

Wireless Streaming: Sound Quality Comparison Among MFi Hearing Aids

 hearingreview.com/2016/11/wireless-streaming-sound-quality-comparison-among-mfi-hearing-aids/

Tech Topic | December 2016 *Hearing Review*

By Jesper Ramsgaard, Petri Korhonen, MS, Tiffany K. Brown, AuD, and Francis Kuk, PhD

Results from this study demonstrate significant differences between the perceived sound qualities of hearing aids using the same baseline 2.4 GHz MFi technology.

The ability to stream from an iPhone to hearing aids using the 2.4 GHz band is becoming popular, and many hearing aid manufacturers carry products in their portfolio with this functionality. Direct streaming provides easy access to phone calls and media content without having to use an intermediary device in the signal chain.

The development of direct-to-hearing-aid-streaming from mobile phones has been pushed by the Made For iPhone (MFi) streaming technology. The MFi streaming utilizes the 2.4 GHz ISM (industrial, scientific, and medical) radio band to send audio wirelessly to the hearing aid. MFi has strong connotations to the Apple/iPhone brand, and one might assume that it is a guarantee of a similar level of sound quality across devices (including hearing aids) carrying the MFi tag. However, behind the MFi label are different implementations of the same streaming protocol—each device having unique sound quality characteristics. This means that the overall quality of the streamed sounds could vary between hearing aids, ultimately influencing how the end users experience the content streamed from their phones.

Sound Quality in 2.4 GHz Streaming

Adding streaming to hearing aids brings functionality resembling consumer Bluetooth headset devices. Watching movies, enjoying music, making phone calls, and listening to audio books are becoming a normal part of hearing aid usage. However, when dealing with a new technology like 2.4 GHz MFi streaming, end users may expect that the progress in connectivity is also supported with a step forward in sound quality. Delivering less-than-ideal sound quality will influence the overall perceived quality of the hearing aid, and could discourage the user from actively using the feature. Sound quality, alongside usability, could be a key contributor to the overall satisfaction for hearing aids with direct streaming capability.

The sound quality in direct-to-hearing aid streaming is limited by a number of factors, which are a consequence of the limited bit rate available for data transfer from the phones to the hearing aids. A typical reference sound quality used in digital audio is a non-compressed audio signal with a sampling rate at 44.1 kHz/16 bit (eg, same as what is used in Compact Disc Digital Audio or CDDA). A signal with these specifications needs a data rate of about 1,411 kbit/s. At this rate, it is assumed that the sound has so few artifacts that it is transparent to the user. Compared to this, and like most other wireless audio technologies, the MFi 2.4 GHz streaming has a more restricted bit rate. While the available bit rate is not a direct indicator for the perceived sound quality, it puts a limit on what can be expected for sounds streamed from phones.

The typical way to work within a restricted bit rate is to encode/compress the audio signal with an audio codec in a “smart” way before transmission. This limits the amount of “space” the audio stream takes up during the transfer so that as much of the critical information can be streamed. The audio signal is then decoded/decompressed into a regular audio bit stream on the receiving end (for a review, see Kuk et al¹). This same procedure is also applied in

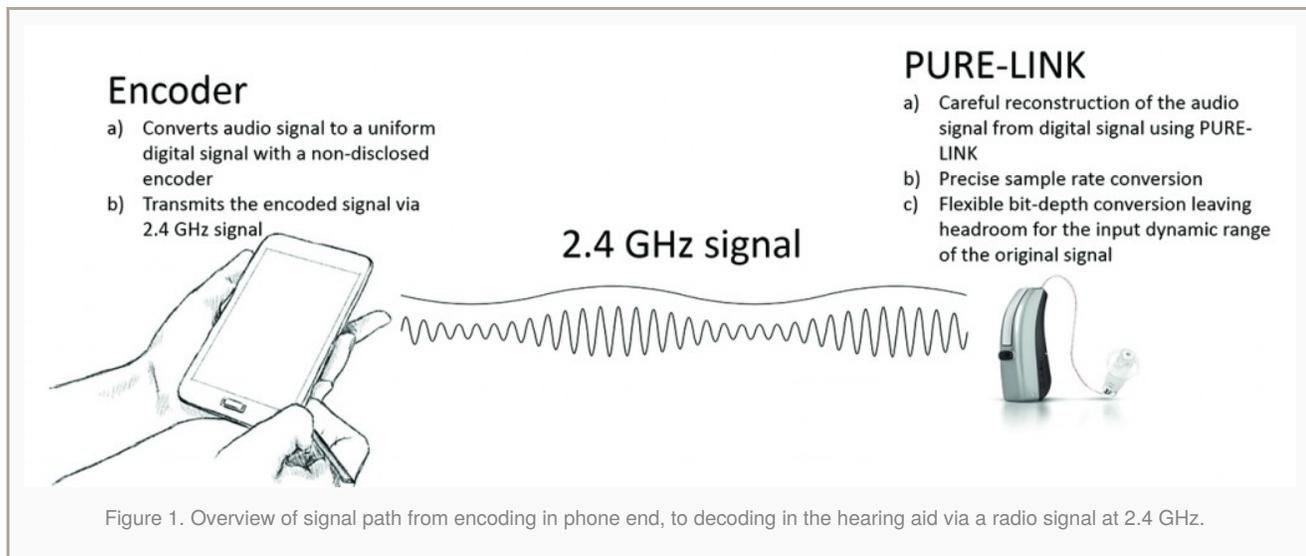
the MFi technology; as such, one may expect different degrees of “goodness” (good replication) from the proprietary encoding/decoding processes employed by different manufacturers. In addition, differences in the stability of the wireless transfer/connectivity, the D/A conversion process, how amplification is applied, etc, could all contribute to the maximum achievable sound quality of the streamed sounds.

The combination of limitations ultimately means that compromises and trade-offs are expected in the quality of streamed sounds. It is fair to assume at the outset that the perceived sound quality of the 2.4 GHz streamed audio will be poorer than the original signal.² It would also suggest, however, that design approaches that minimize the limitations of “good replication” would have the best chance of avoiding artifacts and preserving sound quality.

For some users this means that the sound quality of some wireless streams may be perceived as a degraded audio signal, having artifacts originating from the different steps (eg, encoding-decoding, D/A conversion, signal amplification) in the signal path. Some of the expected perceptual consequences or artifacts that the users may experience include:

- Muffled sounds or lacking clarity as a consequence of skewed frequency response and limited frequency range;
- Sharp or shrill sound from an overemphasis on high frequency amplification;
- Crackling or clipping from insufficient bit-depth allocation in the D/A conversion;
- A lack of spatial precision due to spatial smearing from the audio codec, and
- Hissing related to signal amplification and reproduction.

Thus, a user evaluation of the overall sound quality is the first and foremost point of interest when evaluating the goodness of the streaming products.³



In Widex BEYOND hearing aids, the streamed signal is handled with the Widex Pure-link technology (Figure 1). The Pure-link module utilizes Widex’s knowledge in wireless input handling from the WidexLink technology. This allows the BEYOND to carefully treat and scale the input efficiently through the signal chain and preserve the naturalness of the streamed signals while maintaining a low current drain (<2 mA in the BEYOND for 25% daily streaming). This is achieved through the following two steps:

1) Sample rate conversion: Up-sampling the wireless stream from the default sampling frequency in the 2.4 GHz codec to 33.1 kHz in the Widex hearing aid requires very high precision and care to ensure high sound quality. Doing so avoids adding any additional artifacts from errors in the sampling rate conversion, and ensures as clean a reproduction as possible.

2) Input dynamic range: Pure-link is built on the Widex custom chip, which is designed to allow for optimal scaling of dynamic range through the entire signal chain. Pure-link is not limited to a fixed bit depth, but can use far higher precision where needed for keeping the sound natural and true to the source input. This means no additional clipping of the signal occurs from the limited bit depth that is assigned in the streaming process.

Because each manufacturer has their own proprietary approach to process streamed signals while minimizing artifacts, it is of interest to evaluate how the quality of the streamed sounds may be perceived across various hearing aids. In the following study, we explored the performance of four currently available hearing aids by applying the recognized MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) 4.5 test method. The hearing aids tested all have MFi connectivity and comparable MFi functionality, such as streaming music/speech and phone calls directly to the hearing aid.

Test Methods

As described in a previous article about streaming and sound quality,² the MUSHRA was used in this study due to its test paradigm. Over the last two decades, MUSHRA has been successfully used in assessing sound quality in standardization organizations, academia, and the industry.⁶ In a MUSHRA test, a selection of systems under test (typically audio coding technologies) are rated against a shown reference and hidden bandpass limited anchor systems. The typical test question/parameter applied is “Basic Audio Quality,” which addresses any differences from the reference, using a 100-point scale with associated verbal labels ranging from “Bad” to “Excellent.”

As prescribed in the MUSHRA recommendation, the samples were selected to represent typical materials for the desired application, as well as to show differences in the devices under test (ITU-R 2015).⁴ The 13 samples were a combination of speech and music passages representing daily usage like listening to audio books, music, radio, etc. Samples from the Sound Quality Assessment Material (SQAM)⁷ and high quality recordings were included (see Table 1).

Sample	Description	Duration (sec)
1	Female speech	7.2
2	Male speech	8.2
3	Instrumental sample	14.1
4	Instrumental sample	8.0
5	Instrumental sample	7.1
6	Pop with vocal	20.3
7	Male speech	11.0
8	Pop/Jazz	23.0
9	Classical music	12.7
10	Classical music	16.6
11	Female speech	13.8
12	Male speech	17.0
13	Instrumental sample	11.6

Table 1. List of samples included in the listening test.

In order to allow for a strict double-blind setup, the streamed signals were recorded for the test. The recordings were carried out in an audiometric sound-isolated booth. The dimensions of the booth were 3 x 3 x 2 m (W x L x H). KEMAR was positioned in the center of the room. The receiver-in-the-canal (RIC) hearing aids were coupled with KEMAR using fully occluding ear tips. Each of the four hearing aid were programmed to fit a flat 20 dB HL audiogram. The gain prescription used in each device was based on NAL-NL2 to approximate equal output among devices.

The sound samples were streamed from an iPhone placed in an unobstructed position at 65 cm distance (2.13 ft), giving optimal conditions for the streaming to the hearing aids. All recordings were subjected to quality control, and if any artifacts due to disrupted connectivity were found, the samples were rerecorded for the test (ie, the potential issue of connectivity was not evaluated in this study). To avoid any bias from loudness differences between devices, the recordings were loudness normalized using ITU-R BS.1770-4.⁴

A total of 20 listeners with normal hearing were recruited from Widex headquarters in Lyngø, Denmark. The average

age was 37.4 years (SD = 7.8) with a gender distribution of 11 males and 9 females.

Procedure

The MUSHRA recommendation was applied for the listening test. MUSHRA uses a double-blind multiple comparison paradigm (neither the test leader, nor the listener knows what sound is being presented at any given time) with shown and hidden reference and anchors. Both the sample and systems under test presentation are randomized for each listener in the test.



The possibility for multiple comparisons allows the listener to compare all the hearing aids under test with each other, as well as the reference and anchor systems, making it a well-suited test in exposing differences in sound quality among technologies. Figure 2 shows a picture of the MUSHRA test screen.

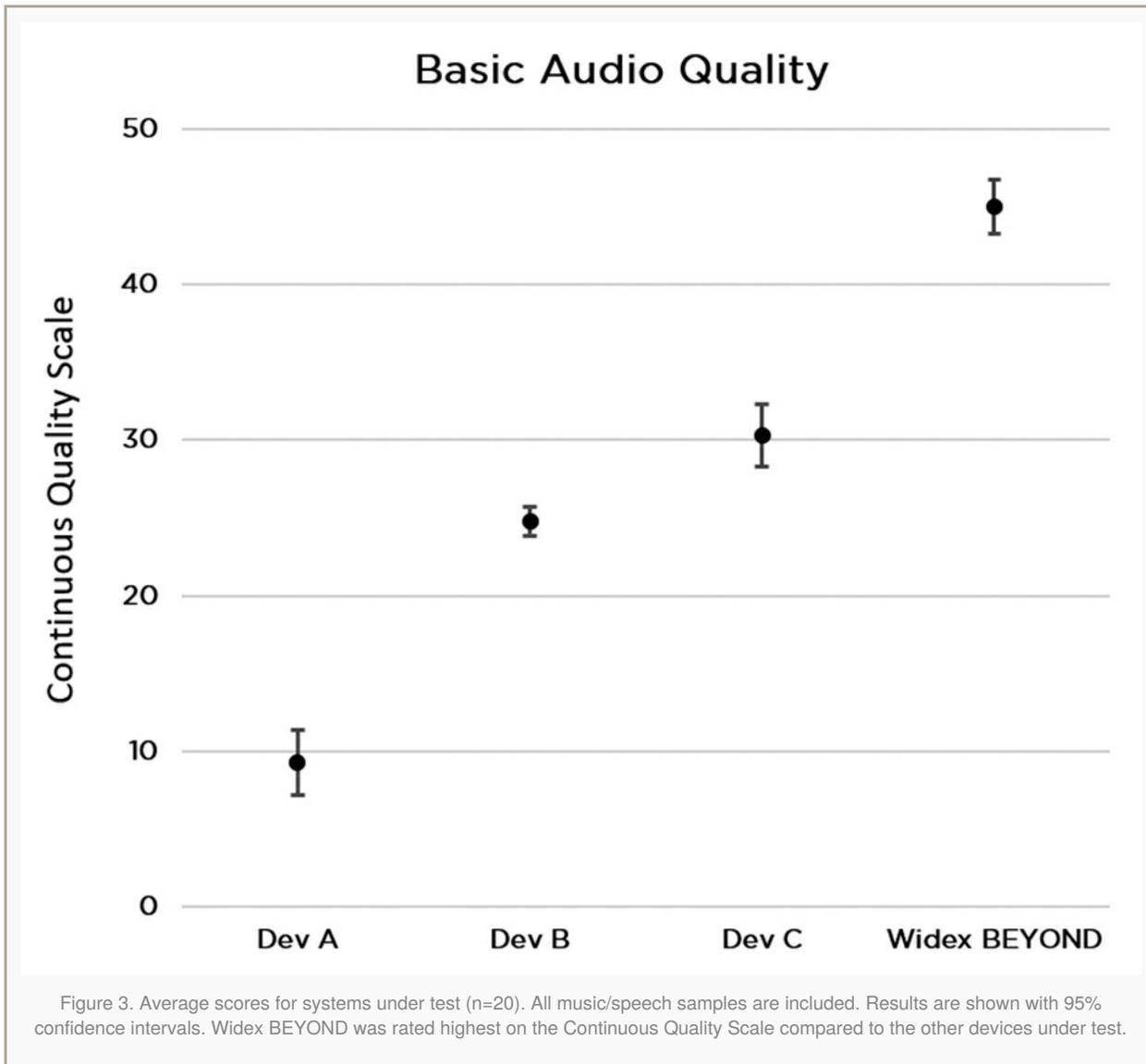
The test session comprised of a training/familiarization phase, allowing listeners to familiarize themselves with the systems under test, as well as the sound samples. After the training/familiarization, the actual test was performed. The total test time was between 30-40 minutes. In the test, two extra samples were included as repetitions, allowing for checking listener repeatability and discrimination, according to the ITU-R BS.2300-0.⁸ The test was performed in listening rooms with a background noise level below NR25.⁹ The test sounds were presented via Beyer Dynamic DT770 (closed) headphones, connected to an external USB D/A headphone amplifier, at a comfortable listening level (set during the familiarization/ training by the listener).

Results

Before analysis, the post screening procedure described in the MUSHRA recommendation was applied (including eGauge analysis). This step is intended to remove any listener who failed to identify the reference system and/or

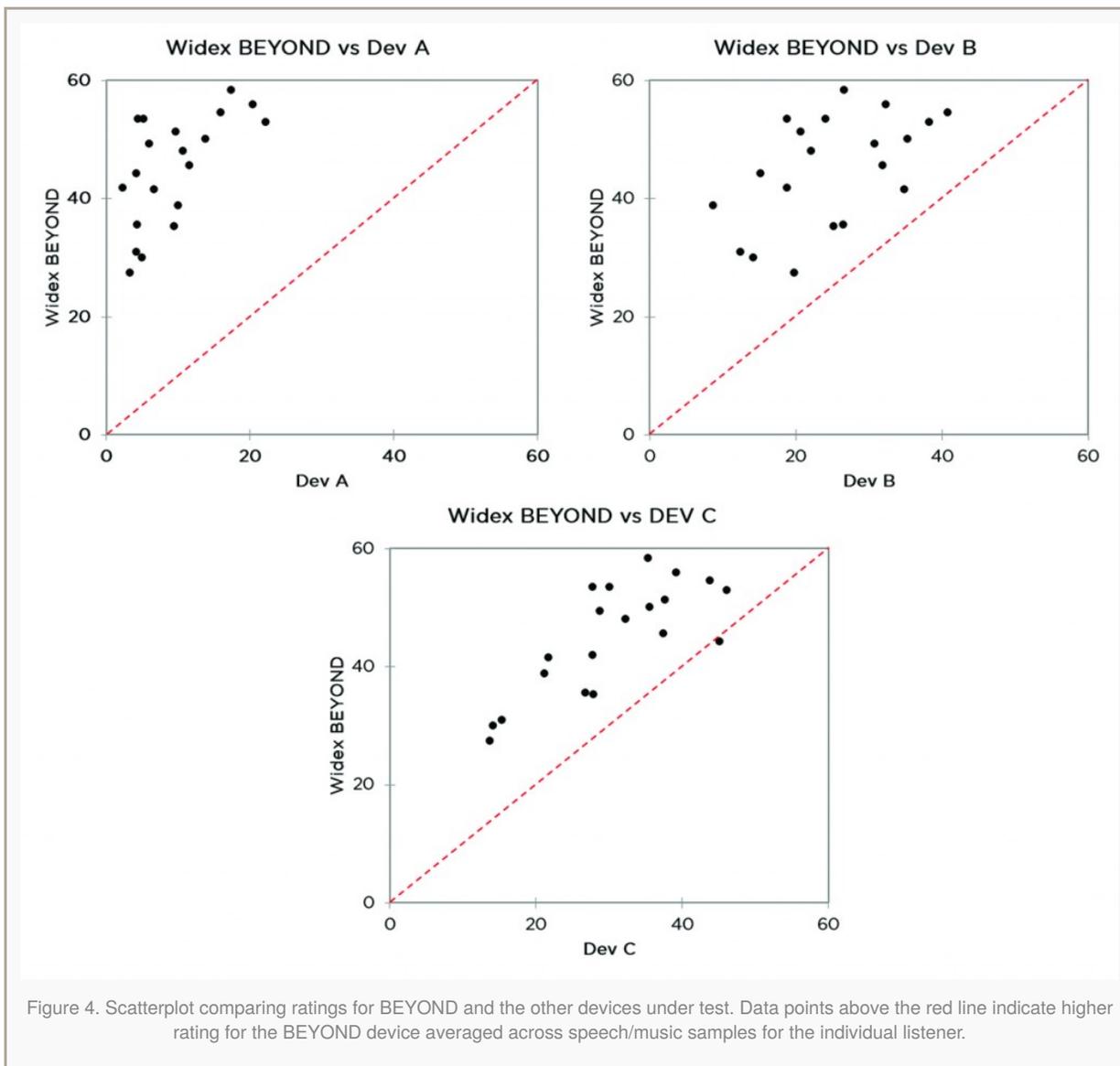
failed to discriminate between the test items.

This measure was taken to provide repeatable ratings. All listeners passed the post-screening procedure, and were included in the final analysis. In addition to checking data quality from the individual listeners, the overall data set was tested for normality to allow for parametric analysis.



Overall, Widex BEYOND was rated as having the best sound quality among the devices under test (Figure 3). The resolution given in the test is observed to be high, with all systems being separated and rated significantly different from each other according to a Tukey post hoc test ($p < 0.001$). This is supported by feedback from the listeners saying that it was “easy” to detect differences in the sounds being presented.

The rating of Basic Audio Quality on the continuous quality scale should be interpreted as a direct parallel to how the user would rate the devices—given the option to compare them directly in real life—with a higher rating reflecting a better perceived sound quality.



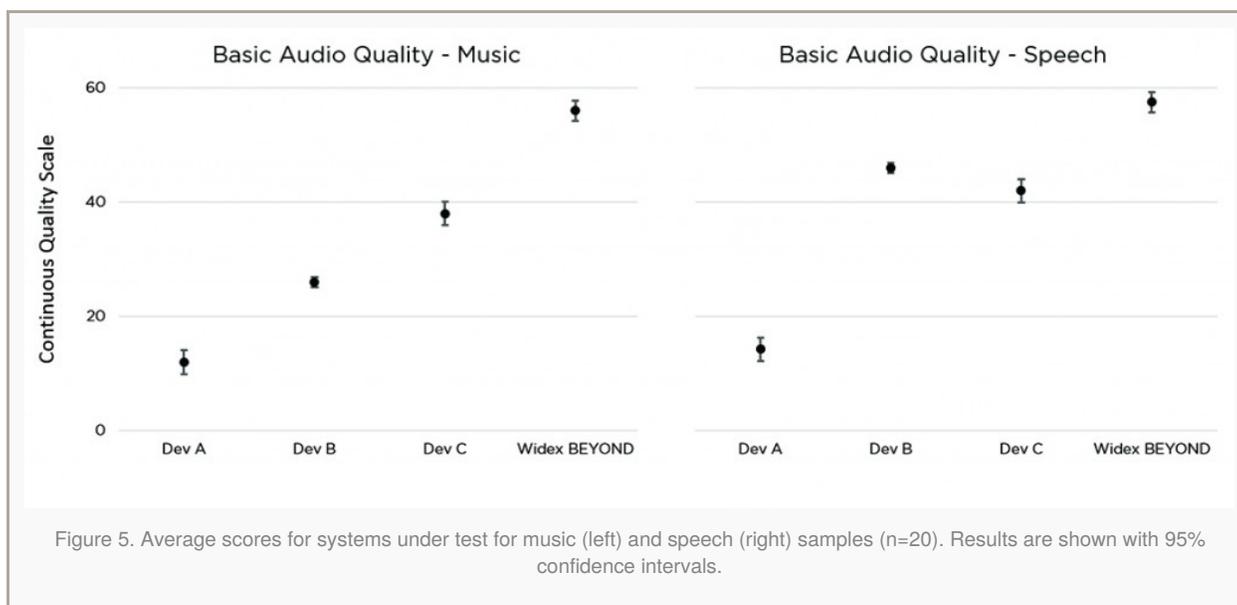
Comparison across devices under test. Figure 4 shows that, across the hearing aids under test, Widex BEYOND was rated as having the highest sound quality with an average of 44.4, compared to Device A (8.7), Device B (24.2) and Device C (29.7). Contributing to this mean difference in scores was that all the listeners rated the Widex BEYOND higher than Device A (20 of 20) and Device B (20 of 20). Almost all the listeners rated the BEYOND to be higher than Device C (19 of 20). The average ratings given by the listeners to the devices were:

- Widex BEYOND (Min. 27.5, Max. 58.4);
- Device A (Min. 2.2, Max. 22.2);
- Device B (Min. 8.6, Max. 40.8), and
- Device C (Min. 13.6, Max. 46.0).

Overall results showed that almost all the listeners rated the sound quality of the Widex BEYOND higher than the other devices tested.

Performance differences between speech and music content. In addition to looking at the overall sound quality averaged across all samples, it is important to make sure that the sound quality performance for different sample categories (eg, music and speech) is consistent. Some devices may be more sensitive to specific music or speech samples due to different frequency content and dynamic characteristics. If a device fails for either of these sample categories, the user will experience lower sound quality for certain types of content, making the device less suited

for streaming either music or speech.



When looking at sound quality differences across music and speech samples for this test, some interactions between devices and sample categories were observed (Figure 5). Device B and Device C showed lower ratings for music than speech samples ($p < 0.001$). This may be an artifact from the D/A conversion in the hearing aids, highlighted by the increased dynamic differences in the music samples. Device A was rated consistently lowest for both music and speech stimuli, while the Widex BEYOND was rated consistently highest for both sample types.

The observed differences in performance between streaming speech or music highlights the importance of selecting appropriate sample materials for testing the streaming sound quality. It is not enough to listen to any single piece of music or speech to assess the overall performance of the device.

Overall, the sound quality of MFi streaming differs widely between the tested devices. The observed sound quality differences are clear, with all devices under test being rated significantly different from each other. Considering that the technology has the same baseline limitations set by the MFi protocol, the differences in outcomes (rated sound quality) would suggest that the differences in signal processing methods used by each manufacturer of MFi hearing aids could lead to real-life differences in subjective ratings, and possibly overall satisfaction.

Summary

We tested the streaming sound quality performance of four hearing aids using a representative selection of music/speech samples according to the MUSHRA recommendation.

System	Description
Reference	PCM, 44.1 kHz, 16 bit.
Mid anchor	7 kHz, low pass filter.
Low anchor	3.5 kHz, low pass filter.
Widex BEYOND	Latest premium RIC hearing aid from Widex
Device A	Latest available premium RIC hearing aid from Manufacturer A
Device B	Latest available premium RIC hearing aid from Manufacturer B
Device C	Latest available premium RIC hearing aid from Manufacturer C

Table 2. List of systems under test, including reference and anchor systems. All hearing aids are models released 2015-2016. Anchor systems are used in the ITU-R BS.1534-3 test⁴ to anchor scale usage and allow for relative comparison of results between tests.

The results showed significant differences between the perceived sound qualities of hearing aids using the same baseline 2.4 GHz MFi technology. The differences are noticeable and significant, with listeners clearly showing higher ratings for the Widex BEYOND with Pure-link technology. In addition, Widex BEYOND showed stable sound quality performance for both speech and music material.

The study also showed that the 2.4 GHz streaming is not a technology that will appear flawless and transparent to the end user. It is potentially the best current method for streaming audio from a phone, but negative user feedback regarding the quality could be expected for the lower performing devices.

Of the devices tested, Widex BEYOND had the best sound quality in streaming directly to the hearing aid from a smart phone applying 2.4 GHz MFi technology—allowing the end user to access the best available quality of music and speech from their iPhone.

References

1. Kuk F, Crose B, Korhonen P, Kyhn T, Mørkebjerg M, Rank ML, Kidmose P, Jensen MH, Larsen SM, Ungstrup M. Digital wireless hearing aids, Part 1: A primer. *Hearing Review*. 2010;17(3):54-67.
2. Ramsgaard J. Sound Quality in Hearing Aid Wireless Streaming Technologies. *Hearing Review*. 2016;23(8) [Aug]:24-27.
3. Bech S, Zacharov N. *Perceptual Audio Evaluation—Theory, Method and Application*. Hoboken, NJ:John Wiley & Sons, Ltd;2006.
4. ITU Radiocommunication Assembly. ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems. October 2015. Available at: <https://www.itu.int/rec/R-REC-BS.1534/en>
5. DELTA. MUSHRA test. SenseLabOnline, 2014. Available at: <http://www.senselabonline.com>
6. Stoll G, Kozamernik F. EBU listening tests on Internet audio codecs. *EBU Tech*. 2000;1[June]:24.
7. European Broadcasting Union. (EBU) Sound Quality Assessment Material Recordings for subjective tests. *EBU-Tech 3296*. Available at: <https://tech.ebu.ch/publications/sqamcd>
8. ITU Radiocommunication Assembly. ITU-R BS.2300-0 (2014). Methods for Assessor Screening. International Telecommunications Union Radiocommunication Assembly. Available at: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BS.2300-2014-PDF-E.pdf

9. International Standards Organization (ISO). ISO 1996-1: 2016, Acoustics–Description and measurement of environmental noise–Part 1: Basic quantities and procedures. Available at: http://www.iso.org/iso/catalogue_detail?csnumber=28633

Jesper Ramsgaard is an Audiological Affairs Specialist, at Widex A/S in Lynge, Denmark.



Jesper Ramsgaard

Petri Korhonen, MS, is a researcher at Widex.



Petri Korhonen, MS

Tiffany K. Brown, AuD, is an audiologist at Widex.



Tiffany K. Brown, AuD

Francis Kuk, PhD, is executive director at the Widex Office of Research and Clinical Amplification (ORCA), in Lisle, Ill.



Francis Kuk, PhD

Correspondence can be addressed to Jesper Ramsgaard at: jram@nullwidex.com.

Original citation for this article: Ramsgaard J, Korhonen P, Brown TK, Kuk F. Wireless Streaming: Sound Quality Comparison Among MFi Hearing Aids. *Hearing Review*. 2016;23(12):36.

