H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, "Acoustic Transparency in Hearables–Perceptual Sound Quality Evaluations"
J. Audio Eng. Soc., vol. 68, no. 7/8, pp. 495–507, (2020 July/August.).
DOI: https://doi.org/10.17743/jaes.2020.0045

Acoustic Transparency in Hearables—Perceptual Sound Quality Evaluations

HENNING SCHEPKER, *AES Associate Member,* **FLORIAN DENK,** *AES Student Member,*
(henning.schepker@uni-oldenburg.de)

BIRGER KOLLMEIER, AND SIMON DOCLO, *AES Associate Member*

Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg, Oldenburg, Germany

In hearing devices, hear-through features that aim to provide the user with acoustic awareness of their surroundings are becoming increasingly popular. In particular, awareness of the user's surroundings can be achieved when the open ear properties can be perceptually restored with the device inserted, typically called acoustic transparency. In this study, we investigate the perceptual sound quality of six commercial consumer hearing devices and two research hearing devices with hear-through features. We conducted two experiments in which normal-hearing participants rated the perceptual sound quality of different audio signals processed by the hearing devices. In Experiment 1, the participants were not provided with an explicit open-ear reference, while in Experiment 2, the open-ear reference was explicitly provided. Results show that most commercial consumer hearing devices are not able to achieve a perceptual sound quality comparable to the open ear. Furthermore, results indicate that a main contributing factor to the overall quality of a hear-through feature is determined by the similarly of the transfer function with the device inserted and the open ear transfer function.

0 INTRODUCTION

In recent years, the availability of consumer hearing device products with so-called hear-through features has steadily increased. Hear-through features allow the user to listen to their environments while at the same time listening, e.g., to music streamed from a smartphone or using mixed-reality applications. When designing a hear-through feature, it is desirable to maintain a natural perception of the sound that closely resembles the perception of the open ear, i.e., is considered to be acoustically transparent [1–6]. Such a hear-through feature can provide the basis for a scalable assistive hearing device by integrating additional hearing support features commonly found in hearing aids, e.g., amplification, dynamic range compression, and noise reduction [7, 8].

Acoustic transparency is achieved when the sound at the aided eardrum, i.e., with the device inserted and the hear-through feature switched on, and the open eardrum, i.e., without the devices inserted, is perceptually equivalent. Typically, a filter is used to modify the signal picked up by the hearing device microphone(s) such that, in superposition with the sound leaking to the eardrum through the (partially) occluded ear canal, the desired characteristics at the eardrum are achieved [1, 8–10]. Since the hearing device output is generally delayed with respect to the leakage component, this superposition may result in comb-filtering effects that degrade the sound quality [11, 12].

Perceptual evaluation of hear-through features in previous studies was often limited to a single research device alone [1, 5, 12, 13] or a comparison with a single commercial device [14]. While some of these studies compared the performance to the open ear recordings [12, 14], others used either simulations [13] or manipulated the playback signal to account for the device being placed in the ear [5]. Most of these studies showed that perceptual transparency compared to the open ear can be achieved to a varying degree depending on the configuration of the hearing device and the stimuli under test.

In a recent study [15], we technically analyzed different commercially available consumer hearing devices and research hearing devices. We found significant differences in the measured response at the eardrum, the processing delay, strength of comb-filtering effects, and binaural cue distortions when the hear-through features were switched on. In order to determine the perceptual relevance of these observed differences, in the present study, we compare the same devices as in [15] with respect to their perceptual sound quality of the hear-through feature. To this end, we perform two listening experiments. In Experiment 1, participants rated the perceived overall sound quality of the different hearing devices without being provided with an explicitly labeled open-ear reference. In Experiment 2, an explicitly labeled open-ear reference was provided, and participants were again asked to rate the perceived overall sound quality, this time compared to the open-ear reference. We hypothesized that in Experiment 1, quality ratings are mainly influenced by preference for the different spectral characteristics or spectral profiles of the considered hearing devices, while in Experiment 2, ratings are mainly influenced by differences between the open ear and the processed sound of the hear-through features of the hearing devices. Thus, we expect that the results from these experiments shed light on the performance of current hear-through features in consumer hearing devices as well as on the impact of providing an explicit reference when comparing hear-through features for acoustic transparency. Furthermore, by comparing the two perceptual experiments, we expect to answer the question of whether perceptual preservation of the open-ear characteristics as measured in Experiment 2 also yields the highest perceived overall sound quality in Experiment 1.

1 METHODS

In Sec. 1.1, we present an overview of the considered hearing devices with hear-through features and the used stimuli in the perceptual sound quality evaluation. We then describe the perceptual sound quality evaluation study in Sec. 1.2.

1.1 Hearing Devices and Stimuli

In the present study, we selected seven commercial hearing devices that were advertised to have hear-through features and were available in the first half of 2019. More specifically, we chose the devices listed in Table 1 and additionally included a prototype based on commodity hardware [16, 17] and a custom prototype, the so-called acoustically transparent earpiece [8, 12]. Hearing devices A–C were advertised as hearing assistive devices, while hearing devices D–G were advertised as wireless earphones with additional functionalities. Note that, in contrast to the accompanying paper [15], we were not able to include the hearing device G, since its sensors detecting the insertion of the device to the ear did not allow for continuous operation in a dummy head's ear. All devices are in-the-ear devices and thus fill considerable parts of the cavum concha. In contrast to the commercially available hearing devices and the commodity hardware-based hearing device, the acoustically transparent earpiece had a semiopen vent, potentially increasing the risk of comb-filtering effects. The hearing devices either had a wired connection between the left and the right ear or are true wireless devices. All commercial hearing devices were controlled by their respective smartphone applications and updated to their latest firmware as of June 2019. Any advanced processing that was not related to the hear-through feature, e.g., noise reduction and directional processing, was disabled, while for some of the hearing devices active noise control was linked with the hear-through and remained switched on, i.e., hearing device C and D.

Fig. 1. Dummy head with inserted acoustically transparent earpiece [12] in the lab with variable acoustics used for stimuli recording. The green absorbing panels can be flipped to make them highly reflective. Note that the floor is highly reflective and some panels on the ceiling and walls were configured to be highly reflective.

The commodity hardware-based hearing device consisted of modified commercial earphones (Roland CS10-EM) connected to a Raspberry Pi und used the open master hearing aid (openMHA) [18, 19] for real-time sound processing. The equalization filter used for the hear-through features was computed using the regularized least-squares procedure presented in [17]. For the acoustically transparent earpiece, we used custom earmolds connected to an RME Fireface UCX soundcard and an Intel NUK personal computer running the openMHA for real-time sound processing [12]. Real-time processing in the acoustically transparent earpiece included a null-steering beamformer-based feedback suppression filter computed according to [20] in combination with an equalization filter for the hear-through feature computed using the regularized least-squares procedure presented in [6]. The acoustically transparent earpiece included two different variants. The first variant (TEP ED) used an equalization filter of the hear-through feature to achieve acoustic transparency which was computed using knowledge of the dummy head eardrum response, while the second variant (TEP IE) used an equalization filter which was computed using the in-ear microphone of the device, i.e., only signals that are accessible within the device are used. Furthermore, we used the open ear as a reference signal and the occluded ear, using the acoustically transparent earpiece switched off, as an anchor signal. For more details on the commodity hardware-based hearing device and the acoustically transparent earpiece, the interested reader is referred to [17] and [12], respectively.

Table 1. Overview on hearing devices used in the experiments. For the first transparent earpiece variant, the equalization filter was computed using knowledge of the eardrum signal of the dummy head (TEP ED), while for the second variant, the equalization filter was computed using knowledge of only the in-ear microphone (TEP IE; see text for additional information).

Identifier	Manufacturer	Model	Venting	Purpose	Feature Name
	Bose	Hearphones	closed	Hearing Support	Neutral
A–C	Nuheara	IQBuds BOOST	closed	Hearing Support	Neutral
	Wear & Hear	BeHear NOW	closed	Hearing Support	Live Music, Neutral
	Bang & Olufsen	BeoPlay E8	closed	Wireless Earphone	Transparency
D–G	Bragi	TheDASHPro	closed	Wireless Earphone	Transparency
	Jabra	Evolve 65t	closed	Wireless Earphone	Hear-through
	Sony	WF-1000X	closed	Wireless Earphone	Environment Normal
H	UOL	Commodity Hardware [16, 17]	closed	Research Device	Amplification off
I	UOL	Acoustically Transparent Earpiece (TEP ED) [12]	semiopen	Research Device	Tranparent, eardrum
J	UOL	Acoustically Transparent Earpiece (TEP IE) [12]	semiopen	Research Device	Transparent, in-ear
OC	UOL	Acoustically Transparent Earpiece (switched off) [12]	semiopen	Research Device	off, anchor
OE	-	-	open ear	-	-

In order to assess the perceptual sound quality of the hearing devices in realistic, but controlled, acoustic conditions, we recorded different audio signals in a lab with variable acoustics using a GRAS 45BB-12 KEMAR Head & Torso with anthropometric pinnae and low-noise ear simulators with the different hearing devices inserted (cf. Fig. 1). All signals were played back at a level that resulted in approximately 82 dB SPL at the open eardrum of the dummy head. An overview of the different acoustic conditions and signals is provided in Table 2. Four different audio signals were recorded in a moderately reverberant setting ($T_{60} \approx 0.45$ s) for three different playback directions. This resulted in a total of 132 recorded stimuli (11 hearing devices $\times$ 4 audio signals $\times$ 3 signal direction). As stimuli (cf. also Fig. 2), we used two speech signals (male and female) taken from [21] and two music signals (an excerpt from a jazz song[1] and an excerpt from a classical piano recording[2]). The loudspeakers were placed at a distance of approximately 2 m from the dummy head wearing the different devices at angles of

[1] J. Redman: Timeless tales for changing times, 1. Summertime

[2] K. Jarret: Bach, Wohltemperiertes Klavier, Book 1, prelude no. 3

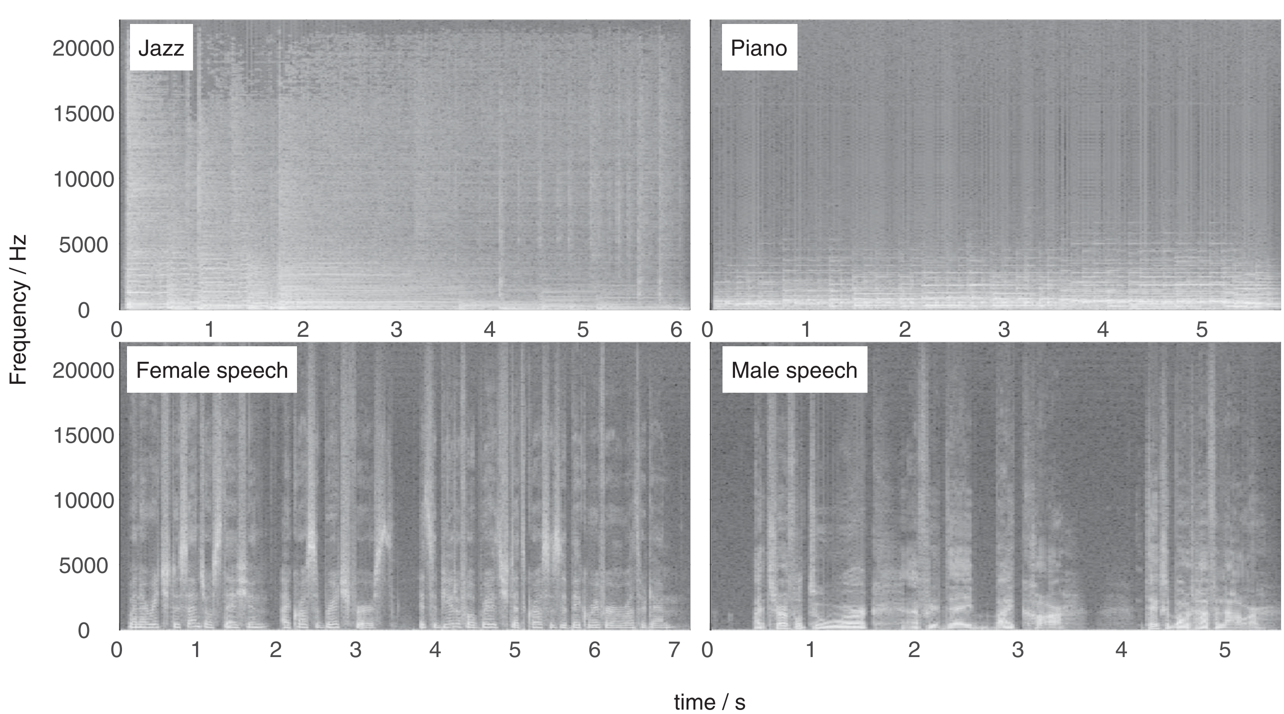


Fig. 2. Spectrograms of the signals used to record the stimuli for the perceptual evaluation.

Table 2. Overview on acoustic conditions and signals.

Reverberation	Signal Direction	Signals
mid $T_{60} \approx 0.45$ s	0° (front)	piano
		jazz
		female speech
		male speech
	90° (left)	piano
		jazz
		female speech
		male speech
	225° (rear right)	piano
		jazz
		female speech
		male speech

0°, 90° (left), and 225° (rear right). To additionally assess the impact of the aided transfer function, i.e., with the hearing devices inserted and hear-through feature switched on, on the perceptual quality, for all acoustic conditions, we recorded the binaural head-related impulse responses. To measure the aided transfer function, we used exponential sweeps [22] with a frequency range of 10 Hz to 22,000 Hz and a duration of 10 s using a playback level that resulted in approximately 72 dB SPL at the open eardrum of the dummy head. Additionally, we measured the transfer function when the hearing devices were switched off as well as for the open ear.

1.2 Perceptual Evaluation

A total of $N = 20$ self-reported normal-hearing participants took part in the study (age: 22.7 ± 2.5). We conducted two separate experiments. In Experiment 1, participants were instructed in writing to rate the perceived overall sound quality of the stimuli recorded with the different hearing devices (cf. Table 1 for the different hearing device settings) and their hear-through feature switched on in a MUltiple Stimulus with Hidden Reference and Anchor (MUSHRA)-like framework [23], similarly as in [24]. This MUSHRA-like framework differs from a conventional MUSHRA framework [25] in that the user interface contains a drag-and-drop feature to rate the stimuli instead of sliders. In order to avoid any influence of the open-ear reference on the sound quality ratings, no reference was explicitly provided to the participants, i.e., the reference button in the MUSHRA user interface was not visible. In Experiment 2, participants were again instructed in writing to rate the perceived overall quality of the same stimuli in the same MUSHRA-like framework. In contrast to Experiment 1, the reference button was now visible in the user interface, and the participants were additional instructed to rate the sound quality compared to this (open-ear) reference. Both experiments were performed twice by each participant to assess the test–retest reliability of the sound quality ratings. Test and retest were performed on separate days, resulting in a total of four sessions (two sessions per experiment) that took about 45 min each. Participants where allowed to take breaks whenever necessary. Each session was preceded by a brief training phase to familiarize the participants with the experiments and with the expected variability of the stimuli using a different set of stimuli than used in the main experiment. Since we expected that Experiment 2 might influence the results of Experiment 1, Experiment 2 was conducted only after Experiment 1 had been performed twice by a participant. The order of conditions, i.e., signal direction and stimulus, were randomized for each participant and experiment. After analyzing the data, three of the participants had to be excluded, since these participants could not reliably identify the hidden reference in the second experiment, resulting in a total of $N = 17$ participants.

All stimuli were digitally stored. Ratings of the stimuli were conducted using a Matlab drag-and-drop graphical user interface [23]. The binaural stimuli were presented over Sennheiser HD650 headphones and amplified to 70 dB SPL using a Tucker Davis HD-7 headphone amplifier connected to a Fireface UCX sound card. Headphones were equalized for a flat magnitude response at the KEMAR eardrum (corresponding to a median human ear) using regularized inversion as implemented in the AKtools toolbox [26]. It should be noted that equalization to the KEMAR eardrum may lead to individual differences compared to equalization to the individual eardrum, especially at high frequencies.

2 RESULT

2.1 Experiment 1

Fig. 3 shows the results of Experiment 1; for each participant, the test and retest results were averaged. Individual panels show the results for the different signal directions. First, consider the condition with a frontal source direction (0°) in the top panel. As can be observed, most participants rated the open ear highest, with ratings generally ranging from good to excellent. The lowest rating is obtained for hearing device F with ratings ranging from bad to poor. In general, the ratings are similar across all different signals; however, the quality of the jazz signal tends to be rated lowest. Similar results can be observed for the remaining two signal directions of 90° and 225°. Statistical analysis of these results was conducted using a three-factor (hearing device, signal, and signal direction) repeated measures analysis of variance (ANOVA) using Greenhouse–Geisser correction for sphericity. ANOVA showed a significant effect of the main factors hearing device ($F(0.56, 9.0) = 60.32$, $p < 0.001$) and signal ($F(0.17, 2.7) = 6.88$, $p = 0.002$) as well as the interaction of hearing device and signal ($F(1.69, 27.0) = 11.1$, $p < 0.001$) and hearing device and direction ($F(1.13, 18.0) = 4.93$, $p < 0.001$). Post hoc analyses were conducted for the factors of hearing device and signal using the Student t test at the Bonferroni correct level of significance. For the factor of hearing device, results of the post hoc analysis are shown in Table 3. For the factor signal, the jazz signal was rated significantly different from the male ($p = 0.0011$) and female speech signal ($p = 0.0068$).

In order to more easily assess the impact of the hearing device on the perceived sound quality ratings, Fig. 4

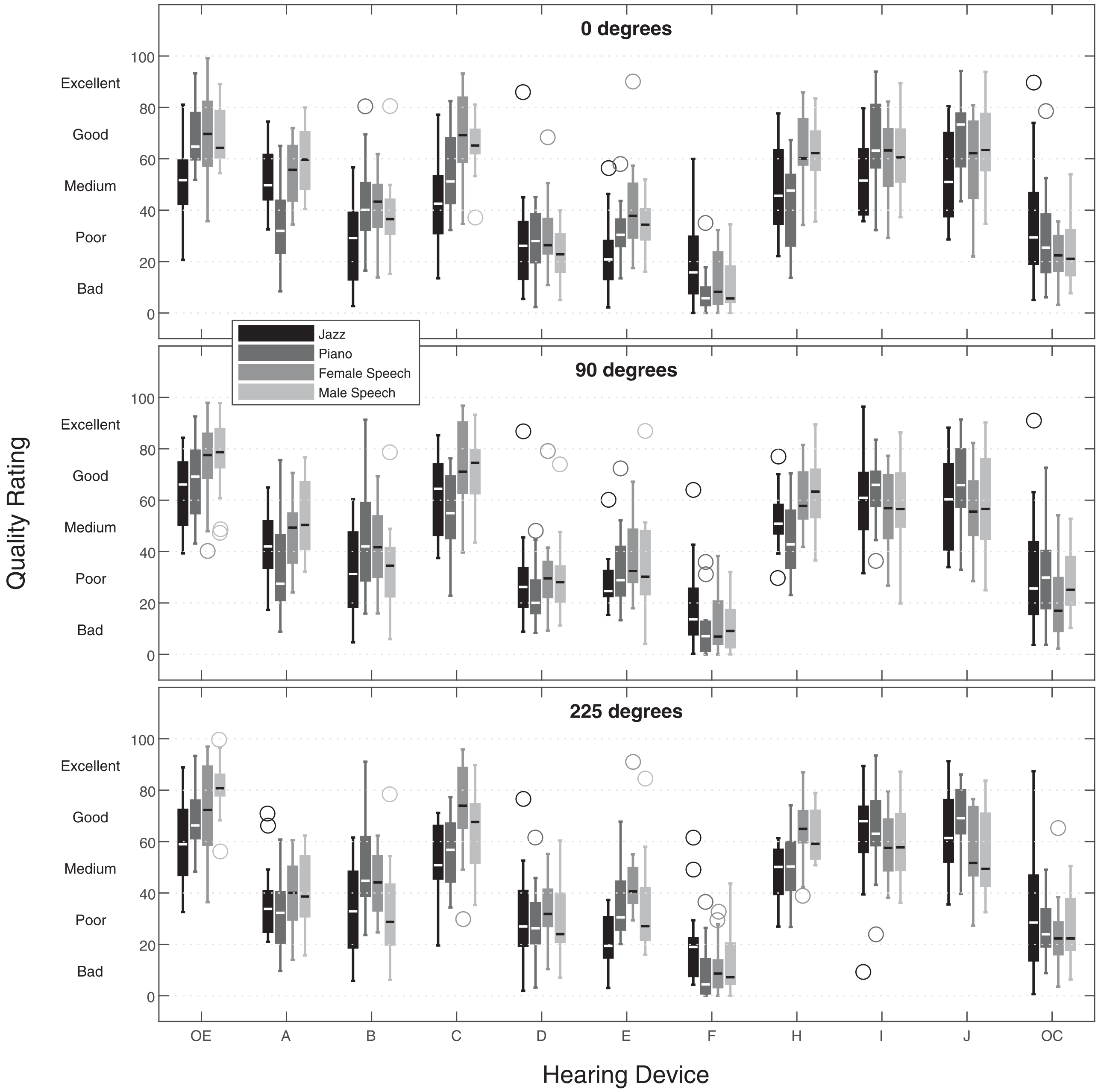


Fig. 3. Sound quality ratings for Experiment 1 without explicit open-ear reference for the different signal directions, hearing devices and audio signals. Boxes show interquartile ranges, points in boxes show the median, whiskers extend to 1.5 times the interquartile range, and circles point show outliers.

shows the aggregated results computed as the median quality rating for each participant across all signals and signal directions. As expected from the statistical analyses, hearing devices C, H, I, and J yield very similar quality ratings. Notably, the quality of devices F was rated lower than the intended anchor signal of the occluded ear (OC). Furthermore, most commercial devices without hearing assistance only achieved poor to medium quality (D–F).

2.2 Experiment 2

Fig. 5 shows the results for Experiment 2; for each participant, the test and retest results were averaged. Individual panels show the results for the different signal directions. First, consider the condition with a frontal source direction (0°) in the top panel. As can be observed, most participants could reliably identify the hidden reference (OE). The highest rating is obtained for hearing device C with ratings generally ranging from good to excellent. The lowest rating is obtained for hearing device F with ratings ranging from bad to poor. In general, the ratings are similar across all different signals; however, as in Experiment 1, the quality of the jazz signal tends to be rated lowest. Similar results can be observed for the remaining two signal directions of 90° and 225°. Statistical analysis of these results was conducted using a three-factor (hearing device, signal, signal direc-

Table 3. Post hoc analysis of the factor hearing device in experiment 1. Asterisks indicate significant differences ($p < 0.05$), while dashed indicate nonsignificant differences.

	OE	A	B	C	D	E	F	H	I	J	OC
OE		*	*	*	*	*	*	*	*	-	*
A	*		-	*	*	-	*	*	*	*	*
B	*	-		*	-	-	*	*	*	*	-
C	*	*	*		*	*	*	-	-	-	*
D	*	*	-	*		-	*	*	*	*	-
E	*	-	-	*	-		*	*	*	*	-
F	*	*	*	*	*	*		*	*	*	*
H	*	*	*	-	*	*	*		-	-	*
I	*	*	*	-	*	*	*	-		-	*
J	-	*	*	-	*	*	*	-	-		*
OC	*	*	-	*	-	-	*	*	*	*	

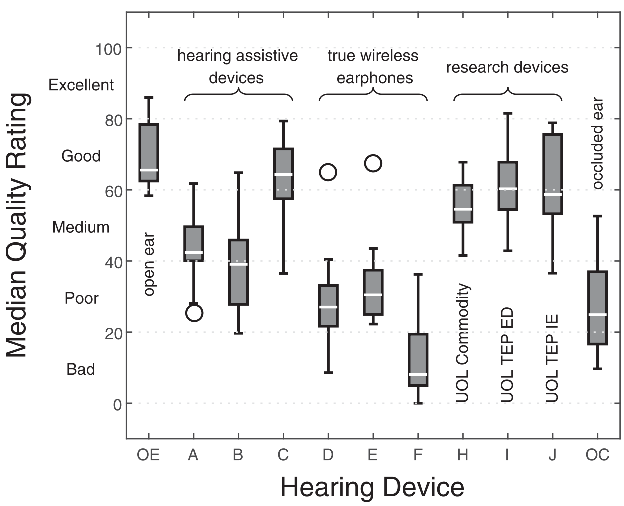


Fig. 4. Quality ratings of Experiment 1, aggregated per participant as median over all directions and signals.

tion) repeated measures analysis of variance (ANOVA) using Greenhouse–Geisser correction for sphericity. ANOVA showed a significant effect of the main factor hearing device ($F(0.74, 11.91) = 179.06$, $p < 0.001$) as well as the interaction of hearing device and signal ($F(2.23, 35.74) = 12.51$, $p < 0.001$), hearing device and direction ($F(1.49, 23.83) = 3.74$, $p < 0.001$), and signal and direction ($F(0.45, 7.15) = 4.86$, $p = 0.001$). Post hoc analysis was conducted for the factor of hearing device using the Student t test at the Bonferroni correct level of significance. The results of the post hoc analysis are shown in Table 4.

In order to more easily assess the impact of the hearing device on the perceived sound quality ratings, Fig. 6 shows the aggregated results computed as the median quality rating for each participant across all signals and signal directions. As expected from the statistical analyses, all hearing devices are rated worse than the open ear. Furthermore, hearing device C yields the highest quality rating compared to the open ear, followed by hearing devices I and J. Notably, again the quality of hearing device F was rated lower than the intended anchor signal of the occluded ear canal (OC). Similarly as in Experiment 1, most commercial

Table 4. Post hoc analysis of the factor hearing device in Experiment 2. Asterisks indicate significant differences ($p < 0.05$), while dashed indicate nonsignificant differences.

	OE	A	B	C	D	E	F	H	I	J	OC
OE		*	*	*	*	*	*	*	*	*	*
A	*		-	*	*	*	*	*	*	*	-
B	*	-		*	*	-	*	*	*	*	*
C	*	*	*		*	*	*	*	-	-	*
D	*	*	*	*		-	*	*	*	*	-
E	*	*	-	*	-		*	*	*	*	-
F	*	*	*	*	*	*		*	*	*	*
H	*	*	*	*	*	*	*		-	-	*
I	*	*	*	-	*	*	*	-		-	*
J	*	*	*	-	*	*	*	-	-		*
OC	*	-	*	*	-	-	*	*	*	*	

hearing devices without hearing assistance only achieved poor to medium quality (D–F).

2.3 Reliability Analysis

In order to determine the reliability of participants to rate either the overall quality without a reference (Experiment 1) or the quality compared to a reference (Experiment 2), we computed the eGauge measure [27, 28]. The eGauge measure performs a statistical analysis of the results obtained for each participant and estimates their statistical significance. Similarly as in [27], we excluded the anchor and reference conditions in Experiment 2 from the eGauge analysis; however, in Experiment 1, we included both conditions since the participants neither had a reference to compare to nor were informed about an anchor signal. Fig. 7 shows the results in terms of reliability and discrimination, where the grey lines indicate the threshold for statistical significance of the performance. Any results left of or below these lines indicate a nonreliable (vertical line) or nondiscriminative participant (horizontal line). While reliability is a measure of how reliable a participant can report similar ratings across different listening sessions, e.g., across test and retest, discrimination is a measure of how well a participants can differentiate the different signals. As can be observed, all participants were able to perform well above the statistical threshold for rejection of these two measures, i.e., they are located in the upper-right quadrant. Furthermore, comparing the two experiments, we observe larger reliability and discrimination in Experiment 2, indicating that providing a reference allows participants to more easily judge differences. However, it should be noted that the trends in both experiments visually appear very similar.

3 DISCUSSION

In this study, we investigated the perceptual quality of hear-through features in commercial and research hearing devices. To the best of our knowledge, this is the first study to investigate the hear-through features of several commercial consumer hearing devices and two different research hearing devices with respect to their perceived sound qual-

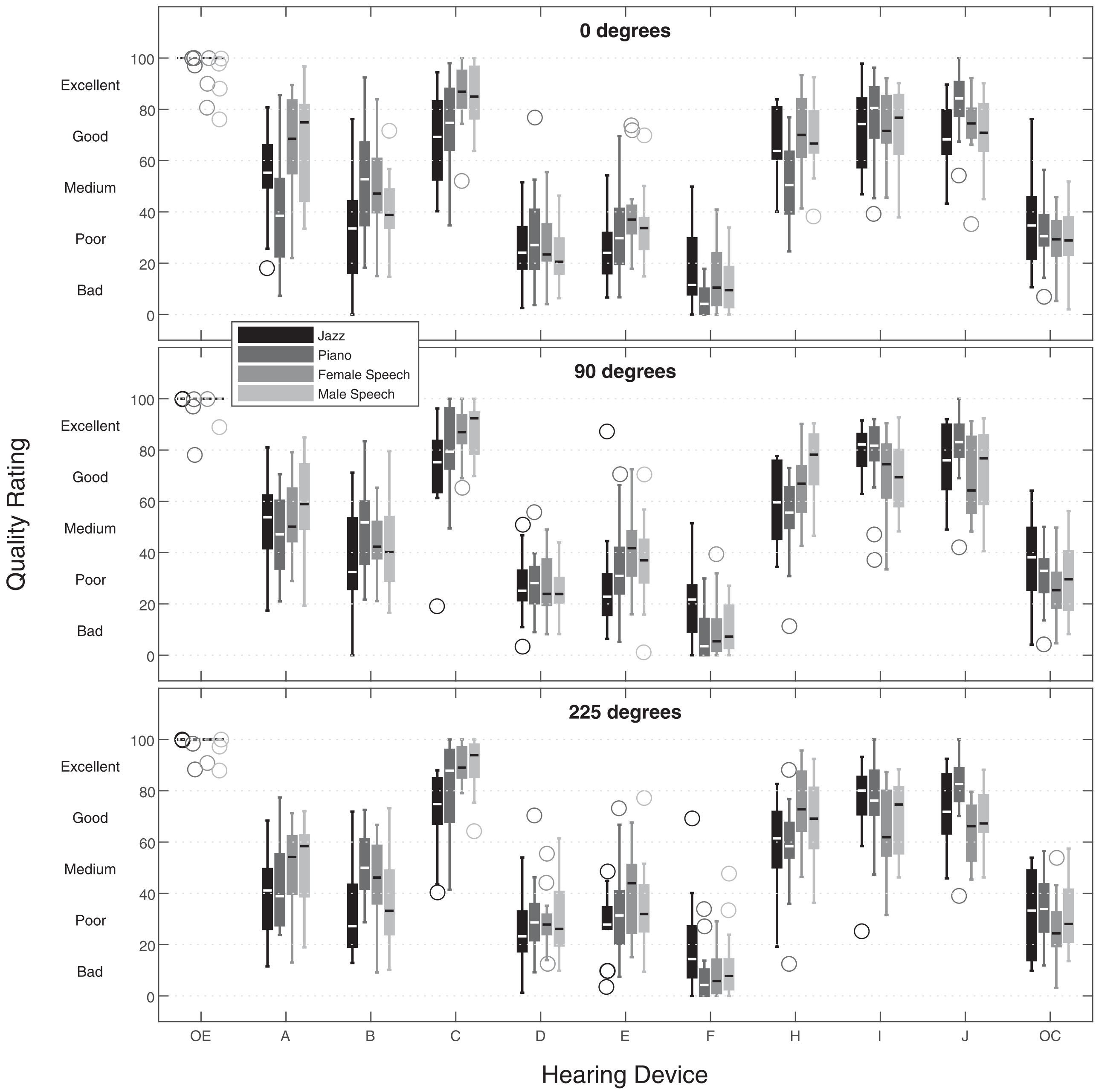


Fig. 5. Sound quality ratings for Experiment 2 with explicit open-ear reference for the different signal directions, hearing devices and audio signals. Boxes show interquartile ranges, points in boxes show the median, whiskers extend to 1.5 times the interquartile range and circles point show outliers.

ity with and without an explicit comparison to the open ear.

3.1 Perceptual Results and Relation With Device Characteristics

For Experiment 1, results showed a significant effect of hearing device and signal as well as their interaction, indicating that the spectral modifications of the hearing devices had different effects on the perceived sound quality depending on the spectral content of the signal. The observed differences in perceived quality between hearing devices can be associated with their technical performance as investigated in [15] and summarized in Table 5. This includes the processing delay, their match to the open-ear response, as well as comb-filtering effects, i.e., spectral ripples due to the superposition of the playback sound and sound leaking into the ear canal, binaural cue preservation and self-noise.

Table 5 shows the summary of these technical results from [15] as well as qualitative ratings of the perceptual quality performed in the present study for the considered devices. On the one hand, hearing device C was rated highest in both experiments and also has a very good match to the open ear, avoids comb filtering effects, and preserves the binaural cues. On the other hand, hearing device F, which

Table 5. Summary of the technical evaluation provided in [15] as well as qualitative ratings for the perceptual quality evaluation. Qualitative ratings are indicated by a five-point scale between "++" (excellent) "+" (good), "+-" (medium), "-" (poor), and "–" (bad) as obtained from the median data in Figs. 4 and 6. Note that the specific configuration of hearing device I was not included in [15]; however, similar ratings would be expected as for hearing device J.

Hearing device	Delay / ms L/R	Response *m*atch to open ear	Comb-filtering *a*voiding ripples	Binaural cues *c*onservation	Self-noise / dBA DF	Quality Exp. 1	Exp. 2
A	9.7	+	-	++	23.9	+-	+-
B	4.5	-	+	++	26.1	-	+-
C	3.1	++	++	++	27.3	+	++
D	<0.1	–	++	+	22.0	-	-
E	1.2 / 0.7	-	++	-	23.2	-	-
F	0.8 / 10.4	–	-	–	18.2	–	–
H	9.0	+	-	++	24.9	+-	+
I	n/a	n/a	n/a	n/a	n/a	+-	+
J	6.3	++	-	+	26.2	+-	+

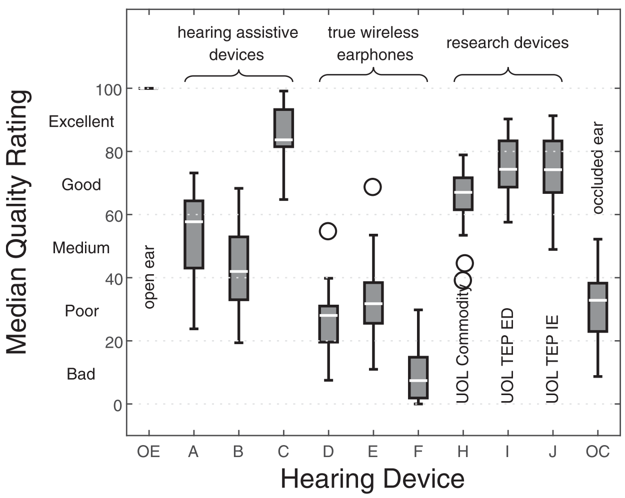
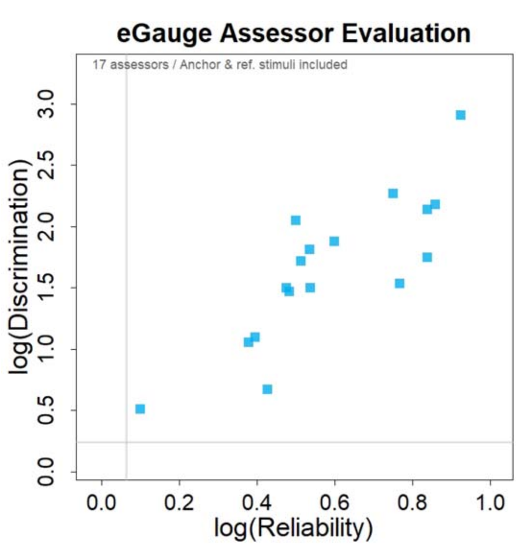


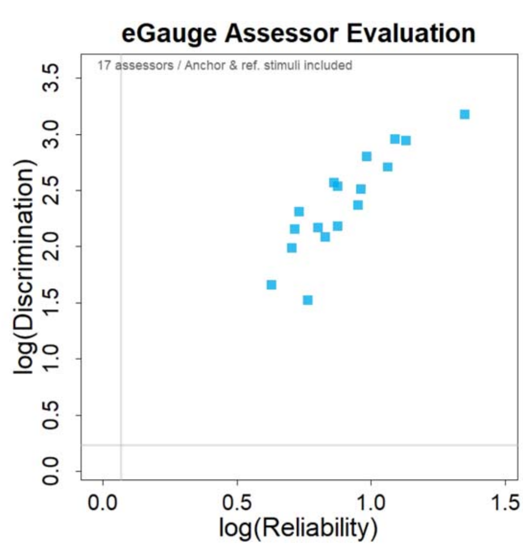
Fig. 6. Quality ratings of Experiment 2, aggregated per participant as median over all directions and signals.

was rated lowest in both experiments, has a very poor match to the open-ear response, does not particularly well avoid comb filtering effects, and has very poor preservation of binaural cues. In the following text, we discuss the influence of each of the technical parameters on the perceived quality.

The delay of the hearing devices did not have a systematic effect on the perceived quality, i.e., no clear relation between delay and perceived quality was observed. The match of the aided response to the open ear seems to be highly indicative for a high perceptual quality in both experiments, i.e., those hearing devices with a good or very good match yield a high perceptual quality. Comb-filtering effects did apparently have less influence on the perceived quality compared to the match to the open-ear response, i.e., when a similar match to the open ear is achieved, the comb filtering effects lead to only minor differences in perceived quality, e.g., comparing hearing devices C, I, and J. However, the extent to which comb-filtering effects of



(a) Experiment 1



(b) Experiment 2

Fig. 7. Results in terms of discrimination and reliability computed using the eGauge measure for (a) Experiment 1 and (b) Experiment 2. Each datapoint corresponds to a participant.

the devices had an effect may also have been masked by strong early reflections due to the acoustic of the recording room, which were then also present in the open-ear recordings. The importance of binaural cue preservation is unclear from the present data. The only hearing devices distorting the binaural cues also had a poor match to the open-ear response (hearing devices E and F). Potentially, for hearing device F, the binaural cue degradation led to perceptual quality ratings lower than those of the intended anchor signal (OC). The level of self-noise apparently did not have any influence on the perceptual sound quality, which can be explained by the fact that the stimulus level generally was much larger than the self-noise levels. Differences between the two transparent earpiece variants (hearing device I and J) were not significant, indicating that, for a dummy head, equalizing the signal at the eardrum or close to the eardrum at the inside of an earpiece is very similar in terms of the perceived quality. However, this may be different when the device is individually fitted to a participant with different ear canal characteristics [5]. In conclusion, the main factor for a high perceptual sound quality of a hear-through feature is the match of the aided response to the open ear. If this can be provided, the next factor contributing to a high perceptual sound quality appears to be the avoidance of comb-filtering artifacts. The impact of binaural cue distortions on the quality, however, remains unclear, since the it was likely masked by other factors.

Considering the different signals, in particular, the jazz signal was rated lower than both speech signals in Experiment 2. One reason for this different rating is the fact that the jazz signal contains many broadband transients (cf. Fig. 2) that may be attenuated in the higher-frequency regions by some of the hearing devices compared to the open ear, yielding a reduced quality in Experiment 2. Therefore, we compared the aided ear responses for the different hearing devices for a frontal signal direction. Fig. 8 shows the magnitude responses of the measured aided ear transfer functions in comparison to the open-ear response for the same direction. Note that the differences between the open-ear response and the aided responses are generally similar to those observed in the anechoic measurements reported in [15]. As can be observed, the hearing devices that yielded the highest quality rating (hearing devices C, H, I, and J) are able to closely reproduce the open-ear response; however, above 10 kHz, most of the devices deviate from the open ear. In contrast, the hearing devices with the lowest ratings (D–F) show a much larger deviation from the open-ear response, especially in the frequency regions above 1 kHz. In conclusion, in many devices, high-frequency content, e.g., above 8–10 kHz is not well matched to the open ear, potentially resulting in the lower perceptual sound quality ratings of the jazz signal, in particular compared to the speech signals.

Furthermore, we did not observe a significant influence of incoming signal direction. However, in Experiment 1, there was a significant interaction between the hearing devices and signal direction, indicating that some of the hearing devices introduce spatial distortions that influence the quality ratings. For example, we observe that hearing device C yields slightly higher quality ratings for the 90° direction compared to the 0° and 225° directions. This effect can be explained by the fact that many devices equalize for an open-ear diffuse field response, and hence, the extent to which distortions affect the quality depend on the matching of the response for a specific direction to the diffuse field [15].

3.2 Influence of Auxiliary Features in Commercial Devices

We aimed at investigating the hear-through features of the hearing devices and hence tried to ensure that any other advanced processing, e.g., dynamic range compression or spatial filtering, was switched off. This also included using flat equalizer settings. In particular, in Experiment 1, using different equalizer settings, e.g., based on individual preferences, could change the quality ratings [29]. However, this was beyond the scope of the present study, in which the main focus was on the ability of the hear-through feature to achieve a high quality compared to the open ear. Additionally, since we used commercially available devices, other advanced signal processing that could not be controlled or deactivated using the proprietary smartphone apps may still have influenced the quality ratings. In particular, for hearing device E, the aided response shown in Fig. 8 is similar to the occluded response in the frequency range between 3 kHz and 6 kHz, indicating that additional processing might have been active during the recordings. One potential candidate of processing that was active here is feedback management, which may have been activated by the exponential sweep measurement signal. In order to verify this, we additionally measured the aided responses using a white noise test signal, for which we did not observe this dip in the frequency response (see dashed black line Fig. 8). This confirms that the observed mismatch between the open-ear response and the aided ear response shown in Fig. 8 and the reduced quality rating in both experiments (cf. Figs. 4 and 6) compared to the open ear are in line with our above reasoning.

3.3 Influence of Providing a Reference on Results

While in Experiment 1 we asked participants to rate the overall sound quality without explicit knowledge about the open-ear reference, in Experiment 2, the participants were asked to rate the quality compared to the open-ear reference. Experiment 1 thus provided insights into the overall quality preference, while by choice of the open-ear reference, Experiment 2 provided insights into the ability of the hear-through feature to achieve a perceptually close match to the open-ear response.

As expected, the eGauge measure showed that participants are more reliable in their quality ratings and small differences could be discriminated better by the listeners when a reference was provided. Nevertheless, even without a reference, participants were able to discriminate between the different devices and provided reliable quality ratings. As we discussed above not all hearing device achieved an aided response at the eardrum comparable to the open-ear

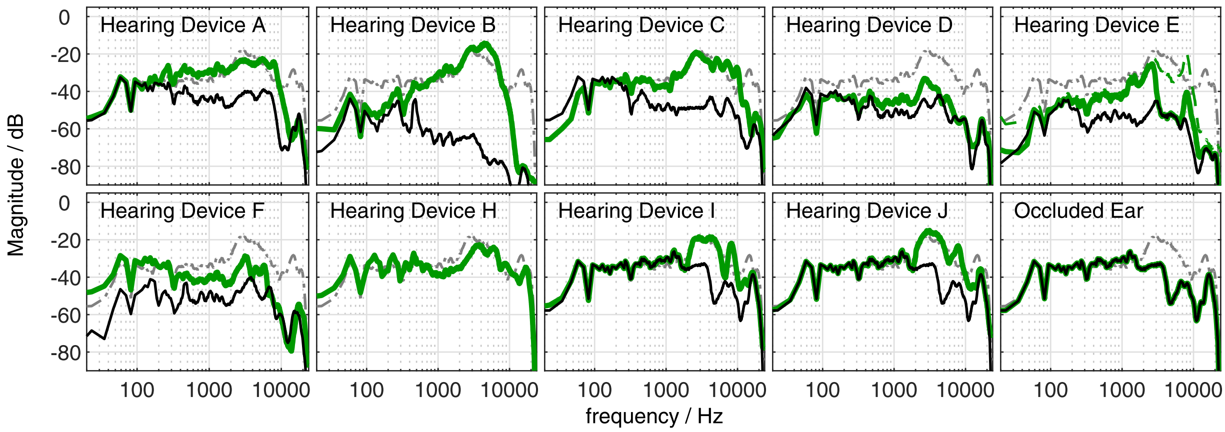


Fig. 8. Aided responses of the left ear for a frontal signal direction (green thick line) and the open-ear response (hearing device OE, dash-dotted grey line) for all considered hearing devices and their corresponding occluded ear responses (thin black line). For hearing device E, we also included the aided response measured using a white noise sequence (dashed thin green line). Note that for hearing device H, the occluded response was not measured.

response. Nevertheless, the perceptual quality ratings for Experiment 1 and 2 visually appear to be very similar. In order to confirm this, we conducted correlation analyses between the results of Experiment 1 and Experiment 2. Pearson's correlation coefficient based on the individual test–retest averages was $r_p = 0.785$, and Pearson's correlation coefficient based on the median data was $r_p = 0.961$. Since providing a reference may have resulted in a nonlinear usage of the rating scale in Experiment 2, we also computed Spearman's correlation coefficient for the individual test–retest averages as $r_s = 0.789$ and for the median data as $\rho = 0.958$ to compute the correlation of the rankings. This shows that the sound quality ratings and their rankings are indeed very similar, supporting the idea that perceptual similarity to the open ear is a good predictor for a high perceptual quality of the hear-through feature of a hearing device.

3.4 Limitations and Implications

The perceptual sound quality ratings obtained in this study are based on recordings made using a dummy head. In general, it is desirable to assess the quality of hearing devices with hear-through features by placing the devices in the participants' ears. However, this requires the participants to manually remove and reinsert the hearing devices. While this is generally feasible in Experiment 1, in Experiment 2, this would require the participants to be able to remember their open-ear reference exactly while being exposed to the reinsertion process, potentially leading to a biased judgement. Furthermore, individual biases favoring devices due to their fit could not be ruled out, since participants had to handle the fitting of devices on their own. Therefore, in order to validate the findings of the present study, this would require a much more involved experimental setup to avoid these biases when focusing only on the perceived sound quality.

Furthermore, we only considered static acoustic scenarios, i.e., neither head movements nor sound source movements were included, and we considered only a single sound source. In more complex scenes, e.g., with multiple talkers and noise sources, the distortion of the binaural cues, as well as delay artifacts, may be more harmful than observed in the present study. For example, in [10], the use of direction-dependent hear-through equalization was investigated, showing that, for over-the-ear headphones, an improved equalization performance can be obtained compared to non-direction-dependent hear-through equalization.

The results of both experiments showed an influence of the aided response at the eardrum on the perceived quality of the hear-through features in hearing devices. In particular, for devices with a good reproduction of the characteristics of the open ear at the aided eardrum, the quality ratings were largest in both experiments. In order to assess the acoustic transparency of a device in the sense of being able to be discriminate a hear-through feature from the open ear, more time-consuming psychoacoustic experiments need to be conducted, e.g., ABx comparisons in [5]. While the experiments in the present study did not directly assess the acoustic transparency of the hearing devices, in Experiment 2, a small difference in perceived sound quality of a hearing device compared to the open ear can be assumed to also indicate a good acoustic transparency.

4 CONCLUSION

In this study, we investigated the perceptual quality of hear-through features in commercial and research hearing devices. We conducted two listening experiments with normal-hearing participants, assessing the perceived overall sound quality of the hearing devices and the perceived sound quality compared to the open ear. Our results showed that both overall sound quality and sound quality compared to the open ear are highly correlated, indicating that in order to achieve a high quality in hearing devices, it is desirable to maintain or recreate the perceptual characteristics of an open ear. Furthermore, for hearing devices with large deviations from these open-ear characteristics, e.g., most importantly the match to the open-ear response, the perceived quality is severely degraded. The extent to which the distortion of spatial cues can limit the perceived sound quality of hear-through features could not be answered by the current study and remains an open issue for future research. A comparison of the discrimination and reliability of the quality ratings of both experiments showed that participants are generally more reliable and can discriminate better when a reference signal is provided.

5 ACKNOWLEDGMENT

The authors thank Feline Malin Barg and Merle Schlender for their support in recording the stimuli and conducting the listening experiments. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project ID 352015383 (SFB 1330 A4 and C1) and Project ID 390895286 (EXC 2177/1).

6 REFERENCES

[1] P. Hoffmann, F. Christensen, and D. Hammershøi, “Insert Earphone Calibration for Hear-Through Options,” presented at the *AES 51st International Conference: Loudspeakers & Headphones* (2013 Aug.), conference paper 3–4.

[2] J. Rämö and V. Välimäki, “An Allpass Hear-Through Headset,” *Proc. 22nd European Signal Processing Conference (EUSIPCO)* (2014 Sep.), pp. 1123–1127.

[3] V. Välimäki, A. Franck, J. Rämö, H. Gamper, and L. Savioja, “Assisted Listening Using a Headset: Enhancing Audio Perception in Real, Augmented and Virtual Environments,” IEEE Signal Processing Magazin, pp. 92–99 (2015 Feb.).

[4] J. Liski, R. Väänäne, S. Vesa, and V. Välimäki, “Adaptive Equalization of Acoustic Transparency in an Augmented-Reality Headset,” presented at the *2016 AES International Conference on Headphone Technology* (2016 Aug.), conference paper 6–4.

[5] F. Denk, M. Hiipakka, B. Kollmeier, and S. M. A. Ernst, “An Individualised Acoustically Transparent Earpiece for Hearing Devices,” *Int. J. Aud.*, vol. 57, no. sup3, pp. S62–S70 (2018 Jun.).

[6] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, “Equalization Filter Design for Achieving Acoustic Transparency in a Semi-Open Fit Hearing Device,” *Proc. ITG Conference on Speech Communication* (2018 Oct.), pp. 226–230.

[7] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, “Augmented Reality Audio for Mobile and Wearable Appliances,” *J. Audio Eng. Sec.* (2004 Jun.).

[8] F. Denk, S. Vogl, H. Schepker, B. Kollmeier, M. Blau, and S. Doclo, “The Acoustically Transparent Hearing Device: Towards Integration of Individualized Sound Equalization, Electro-Acoustic Modeling and Feedback Cancellation,” *Proc. 1st Int. Conference on Challenges in Hearing Assistive Technology (CHAT-17)* (2017 Aug.), pp. 89–94.

[9] J. Fabry, F. König, S. Liebich, and P. Jax, “Acoustic Equalization for Headphones Using a Fixed Feed-Forward Filter,” *Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019 May), pp. 980–984.

[10] R. Gupta, R. Ranjan, J. He, and W. S. Gan, “Parametric Hear Through Equalization for Augmented Reality Audio,” *Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019 May), pp. 1587–1591.

[11] M. A. Stone, B. C. J. Moore, K. Meisenbacher, and R. P. Derleth, “Tolerable Hearing Aid Delays. V. Estimation of Limits for Open Canal Fittings,” *Ear and Hearing*, vol. 29, no. 4, pp. 601–617 (2008 Aug.).

[12] H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, “Subjective Sound Quality Evaluation of an Acoustically Transparent Hearing Device,” presented at the *2019 AES International Conference on Headphone Technology* (2019 Aug.), conference paper 18.

[13] J. Rämö and V. Välimäki, “Digital Augmented Reality Audio Headset,” *Journal of Electrical and Computer Engineering*, pp. 1–13 (2012 Oct.).

[14] R. Gupta, R. Ranjan, J. He, and G. Woon-Seng, “On the Use of Closed-Back Headphones for Active Hear-Through Equalization in Augmented Reality Applications,” presented at the *2018 AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), conference paper P8-2.

[15] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, “Acoustic Transparency in Hearables - Technical Evaluation,” *J. Audio Eng. Soc.*, vol. 68, no. 7/8 (2020), pp. 508–521.

[16] M. R. Schädler, “Instructions for Building an Almost Consumer Hardware Based Prototype of a Hearing Aid,” https://github.com/m-r-s/hearingaidprototype (2017).

[17] M. Buhl, F. Denk, M. Bodenstein, N. Wiedemann, S. Jacobsen, M. R. Schädler, and B. Kollmeier, “Calibration of a Mobile Hearing Aid Prototype and Its Validation: Towards Transparent Listening With Commodity Hardware,” presented at the *German Annual Conference on Acoustics (DAGA)* (2019).

[18] G. Grimm, T. Herzke, D. Berg, and V. Hohmann, “The Master Hearing Aid: A PC-Based Platform for Algorithm Development and Rvaluation,” *Acta Acustica united with Acustica*, vol. 92, no. 4, pp. 618–628 (2006 Jul.).

[19] T. Herzke, H. Kayser, F. Loshaj, G. Grimm, and V. Hohmann, “Open Signal Processing Software Platform for Hearing Aid Research (openMHA),” *Proc. Linux Audio Conference* (2017 May), pp. 35–42.

[20] H. Schepker, S. Nordholm, L. T. T. Tran, and S. Doclo, “Null-Steering Beamformer-Based Feedback Cancellation for Multi-Microphone Hearing Aids With Incoming Signal Preservation,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 4, pp. 679–691 (2019 Apr.).

[21] E. Vincent, S. Araki, F. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, V. Gowreesunker, D. Lutter, and N. Q. K. Duong, “The Signal Separation Evaluation Campaign (2007–2010): Achievements and Remaining Challenges,”

Signal Processing, vol. 92, no. 8, pp. 1928–1936 (2012 Aug.).

[22] A. Farina, “Advancements in Impulse Response Measurements by Sine Sweeps,” presented at the *122nd Convention of the Audio Engineering Society* (2007 May), convention paper 7121.

[23] C. Völker, T. Bisitz, R. Huber, B. Kollmeier, and S. M. A. Ernst, “Modifications of the MUlti Stimulus Test with Hidden Reference and Anchor (MUSHRA) for Use in Audiology,” *Int. J. Aud.*, vol. 57, sup. 3, pp. S92–S104 (2018 Jun.).

[24] F. Denk, B. Kollmeier, and S. M. A. Ernst, “High-Fidelity Hearing Instruments: Evaluating Listening Quality of a New Prototype Using a Method for Evaluating Modified Listening (MEML),” presented at the *2016 AES International Conference on Headphone Technology* (2016 Aug.), conference paper 7-3.

[25] ITU-R, “Recommendattion ITU-R BS.1534-3 (10/2015): Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems” (2015 Oct.).

[26] F. Brinkmann and S. Weinzierl, “AKtools – An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics,” presented at the *142nd Convention of the Audio Engineering Society* (2017 May), eBrief 309.

[27] G. Lorho, G. L. Ray, and N. Zacharov, “eGauge - A Measure of Assessor Expertise in Audio Quality Evaluations,” presented at the *AES 38th International Conference: Sound Quality Evaluation* (2010 Jun.), conference paper 7-2.

[28] ITU-R, “Report ITU-R BS.2300-0 (04/2014): Methods for Assessor Screening” (2014 Apr.).

[29] N. S. Jensen, O. Hau, J. B. B. Nielsen, T. B. Nielsen, and S. V. Legarth, “Perceptual Effects of Adjusting Hearing-Aid Gain by Means of a Machine-Learning Approach Based on Individual User Preference,” *Trends in Hearing*, vol. 23, pp. 1–23 (2019).

THE AUTHORS

Henning Schepker

Florian Denk

Birger Kollmeier

Simon Doclo

Henning Schepker received a B.Eng. degree from Jade University of Applied Sciences, Oldenburg, Germany, in 2011, and an M.Sc. degree (with distinction) from the University of Oldenburg, Oldenburg, Germany, in 2012, both in hearing technology and audiology. In 2017, he received a Dr.-Ing. degree from the University of Oldenburg. Since 2018, he has been a Postdoctoral Researcher with the Signal Processing Group, Department of Medical Physics and Acoustics, University of Oldenburg. His research interests are in the area of signal processing for hearing devices and speech and audio applications, as well as speech perception.

•

Florian Denk studied Engineering Physics and Physics at the University of Oldenburg, Germany, where he received his B.Eng. and M.Sc. degrees in 2014 and 2017, respectively, and his Ph.D. in 2019. He is currently a postdoctoral researcher in the Medical Physics group, Department of Medical Physics and Acoustics, at the University of Oldenburg. His work centers around improving the performance and sound quality of hearing devices especially for (near-to) normal hearing users, for which he conducts research in the fields of external ear acoustics, sound perception and localization, signal processing, and acoustic measurement technology. He published several datasets and was involved in the development of hearing device research platforms that are available to the public. In 2019, he received the Best Student Paper Award at the AES Conference on Headphone Technology.

•

Birger Kollmeier received a Ph.D. degree in Physics and an M.D.-Ph.D. degree in Medicine in Göttingen, Germany. Subsequent to his habilitation in Physics in 1991, he was appointed full Professor of Physics at the University of Oldenburg and the head of the Medical Physics Group in 1993. He is the scientific director of the Hörzentrum Oldenburg, Oldenburg, Germany, the speaker of the Center of Competence "Hearing Aid System Technology (HörTech)," head of the Fraunhofer Project Group "Hearing, Speech and Audio Technology" since 2008, and speaker of the Cluster of Excellence "Hearing4all" since 2012. He has supervised more than 75 Ph.D. theses and been awarded several scientific prizes, including the Alcatel-SEL research prize for technical communication and the German President's prize for technology and innovation (Deutscher Zukunftspreis) in 2012.

•

Simon Doclo received an M.Sc. degree in electrical engineering and a Ph.D. degree in applied sciences from KU Leuven, Belgium, in 1997 and 2003, respectively. From 2003 to 2007, he was a Postdoctoral Fellow at the Electrical Engineering Department (KU Leuven) and the Cognitive Systems Laboratory (McMaster University, Hamilton, Canada). From 2007 to 2009, he was a Principal Scientist with NXP Semiconductors in Leuven, Belgium. Since 2009, he has been a full professor at the University of Oldenburg, Germany, and scientific advisor for the Division Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical and biomedical applications, more specifically microphone array processing, speech enhancement, active noise control, acoustic sensor networks, and hearing aid processing. He has received several best paper awards (EURASIP Signal Processing 2003, IEEE Signal Processing Society 2008, VDE Information Technology Society 2019). He is a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, the EURASIP Technical Area Committee on Acoustic, Speech and Music Signal Processing, and the EAA Technical Committee on Audio Signal Processing. He served as guest editor for several special issues (IEEE Signal Processing Magazine, Elsevier Signal Processing) and was associate editor for IEEE/ACM Transactions on Audio, Speech and Language Processing and EURASIP Journal on Advances in Signal Processing.