets on a manikin's HRTF over 200 azimuth and 6 elevation angles around the head. The analysis shows that non-negligible relevant distortions caused by the HMD are mostly present at the contralateral anterior quadrant (azimuth 270°-360°) and ipsilateral posterior (azimuth 90°-180°), starting from approximately 3 kHz up to 8 kHz. The effect of other HMDs was very recently published for a Manikin head [7].

Regarding virtual spatial audio perception while wearing an HMD, Axel et al [2] have recently analyzed the effect of the HMD on localization accuracy in loudspeaker-based virtual sound environments. Their results show a reduction in the localization accuracy while wearing the HMD; however, once visual information was added to the experiment the localization error induced by the HMD was negligible with respect to the localization accuracy of the presented experiment. These studies show that the effect of the HMD in general, and, more specifically the Rift, is not negligible and, for specific angles, this should be compensated for [8, 9]. However, the studied angles are mostly in the horizontal plane with a limited number of elevation angles (between 1 to 6) and with a limited number of subjects. Moreover, in order to gain an insight into how substantial is the effect of the HMD on an individualized HRTF, it is valuable to evaluate this effect with respect to the difference between an individualized HRTF and the more commonly used generic (e.g. manikin head's) HRTF. Both generic and individualized HRTFs have to be evaluated using the same metrics and with the same system. Such a comparison is missing in current literature.

In this paper we study the effect of the Rift HMD on acoustically measured individualized HRTFs for 24 subjects. The study compares the individual HRTF (without the effect of the Rift) with the same HRTF with the addition of the Rift, and also these two individual HRTFs are compared with a generic HRTF. The 24 human subjects and one KEMAR manikin HRTFs were measured in an acoustically treated chamber, over 612 different directions, (36 azimuth angles $\times$17 elevations angles), without and with the Rift. The effect of the HMD is objectively evaluated based on the Spectral Difference Error (SDE) and on discrepancies in the ITD. Analysis of the SDE shows that the distortion of the HRTF magnitude is both frequency and directional dependent, and bigger disturbances are introduced by the headset at the contralateral directions. Analysis of the ITD reveals that ITD errors are usually larger around the front side of the head. Finally, a MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) test [1] is used to perceptually evaluate the effect of the Rift on the measured HRTF based on the assessment of 15 subjects. In this test, subjects were asked to evaluate four different HRTFs: the subject's individualized HRTFs, without and with the Rift, and generic HRTFs, also without and with the Rift. The results validate the objective evaluation and show that the effect of the Rift is perceptible and is significantly smaller than the difference between individualized and generic HRTFs.

2 HRTF MEASUREMENT WITH AND WITHOUT RIFT

HRTFs were measured using a rotating arc-shaped array of loudspeakers that is housed in an acoustically treated chamber (anechoic down to a frequency of approximately 500 Hz). To conduct the HRTF measurements, the subjects were positioned in the center of the array in a seated position. Wall-mounted lasers were used to accurately position the subjects at the center of the apparatus, with the center of the head taken to be the mid-point of a line connecting the 2 ears. Knowles FG-23329 microphones were placed at the entrance to the ear canal and embedded in foam earplugs. A chinrest mounted on the base plate of the apparatus was used to help the subject remain still for the duration of the measurements. The entire platform can be electronically raised and lowered to make sure that the subjects' head is positioned at the correct height.

During the measurement, the subjects remained still, and a semi-circular arc mounted with 17 Meyer Sound MM-4XP loudspeakers was moved manually around them. The loudspeakers range in elevation angle from $-66°$ to $+85°$ (with spacing of 9°-10° degrees between adjacent loudspeakers) and were sequentially positioned every 10° in azimuth around the full sphere, yielding a total of 612 measurement directions. The distance from the center of the apparatus to the front of each loudspeaker was 1.53 meters. 1-second logarithmic sine sweeps were played from each loudspeaker, ranging in frequency from 200 Hz to 20 kHz at a sampling frequency of 48 kHz, with a level of 94 dB SPL at the center of the array. During the measurement, subjects wore a wig cap with several attached infrared-reflective markers, which can be tracked by 10 OptiTrack™, Prime 17W, cameras mounted on the room's walls. The subjects' head position and orientation were tracked during the measurement, and a warning dialog box appeared in the measurement software (written in Matlab, communicating with OptiTrack's Motive software for motion tracking) if the subject exceeded certain movement thresholds. The measurement sessions described in this study used a three-stage process. Subjects' HRTFs were measured first without wearing the Rift. Then, the Rift was placed on the subject (while they tried to remain as still as possible, keeping their chin on the chinrest), and alignment was checked again using the alignment lasers. HRTFs were then measured while the subject was wearing the Rift. Finally, the Rift was removed, the subject's position was checked with the lasers once again, and HRTFs from all elevation angles at 0° azimuth were measured for validation. After the HRTF measurements were completed, the chair and chinrest were removed, and free-field microphone measurements were taken from all loudspeakers in the array (positioned at 0° azimuth) to both the left- and right-ear microphones, which were aligned to the center of the array with the alignment lasers and supported with a microphone stand. In post-processing, these free-field sweep responses were deconvolved from the measured head-related sweep responses to yield head-related impulse responses (HRIRs) and remove the individual frequency response of the microphones and the loudspeakers. The HRIRs were then windowed in time to 256 samples, corresponding to a length of 5.3 ms at a sampling frequency of 48 kHz.

3 EFFECT OF RIFT ON HRTF - OBJECTIVE EVALUATION

This section presents an objective evaluation of the effect of the Rift on the measured HRTF of 24 human subjects and KEMAR. At the beginning of the section a validation of the measurements is presented. Then, the SDE and ITD error are used as objective metrics to compare the individual HRTFs, with and without the Rift, as well as with the generic HRTF. All the computations and graphs in this study use the coordinate system defined by the AES standard 69-2015 [4], where the azimuth angle $\in (-180°, 180°]$, is the horizontal angle in degrees measured counterclockwise from the listener look direction, the elevation angle $\in [-90°, 90°]$ is the vertical angle in degrees measured from the horizontal plane.

3.1 Repeatability of HRTF measurement

As mentioned in the previous section, subject's movement was limited by using a chinrest and alignment lasers. To quantify the effect of subjects movement and the effect of the Rift on the measured HRTF, the HRTF without the Rift ($HRTF_{WO}$) and the HRTF with the Rift ($HRTF_{Rift}$) are compared with the second measurement of the HRTF without the Rift ($HRTF_{WO2}$), measured after the $HRTF_{Rift}$.

Figure1 shows the three compared HRTF conditions over the front direction for (a) KEMAR and (b) a human subject. In Fig.1(a) it is noticeable that there is a remarkable effect of the Rift on the HRTF, and that the difference between $HRTF_{WO}$ and $HRTF_{WO2}$ is negligible, which implies that the measurement repeatability error is relatively small. As expected with human subjects, in Fig.1(b) it is noticeable that the difference between $HRTF_{WO}$ and $HRTF_{WO2}$ is larger than in the previous case using KEMAR, but it is interesting to note that the effect of movement seems smaller than the effect of wearing Rift.

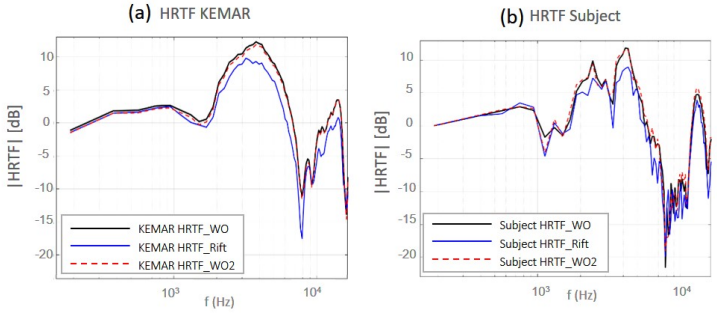

Figure 1. HRTF magnitude over the front direction for (a)KEMAR and (b)human subject for the left ear.

3.2 Effect of the Rift on HRTF magnitude

To determine the differences in the HRTF magnitude for the cases without and with the Rift HMD, a comparison between the spectrum of both signals has been carried out. The SDE between two HRTFs can be computed as (Eq. 5.19 in [15]):

$$SDE_{WO-Rift}(\Omega, f) = |20\log_{10}\frac{|HRTF_{WO}(\Omega, f)|}{|HRTF_{Rift}(\Omega, f)|}| \quad (1)$$

where Ω is the direction angle (azimuth and elevation) and f is the frequency. To study the effect of the Rift across subjects, the $SDE_{WO-Rift}(\Omega, f)$ was computed for each subject and then averaged across subjects and across frequencies, denoted by $\overline{SDE}_{WO-Rift}(\Omega)$. All the computations have been done for both two ears, however, figures in the paper show only the left ear.

Figure 2 shows the $\overline{SDE}_{WO-Rift}(\Omega)$ averaged across a wide frequency bandwidth ($0-16$kHz). The horizontal and vertical axis indicate the azimuth and elevation angles in degrees, respectively. The color map indicates the $\overline{SDE}_{WO-Rift}(\Omega)$ from 0 dB (denoted by the blue color) to 6 dB (denoted by the red color). A higher SDE value implies larger differences between the $HRTF_{WO}$ and the $HRTF_{Rift}$. Figure 2(a) shows the average across all human subjects and Fig.2(b) shows the SDE for the KEMAR manikin. In these graphs larger errors can be observed in contralateral directions. Also, comparing the two graphs on the contralateral side for low elevation angles, larger errors are found for human subjects (Fig.2(a)) than for the KEMAR (Fig.2(b)), which might be partly explained by subjects' movement.

In order to analyze the $\overline{SDE}_{WO-Rift}(\Omega)$ across different azimuth and elevation angles, while taking into account that the SDE is also frequency dependent, the averaged SDE was calculated over three different frequency bands. Figure 3 shows the averaged SDE across three frequency bands: low frequencies (lower than 1 kHz), mid frequencies (between 1 and 5 kHz) and high frequencies (higher than 5 kHz). In the low frequency band the HRTF is less affected by the presence of the Rift, due to the fact that the wavelengths are large relative to the Rift form factor. The effect of the Rift is larger at mid frequencies, and most dominant at high frequencies.

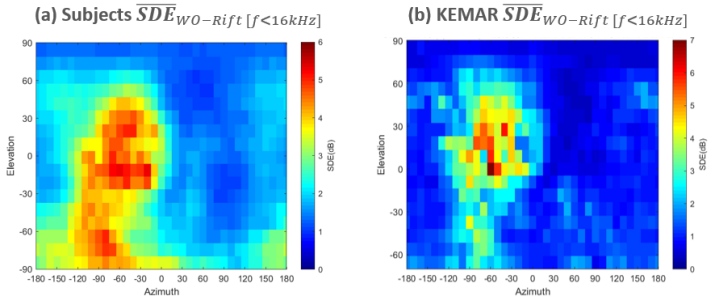

Figure 2. $\overline{SDE}_{WO-Rift}(\Omega)$ averaged across frequency for (a) all human subjects and (b) KEMAR manikin.

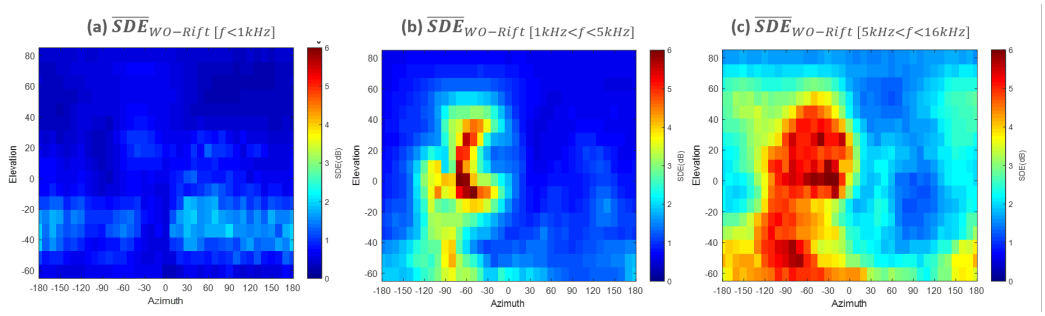

Figure 3. $\overline{SDE}_{WO-Rift}(\Omega)$ averaged across all subjects and KEMAR for three different frequency bands (a) low frequency, (b) mid frequency and (c) high frequency.

For mid and high frequencies, the differences between both sides of the listener's head is easily appreciable. At the ipsilateral ear (positive azimuth angles range for left HRTF) the errors are lower than at the contralateral ear. This effect is caused because the sound wave will be obstructed by the headset to get to the contralateral ear, which is not the case with the ipsilateral ear, where the direct path is not affected. A symmetrical effect also occurs with the right-HRTF.

Different factors, such as the effect of the Rift's presence, the effect of subject movements, and the measurement system repeatabilily contribute to the SDE values that are presented in Fig. 3. In order to evaluate the part of the effect due to the Rift in this overall error, $\overline{SDE}_{WO-Rift}$ is compared to $\overline{SDE}_{WO-WO2}$.

Figure 4 shows the *SDE* averaged across all directions for a specific subject. The vertical axis shows the SDE value and the horizontal axis the frequency value. Even taking into account the errors introduced by the movement of the subjects, the graph shows that the $\overline{SDE}_{WO-Rift}$ is larger than the error caused by the subject movements ($\overline{SDE}_{WO-WO2}$). In addition, the difference between the Individualized-$HRTF_{WO}$ and the Generic-$HRTF_{WO}$ suggests that this difference is larger than the effect caused by the Rift. This implies that, although the values of $\overline{SDE}_{WO-Rift}$ are relatively large, it still preserves some of the individualized characteristics of the HRTF. Similar results were obtained for other subjects as well.

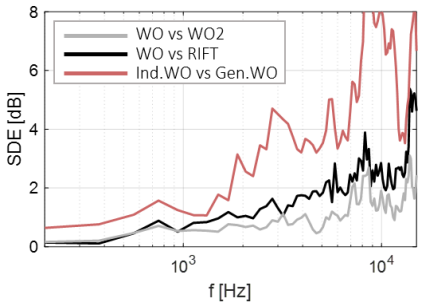

Figure 4. SDE averaged across all direction for one subject. SDE was calculated for three pairs of HRTFs

3.3 Effect of Rift on ITD

ITD was calculated using the method presented by Katz [11], where it is estimated by determining the time when the result of the correlation between the right and the left HRIRs reaches its maximum. For each subject,

the ITD error was calculated as the absolute value of the difference between the ITD of the HRTF without the Rift and with the Rift for each direction. Then, to get an estimation of the effect across subjects, the average was estimated as shown in Equation 2, where N is the number of subjects:

$$\overline{ITD}_{error}(\Omega) = \frac{1}{N}\sum_{i}^{N} |ITD_{WO_i}(\Omega) - ITD_{Rift_i}(\Omega)|. \quad (2)$$

Figure 5 shows the average ITD error for different azimuth angles (horizontal axis) and different elevations (vertical axis), where the color map shows the $\overline{ITD}_{error}(\Omega)$ values in μs. The figure shows how the Rift

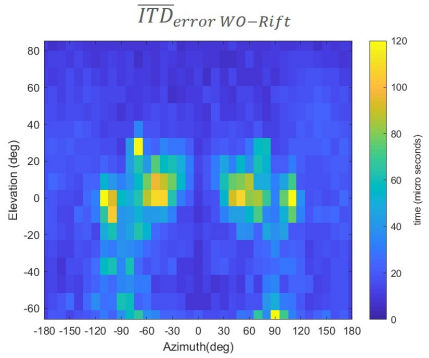

Figure 5. $\overline{ITD}_{error}$ averaged for all subjects and KEMAR

affects the ITD and how the distortion is larger for frontal directions, where $\overline{ITD}_{error}(\Omega)$ is above 50 μs. More specifically, this error can be found in two regions, one on the front-left side of the listener's head for angles with azimuth between 30° and 80° and elevation between 0° and 30°, and the other on the symmetrical one on the right-front. These regions correlate with the headset size and shape. There is also a non-negligible error for angles at $-100°$ and $-110°$ azimuth and 0° and $-10°$ elevation. In addition, a symmetrical effect is evident on the other side of the listener's head. These discrepancies between the ITD_{WO} and ITD_{Rift} can be explained, by shadowing, diffraction and by reflections caused by the presence of the Rift.

4 PERCEPTUAL EVALUATION

To investigate the effect of the Rift on the perception of sound by a listener, a subjective evaluation was performed. A MUSHRA test, defined in ITU-R BS.1534-3 recommendations for advanced sound systems [1], was carried out, where participants evaluated their individualized HRTF and the generic HRTF, both without and with the Rift, denoted as: Individualized-$HRTF_{WO}$, Individualized-$HRTF_{Rift}$, Generic-$HRTF_{WO}$, and Generic-$HRTF_{Rift}$. The Individualized-$HRTF_{WO}$ was chosen to be the reference (labeled and hidden reference) as it is considered to produce the highest perceived quality. An anchor was not used in this study. The main reason for this was to avoid the consequences of biasing or compressing the differences between test signals due to an inappropriate anchor. The MUSHRA test was chosen because it allows the subjects to compare all tested conditions on the same trial, and provide feedback regarding their perceptible similarities and differences. Moreover, this type of perceptual test is suitable for evaluation of intermediate audio quality [1].

4.1 Material and methods

The MUSHRA test recommendation specifies that the experiment population must be composed of experienced listeners, due to their ability to detect subtle artifacts in different sounds. Hence, 15 experienced participants (13 males and 2 females) were chosen from the pool of 24 subjects whose HRTF was measured and analyzed in the previous sections. Two subjects were excluded in a post-screening due to frequently rate of the hidden reference as though it were significantly impaired [1]. The GUI used during the test was developed in Matlab and the audio was played back via an RME Fireface and DT 990 PRO-250 OHM headphones (headphones compensation filter was not applied).

4.1.1 Tested Conditions

The tested conditions in this perceptual evaluation include all combinations of three independent variables: HRTF type, stimulus (i.e. audio material), and source direction. For each condition the sound quality was evaluated based on two attributes (dependent variables).

The 4 tested HRTF types were: Individualized-$HRTF_{WO}$, Individualized-$HRTF_{Rift}$, Generic-$HRTF_{WO}$, and Generic-$HRTF_{Rift}$. All were measured with the same system, described in section 2.

In order to evaluate signals with different spectral content, 3 different stimuli were tested: female speech recording [10]; pink noise - a sequence of 4 broadband noise bursts with a length of 750 ms with 500 ms of silence between them, each noise pulse was faded in and out with a 50 ms raised cosine window; and Guitar recording [10].

As the effect of the Rift on the HRTF was found to be direction dependent, 3 different directions were tested: [*azimuth*,*elevation*]=[60°, 0°], [60°, 30°], and [−15°, 45°]; these directions where chosen considering the directions with large and medium errors obtained in the objective evaluation (see Fig. 2).

Following the MUSHRA test recommendations[1], two attributes were evaluated: (1) *Timbre quality*, which is used to judge all detected differences in timbral impression, e.g. brightness, tone colour, coloration, clarity, hardness, equalization, or richness between the reference and the evaluated condition; (2) *Localization quality*, which is used to judge whether the perceived location of the virtual sound source is the same as the one of the reference.

4.1.2 Procedure

The MUSHRA test procedure consists of two sessions: a training session and the assessment session.

At the beginning of the test subjects perform a training session. In this session, all combinations of HRTF type and stimulus that are experienced during the assessment session are exposed to the subjects. Subjects can spend as much time as they wish in this part, until they feel confident with the signals. In this session, the four HRTF type conditions for the three stimuli are presented, for direction [60°, 30°]. This direction was chosen because, according to the objective evaluation, it is the one that shows the largest differences between the conditions. Subjects were asked to modify the volume of the signals at the beginning of the test at their convenience.

The assessment session is divided into two parts, one part for each attribute, presented sequentially. During the first part, the *timbre quality* attribute is evaluated and in the second part, the *localization quality* attribute is evaluated. In each MUSHRA-screen the differences between the four HRTF type conditions were evaluated, for a single pair of stimulus and direction, ordered randomly. In total 18 MUSHRA-screens were evaluated by each subject (2 different attributes, 3 different stimuli and 3 different source directions).

During each MUSHRA-screen, subjects are presented with the labeled reference and four HRTF type conditions: three test samples, and a hidden version of the reference, all of them presented in random order. The labeled and hidden reference corresponds to the Individualized-$HRTF_{WO}$, while the other three correspond to the Individualized-$HRTF_{Rift}$, Generic-$HRTF_{WO}$, and Generic-$HRTF_{Rift}$. Each subject is asked to provide a score between 0 and 100 for the four conditions, according to how similar each condition is to the reference. When assessing the conditions, subjects can switch instantaneously between conditions and the labeled reference while listening.

To allow the subjects to become familiar with the test setup, a practice MUSHRA-screen test is carried out at the beginning of each part, where the same format as the actual assessment part is presented. In this case, pink noise placed at 60° azimuth and 30° elevation is used as the stimulus, since this signal aids the assessment of small impairments and helps to find the hidden reference. Subjects are not able to pass this practice MUSHRA-screen until they identify the hidden reference by ranking it with a score of 100.

4.2 Results

In total, 36 conditions were statistically tested in a three factorial repeated measure multivariate ANOVA (analysis of variance) design, with factors: HRTF type (Individualized-$HRTF_{WO}$, Individualized-$HRTF_{Rift}$, Generic-$HRTF_{WO}$, and Generic-$HRTF_{Rift}$); stimulus (Female speech, Pink Noise, Guitar); and direction ([60°, 0°], [60°, 30°], and [−15°, 45°]). These conditions were evaluated using two dependent variables, which are the tested attributes (*timbre quality* and *localization quality*).

Figure 6 shows the listening test results for all tested conditions, for all subjects and for the two tested attributes by means of box plot. Ratings reveal that, in most cases, subjects rate the Individualized-$HRTF_{WO}$ with 100 (as expected for the hidden reference). The median rate for Individualized-$HRTF_{Rift}$ is lower than the reference Individualized-$HRTF_{WO}$ for all tested conditions, which implies a detrimental effect of the Rift presence on the individualized HRTF. In addition, it is interesting to note that the ratings for generic HRTF conditions (both Generic-$HRTF_{WO}$ and Generic-$HRTF_{Rift}$) are consistently lower than the ratings for Individualized-$HRTF_{Rift}$ when comparing the median rating of the same direction-stimulus pair. This result may imply that the Individualized-$HRTF_{Rift}$ is perceived by subjects to be closer to the reference Individualized-$HRTF_{WO}$ than the generic HRTF conditions. It is also noticeable that with the pink noise stimulus the differences in median ratings between Individualized-$HRTF_{WO}$, Individualized-$HRTF_{Rift}$ and the generic HRTF conditions were the greatest.

To analyze the significance of the main effects and interaction of the tested factors on *timbre quality* and *localization quality*, the multivariate repeated measure ANOVA approach was used. The reason for not using the univariate approach with Huynh-Feldt correction was that, for some factors, the Huynh-Feldt $\tilde{\varepsilon}$ was lower than

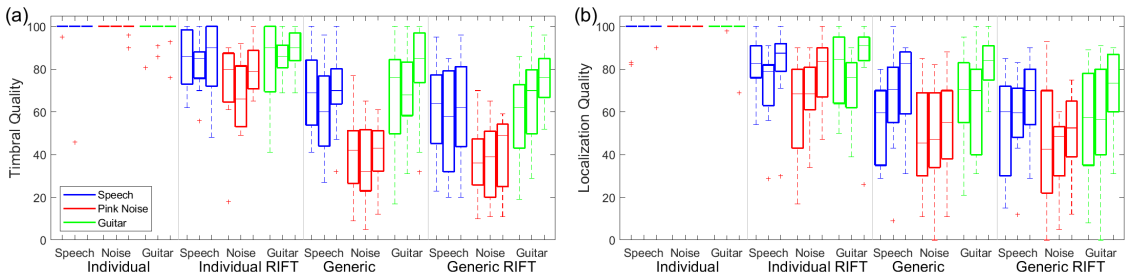

Figure 6. Ratings for all test conditions. Ratings are grouped by HRTF type and stimulus as indicated in the labels. For each group three tested directions are presented: left to right - $[60°, 0°]$, $[60°, 30°]$, $[-15°, 45°]$. The box bounds the interquartile range (IQR) divided by the median, and Tukey-style whiskers extend to a maximum of 1.5IQR beyond the box [13]. Ratings for (a) *Timbre quality* (b) *Localization quality*

the recommended threshold of $\tilde{\varepsilon} > 0.85$ [1]. The multivariate tests show a highly significant ($p < 0.001$, Pillai's trace) main effect of the HRTF type and the stimulus, on both *timbre quality* and *localization quality*. The first-order interaction between stimulus and HRTF type is also highly significant with ($F_{(6,7)} = 19.14$, $p = 0.001$) and ($F_{(6,8)} = 7.57$, $p = 0.006$) for *timbre quality* and *localization quality*, respectively. A significant effect was also found for the direction ($F_{(2,11)} = 6.7$, $p = 0.013$) and ($F_{(2,12)} = 9.4$, $p = 0.003$) for *timbre quality* and *localization quality*, respectively, but no significant effect for the interaction of direction with other factors, except for the interaction between direction and HRTF type on *localization quality* ($F_{(6,8)} = 5.862$, $p = 0.013$). Requirements for normality were met for most of the factor combinations. In some cases with Individualized-$HRTF_{Rift}$ a significant deviation from normality was observed, which was probably due to a large number of high rankings (e.g. 100), which resulted in substantial negative skew of both the data and the residuals (extreme values were skewness < -1.5, kurtosis > 4). Attempting to correct for this skewness by means of $Log10$ correction was not successful (resulted in other conditions that did not meet the normality requirements); the raw data was, therefore, left unchanged. For a detailed analysis of the main effects, pairwise comparisons with Bonferroni correction were performed.

The result of main interest is that the perceived difference between Individualized-$HRTF_{Rift}$ and all other HRTF types was highly significant ($p \leq 0.001$) for both *timbre quality* and *localization quality*. These results may suggest that (1) the effect of the Rift on an individualized HRTF is perceived as a significant difference in source timbre and source direction (with a score lower by more than 17 points than for the Individualized-$HRTF_{WO}$); (2) although significant, the perceived effect of the Rift on an individualized HRTF is smaller compared to the perceived difference between and individualized HRTF and a generic HRTF (Individualized-$HRTF_{Rift}$ received a mean score that was higher than the generic HRTF conditions by more than 14 points). These results may imply that, although the detrimental effect of the Rift on the individualized HRTF is significant, some characteristics of the individualized HRTF are still preserved, which makes the Individualized-$HRTF_{Rift}$ to be perceived closer to the true Individualized-$HRTF_{WO}$ than the Generic-$HRTF_{WO}$.

Regarding the effect of the stimulus (type of audio material), it was found that when using pink noise the perceived quality of the tested conditions was significantly lower ($p < 0.001$) than when using the speech or guitar stimuli. In addition, it was found that the interaction between stimulus and HRTF type was ordinal (i.e. the main effect remains unambiguous) and that the perceived quality of the tested HRTF types were lower with pink noise than with the speech or guitar stimuli. The interaction between stimuli and HRTF type, and the fact that with pink noise scores were lower, can be explained by the fact that stimuli with richer frequency content may assist in detecting differences in both timbre quality and localization quality.

The only direction that yielded significantly higher score than the other directions was $[-15°, 45°]$ for localization quality. For the other two directions no significant difference was observed. This can be explained by the fact that ITD for directions [60°, 0°] and [60°, 30°] was found to be much larger than for $[-15°, 45°]$. It is not possible to generalize this result to other directions, but it gives some indication of the relation between the objective evaluation and the subjective results.

In addition to the MUSHRA test results, an informal questionnaire was filled by the subjects at the end of the test. Subjects rated from 0 to 5 the two following questions: (a) when rating *timbre quality*, how dominant was the difference in low, mid and high frequencies?; and (b) when rating *localization quality*, how dominant was the lateralization, the elevation, and the externalization? Average rating for each question indicates that subjects selected high frequency as dominant to the assessment of *timbre quality* (4.4), followed by mid frequency (3.5) and low frequency with the lowest score (2.3), which fits the objective analysis, where the largest differences were observed at high frequencies. Regarding *localization quality* ratings, subjects found the lateralization to be the most dominant cue (3.9), followed by the elevation (3.3), and finally, the externalization (2.3).

CONCLUSIONS

This paper investigates the effect of an HMD, more specifically the Oculus Rift™, on individualized HRTFs for multiple directions. The HRTFs were acoustically measured on 24 subjects and one KEMAR manikin. The effect was studied, comparing the HRTF for each subject with and without the Rift. From the objective evaluation we can map the directions where the effect of the Rift on the HRTF is larger, taking into account the ITD error and the SDE. A subjective evaluation validates the objective study and reveals that the effect of the Rift on the individualized HRTF is significant in the tested directions. However, according to the subjects' assessments, they find more similarities, in terms of the timbre and localization quality, between their individual HRTF and their individual HRTF while wearing the Rift than between their individual HRTF and the generic HRTF. This may imply that, although the effect of the Rift on an individual HRTF is significant, there are still some characteristics of the individual HRTF that are preserved.

These findings suggest that compensation for the effect of the Rift on an individual HRTF should be applied in experiments that include a subject that is wearing a Rift and a real sound source (such as an external loudspeaker). It would be very interesting to extend the subjective study with a comprehensive localization experiment, to see how the distortion introduced by the Rift affects the perceived location of a source in specific directions.

ACKNOWLEDGEMENTS

We thank our colleagues from Facebook Reality Labs, Alex Gustafson, Chris Hanson, Kamilah Uddin, Owen Brimijoin, Pablo F. Hoffmann, Paul Solis and Vamsi Krishna Ithapu, who provided insight and expertise that greatly assisted the research and improved the manuscript.

References

[1] ITU-R BS. 1534-3: Method for the subjective assessment of intermediate quality level of audio systems, October 2015.

[2] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, 14(3), 2019.

[3] F. Asano, Y. Suzuki, and T. Sone. Role of spectral cues in median plane localization. *The Journal of the Acoustical Society of America*, 88(1):159–168, 1990.

[4] I. AUDIO ENGINEERING SOCIETY. AES standard for file exchange - Spatial acoustic data file format Users. Technical report, Audio Engineering Society Inc., New York, New York, USA, 2015.

[5] M. D. Burkhard and R. M. Sachs. Anthropometric manikin for acoustic research. *The Journal of the Acoustical Society of America*, 58(1):214–222, 1975.

[6] C. I. Cheng and G. H. Wakefield. Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space. In *Audio Engineering Society Convention 107*. Audio Engineering Society, 1999.

[7] A. Genovese and A. Roginska. Hmdir: An hrtf dataset measured on a mannequin wearing xr devices. In *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, Mar 2019.

[8] A. Genovese, G. Zalles, G. Reardon, and A. Roginska. Acoustic perturbations in hrtfs measured on mixed reality headsets. In *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[9] R. Gupta, R. Ranjan, J. He, and G. Woon-Seng. Investigation of effect of vr/ar headgear on head related transfer functions for natural listening. In *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[10] V. Hansen and G. Munch. Making recordings for simulation tests in the archimedes project. *J. Audio Eng. Soc*, 39(10):768–774, 1991.

[11] B. F. Katz and M. Noisternig. A comparative study of interaural time delay estimation methods. *The Journal of the Acoustical Society of America*, 135(6):3530–3540, 2014.

[12] J. Licklider and J. Webster. The discriminability of interaural phase relations in two-component tones. *The Journal of the Acoustical Society of America*, 22(2):191–195, 1950.

[13] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley Pub. Co., 1977.

[14] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–1661, 1992.

[15] B. Xie. *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.