WHITEPAPER
2021

Oticon MyMusic™ – Clinical Evidence

SUMMARY

Oticon is introducing the MyMusic™ dedicated music program in Oticon More™. MyMusic is the result of developing a music rationale in its own right, based on current evidence on music perception in people with hearing loss and hearing aids and recommendations for optimal music amplification and listening.

In this white paper, we share with you the clinical study performed with 23 test participants who compared Oticon MyMusic to the previous music program and the General program in Oticon More. The participants represented a wide range of hearing impairments and both avid and casual music listeners were represented. Prior to the study, a total of 8 music and speech sound scenes were recorded in order to include different types and styles of music, as well as different listening modalities (live music sound scenes, stereo/living room sound scenes, and streaming sound scenes). Each participant rated their preference in a blind and randomized setup where they listened to the sound recordings using headphones. The method used was a modified sound preference test similar to Man et al (2021). Results showed a significant preference for Oticon MyMusic over both the previous music program and the General program, for all music sound scenes. In fact, Oticon MyMusic was rated 72% higher than the previous music program, on average.

This result is a testament to the new game-changing music rationale in Oticon More hearing aids that improves the music listening experience for people with hearing loss.

AUTHORS

Brian Kai Loong Man, Regitze Kjeldal, Matilde Sørup Yssing, Marie Frederikke Garnæs and Susanna Løve

Centre for Applied Audiology Research, Oticon A/S

oticon
life-changing **technology**

Introduction

Rewards are at the root of all human actions. These actions are made in order to obtain something else, usually a predominantly pleasant emotion (Zatorre & Salimpoor, 2013). Music can bring such value through the subjective experience that it evokes. This encompasses a wide variety of conceptualizations that can be further characterized as emotions. This includes, but is not limited to, joy, sadness, and anger. In several studies, questionnaires have been administered in order to uncover test participant's reasons for listening to music. This has shown that many people use music to regulate their emotions in addition to simply relaxing. Music has also been shown to play a part in feeling of identity and belonging, both of which have been posited as essential to our well-being (Laukka, 2006).

The Differences Between Music and Speech

Hearing aids (HAs) enable listeners with hearing loss to hear sounds that are otherwise inaudible to them. Both proprietary and generic fitting rationales in hearing aids are largely based on speech models, defining a frequency and its respective dynamic range so as to prescribe a certain "target gain" to be reached by the hearing device. The result is a level of amplification that is optimized for both intelligibility and comfort to speech sounds.

Music and speech are, however, very different. Owing to the limited structural variation across vocal tracts in humans relative to musical instruments, music has a larger dynamic and frequency range than speech, and also contains large, dramatic changes uncharacteristic to speech (Chasin and Russo 2004). Moreover, music differs widely across types of instruments, ensemble (practicing alone or in an orchestra) and compositional style. The visual representation of the resulting differences in bandwidth and dynamic range can be seen in Figure 1.

As can be seen in Figure 1, music encompasses a wider range, both in terms of loudness (dynamic range) and frequency. Thus, fitting rationales optimized for speech may not necessarily translate to good musical listening experience. In fact, hearing aid users commonly report a lower perceived sound quality of music, especially for live music (Madsen and Moore 2014). Given the acoustical differences and reported difficulty with musical enjoyment from hearing aid users, there is a definite need for a music program that overcomes the challenges above.

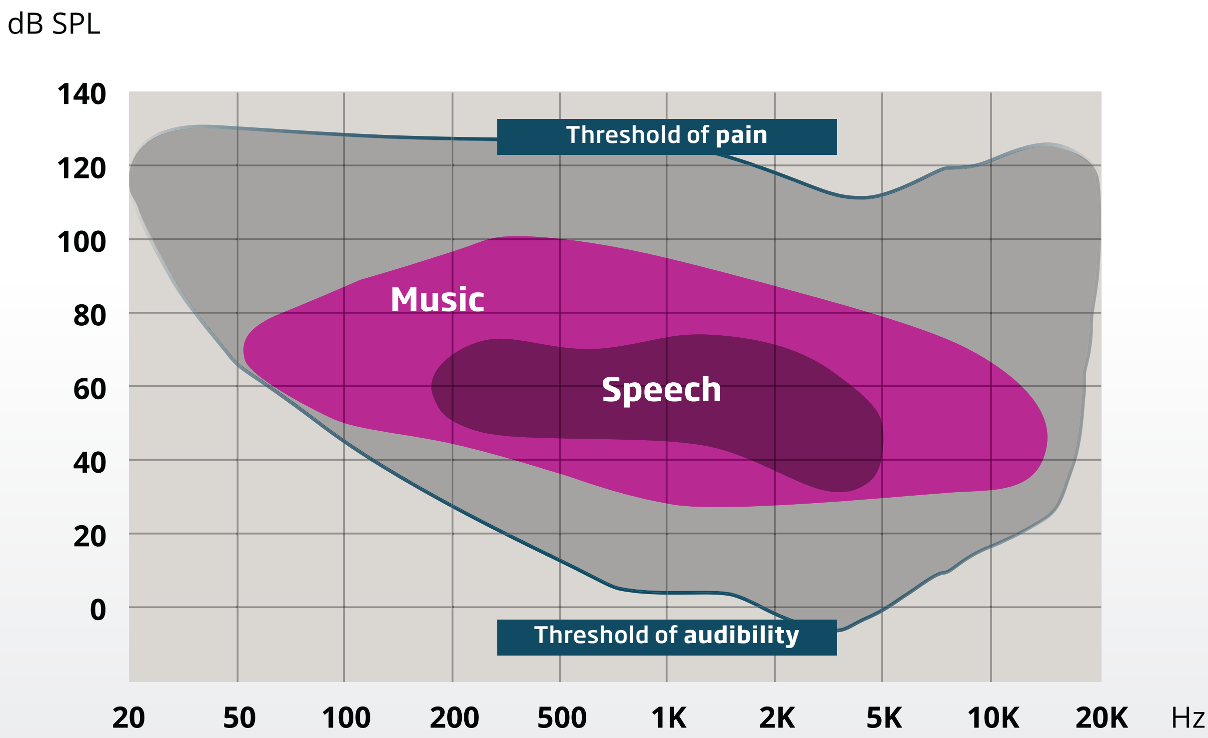


Figure 1: Frequency-intensity range of speech and music within the audibility of the human auditory system. Adapted from Vaisberg et al. (2017)

Taking in the principles that have just been discussed, Oticon now offers a new music program – Oticon MyMusic – that is designed to enhance the music listening experience of its users (for more information see Brængaard, 2021). Therefore, in order to evaluate its performance, a modified MUSHRA paradigm (see Man et al., 2021 for a detailed description of the paradigm) was used to compare the new Oticon MyMusic program to the original music program and the general speech program.

Methodology

Given that the main purpose was to evaluate the subjective sound quality of music across different hearing aid programs, the modified MUlti Stimulus with Hidden Reference and Anchor (MUSHRA) paradigm from Man et al (2021) was used. However, certain learnings were also obtained from the limitations of the previous methodology and as such, this section will briefly explain repeated concepts while focusing more on differences between the two experimental setups. In short, the order of the study was first to create a number of music sound recordings for comparison. Following this, 23 test participants with varying degrees of hearing loss were recruited to compare and rate their preferences for the music recordings. The following sections will go through this methodology in more detail:

We start by dissecting the interface of the experiment (Figure 2):

(A) A description of one of the given sound scenes was provided at the top of the interface. The objective was to include a wide variety of music sound scenes in order to capture the variety of musical environments that hearing-impaired listeners may be exposed to daily. Therefore, a total of 8 sound scenes were recorded, six of which were purely music. Two of the scenes were speech, in order to determine how Oticon MyMusic handled speech signals if a hearing aid wearer were to use this program for more general purposes, either deliberately or by mistake. The 8 sound scenes were further segregated into live, stereo and streaming scenes. The live scenes contained a choir and rock concert where ambisonic reproductions were played back from a 16-channel array to simulate the surround sound of being in a live concert. The stereo scenes, consisting of a pop song, classical music, and clean speech were used to simulate a typical "listening at home" environment where people have access to stereo systems. Finally, streaming sound scenes involved the same three stimuli as the stereo scenes, which were streamed directly from an iPhone to the hearing aids. However, in order to simulate an environment such as listening to a music or podcast on a train or café, background noise was presented in a sound field roughly 3-7 dB lower than the target stimuli. The resulting streaming conditions were pop streaming, classical streaming, and speech streaming respectively. Table 1 below provides a summary.

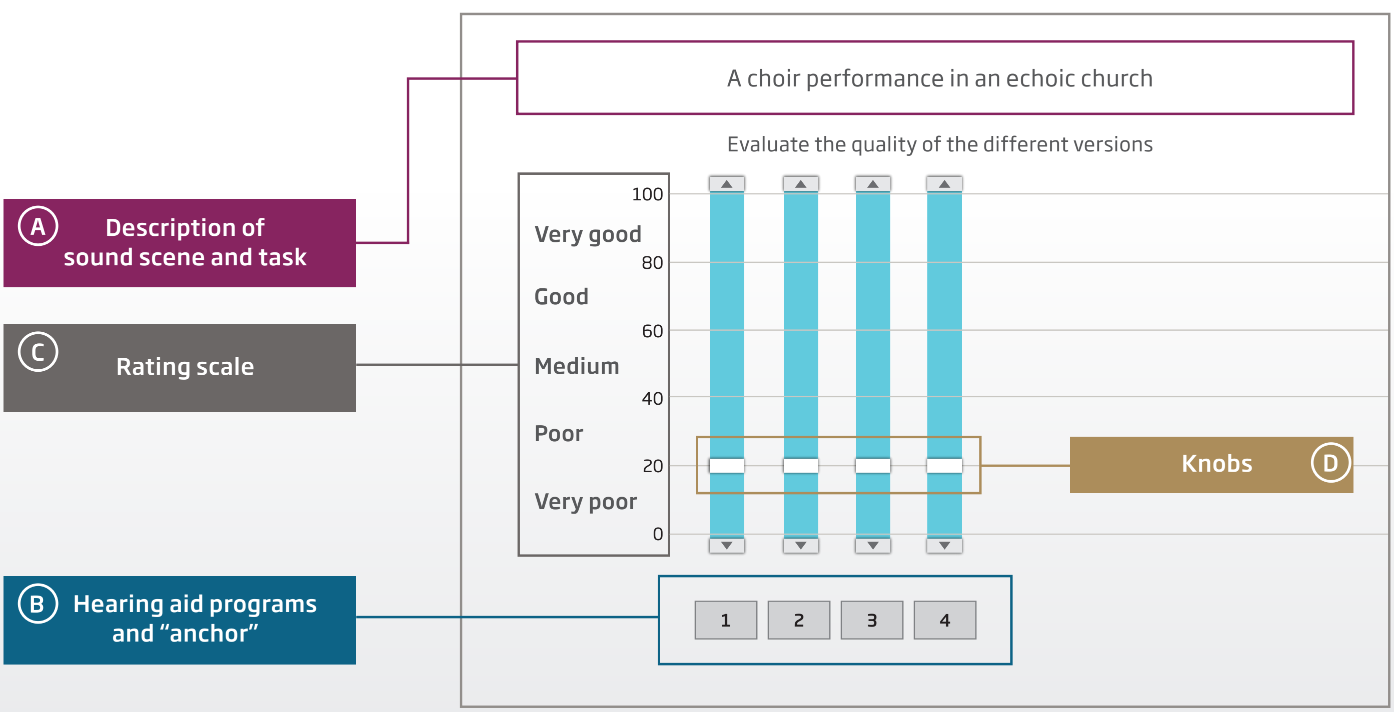


Figure 2: The experimental interface. Due to blinding of the experimental conditions, the subject would not have known that there were two identical hearing aid programs among the conditions and which button corresponded to which experimental condition

The outputs of the three hearing aid programs were recorded using a pair of Oticon More hearing aids with a closed micro-mold fitted to a head-and-torso simulator (HATS) placed in the center of a 16-loudspeaker array. The justification for using occluding acoustics was to ensure that the recordings obtained were as a result of hearing aid signal processing, instead of sound that has directly passed into the ears of the listener. Moreover, including recommended acoustics would mean introducing degree of occlusion as a variable which may again reduce the validity of attributing perceived differences as a result of the music handling algorithms within the programs.

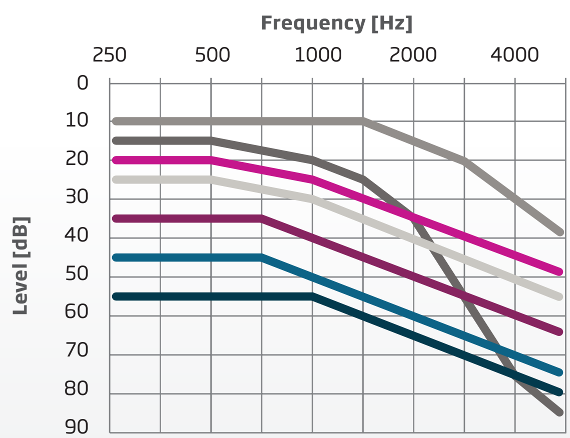
B Test subjects listened to recordings from three different hearing aid programs: MyMusic, the previous music program (Prev), and the Oticon General program (Gen). The Prev music program refers to the dedicated music program before MyMusic while the general speech program (Gen) is the default program. Once again, the experiment had to account for the range of hearing sensitivities of the test participants. Hence recordings were performed with the hearing aids fitted to the same audiograms as Man et al (2021, see figure 3). Consequently, for a given test participant, they would've only listened to recordings from hearing aids that were fitted to the standard audiogram closest to their own personal audiogram.

Unlike Man et al (2021), a change to the hidden anchor (ITU 1534-1, 2015) was made. Instead of a low-level anchor which should sound the poorest of the experimental conditions, a mid-level anchor was used. This was defined as a condition in which is known or assumed to be rated in the middle, between the best and worst conditions. An assumption was made that the Prev music program would fill this role. As a result, an additional copy of the Prev music program was added as the mid-level anchor. Effectively, this meant that the subjects had to rate four experimental conditions blindly, two of which were in fact identical. The mid-level anchor was chosen over the low-level variant based on the following reasons: Firstly, results from Man et al (2021) revealed that although one condition was rated considerably higher than the others, the more poorly rated conditions differed little in terms of ratings and their differences could not be captured. That was unexpected and hence by placing the anchor in the middle, rating differences between each experimental condition should be more pronounced. Secondly, as two of the buttons corresponded to the same recording (Prev and anchor), listeners should theoretically rate both at least very similar to each other. This acted as a post-hoc method to assess whether participants understood the task well enough and were not simply rating randomly.

C **D** Each participant was asked to listen to the four conditions (3 HA programs + 1 anchor) for each sound scene, and rate them on the rating scale presented on their screen ranging from 0 (very poor) to 100 (very good). This was done by adjusting the four knobs on the screen for every experimental condition. The

Group Condition	Sound Scene	Level Target	Level Noise
Live music	Choir	78 dB SPL	-
	Rock (Eagles)	80 dB SPL	-
Streaming with background noise	Pop streaming	78 dB SPL	71 dB SPL
	Classical streaming	76 dB SPL	71 dB SPL
	Speech streaming	74 dB SPL	71 dB SPL
Stereo (listening at home)	Pop Stereo	75 dB SPL	-
	Classical Stereo	75 dB SPL	-
	Speech Stereo	65 dB SPL	-

Table 1: Summary of all sound scenes and corresponding presentation levels



Distribution of participants in each Audiogram						
N1	N2	N2.5	N3	N3.5	N4	S1.5
1	3	5	4	2	7	1

N1 N2 N2.5 N3 N3.5 N4 S1.5

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

participants were able to switch between the conditions as many times as they wished before moving forward to the next sound scene. Much like Man et al (2021), there was no particular focus on specific attributes such as clarity or comfort, as the focus was on sound quality as a whole.

For this experiment, 23 test participants with age-related sensorineural hearing losses ranging from mild to severe were recruited. Their ages ranged from 48 to 83 with a mean age of 68 (standard deviation = 9.2). Among them, seven were avid listeners while sixteen were casual listeners of music. This was determined by a questionnaire regarding their music listening habits before the experiment. Each participant was asked to do a training round to familiarize themselves with the method, afterwards the test was repeated twice again for actual data collection.

Results

After gathering all the data obtained, a useful place to start is by presenting the data as it is. The ratings belonging to the anchor and Prev were merged, as they are in fact the same program and were rated to be very similar to each other based on observations. The distribution of all the ratings is summarized in the probability density plot below (Figure 4).

In order to understand the probability density plot, one must read in terms of the area under a given curve. Under each colored curve is a total area of 1, or 100% of all the datapoints resulting from a given experimental condition. From there, one can derive the area belonging to a certain range along the x-axis to determine the proportion of the samples who provided the corresponding ratings. For example, under the blue curve (MyMusic) a large proportion of its area is within the range of ratings between 75-85. This means that MyMusic has been given such rating a lot of times proportionally. Meanwhile, a large proportion of the area under the dark grey curve (Gen) belongs within the range of ratings between 20-30 and the area underneath the light grey curve (Prev) is located primarily between the ratings of 20-45. Comparing the results of the avid and casual listeners did not prove to highlight any meaningful differences between them. Therefore, it was decided to keep them

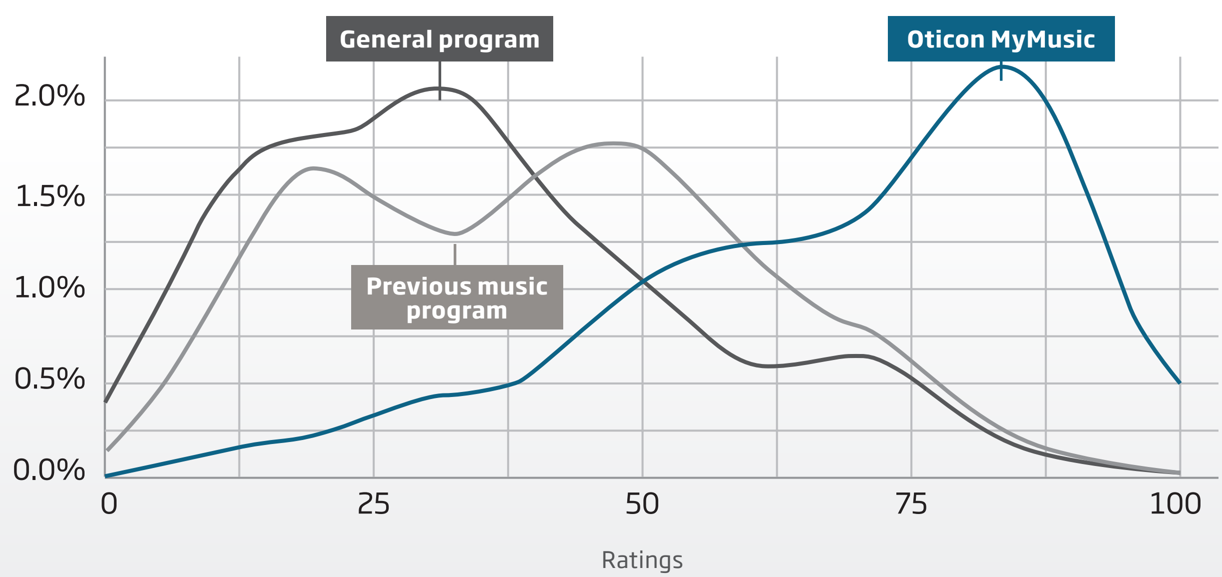


Figure 4: Probability density plot for all obtained data. Data from the old music program and anchor were combined, given they did were not rated differently after post-hoc observations

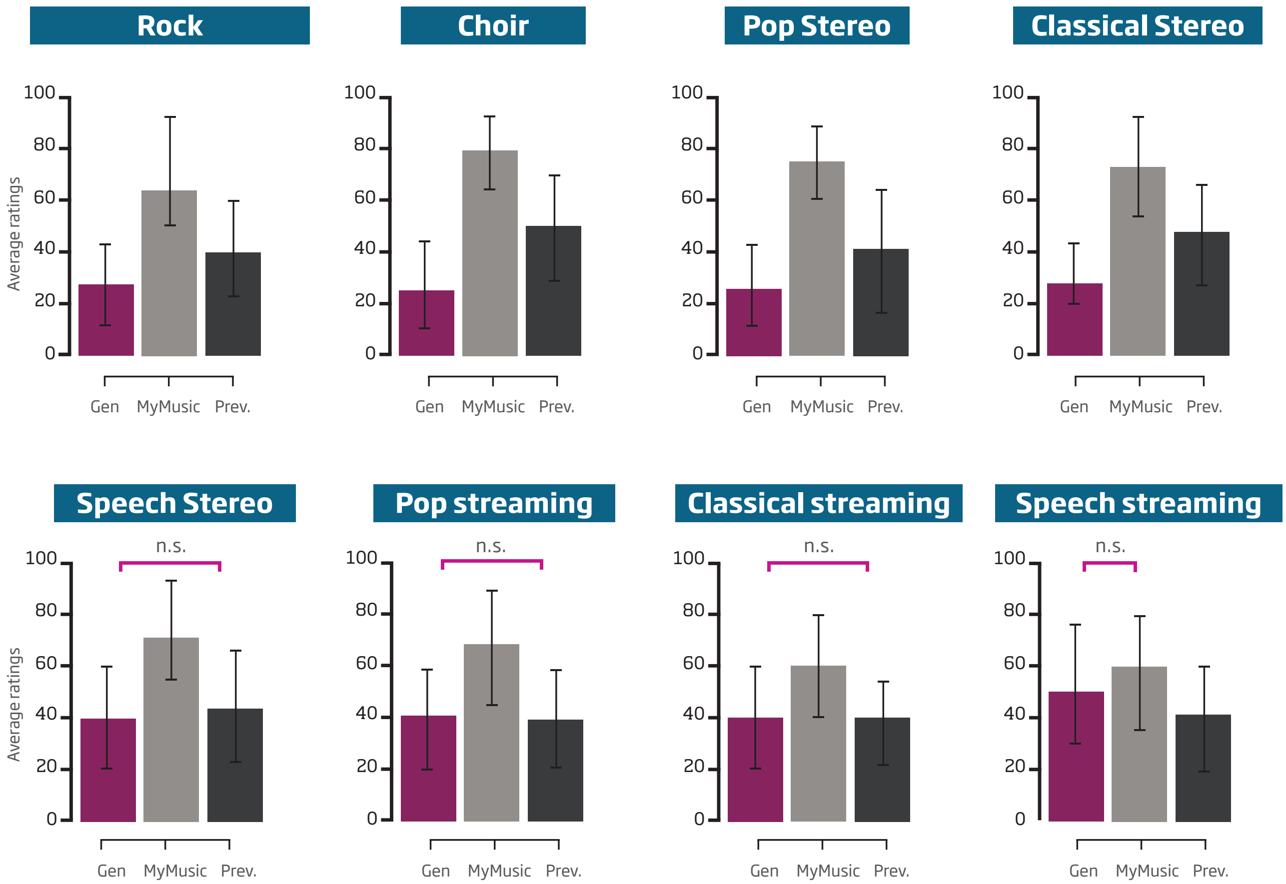


Figure 5: Pairwise comparisons (p = 0.05) between each experimental condition across the different LME models from each sound scene. Magenta lines above each pair of bar plots indicate non-significant differences. This was done so as not to overpopulate the graphics with too many lines above the bars.

within the same sample group for subsequent analyses.

So far, though seemingly pointing to a clear conclusion, the above figure is only a qualitative visualization of the data. In order to statistically quantify whether the MyMusic program was truly preferred over the other two programs, a statistical analysis had to be performed. To do so, a general linear mixed effect (LME) model with hearing aid program as the fixed effect and the participant identification as the random effect was fitted to the data of each sound scene independently. This resulted in 8 LME models, each revealing the effect of different hearing aid programs on rating scores. By treating hearing aid programs as the only fixed effect, the model has the advantage of accounting for large inter-subject variability inherent to sound quality ratings (Man et al 2021). Subsequently, Tukey's honest significant differences test (Tukey 1949) was used to quantify all possible pairwise comparisons at a 0.05 significance level. Figure 5 summarizes the pairwise differences of the sound scenes.

Each of the 8 bar charts shown displays the average ratings and corresponding standard deviation of the ratings for each hearing aid program. The brackets above any two bars indicate two conditions in which no significant differences were found. In all cases but one (speech streaming), MyMusic was rated significantly higher than both Prev and Gen ($p < 0.05$). Finally, by averaging the ratings for the six music sound scenes, MyMusic, Prev and Gen were 70.63, 41.01 and 30.76 respectively. Therefore, it was concluded that MyMusic was rated 72% higher than the previous music program.

Discussion

The results of this clinical preference test comparing Oticon MyMusic to the General program and the previous music program showed outstanding performance and this can be attributed to our new and fundamentally different approach to amplifying music in hearing aids. The results followed our prediction that the General program would be rated the lowest, followed by a higher rating for the previous program (thus validating our assumptions regarding the suitability of the anchor), and finally a very high rating for Oticon MyMusic.

Some readers may be surprised by the findings from the speech scenes – why was MyMusic rated higher than the General program for the Clean Speech scene? To answer this, we may look into the effects of compression across scenarios of increasing complexity. For the purpose of perceived sound quality, listeners tend to prefer lower compression ratios (or more linear amplification) for speech in quiet without compromising speech recognition (Boike and Souza, 2000; van Buuren, Festen and Houtgast 1999). The relationship becomes more ambiguous in challenging speech environments with noise, as compression may be required to suppress noise that is too loud in order to avoid loudness discomfort. This interaction is demonstrated clearly by the divergent findings between Speech Stereo and Speech Streaming. In Speech Stereo, MyMusic was reported to be significantly better than the General program. This finding was however, not repeated in Speech Streaming where noise was introduced. Thus, one may speculate that the General program might perform even better as the noise level increases even more. The two speech scenes in the study were included to ensure that even hearing aid listeners who use MyMusic for speech listening at times, can still use this program in a satisfactory way. The study results confirm this, but also show why it is still widely recommended to use the General program for listening in diverse, everyday sound environments where speech is present and all the available sound processing in Oticon More is utilized to the fullest extent when needed. For more technical information regarding changes to compression and other features, see Brændgaard (2021).

Conclusion

Across 23 test participants with a wide range of hearing loss and music preferences, the Oticon MyMusic program was, on average, rated significantly higher than the previous music program and the Oticon General program for all music sound scenes. This was true for live music listening, stereo music listening, and music streaming scenarios. For the six music sound scenes, Oticon MyMusic was rated 72% higher than the previous music program and higher still when compared to the General program. This is attributed to the fundamentally different amplification approach applied in Oticon MyMusic, where the unique properties of music are taken into account in every aspect of the signal processing strategy. It is quite clear that people with hearing impairment have very different listening needs for speech and music, and it is therefore highly recommended to consider adding Oticon MyMusic as a program in Oticon More hearing aids to enhance the quality of life for any listener who enjoys music.

21-217351 15500-0236 / 10.21

References

1. Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard Audiograms for the IEC 60118-15 Measurement Procedure. Trends in Amplification, 14(2), 113–120.
2. Boike, K. T., & Souza, P. E. (2000). Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification. Journal of Speech, Language, and Hearing Research, 43(2), 456-468.
3. Brændgaard, M. (2021). The development behind Oticon MyMusic. Oticon Techpaper.
4. Chasin, M., & Russo, F. A. (2004). Hearing Aids and Music. Trends in Amplification, 8(2), 35–47.
5. ITU-R. (2015). Recommendation ITU-R BS.1534-3: Method for subjective assessment of intermediate quality level of audio systems. Geneva: International Telecommunication Union.
6. Laukka, P. (2006). Uses of music and psychological well-being among the elderly. Journal of Happiness Studies, 8(2), 215–241.
7. Madsen, S. M., & Moore, B. C. (2014). Music and Hearing Aids. Trends in Hearing, 18, 2331216514558271.
8. Man B.K.L., Garnæs M.F., Løve S (2021). Oticon More Competitive Benchmark Part 2 – Clinical Evidence. Oticon Whitepaper.
9. Pallavi, J., Perumal, R. C., & Krupa, M. (2018). Quality of Communication Life in Individuals with Broca's Aphasia and Normal Individuals: A Comparative Study. Annals of Indian Academy of Neurology, 21(4), 285–289.
10. Tukey, J. W. (1949). Comparing individual means in the analysis of variance. Biometrics, 99-114.
11. Vaisberg, J., Folkeard, P., Parsa, V., Macpherson, E., Froehlich, M., Littmann, V., & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. Audiology Online.
12. van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. The Journal of the Acoustical Society of America, 105(5), 2903-2913.
13. Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. Proceedings of the National Academy of Sciences, 110(2), 10430–10437.