WHITEPAPER
2021

Oticon More™ competitive benchmark Part 2 – Clinical evidence

SUMMARY

This white paper is Part 2 of a comparison between attributes of the Oticon More hearing aid and two top competitors. In Part 1, a comparison was made using technical measurements performed using a controlled and evidence-based methodology. Using the hearing aid sound recordings from Part 1, Part 2 explores the subjective perception of sound quality, as experienced by 22 test participants.

The participants completed a self-led set of listening and rating tasks for 7 sound recordings, where they listened blinded to the recordings through headphones and rated sound quality on a scale, using a modified version of the well validated subjective sound quality test (MUlti Stimulus with Hidden Reference and Anchor, MUSHRA). The MUSHRA focus is less on the ends of the scale (0-100), but rather on the sound quality rating differences between the hearing aid recordings. Results showed two sides of the same story: 1) Oticon More was rated significantly higher than its competitors for all sound scenes and 2) Up to 8 out of 10 people preferrred the Oticon More over its competitors in scenes with music and complex speech scenes. This not only confirms the key benefits shown from the technical measurements in Part 1 of the study, but also confirms that hearing impaired listeners perceive these benefits as overall better sound quality in a variety of different situations, bringing the findings from the lab to human experiences.

EDITORS OF ISSUE

Brian Kai Loong Man,
Clinical Research Audiologist, Centre for Applied Audiology Research, Oticon A/S

Marie Frederikke Garnæs,
Student Assistant, Centre for Applied Audiology Research, Oticon A/S

Susanna Løve,
Director of Clinical Audiology, Centre for Applied Audiology Research, Oticon A/S

oticon
life-changing **technology**

Sound Quality is a Multi-Layered Word

As consumers of headphones and loudspeakers in our everyday lives, think about what made you purchase the product in the first place. Is it the comfort? How attractive it looks? There are many factors that may influence why you chose one of them out of all the other potential candidates, but one characteristic that we can quite confidently say you considered was sound quality. This may sound trivial at first, as you may take off your headphones or turn off your speakers whenever you are done listening to a song. Unfortunately, this is not the case for users of hearing aids, as they will be reliant on such devices to listen to all the everyday sounds that surround them. From the casual conversation in a café, to the lively chatter at a party and even to the melodic performance of an orchestra. Sound quality quickly becomes a crucial factor when selecting the correct hearing aids, as that will be the type of sound users will be exposed to from the moment they get out of bed, to the moment they return. In fact, sound quality-related issues are the greatest contributors to why hearing aid users are satisfied or unsatisfied with their hearing aids (Marke Trak X, 2019). And yet, such a critical factor is to this day, highly debated upon. What components constitute sound quality, and can such an abstract, subjective concept even be measured?

In general, “quality” refers to the character of an object or a merit of its superiority to something in the same category. In the auditory domain, sound quality, timbre, tone colour, timbral colour, spectral colour and many other terms are commonly regarded as synonymous in psychoacoustical literature. This has caused a great degree of confusion among researchers, and as a result, often renders comparisons between findings invalid. Scientists have attempted to “decompose” the word into sizeable and measurable chunks, proposing that sound quality is indeed an “umbrella term” in which many psychoacoustical and physical elements branch out from (Figure 1). One unique element of the term is that apart from physical attributes that can be distinguished such as brightness and sharpness, is its emotional aspect. It is for this reason that in order to assess sound quality, the human factor becomes a necessary component. Such subject-oriented approaches have the important benefit of being able to capture the real-world experiences hearing aid users may have, thus ensuring that the More sound approach not only provides measurable benefits in the lab (Santurette et al 2021), but outside of it as well.

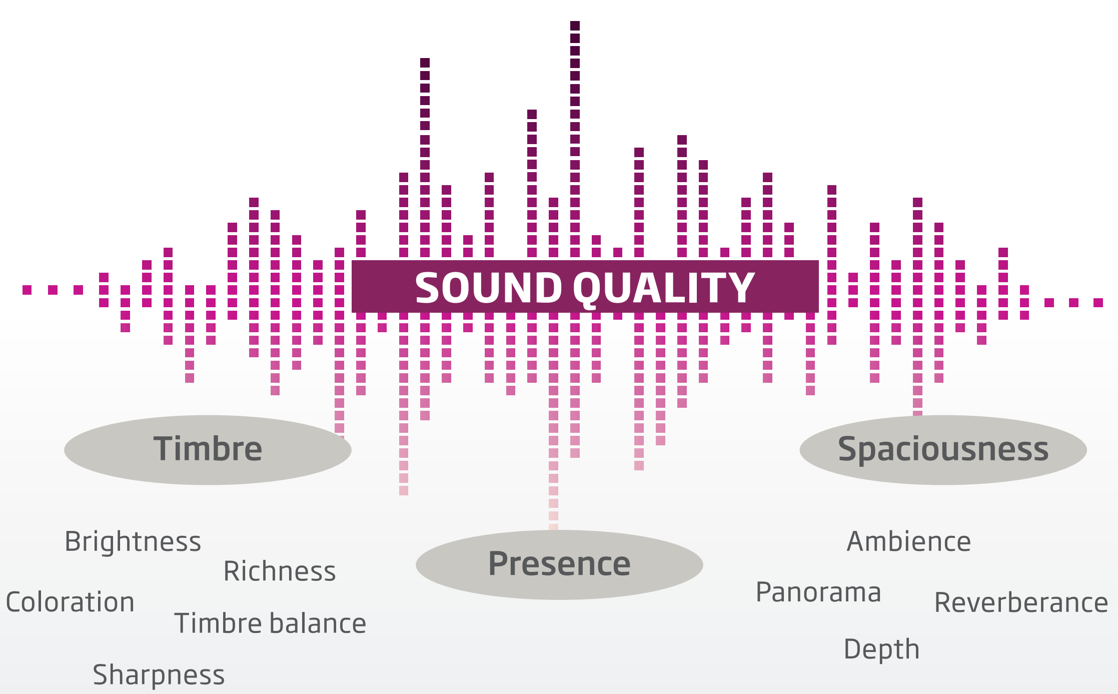

Figure 1. Sound Quality is a term that encompasses a large variety of smaller physical and psychological attributes.

The Multi Stimulus Experiment Comparison Test

Experiments that involve humans, particularly when it comes to assessing emotional qualities such as sound quality often comes with its inherent downfall – variability. This is because when judging sound quality, different physical attributes are weighed differently across listeners. Some may prefer a more balanced presentation of the spectral tones for comfort, some may only want to focus on the salience of the talker in front of them, others may simply want more richness and thus become fully immersed in their physical environment. As a result, everyone seems to have an opinion on what makes good sound quality, and this is further complicated with the wide range of sound scenes and instruments that listeners are exposed to daily.

To understand the challenges of assessing sound quality, a wide variety of questionnaires and experimental procedures have been developed to tackle the different levels and components of sound quality. Some researchers seek to take advantage of the high ecological validity of questionnaires, while others prefer a well-controlled lab setup. In this case, we wanted to strike a balance between the benefits of the two extremes, and thus a modified, double-blinded MUSHRA (ITU 1534-1, 2015), hereby referred to as the Multi Stimulus Comparison Experiment was used. Briefly put, it has the following advantages:

- It is a well validated test method that has been used in assessing other audio systems as well as hearing aids in scientific literature
- It allows quickly switching between each condition, giving a great amount of flexibility for the listener to compare between any of the conditions
- It can be easily administered to listeners in an easy to control manner

By leveraging the above advantages, the paradigm was used to compare the perceived sound quality of the Oticon More against two premium hearing aids, hereby referred to as Competitors A and B. Test participants listened to realistic recordings made in a similar manner to Santurette et al (2021): A head-and-torso simulator (HATS) mannequin fitted with the different test hearing aids was placed in the centre of a 29-loudspeaker array. A range of different sound scenes were presented and recorded by the highly sensitive microphones located at the end of the two contralateral ear canals. This produced output recordings of the hearing aids for each sound environment. The test participants subsequently rated how well each of them sounded using a scale. To further explain this, we may dissect the test interface shown below (Figure 2):

A As can be seen from the top, the description of a sound scene was displayed to the listener. There were 7 total sound scenes divided into three main categories – Music, Complex Speech and Quiet Speech. The Music

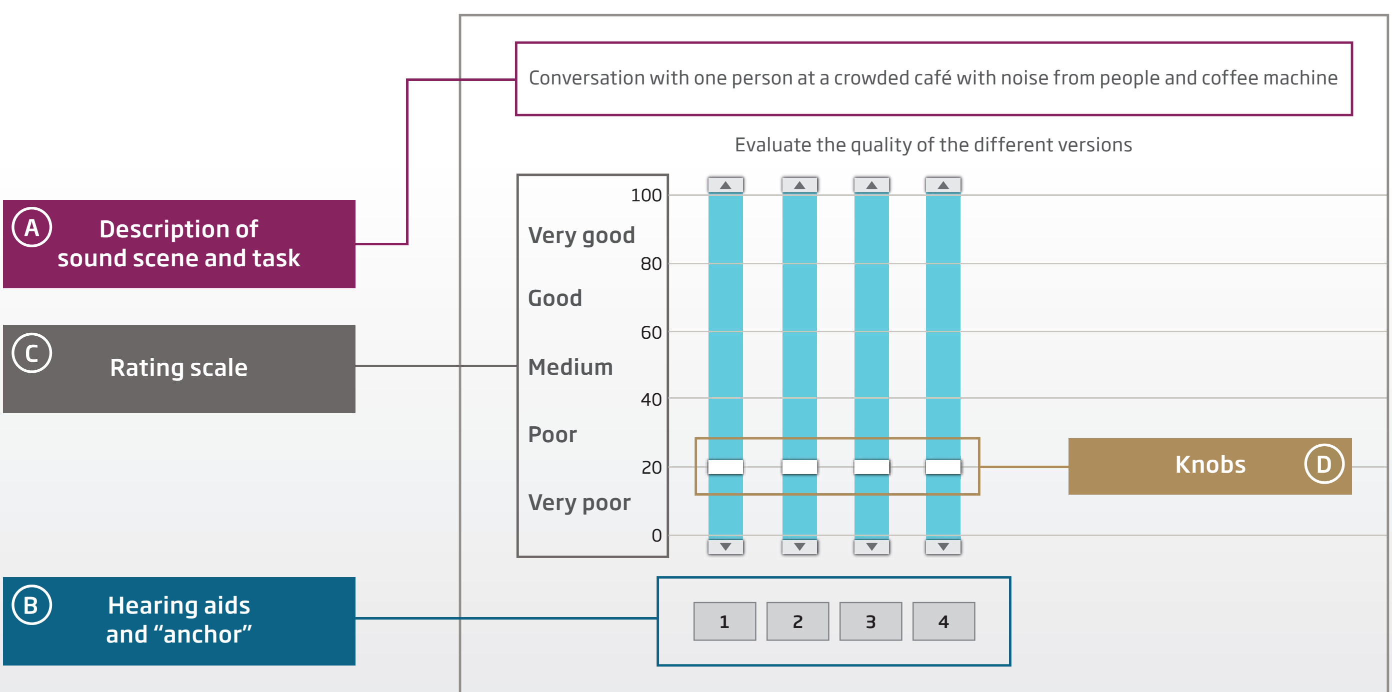

Figure 2: User Interface of the test setup

scenes contained an Orchestra, Choir and a Rock Concert; The Complex Speech scenes consisted of speech in a Café, Canteen and with a Facemask: The Café and Canteen scenes contained speech in a noisy environment while the Facemask scene involved an American male talking while wearing facemask; Finally, Quiet Speech containing the Car scene consisted of speech inside a car at low stimulus levels. All presentation levels were based on recommendations from the ARTE database (Buchholz and Weisser 2019).

B Test subjects listened to the recordings of each hearing aid by clicking on the buttons. The hearing aids tested here were identical to Competitors A and B from Santurette et al (2021) but were kept at First Fit. These were the default settings that each manufacturer prescribed, after an audiogram was inputted into the fitting software. Hence, the output recordings best reflected how each hearing aid would sound the first time a hearing aid user gets their hearing aids from the clinic. Of course, not everyone had the same level of hearing loss, and recordings had to provide sufficient gain for the wide range of hearing losses. To overcome this obstacle, recordings were done by inputting a wide range of standard audiograms (N1 to N4; Bisgaard et al 2010) into each hearing aid. To provide additional flexibility, intermediate hearing losses N2.5, N3.5 and S1.5, calculated by the mid-way point between two adjacent audiograms were also obtained. Subsequently, a given listener would have only listened to recordings that contained recordings from a hearing aid that was fitted to the standard audiogram closest to their own personal audiogram.

Adhering to the ITU 1534-1 (2015) guidelines, the hidden anchor was also added as the final condition. In this case, the hidden anchor was a "hearing aid" which sounded the poorest and was obtained by heavily distorting one of the recordings. This is critical as the anchor ensured that the listener used the full range on the rating scale, highlighting the differences between the conditions we were interested in. The reference condition was omitted, as there was no way of telling what "perfect" sounds like to hearing-impaired listeners. The same has been done in audiological scientific literature (Sanchez-Lopez et al 2020). This results in a total of 4 conditions (3 hearing aids + 1 anchor), corresponding to the four available buttons in B which the listener can switch instantly at will as many times as they want.

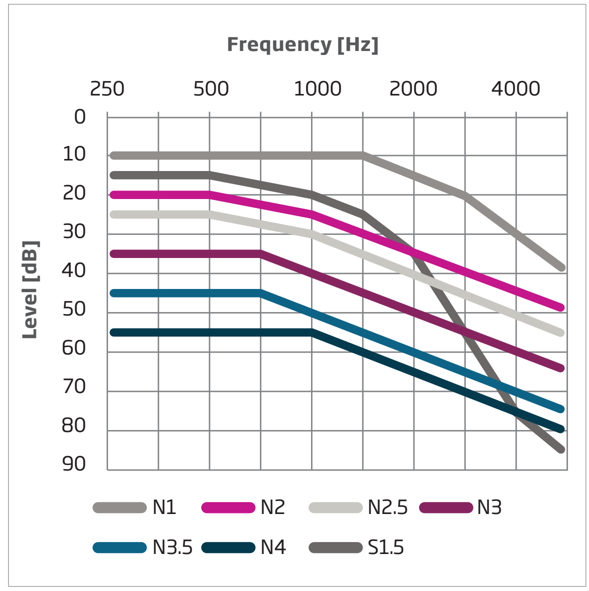

Figure 3: Standard Audiograms used for recordings (Bisgaard et al 2010) and intermediate audiograms.

C D Listeners were instructed to read the description and evaluate the sound quality and preference for each of the hearing aids for every sound scene. There was no particular focus on specific attributes such as comfort or speech clarity, as the focus was on sound quality preference. This means that the focus is at the "top of the umbrella", where it is the summation of all the different attributes (Figure 1). They were not informed of the existence of an anchor and should have thought that there were 4 hearing aids in total. The rating scale ranges from 0 to 100, with a higher score indicating a stronger preference. This was carried out by adjusting the knobs which could be done as much as needed by the listener.

22 hearing aid users were recruited to take part in the experiment. Both Oticon (n = 17) and non-Oticon (n = 5) hearing aid users took part in the study to avoid potential bias towards the Oticon devices due to familiarity to manufacturer-specific settings. The participants repeated the exact same test procedure, rating each hearing aid for the 7 sound scenes, a total of three times: The first round acted as a training phase while the other two were to collect data. The purpose of training was to verify that the test method was reliable and to familiarise them with the procedure.

Results

Overall, the ratings obtained from the experiments revealed that Oticon More performed very well compared to its competitors. Figure 4 below demonstrates an overall view of the data distribution obtained from the experiments:

It can be seen from the density plot that Oticon More populates the high end of the rating spectrum with a peak between 75 and 80. On the other hand, both Competitors A and B largely populate the lower end of the rating spectrum. This indicates a general preference towards the More sound quality. However, such is only a qualitative way of viewing the results and statistical evidence is necessary to objectively support this claim. This is explored in detail in the following two sections.

More is Rated Higher than Competitors

Statistical analyses of the ratings between each of the hearing aids were carried out. Since each sound scene was inherently different from each other, it only made sense to treat each sound scene indepedently. General linear mixed effects models were fitted to the data of each individual scene, aiming to determine what effect the type of hearing aid had on resulting ratings. This had the advantage of comparing each condition to another (More vs Competitor A, More vs Competitor B and Competitor A vs Competitor B) while also accounting for the large inter-subject variability mentioned in the beginning. In practice this was carried out using a post-hoc Tukey's honest significant difference test on the fitted model to evaluate the pairwise differences in ratings across the hearing aids with a significance level of 0.05 ($p < 0.05$) (Tukey, 1949). Figure 5 on the next page demonstrates the results.

In most cases, the hearing aids were found to have statistically different ratings, with More consistently being rated higher than the two competitors as can be seen from Figure 5. The red horizontal bars in the top of each figure indicates that no significant differences were found between two hearing aids, which was only the case between Competitor A and Competitor B in a few scenes. From these results it can be inferred that More has a perceptually and measurably better sound quality compared to its competitors.

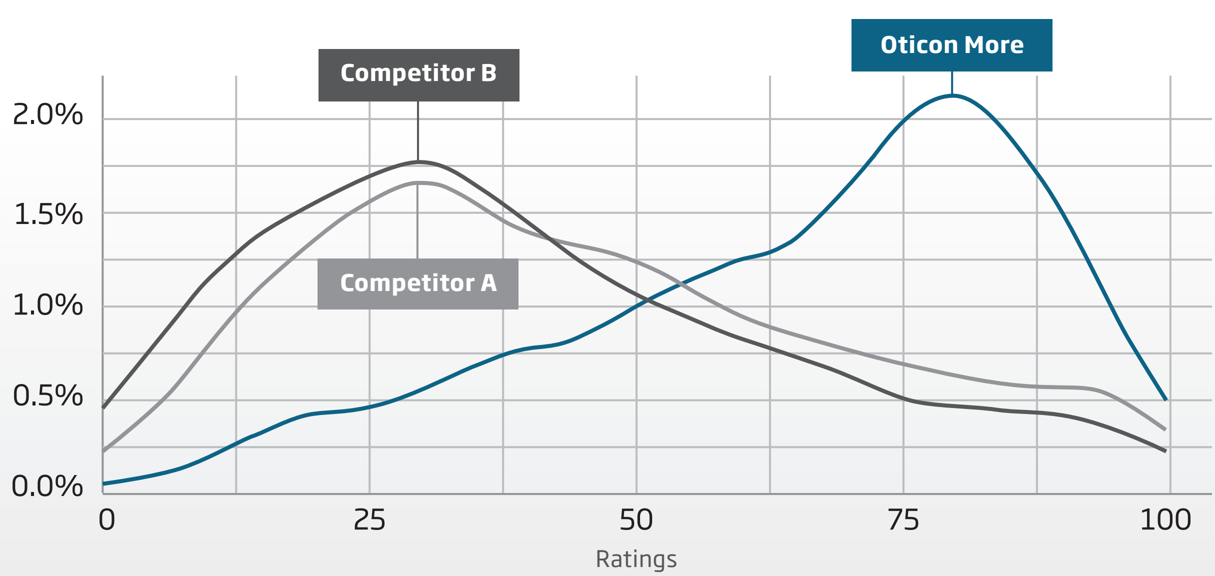

Figure 4: Density plot visualizing the distribution of ratings given to the different hearing aids in all sound scenes by the hearing-impaired listeners

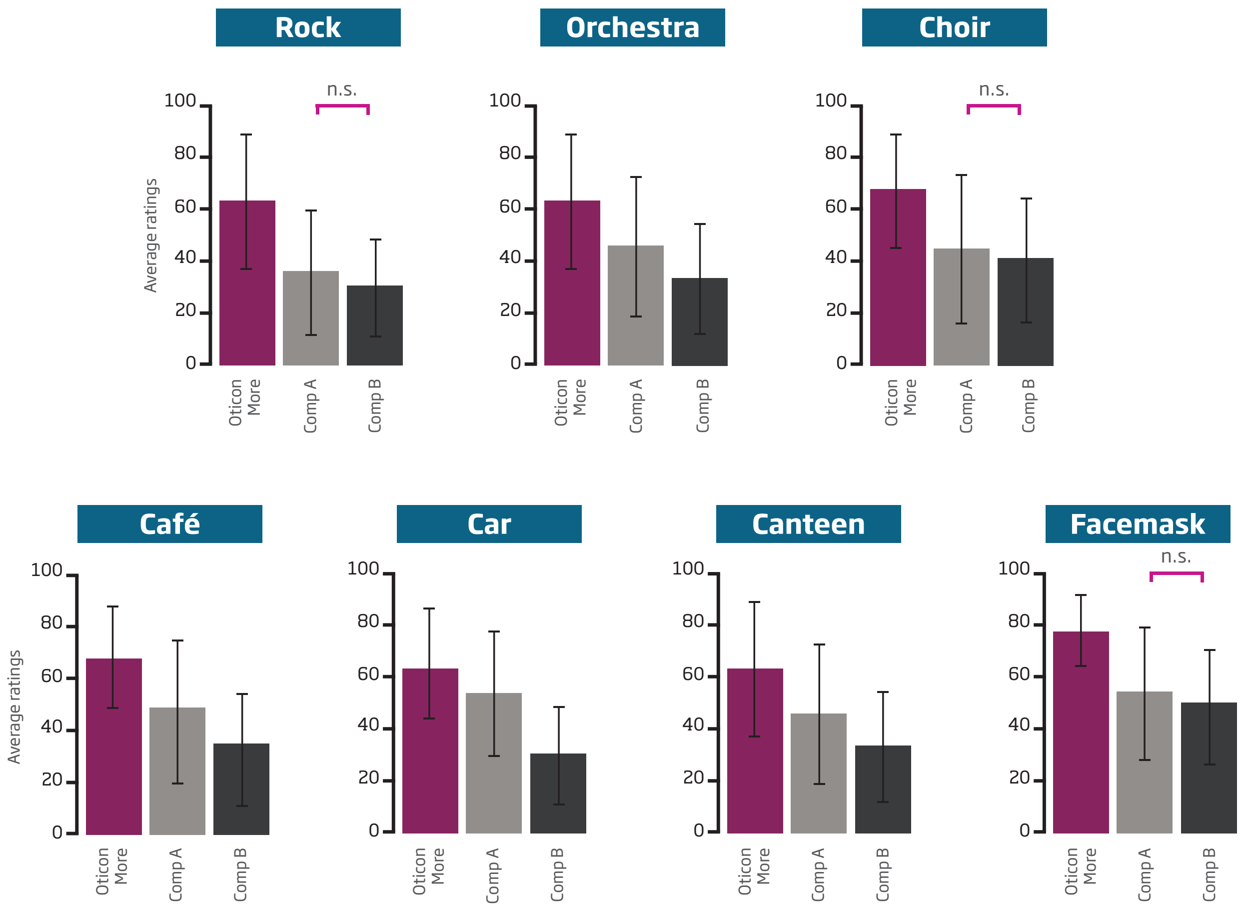

Figure 5: Overview of ratings given in the scenes displayed for More, Competitor A and B. The grey bars indicate the average value of ratings given to either of the 3 hearing aids. The black error bars indicate 1 standard deviation of the data. Horizontal pink lines above two bars indicate no significant differences between two conditions.

More is Preferred Over Competitors

The percentage preference for More over its competitors was also made. This was carried out by extracting the number of participants out of the 22 that rated More as the highest. From these a percentage score could be calculated, and the results are displayed in the Table 1 on the next page.

More was rated highest out of the three by at least 59% of the participants, and higher by at most 82%, with the majority of preference scores at 77%. Therefore, the results here reveal a preference towards More compared to both competitors, across all the sound scenes. In terms of each group, the complex speech scenes, with an average stimulus level of 73 dB SPL had an overall 75.6% preference for More. While the Choir, Orchestra and Rock scenes (average 79 dB SPL) had a mean preference of 80.3%. Finally, for Quiet Speech, possibly due to very low stimulus levels (55 dB SPL) the preference was 59%.

Conclusion

In Part 1, Santurette et al (2021) demonstrated that the new audiological perspective in Oticon More (Santurette and Behrens 2020) outperforms the traditional directionality, noise reduction and compression approaches of the two latest premium competitor hearing aids in terms of: 1) making speech stand out from the background 2) preserving speech cues important for understanding and 3) adaptation speed to changing sound scenes.

In continuation, this experiment set out to achieve two major objectives: 1) to validate the findings from Santurette et al (2021) by bringing findings from the lab to get closer to the real world, where humans are involved and 2) assess sound quality in a well validated manner, where a balance has been struck between the benefits of a well-controlled lab setup and the emotional aspects of the listening experience. Both goals were realised by utilising an adapted MUSHRA paradigm (Sanchez-Lopez et al 2020; Simonsen and Legarth 2010).

Evidence from this study shows that Oticon More outperforms two of the latest premium competitor hearing aids across a wide range of sound scenes:

- In a test of sound quality preference, More is rated to be significantly higher for a wide range of sound scenes, from music to complex speech scenes
- More is also preferred proportionally more than the two premium competitor hearing aids over a wide range of sound scenes

Group	Scene	No. of Participants who rated Oticon More is the highest (out of 22)	Percentage Preference
Complex speech	Café	16	73
	Canteen	17	77
	Facemask	17	77
Quiet speech	Car	13	59
Music	Choir	17	82
	Orchestra	17	77
	Rock	18	82

Table 1: The proportion in which Oticon More is rated as the highest compared to Competitors A and B for each of the sound scenes.

References

1. Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. Trends in Amplification, 14(2), 113–120. https://doi.org/10.1177/1084713810379609
2. Buchholz, J. M., & Weisser, A. (2019). Ambisonic Recordings of Typical Environments (ARTE) Database. https://doi.org/10.5281/ZENODO.2261632
3. International Telecommunication Union (2015) Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems. https://www.itu.int/rec/R-RECBS.1534/en.
4. International Telecommunications Union Radiocommunication Assembly. ITU-R Recommendation BS.1116-1 (1997). Methods for the subjective assessments of small impairments in audio systems including multichannel sound systems. Geneva, Switzerland: ITU-R; 1997.
5. ITU-R BS.1534-1 (2003). Method for the subjective assessment of intermediate quality levels of coding systems. International Telecommunications Union Radiocommunication Assembly.
6. Sanchez-Lopez, R., Fereczkowski, Michal, Santurette, Sébastien, Dau, Torsten, & Neher, Tobias. (2020). Data and materials from: 'Towards Auditory Profile-based Hearing-aid Fitting: Fitting Rationale and Pilot Evaluation' (1.1) [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.4421553
7. Santurette S. and Behrens T. (2020). The audiology of Oticon MoreTM. Oticon Whitepaper
8. Santurette S., Lu X., Ermert C.A., Man B.K.L. (2021). Oticon MoreTM competitor benchmark Part 1 – Technical Evidence. Oticon Whitepaper
9. Man B.K.L., Garnæs M.F., Løve S. (2021). Oticon MoreTM competitor benchmark Part 2 – Clinical Evidence. Oticon Whitepaper
10. Simonsen CS, Legarth SV (2010). A procedure for sound quality evaluation of hearing aids. Hearing Review. 17(13):32-37.
11. Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. Biometrics, 5(2), 99–114. https://doi.org/10.2307/300191
12. Völker, C., Bisitz, T., Huber, R., Kollmeier, B., & Ernst, S. M. (2018). Modifications of the MUlti stimulus test with Hidden Reference and Anchor (MUSHRA) for use in audiology. International journal of audiology, 57(sup3), S92-S104.

72697UK / 2021.06.07 / v1