FEBRUARY 2016

BEFLEX: IMPROVING THE SELECTION PROCESS

BARBARA SIMON, AuD
CHRISTOPHE LESIMPLE, B.SC.
JULIE TANTAU, AuD

ABSTRACT

Selecting the correct hearing instrument can be daunting for a client. How can the hearing care professional (HCP) help them make the right decision, a decision that will best fit their lifestyle and hearing needs? This article describes a new device offered by Bernafon called BeFlex, which allows the client to compare up to three performance categories in one instrument. BeFlex was tested internally in two sessions. The first test evaluated the different performance categories by direct comparison in a lab test. The second was a field test that allowed the clients enough time to use the devices outside of the clinic in their everyday environments. Overall, the testing showed that nearly all clients could distinguish between the categories and found that BeFlex would be helpful when making a purchase decision. In addition, the lab test showed that having advanced features helped reduce listening effort and provided better sound quality. BeFlex's direct comparison is a good option for clients who are uncertain and need more support before making a final purchase decision.

bernafon
Your hearing · Our passion

The experience of wearing instruments and trying different categories will facilitate their purchase decision and give them an important role in their potential success with hearing instruments.

INTRODUCTION

With the wide variety of hearing instruments available, clients may have difficulty deciding which instrument to purchase, especially when they are first time users. Typically, the HCP makes recommendations about technology level and style based on various information including the client's hearing loss, dexterity level, and lifestyle. Clients may even have the opportunity to try a pair of demo hearing instruments in the clinic during their appointment. However, what if the client is still unsure about which technology level or style is best for them or whether they would even benefit from hearing instruments? Instead of making a purchase with which they're not confident or not making a purchase at all, the HCP now has a new solution to offer. This solution is called BeFlex. It is a new trial hearing instrument from Bernafon that allows the HCP to program up to three different performance categories into one device. The client can then wear the instruments for a specified amount of time at home and in their typical daily environments. The experience of wearing the instruments and trying different categories will facilitate their purchase decision and give them an important role in their potential success with hearing instruments.

An internal study was conducted at Bernafon in Bern to determine:
1) whether word recognition testing would show any significant results between the performance categories, and 2) whether the clients could perceive a difference between the performance categories when worn over an extended period of time. To answer question (1) a lab test was conducted, and to answer question (2) a field test was conducted.

TEST CLIENTS

Eight test clients participated in the testing. There were 2 women and 6 men between the ages of 53 and 80 years of age. The test clients had, on average, a mild to severe sloping bilateral hearing loss. Figure 1 shows the averaged audiograms for the eight test clients. The error bars show the standard deviation for each frequency. All were experienced bilateral hearing instrument users.

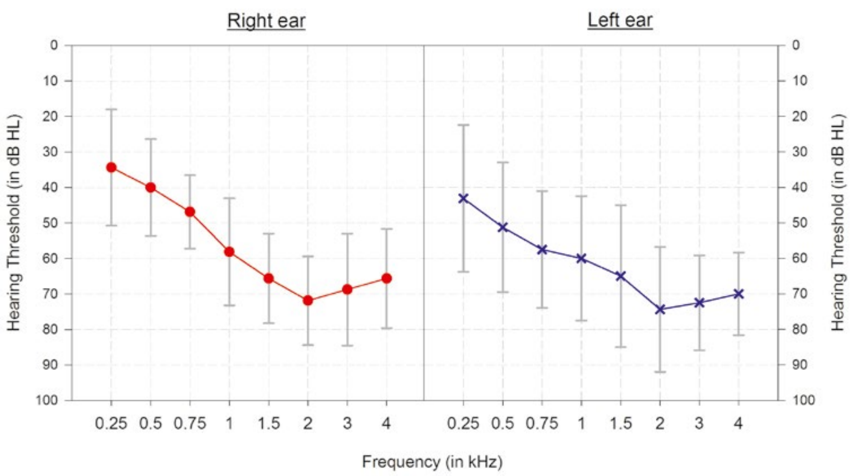

Figure 1. Average thresholds with the standard deviation for each frequency.

LAB TEST

Methods. The test instruments used were BeFlex Power BTEs and were programmed using the NAL-NL2 rationale. The first fits were matched to NAL-NL2 targets on the Verifit real-ear measurement equipment using the REM Speechmap function with the ISTS input signal at 65 dB. Fine-tuning of the gain was made when necessary for each individual subject.

For the lab test, the BeFlex devices were programmed with three performance categories: Category 9 (P1), Category 1 (P2), and a Control Omni (P3). For P1 and P2 the default settings were used so that all features were programmed as they would be for normal use. As the control, P3 was programmed as a Category 9 but with omni-directional directionality and with all features turned off.

For the lab test, the BeFlex devices were programmed with three performance categories: Category 9 (P1), Category 1 (P2), and a Control Omni (P3).

The test stimulus was the current implementation of the "WaKo" Monosyllabic Rhyme Test (1989). This test is designed to measure word recognition and may be conducted in quiet or with background noise. It is a closed set test, which offers the test items either on a touch screen or on paper. Each tested item includes the correct test response and 4 additional rhyming (or similar) words. The test is automatic, meaning that each test item is automatically presented once a response is selected, and requires no additional support from the test audiologists.

The test stimulus consists of various lists of 47 original WaKo words spoken by a male speaker. Each list is built with monosyllabic words that are phonetically balanced and close to the German phoneme occurrence distribution (Wallenberg & Kollmeier, 1989). For the lab test, WaKo noise was presented continuously for each list. To engage the directional microphones and environment detection, a 5-second steady-state noise precursor was played before the start of each list. The noise is a steady-state noise that was developed based on the long-term spectrum of all the tested words. Speech was presented via a single loudspeaker at zero-degree azimuth in the sound field, 110 cm from the ear, and noise was presented via a single loudspeaker at 180 degrees. The touch screen (Elo EL1529) used to collect the answers from the test clients was placed at a reachable distance in front of the listener to ensure comfort and to maintain the distance from the loudspeaker during the test. The stimuli were routed through a GSI 61 two-channel clinical audiometer in order to control speech and noise presentation levels.

To avoid any SNR confounding bias, word recognition was tested at a fixed 0 dB SNR in the following conditions:

i. Unaided
ii. Aided Omni Directional (Control Omni)
iii. Aided with True Directionality (Category 9)
iv. Aided with Automatic Directional (Category 1)

Each presentation of a word included 5 word choices; therefore, it was a forced choice even when they were unsure. The test clients indicated their selection via the touchscreen, using the Praat software (Boersma, 2001). Word recognition was measured as the number of correct responses for each list. In addition to recording answers, the software measures the response time and sound quality. The response time is the time in seconds between the stimuli and the client's answer and is used as an indication of listening effort. Lastly, sound quality was measured by the subjective rating of the clarity of each word using a 1 (poor) to 6 (excellent) categorical scale. This rating was indicated via the touchscreen after the selection of each word.

Results. For the data analysis, the software package Sigmaplot 12.5 was used to perform analyses of variance and any further paired tests or post hoc tests. The re-sampling procedure, bootstrapping, was made with Visual Inference Tool (on iNZight 2.0.3) that runs under R (R Core Team, 2012). These tools were used to evaluate the effect of the test conditions (Unaided, Omni, C9, and C1) on the variables (word recognition, response time, and sound quality) and determine whether the effect was significant.

Within all the aided conditions, only the difference between Category 9 and Omnidirectional was significant.

Word Recognition. The percentage of correctly identified words was transformed into rationalized arcsine unit (rau) before the analysis of variance (Sherbecoe R.L. & Studebaker G.A., 2004). The value in rau is comparable to the percentage for scores between 20 and 80%.
Word recognition scores were significantly affected ($F_{3, 21} = 15.63$, $p<0.001$) by the test condition. Bonferroni post hoc tests revealed a significant improvement in word recognition between the unaided condition (average = 35.3 rau and SE = 7.8 rau) and all aided conditions: Omnidirectional (average = 51.0 rau and SE = 4.4 rau), Category 1 (average = 56.1 rau and SE = 4.0 rau), and Category 9 (average = 62.5 rau and SE = 3.9 rau). Within all the aided conditions, only the difference between Category 9 and Omnidirectional was significant ($p = 0.03$).

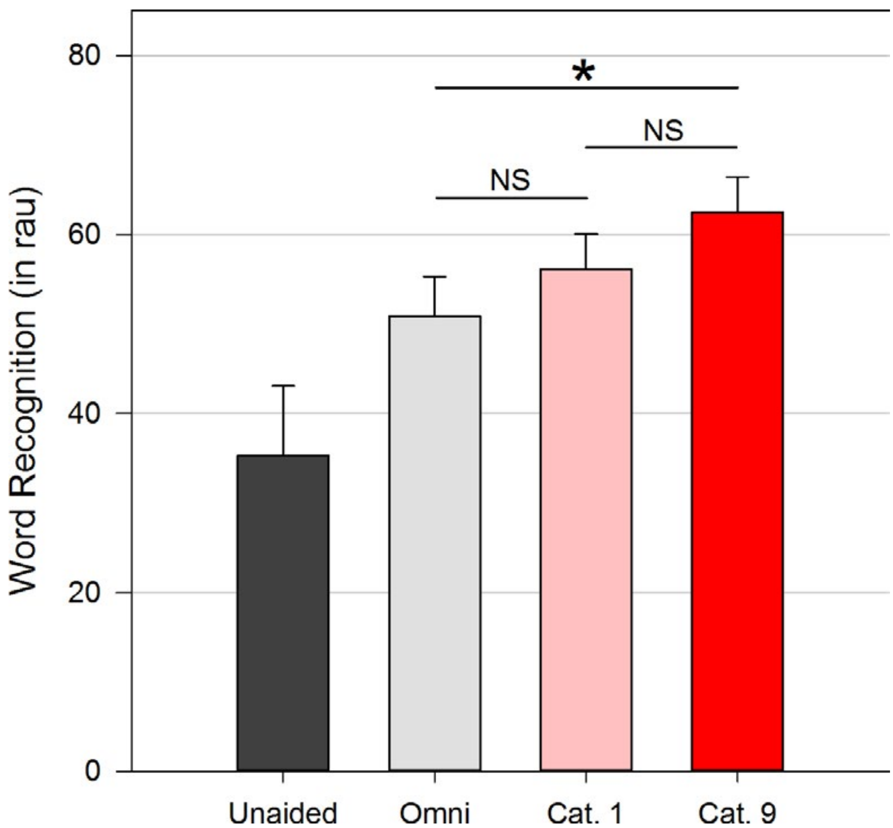

Figure 2. Average word recognition scores (in rau) for each test condition.

Figure 2 shows the average word recognition scores. The star between the Omni and Category 9 represents a significant difference between those two aided conditions (* = p<0.05). In addition, although not marked, a significant difference was found between the unaided and all three aided conditions. No significant difference was found between Omni and Category 1 or between Category 1 and Category 9.

Response Time. The response time (RT) data are not symmetrically distributed (Whelan, 2008). They have a positive skew that can affect the mean RT in any one test condition for a test participant. To find the best estimation for the central tendency, the median RT was resampled 1000 times using a bootstrapping procedure. Bootstrapping is an appropriate method for analyzing measurement statistics in situations where observed values do not fulfill normality.

The test condition had a significant effect on response time (F 3, 21 = 9.04, p<0.001). Bonferroni post hoc tests revealed a significant improvement (p = 0.002) in response time between the Unaided condition (average = 2.04s and SE = 0.22s) and Category 1 (average = 1.21s and SE = 0.08s) and a similar improvement (p < 0.001) between the Unaided condition and Category 9 (average = 1.09s and SE = 0.08s).

A significant difference in response time was found between the Unaided condition and Category 1 and between the Unaided condition and Category 9.

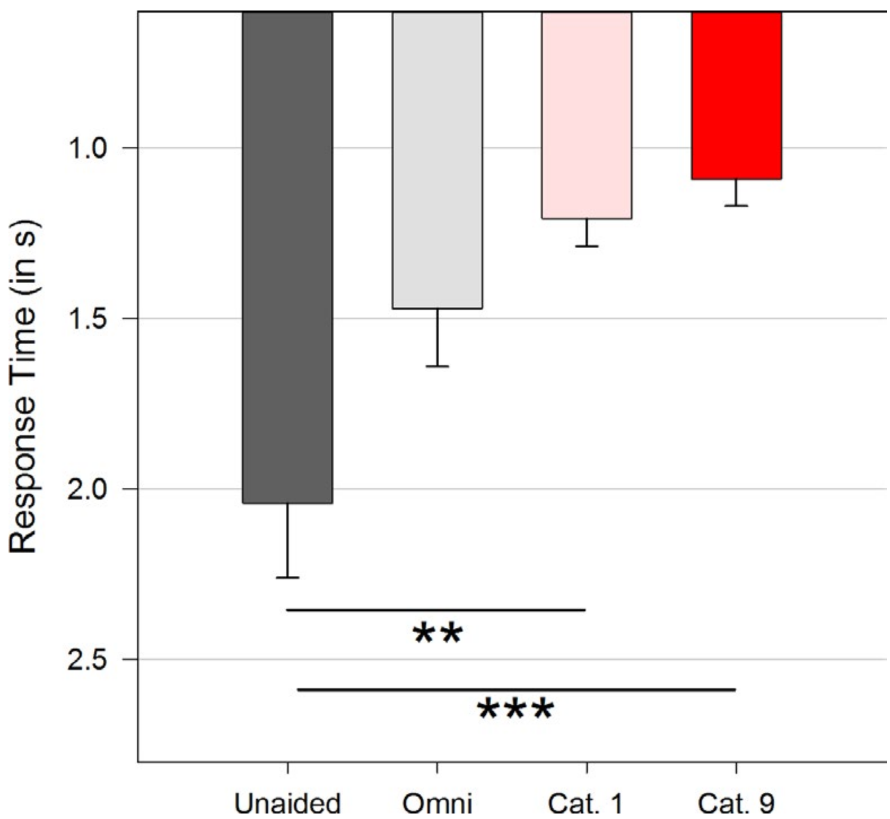

Figure 3. Average response time (in seconds) for each test condition.

In Figure 3, the average difference in response time is shown for the Unaided, Omni, Category 1, and Category 9 conditions. A significant difference in response time was found between the Unaided condition and Category 1 and between the Unaided condition and Category 9. The number of stars signifies the level of significance (** = p<0.01 and *** = p< 0.001). There was no significant difference between the Unaided condition and Omni or between any of the aided conditions.

Sound Quality. The sound quality was evaluated using a subjective rating (1-6) of each word. The subjective sound quality rating scale uses a scale that is similar to the Swiss school system notation scale. It was believed that all the test clients would share a common understanding of the scale without having to be instructed multiple times (i.e., 1 is the worst score, 4 the average, and 6 stands for the best score). Repeated ordinal data from the subjective sound quality rating can be used to rank sound quality between different test conditions. However, the ranking does not specify the magnitude of the differences; for this reason, the raw averages across each list cannot be used. To address this point, resampling using bootstrapping was applied, as suggested by Svensson (2003), to estimate the central tendency value for each test condition.

The Shapiro-Wilk test did not find any deviation from a normal distribution ($p = 0.346$). Testing for equal variance between the test conditions ($p = 0.265$) indicated that the assumptions for parametric analysis of variance were met. The test condition had a significant effect on the subjective sound quality rating (F 3, 21 = 30.7, $p<0.001$). Bonferroni post hoc tests revealed that the differences between all test conditions were significant with the exception of the difference between the Omni and Category 1 conditions.

The star represents a significant difference between Category 9 and Omni and between Category 1 and Category 9.

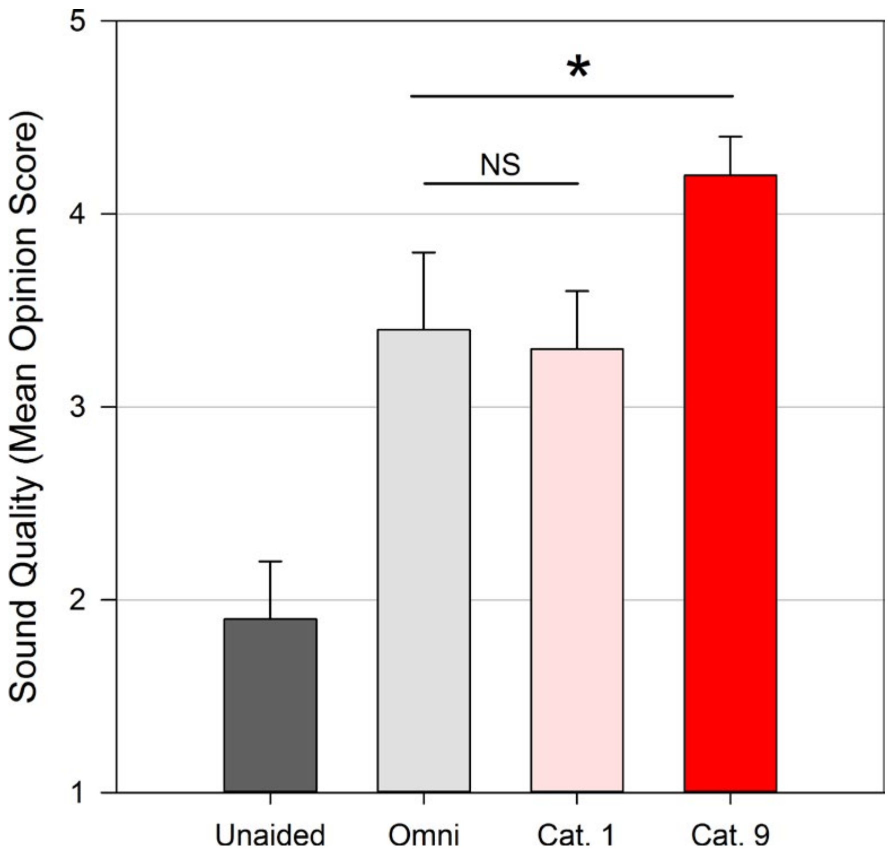

Figure 4. Average sound quality scores (rating system) for each test condition.

In figure 4, the average difference of sound quality ratings for each test condition is shown. The star represents a significant difference between Category 9 and Omni and between Category 1 and Category 9.

These results suggest that word recognition can be improved simply by having a hearing instrument.

Although not marked on the graph, a significant difference was also found between the Unaided condition and all the aided conditions. The only non-significant difference was between Omni and Category 1 (as shown on the graph).

Overall Results. In order to summarize the tested effects (RT, SQ and WR) and the tested conditions, the overall findings are shown in Table 1 and represented visually in the multidimensional paradigm in Figure 5 below. It is clear that the unaided findings have the lowest rating of all three tested effects. When the test clients were aided with an Omni-directional system, the scores for all three effects significantly improved over the unaided condition. When features were added with Category 1, a small but non-significant improvement was seen in WR and RT, but not in SQ. When the more advanced features from Category 9 were added, a significant improvement was seen in WR and SQ effects as compared to the Omnidirectional, but only in SQ, when compared to the Category 1 program.

	Word Recognition (in rau)		Response Time (in s)		Sound Quality	
	Average	*SE*	Average	*SE*	Average	*SE*
Unaided	35.3	*7.8*	2.04	*0.22*	1.9	*0.3*
Omnidirectional	51.0	*4.4*	1.47	*0.17*	3.4	*0.4*
Category 1	56.1	*4.0*	1.21	*0.08*	3.3	*0.3*
Category 9	62.5	*3.9*	1.09	*0.08*	4.2	*0.2*

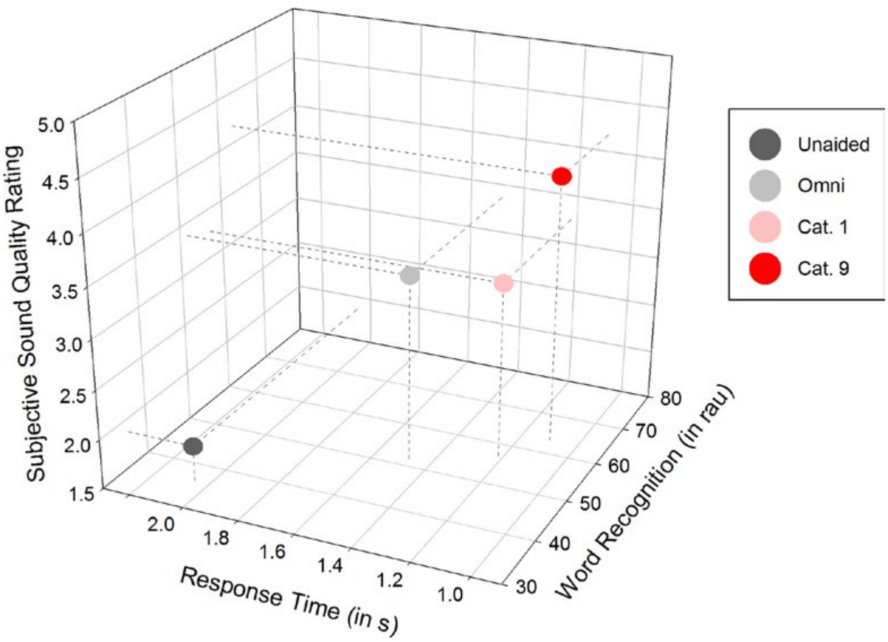

Figure 5. Multidimensional graph of all three tested variables.

These results suggest that WR can be improved simply by having a hearing instrument. In more complicated situations, having comfort features and a better directional microphone system (as seen with Category 1 and Category 9) can provide additional WR benefit and improved SQ. These features demonstrated lower response times as well, suggesting that these added features have the possibility to reduce listening effort in difficult listening situations.

This consistency suggests that the preferred selection was based on the test client's specific needs and not on the overall program usage.

FIELD TEST

Methods. The same 8 test clients that participated in the lab test participated in the field test. For the field test, the devices were programmed with three performance categories as follows: Juna 9 (P1), Nevara 1 (P2), and Saphira 5 (P3). All performance categories were programmed with their default settings. The test clients wore the devices for approximately 2 weeks. They were instructed to wear the devices in as many environments as possible and to use each program in those environments for an equal amount of time. After using the devices, the test clients completed a questionnaire and interview regarding their experience with BeFlex. The questionnaire was created specifically to investigate the ability of test clients to differentiate between categories and to quantify the benefit received from wearing BeFlex. Additionally, the data logging information from Oasis was saved in order to verify the clients' usage of the devices.

Results. Four test clients chose the Juna 9, three chose the Saphira 5, and one test client chose the Nevara 1 as their preferred performance category. The data logging revealed that all test clients remained in P1 for more than 50% of the time. It was anticipated that the category in the P1 slot, as it is the start-up program, would be the most used program and might affect preference. However, despite the fact that P1 had the highest usage for each client, it did not impact overall preference as it was not unanimously chosen as the preferred category. This is further supported by the fact that the reasons given by the test clients for their preferences were consistent for each price category. Clients that chose the Juna 9 (Program 1) cited better speech understanding, and clients that selected Saphira 5 (Program 3) cited more comfort especially with sharp sounds as the reasons for their selection. This consistency suggests that the preferred selection was based on the test client's specific needs and not on the overall program usage.

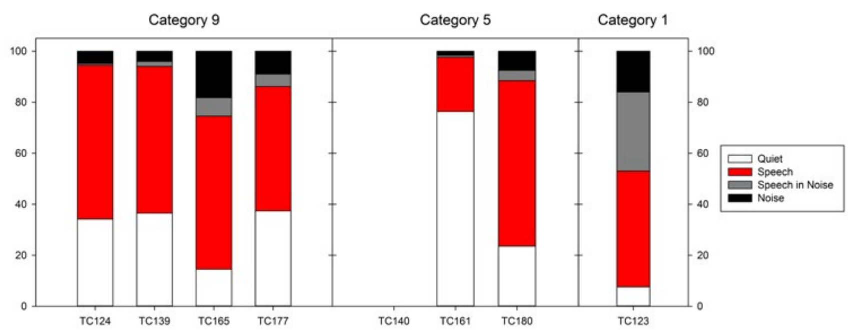

Figure 6. The graph shows the percent of time spent in different environments for each for each Test Client (TC).

Figure 6 above shows the percentage of time that clients spent in quiet (white), speech (red), speech in noise (grey), and noise (black) environments. They are grouped to indicate the preferred device category for each test client. As the graph shows, four test clients chose Category 9, three chose Category 5, and one chose Category 1 as their preferred category. The preference for test client 140 is shown, but the data logging was unable to be saved due to technical issues.

Those same test clients found that the ability to compare price categories would help them to make a purchase decision.

Conclusions and Discussion. Those test clients that preferred Category 9 were in speech environments more than quiet environments. These findings are consistent with their reasons for choosing Category 9 as their preferred program. For those test clients that preferred Category 5, there was not a consistent environment where more time was spent. Since they selected their program preference based on comfort, the environment may have been less important to them than the events occurring within the environment.

Based on the answers to the questionnaire and the interview, all except one test client were able to notice a difference between the performance categories. Those same test clients found that the ability to compare price categories would help them to make a purchase decision. These test clients found that BeFlex fulfilled its intended purpose of helping end users compare performance categories in order to determine which one to purchase. The inability to determine a difference between price categories can also provide useful information to the client concerning their device selection. It may be an indication that any price category will fulfill their needs. The overall experience of the BeFlex trial should be considered when counseling clients toward their purchase decision.

The data from these tests supports the BeFlex trial device concept, as the test clients were able to objectively and subjectively distinguish a difference between the programs they used with their BeFlex devices. Not only did they notice a difference between performance categories, but they were also able to choose a preferred category. BeFlex is currently available so HCPs can already put BeFlex to use and guide their clients toward confident hearing instrument purchases.

References

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International, 5* (9/10), 341-345.

Gustafson, S., McCreery, R., Hoover, B., Kopun, J. G., & Stelmachowicz, P. (2014). Listening effort and perceived clarity for normal-hearing children with the use of digital noise reduction. *Ear & Hearing, 35*(2), 183–94.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sherbecoe R.L. & Studebaker G.A. (2004) Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units. *International Journal of Audiology, 43*(8), 442-8

Svensson, E. (2003). Statistical methods for repeated qualitative assessments on scales. *International Journal of Audiology, 42* Suppl 1, S13–22.

v. Wallenberg, E.L. & Kollmeier, B., (1989). Sprachverständlichkeitsmessungen für die audiologie mit einem reimtest in deutscher sprache: erstellung und evaluation von testlisten. *Audiologische Akustik, 28*(2), 50-65.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482.

World Headquarters

Switzerland
Bernafon AG
Morgenstrasse 131
3018 Bern
Phone +41 31 998 15 15
Fax +41 31 998 15 90

bernafon
Your hearing • Our passion