University of Alberta

Analyzing Sound Quality of Advanced Bone Anchored Hearing Aids

by

Tyler Wetter

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech Language Pathology

Department of Speech Pathology and Audiology

©Tyler Wetter

Spring 2013

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission

Abstract

In this study we explored whether differences in sound quality existed between new advanced Bone Anchored Hearing Aids (BAHA). Three groups of subjects were tested. Two groups, those with normal hearing and those with sufficient residual cochlear hearing, were tested with Oticon's Ponto Pro to Cochlear's BP100. The third group had either mixed hearing loss or single-sided deafness and they compared the more powerful devices (Oticon's Ponto Pro Power and Cochlear's BP110). Some differences emerged within the normal hearing listeners and the power user's. However, no significant differences in sound quality were revealed in the non-power users of the Ponto Pro and BP100. These limited results most likely reflect a lack of power in the data due to the limited number of subjects per group. Continued data collection is warranted.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Hodgetts. Thank you for your patience and support throughout this process. You told me early on that a great supervisor is one who will be there come crunch time. You most definitely were and I know this project could not have been successful without all your contributions. I also experienced a huge learning curve with this thesis and you were always there to answer any questions or concerns that I had. It was a pleasure working with you and getting to know you.

I would also like to thank Amberley Ostevik. You were so patient with me and willing to answer questions I had about the project. I thank you for all your hard work with data collection and teaching me what it means to do research. You were such a pleasure to work with and I feel fortunate and grateful that I had the opportunity to work with you.

In addition, I have to thank Herman Lundgren and Dylan Scott. I know this project would not have been possible without all the hard work and dedication you put into it. Thank you so much for answering my questions with diagrams. You both were very supportive and were always good for a few laughs. Thanks to both of you.

I also want to thank my committee for the time and effort they put into this project. Your guidance and suggestions are greatly appreciated.

Lastly, I would like to thank my family and my girlfriend for always encouraging me to keep going when times were tough. I know that this journey would not have been possible without all your support. I'm very fortunate to have such wonderful people in my life.

Table of Contents

List of Tables

List of Figures

Introduction

Bone anchored hearing aids (BAHAs) have become increasingly popular for individuals who have chronic otitis media (ear infections), atresia (absence of the ear canal), and single sided deafness. BAHAs can be fitted for individuals with conductive or sensorineural hearing loss. Traditionally, bone conduction hearing aid devices (BCHAs) were fitted to headbands or eyeglasses. These techniques had limitations due to pressure, causing users to experience discomfort, pressure sores and headaches (Bance et al., 2002). Bance and colleagues (2002) also noted that even when these traditional devices had good skin contact, the skin and subcutaneous tissue could dampen the signal by as much as 20 dB. This dampening, accompanied with the need to continuously loosen the headband to avoid discomfort, resulted in significant sound quality limitations. Others (Hakansson et al., 1994) noted that the appearance of the BCHAs on a headband hindered these devices further and concluded they should be considered a last resort.

In contrast, the BAHA uses direct bone conduction. In order to achieve this, a surgically implanted titanium screw is placed in the parietal/temporal region of the skull. Titanium has a unique ability to promote osseointegration when placed into bone (Hakansson et al., 1994). When enough time has passed (typically 6 weeks to 3 months) the screw will be stable and capable of anchoring a BAHA. The BAHA is connected to the screw via a titanium abutment. With this system, the BAHA delivers sound directly to the bone without any loss of energy to the skin and subcutaneous tissue (Hakansson, 1994). Although the BAHA device is a more expensive solution, Snik et al. (2005) claim that this can be justified by fewer trips to the clinic and ultimately, improved communication performance. Snik and colleagues (2005) also indicated that implantation of the screw is achieved in 90 - 98% of cases, with high degrees of safety.

Figure 1. Bone anchored hearing aid and titanium implant in the skull.

Image source: http://ent.uci.edu/BAHA.htm

Reports indicate superior performance of BAHAs over conventional bone conduction hearing aids because sound quality remains steady even at higher volume settings (Snik et al., 2005). Snik and colleagues (2005) point to the fact that BAHAs are capable of more efficient sound transmission (especially in the high frequencies) through direct bone conduction, which ultimately results in better speech perception.

Bance et al. (2002) conducted a study that compared the audiometric performance of BAHAs to air conduction hearing aids. They concluded that regardless of which hearing aid was used, it never restored or rehabilitated the patients to the level of normal listeners. In all cases aided hearing-impaired subjects performed worse than normal hearing individuals. In addition, they concluded that both hearing aids produced very similar audiometric results. Sink et al. (2005) echoed this finding, stating that ambiguous results were found for those individuals who switched from air conduction devices to bone anchored hearing aids. They concluded that BAHAs were not superior to air conduction aids. Mylanus et al. (1998) however, states that once the air-bone gap is greater then approximately 30 dB SPL, better results should be expected with BAHA devices. This is because air conduction hearing aids require higher gains, which ultimately result in increased opportunities for feedback and poorer sound quality. Mylanus et al. (1998) also stated that the majority of the patients in the study preferred the BAHA device, when compared to their well-fitted air conduction aids. This appeared to be due to a decreased number of visits to the outpatient clinic and a reduction in ear infections. The study did indicate that neither type of hearing aid was preferred over the other when speech recognition in noisy environments was evaluated.

Flynn et al. (2009) conducted a study that examined solutions available to those individuals with severe mixed hearing loss. Their results confirmed those of Mylanus et al (1998), concluding that BAHAs were superior devices once the air-bone gap exceeded 30 dB SPL. The researchers examined ten subjects with mixed hearing loss who had worn a BAHA device for at least a year. Mixed hearing loss was defined as an average sensorineural loss greater than 25 dB SPL, in addition to an air-bone gap exceeding 30 dB SPL. In the study a Baha Intenso was compared to an air conduction device, Oticon Sumo DM, which is a digital superpower hearing aid. The results of the study indicated that all subjects increased speech understanding in noise with the BAHA. Furthermore, all subjects reported that they preferred the sound quality of the BAHA to that of the air conduction hearing aid. Although this study indicates that BAHA users preferred the bone conduction device to an air conduction aid, it is important to remember that this was only for people with severe mixed hearing loss and only investigated one particular BAHA device. Further research is still needed in order to determine what sound quality differences exist between various BAHAs.

Even though BAHA devices have been proven to be effective for individuals with chronic otitis media (COM) and external auditory canal malformations, limited research has looked specifically at sound quality in these devices. Hodgetts, Chen & Parsa (2004) conducted a study that examined perceptual differences between BAHA recordings using three BAHA devices: the Compact, Classic 300 and the Cordelle. The researchers were also interested in determining whether or not these perceptual judgments correlated with objective measures of sound quality. 14 normal hearing listeners were asked to listen to 18 different BAHA recordings and make subjective sound quality judgments using a visual analog scale. Each BAHA device was set to three different frequency responses: F1= manufacturers' default setting at full volume, F2= manufacturers' default setting at volume 2 and F3= potentiometers adjusted to the opposite of manufacturers' default setting at full volume. A standard passage read by a male speaker (carrot passage from the Audioscan Verifit, Dorchester, ON) was delivered at two different input levels (65 dB SPL and 75 dB SPL). The BAHA devices were placed on a skull simulator and the recordings were digitized using an audio interface and stored as .wav files. These authors obtained two different objective measures of sound quality. The results of the study indicated that there was a significant main effect for the device, with the Compact receiving the highest sound quality ratings. In addition, all listeners preferred the second frequency response (manufacturer volume 2) regardless of the BAHA device or input level. The researchers also found a significant interaction effect; the Classic and the Cordelle ratings decreased as the level of input increased. Furthermore, the study concluded that a small correlation was found between both objective and subjective measures of sound quality, indicating that revisions are needed for the instrumental measures to match with the subjective ratings. This indicates that although instrumental measures evaluate sound quality, normal listeners tend to disagree and rate sound quality differently. Further research needs to be conducted in order to determine why there is such a small correlation and what characteristics of the .wav files cause them to deemed more acceptable by normal listeners.

More recently, Kompis et al. (2007) examined sound quality differences between a BAHA Divino and a BAHA Compact, with the Divino being a modern digital bone anchored aid that utilizes more complex digital processing. Seven subjects were recruited to participate in the study (3 males and 4 females). The participants ranged in age from 19-66 years and had all been wearing a BAHA for at least 2 years. All the subjects suffered from bilateral conductive hearing loss, with some having an additional mild to moderate sensorineural component. The aim of the study was to determine if one device was superior to the other in terms of speech understanding in quiet and in noise. 5 of the subjects were currently using the BAHA Compact, so their devices were not changed, as they were already set optimally. As for the other 2 subjects, BAHA Compact devices were fitted and adjusted to mimic their current BAHA Classic devices as closely as possible. Speech recognition thresholds in quiet and in noise were measured while the subjects were wearing the BAHA Compact. Subjects were also asked to complete a questionnaire (Abbreviated Profile of Hearing Aid Benefit – APHAB) reflecting on how the hearing experience compared with their own BAHA. Following the testing, the subjects were all fitted with BAHA Divinos, which they were asked to wear for 3 months. Participants also had the option of returning after the first month if further adjustments were needed. Following the 3-month period, the subjects were invited back and speech audiometry thresholds were again completed in quiet and in noise. Again, the subjects were asked to complete the APHAB questionnaire reflecting on their hearing experience. In addition, the subjects completed a custom questionnaire regarding their experience with the new BAHA Divino.

The results of the study indicated that there was no statistically significant difference between the understanding of monosyllabic words in quiet with the Compact and the Divino. In the noise condition, the BAHA Divino was superior when using the directional microphone noise reduction mode. This was true for noise being presented from both the front and the rear, and resulted in an average improvement of 1.9 dB. According to the researchers, the analysis of the APHAB questionnaires was complicated and ultimately revealed small, statistically insignificant differences. In terms of the custom questionnaire, the participants rated the BAHA Divino as more favorable then their own device for all aspects (single speaker in noise, several speakers in noise, single speaker in quiet, listening to radio or television and overall sound quality). The results were however, only statistically significant for single speaker in quiet and overall sound quality judgments.

Although the new digital device was rated as having better overall sound quality, the results need to be interpreted with caution. The individuals in the study were not blinded as to what device they were rating. There was also a risk that order effects were present. It is possible that the subjects in the study rated the Divino as being superior because it was a newer device and they felt that the technology should be better. Furthermore, the study only evaluated the ratings of 7 individuals, making it difficult to generalize the findings.

A review of the literature reveals very little is known about sound quality in BAHA. In most studies BAHA users judged the sound quality of a particular device to which neither the subject, nor the examiner was blinded. There are often confounds such as order effects and lack of counterbalancing as well. Additionally, the study by Hodgetts et al., (2004) delivered these recording to listeners by air conduction. Perhaps there are differences in the sound quality ratings in the same users depending on whether the recordings were delivered via air or bone conduction transducers.

In recent years, the technology available in BAHAs has increased dramatically. Devices contain full digital processing, are computer programmable and have multi-memory and multi-microphone technology. There are 2 manufacturers of BAHAs, and at present, 4 new devices. While the devices are similar in terms of features, we have measured difference in terms of their prescriptive capabilities and output settings for a given hearing loss. They differ in terms of compression characteristics and in terns of frequency response/audibility. To date, no direct systematic measures of sound quality differences have compared these new devices. In the present study we will examine whether there are significant sound quality differences between different BAHA devices. Three groups of subjects were examined in this study. The first group, normal hearing listeners rated recordings from the 2 non-power BAHAs, Oticon's Ponto Pro and Cochlear's BP100. The second group of subjects was non-power BAHA users and they also rated the sound quality of the non-power Ponto Pro and BP100. The final group of subjects was power BAHA users and they rated the sound quality of the power devices, Oticon's Ponto Pro Power and Cochlear's BP110.

Figure 2. Devices used in the study. The top two are from Cochlear Corporation and the bottom two are from Oticon Medical.

The experimenters set out to examine the effects of three different passages, male speech, female speech and music, at two different input levels (60 and 75 dB SPL), on each BAHA device. The following research questions were of interest within each of the groups of subjects.

Research Question(s)

Normal Hearing Listeners

1. Are there significant differences between sound quality ratings by air vs. bone conduction? (Main Effect)
2. Are there significant differences between devices? (Main Effect)
3. Are there significant differences between devices at different levels? (Main effect)
4. Are there significant differences between devices for different stimuli? (Main effect)
5. Are there any significant interactions between any of the Main effects? (Interaction effects).
6. If there are significant interactions, wherein do the differences lie? (Paired-contrasts)

Non-Power BAHA Users

1. Are there significant differences between devices? (Main Effect)
2. Are there significant differences between devices at different levels? (Main effect)
3. Are there significant differences between devices for different stimuli? (Main effect)
4. Are there any significant interactions between any of the Main effects? (Interaction effects).
5. If there are significant interactions, wherein do the differences lie? (Paired-contrasts)

Power BAHA Users

1. Are there significant differences between devices? (Main Effect)
2. Are there significant differences between devices at different levels? (Main effect)
3. Are there significant differences between devices for different stimuli? (Main effect)
4. Are there any significant interactions between any of the Main effects? (Interaction effects).
5. If there are significant interactions, wherein do the differences lie? (Paired-contrasts)

Method

Subjects

Group 1 – Normal Hearing Subjects

Seven normal hearing subjects ranging from 28 to 45 (Mean = 32.4) were recruited at the Institute for Reconstructive Sciences in Medicine (iRSM) in Edmonton, Alberta Canada. The normal hearing subjects were all colleagues at the iRSM who agreed to participate.

Groups 2 and 3 – BAHA Users.

Eighteen adult BAHA users were recruited from the Bone Conduction Amplification Program at the Institute for Reconstructive Sciences in Medicine (iRSM) in Edmonton, Alberta. Subjects were fitted with the BAHA for a minimum of 3 months prior to participating in the study, were greater than 18 years of age and had no known cognitive deficits. Eight of the subjects fit the criteria for the non-power BAHA tests (Ponto Pro and BP100) and ten subjects were BAHA power users (Ponto Pro Power and BP110). Subject details can be found in Table 1. All subject were consented according to the ethics procedures at the University of Alberta and Covenant Health.

Table 1. Subject Details.

Subject	Gender	Age	Time Implanted	Ear	Eitology	Type of Hearing Loss
S1	F	71.7	24-Jan-12	L	Bilateral chronic otitis media (COM)	Bilateral mixed
S2	F	71.3	02-Aug-12	R	Right COM	Right mixed loss
S3	F	63.4	06-Nov-06	L	Recurrent left otorrhea/COM	Left severe mixed
S4	F	72.3	05-Jun-12	R	Mastiodectomy - COM	Right conductive
S5	F	61.8	09-Jan-09	R	Bilateral COM	Bilateral mixed
S6	F	27.1	29-Sep-06	L	Right stenosis; Left atresia (translocated 18 chro.)	BilateraL - Right mixed Left conductive
S7	F	55.3	08-Jan-10	R	Right otosclerosis	SSD
S8	M	54.4	12-Aug-11	R	Bilateral COM	Bilateral - conductive
S9	M	66.1 1	08-Apr-94	R	Bilateral COM	Bilateral mixed
S10	F	29.7	24-May-12	R	Cholesteatoma; Right COM	Right conductive
S11	M	64	17-Dec-09	R	Right otosclerosis	Bilateral - Right mixed, Left conductive
S12	F	57.9	26-Jan-07	L	Left otorrhea; COM	Left mixed
S13	M	56.5	21-Apr-06	R	Right COM	Right mixed
S14	M	70	01-Feb-08	L	Left sudden deafness	SSD
S15	M	58.2	31-May-11	L	Menieres disease - left asymmetrical SNHL	SSD
S16	M	74.7	16-Jun-97	L	Left COM	Left conductive
S17	F	54.2	03-Feb-05	R	Bilateral COM	Bilateral - conductive
S18	M	64.4	06-Jan-05	L	Aoustic neuroma	SSD

Procedure

Phase 1 - Hearing Testing, Device Setting and Recording

For the normal hearing subjects, each device (Ponto Pro and BP100) was set to the same hearing loss (a flat conductive loss with 0 dB HL thresholds in each manufacturer's software with a 30 dB SPL air-bone gap). The non-power and power BAHA users were tested using each manufacturer's software to obtain their in-situ thresholds through each device (Ponto Pro and BP100 for the non-power group and Ponto Pro Power and BP110 for the power group). Each manufacturer plots an "HL" value for these thresholds. However, there is no standard for HL measured by direct bone conduction. Therefore, all thresholds were converted to force level thresholds (ref: 1 μN) as measured on a skull simulator. For all BAHA subjects the air-bone gap is irrelevant since the device and all recordings deliver sound directly to the skull via the abutment. We measured a reference equivalent threshold force level for each HL value in each manufacturer's software and used that to convert the "HL" values to force on the skull simulator in dB FL (ref: μN). The thresholds for the 2 groups (measured with each device) can be seen in Figure 3.

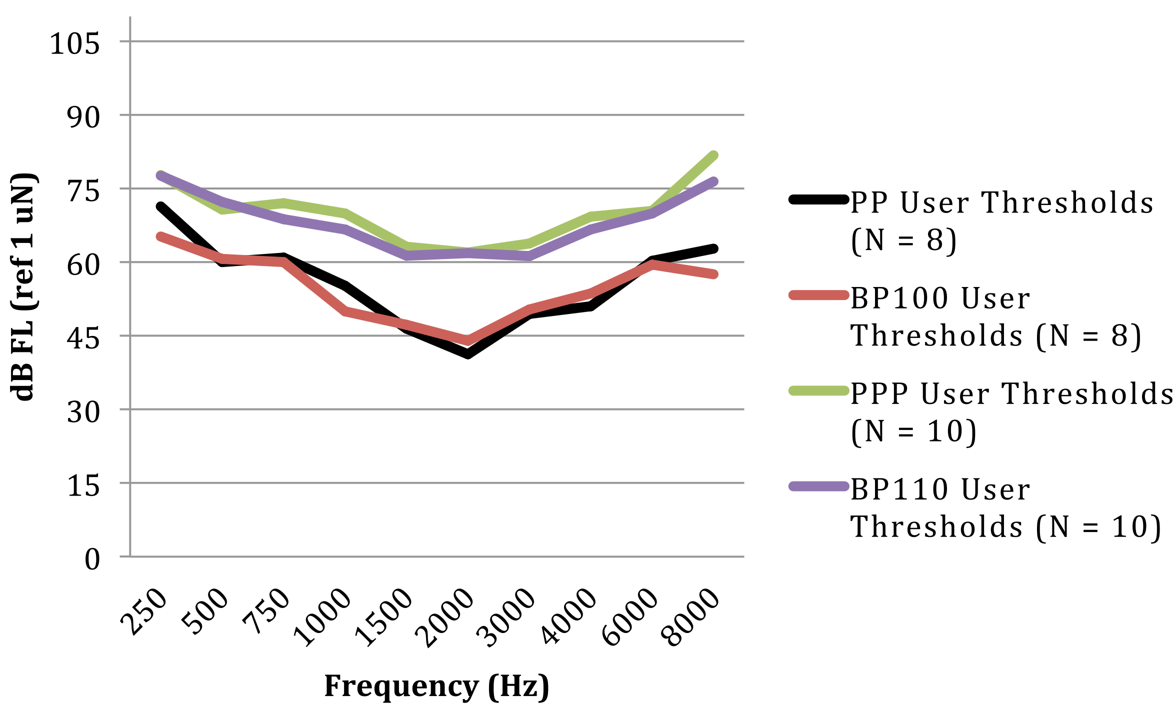


Figure 3. Mean force level thresholds on a skull simulator for each subject group (non-power and power users) obtained with each of the manufacturer's devices.

Following the in-situ threshold measurements, each manufacturer's software prescribed the settings according to their proprietary fitting rationales. These fittings were not altered for this study. It is likely that a change in prescription would have altered the sound quality of the devices, however, no other prescriptive algorithm is yet commercially available within each manufacturer's software. Therefore, in order to best approximate true clinical differences (if any) in sound quality, the settings were left as prescribed by each manufacturer, from the in-situ audiograms.

Recordings

Once the hearing tests were complete, each of the devices was set accordingly on an individual basis. Only 2 devices were set for each patient. If the individuals presented with conductive hearing loss, the Ponto Pro and BP100 were used. If the subjects presented with mixed hearing loss or single sided deafness, the power devices were used (Ponto Pro and BP110). This constituted the first phase of study involvement for subjects, and they were rebooked for Phase 2 at a later date.

The Interfacial Biomechanics Laboratory, in conjunction with the Bone Conduction Amplification Laboratory at iRSM, developed a test box to be used for audiological measures. The test box included a National Instruments data acquisition system (DAQ) and a signal amplifier. The data acquisition system (Figure 4) is compatible with system design software, LabVIEW. LabVIEW uses a graphical programming language, which is organized in a block diagram form. The recording software was programmed in LabVIEW to be used in conjunction with the test box.

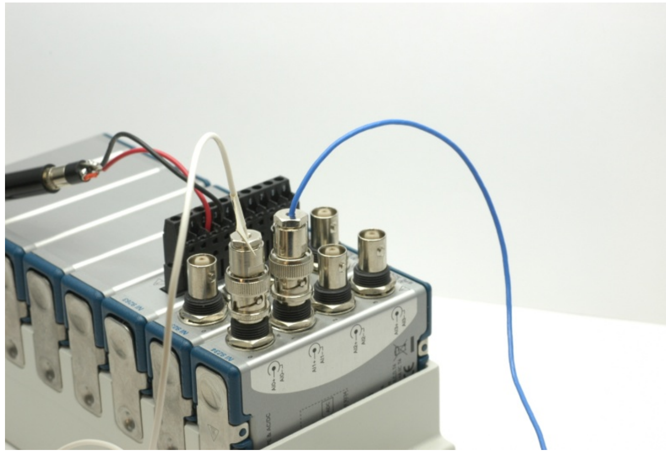
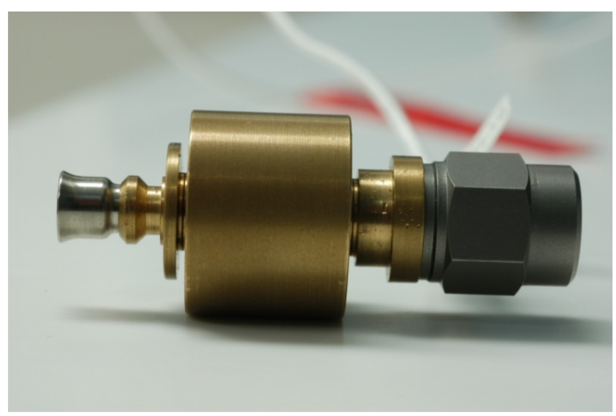


Figure 4. National Instruments data acquisition system (this is the same DAQ that has been integrated into the test box).

Each of the set devices were connected to a skull simulator for recording purposes (Figure 5a and 5b). The skull simulator mimics the human skull from a force level perspective and therefore the sound recorded through the skull simulator was a very close approximation to what would be delivered to the patients' abutments. (Additional corrections between each individual's skull and the skull simulator were made and will be discussed later.)

(a)



(b)

Figure 5. (a) Skull simulator and (b) internal components of skull simulator consisting of an abutment, mass and accelerometer.

The skull simulator was connected to the DAQ. A skull simulator was used because it is more difficult to record the BAHA vibrational outputs from a subject's head. These sound vibrations were transmitted from the BAHA transducer to the abutment on the skull simulator and then processed as acceleration by the accelerometer. The accelerometer within the skull simulator connected to the DAQ, which subsequently transferred the acceleration signals to the software (see Figure 6). As the passages were played through the BAHA, they were recorded as force output from the skull simulator (since we know both the acceleration and the mass, the force is easily derived). Thus, the software read the acceleration inputs from the DAQ and converted them to .wav files. For the scope of this study, the software was required to start/stop recording on demand, save each recorded file in a systematic way, and record the frequency content (force output) of each device.

Figure 6. SoundQual.vi recording software, designed using Labview.

The researchers were interested in the influence of type of input (3 levels – male speech vs female speech vs music) as well as level of input (2 levels - 60, 75 dB SPL) on the sound quality of the different BAHAs. The speech files were delivered to each device using an AudioScan Verifit at zero degrees azimuth. The Verifit allowed speech to be delivered to both BAHAs using external speakers. The speakers and the devices were contained within a sound proofed test box to eliminate any external noise that could bias sound quality recordings. Furthermore, all the recordings were done within a sound booth. Lastly, a reference microphone was calibrated prior to each subject's research device being tested. This ensured that all the files were delivered at 60 and 75 dB SPL. Of note, by using the Verifit to deliver 60 and 75 dB SPL inputs we not only changed the overall level of the speech signals but also the spectral weighting on the signals to approximate a normal conversational (60 dB SPL) level and the spectral shaping of loud or shouted speech (75 dB SPL) (Pearsons et al., 1977).

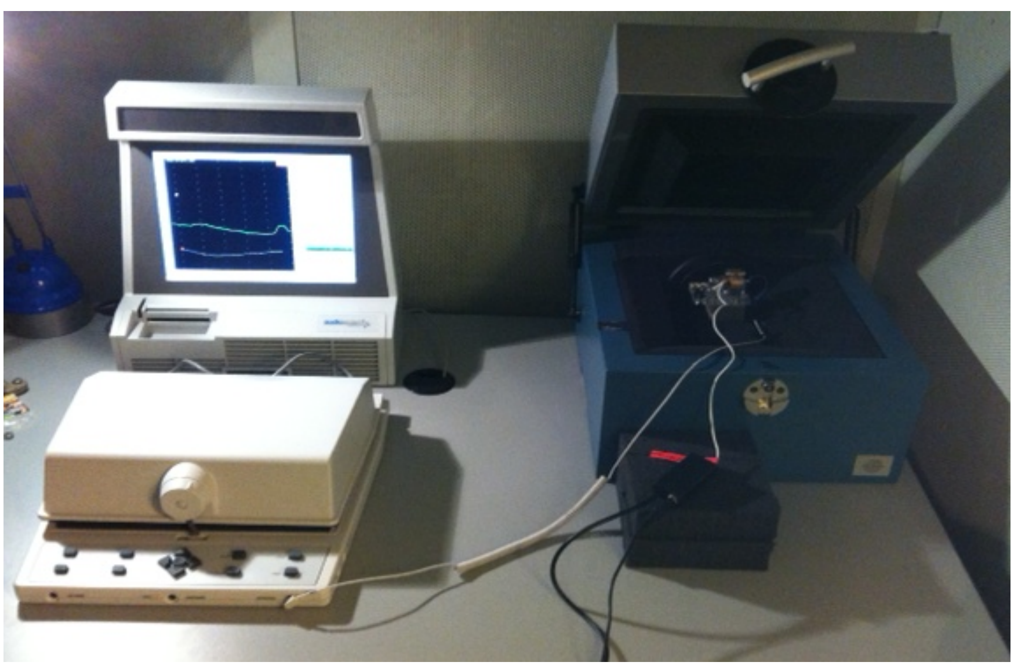


Figure 7. Audioscan Verifit, Skull Simulator and Sound Proofed Test Box.

The same skull simulator and LabVIEW program were used to record the music files, however the Verifit was not. The skull simulator was placed on a stand within the sound booth and music was delivered to the BAHAs at both 60 and 75 dB SPL. The standard music file (the first 10 seconds of Gilbert O'Sullivan's 'Alone Again, Naturally') was delivered through speakers attached to a computer. The stand, skull simulator, and speakers were set up in a manner that was universal across all the recordings. This allowed the researchers to be confident that the music files were delivered to the BAHA devices at 60 and 75 dB SPL.

The speech passages and a music passage were delivered to each BAHA, at each input level, for each subject's unique setting. In other words, each user had 2 devices, with 3 passages and 2 input levels for each, for a total of 12 recordings.

Once all the recordings were complete, the files were clipped, normalized and de-convoluted using a software program titled *DeconvoluteMe.vi*.

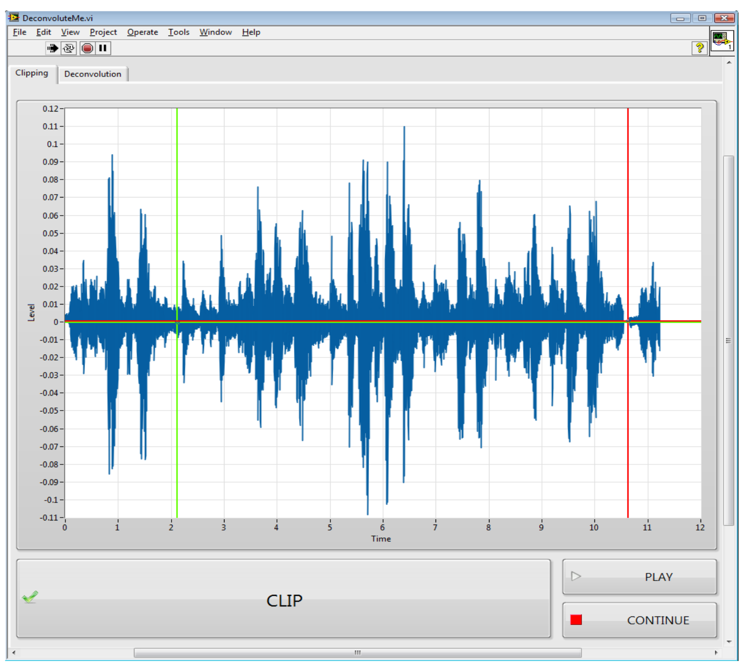


Figure 8. DecovoluteMe.vi software program.

This software program was developed for use in the Bone Conduction Amplification Laboratory at iRSM. The files were clipped to remove any parts that were not directly related to the actual speech or music file. They were normalized to ensure playback at the same volume, accounting for the difference in loudness between the 60 and 75 dB SPL recordings. In addition, it allowed the subjects to choose a comfortable listening level, prior to beginning the ratings. This minimized any influence of loudness accounting for differences in sound quality judgments. Lastly, the files were de-convoluted to correct any influence the skull simulator had on the files. Although similar, the skull simulator and the human skull are known to have different properties and the de-convolution allowed those to be corrected.

De-convolution Process

The original .wav files were altered by the transfer function of the transducer-skull simulator system; this being the first electromechanical system influencing the original .wav file. This ultimately resulted in recordings that included a resonant peak around 700 Hz, which was not included in the original file. If the patient were to listen to the files at this point, the combined influence of the 'naked' transducer and the actual frequency response of the individuals skull would further influence the recordings (second mechanical system to influence recordings). This would further transform the original file, compounding the influence of the resonant peaks. In order to compensate for this problem, all the recordings were de-convoluted using the impulse response of the transducer-skull simulator system. The de-convolution process removed the effect of the BAHA transducer and the skull simulator (first mechanical system). This process can be described as inverse filtering (matching the frequency response created by the first mechanical system to a frequency response which is the exact opposite, and essentially eliminating it altogether). This resulted in nearly identical frequency spectra of the original and de-convoluted sound files. Once the files were played back to the patient, the only influence affecting the recordings would be that of the combined mechanical influence of the 'naked' transducer and the individual's real head frequency response (refer to Figure 9). This ensured that the recordings were done in a manner that replicated exactly how a BAHA patient would hear the speech and music if they were listening to the prescribed device on their head.

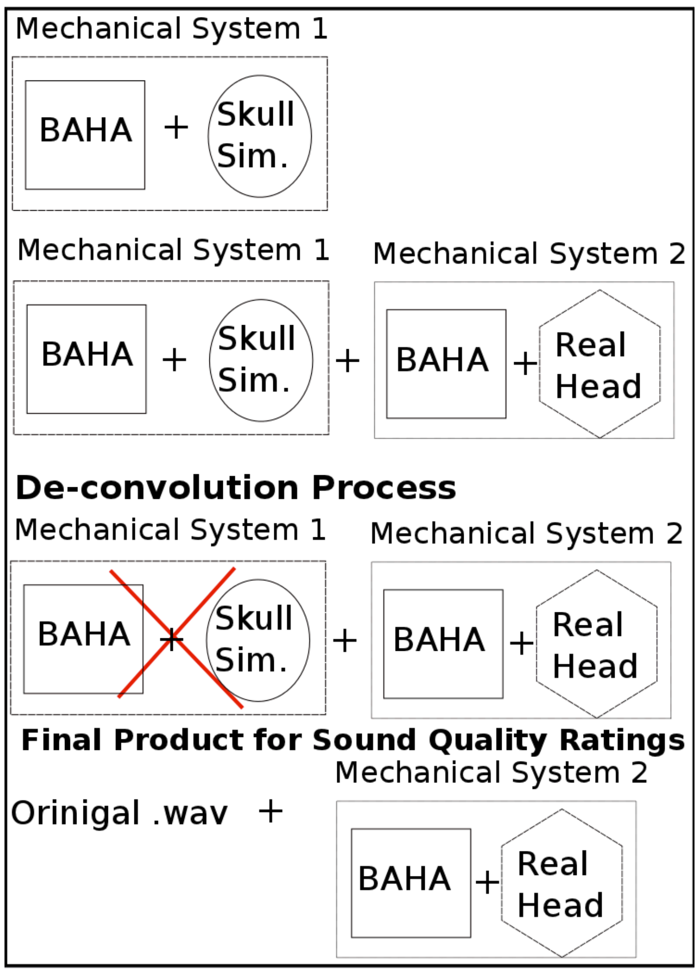
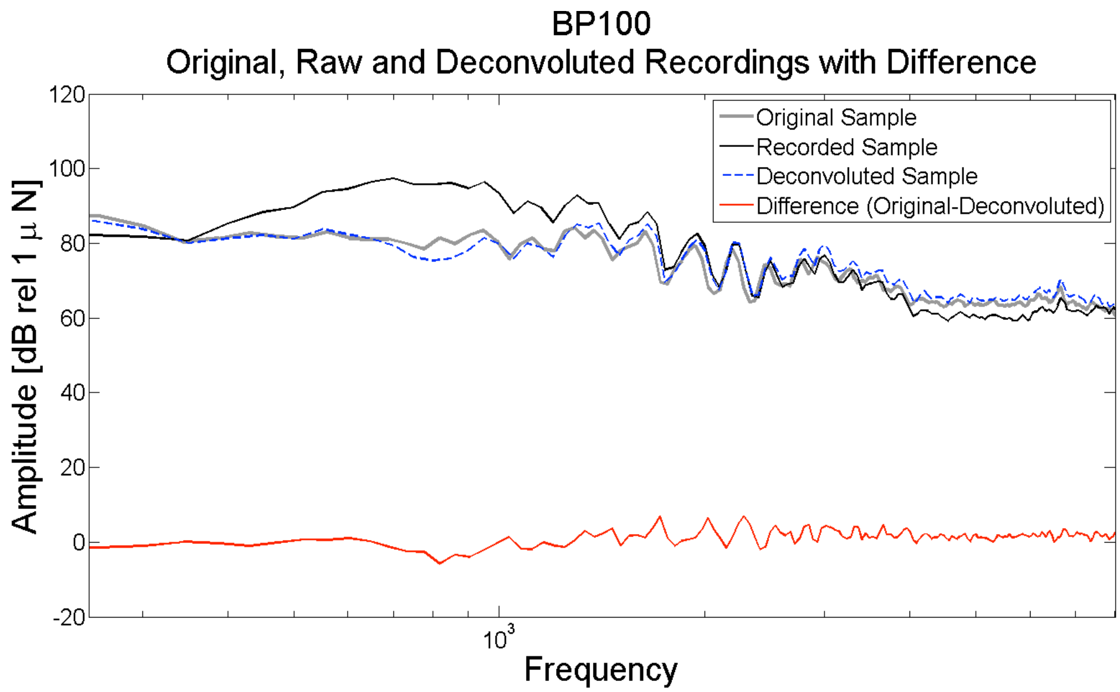
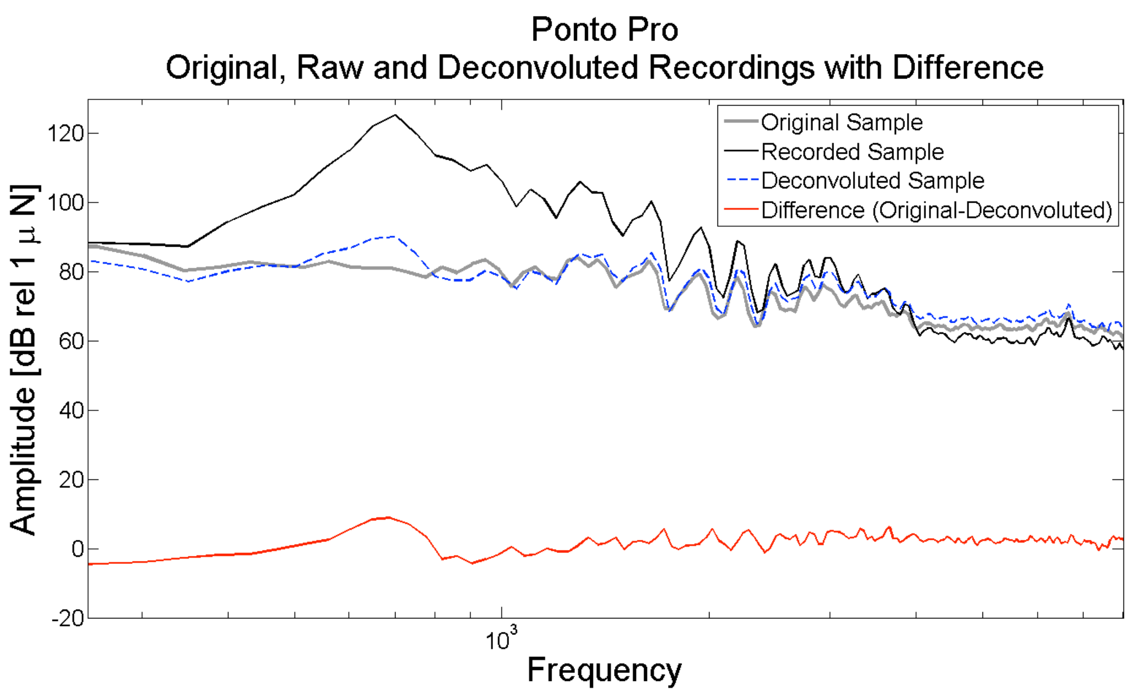


Figure 9. Schematic of de-convolution process.

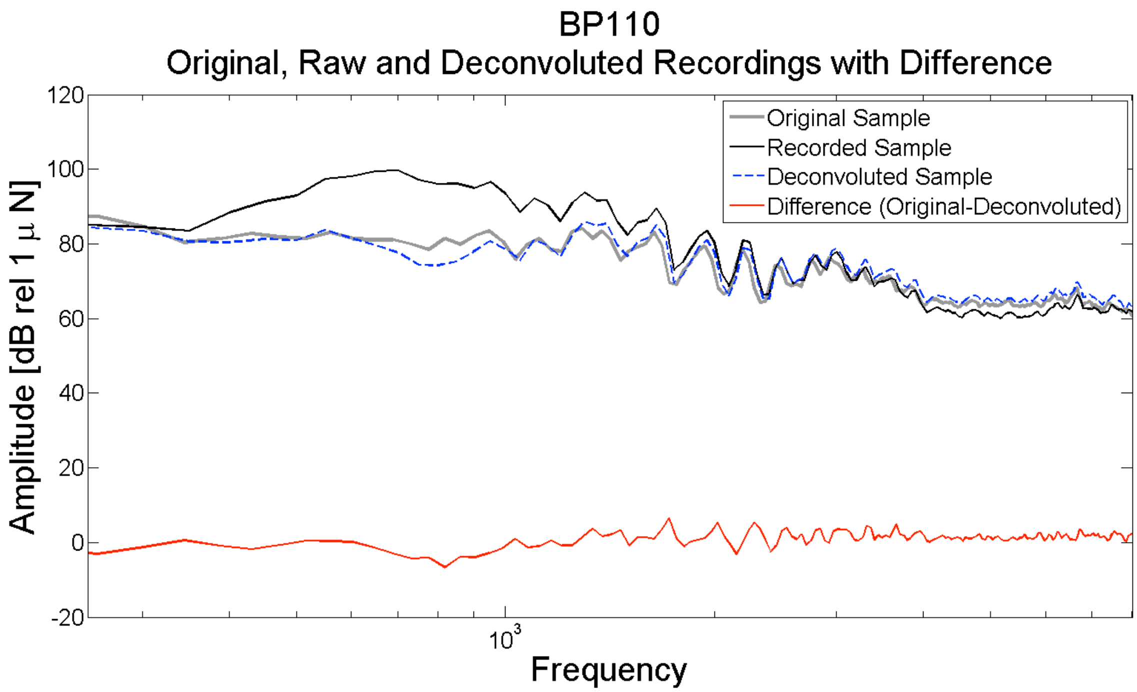
The following figure demonstrates the waveforms for the original, the recorded and the de-convoluted music .wav files for each BAHA device. In addition, each graph displays its own waveform depicting the difference between the original .wav and the final de-convoluted .wav.



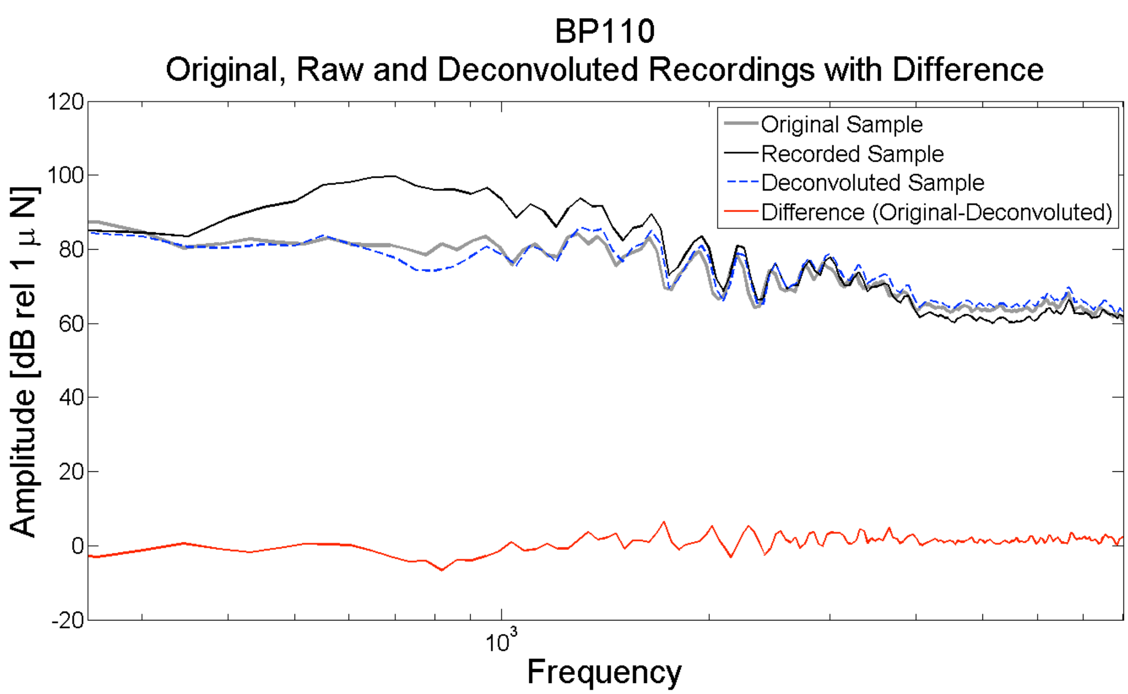
(a)



(b)



(c)



(d)

Figure 10. Comparisons of the original music .wav file to the recorded and the de-convoluted music .wav file for each processor. In addition, a difference waveform between the original file and the final de-convoluted file. (a) BP100, (b) Ponto Pro, (c) BP110 and the (d) Ponto Pro Power.

As expected, minimal differences were found to exist between the original sound files and the final de-convoluted product. Figure 11 compares the original music sample to normalized, de-convoluted spectra of each research device. Again, as expected the differences were minimal.

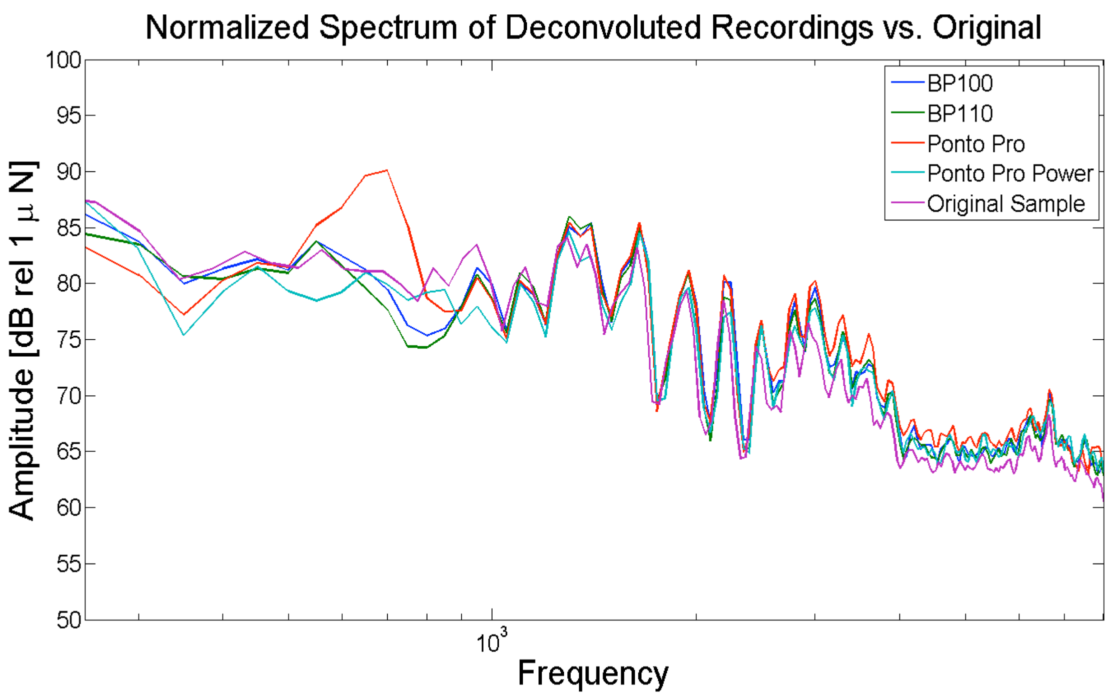


Figure 11. Normalized, de-convoluted spectra of each BAHA device compared to the original music sample.

Phase 2 – Sound Quality Assessment

Once all recordings were clipped, normalized, and de-convoluted the .wav files were loaded into a program called *SoundQualTesv2.vi* (a software program similar to that developed by Dr. Parsa at UWO). *SoundQualTesv2.vi* allowed the recordings to be presented to the user on a single screen (based on what they selected: male speech, female speech or music) with an accompanying digital slider. The slider was equivalent to a Visual Analogue Scale, which allowed the subject to slide the scale between 1 (low quality) to 10 (high quality) for each recording.

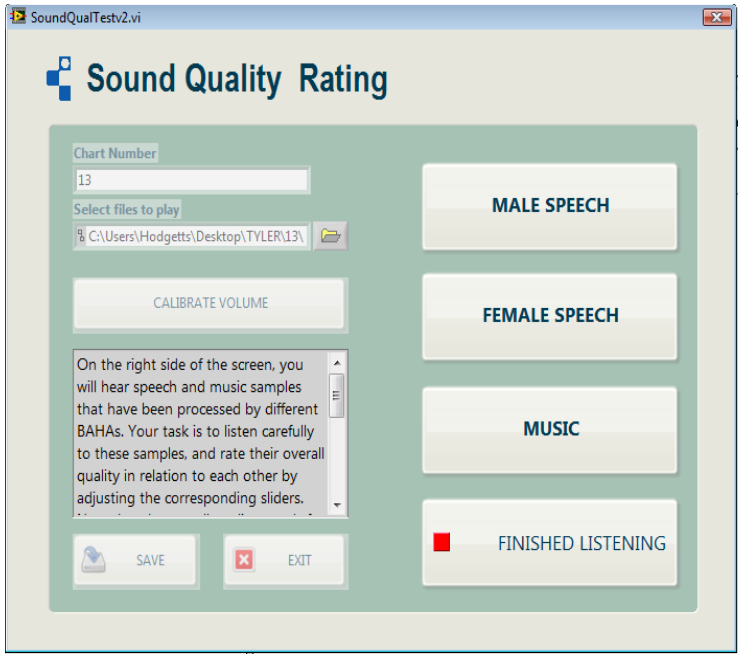


Figure 12. SoundQualTestv2.vi software.

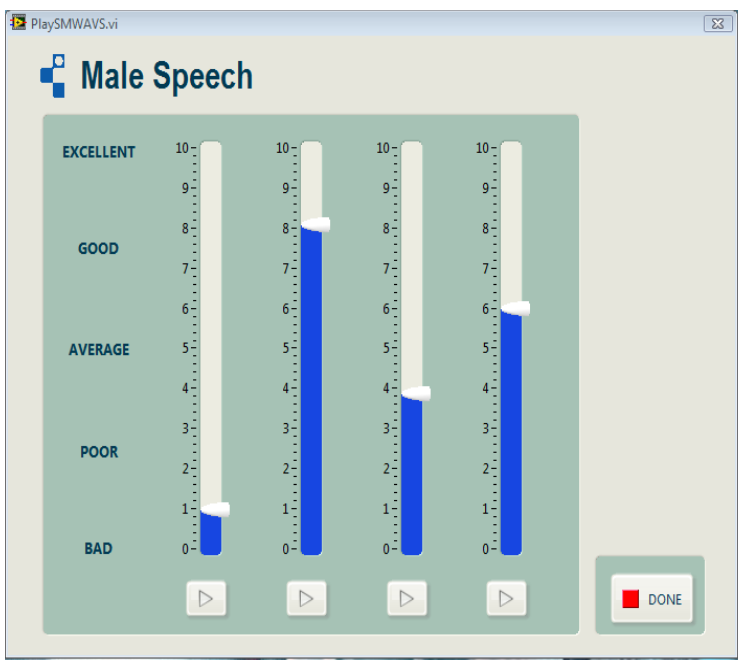


Figure 13. Visual analog scale used to judge sound quality.

The subjects were then invited back for Phase 2 of the study. In order to deliver the recordings to the subjects, 2 isolated (naked) BAHA transducers were used to send the .wav files to the user's abutment for listening. The bare transducers were housed in a special casing developed at iRSM so that the subject could not tell which transducer was being used. The bare transducers were determined based on which devices were used to make the recordings (if the power devices were used for the recordings, then the bare power devices were used to play the recordings back).

The subjects were able to set the level of the recording to their most comfortable listening level, again, ensuring that audibility had a limited influence on the ratings of sound quality. Participants were blinded as to what condition they were experiencing to avoid any potential biases. Subjects simply listened to the recordings as often as they wished until they felt they could provide a reasonable estimate of the sound quality with the digital slider. Order of the recordings was randomized on screen, across subjects to minimize any biases associated with fatigue or order effects. All subjects listened to the 12 recordings (4 male speech, 4 female speech, and 4 music) and rated each one. Once they completed the first 12 ratings, the 'naked' transducer was changed and they then repeated the steps described above and rated the same 12 recordings. The order of the naked transducers was counterbalanced across subjects. See Figure 14.

Figure 14. Isolated (i.e. ,"naked") Ponto Pro and BP100 transducers.

Subjective sound quality judgments were made using both manufacturers 'naked' transducers to ensure that both devices' transfer functions (mechanical responses of the bare BAHAs) were accounted for and given opportunity to deliver the recordings. It also provided the researches with a test re-test reliability measure. This was beneficial in determining if the individuals in the study could consistently rate the sound quality of the devices across 2 trials and if the results were truly repeatable. Figure 15 shows the frequency/output responses from each of the 4 transducers driven with the same electrical sweep signal (100 nV stimulus). While each of the transducers shares a similar design, mass and development architecture, there are some notable differences. Firstly, it is clear that the power devices have a higher resonant peak. There are also differences in the resonant peak frequency associated with each transducer. These differences underscore the importance of testing each subject twice once using each of the naked transducers to ensure that these differences were at least accounted for.

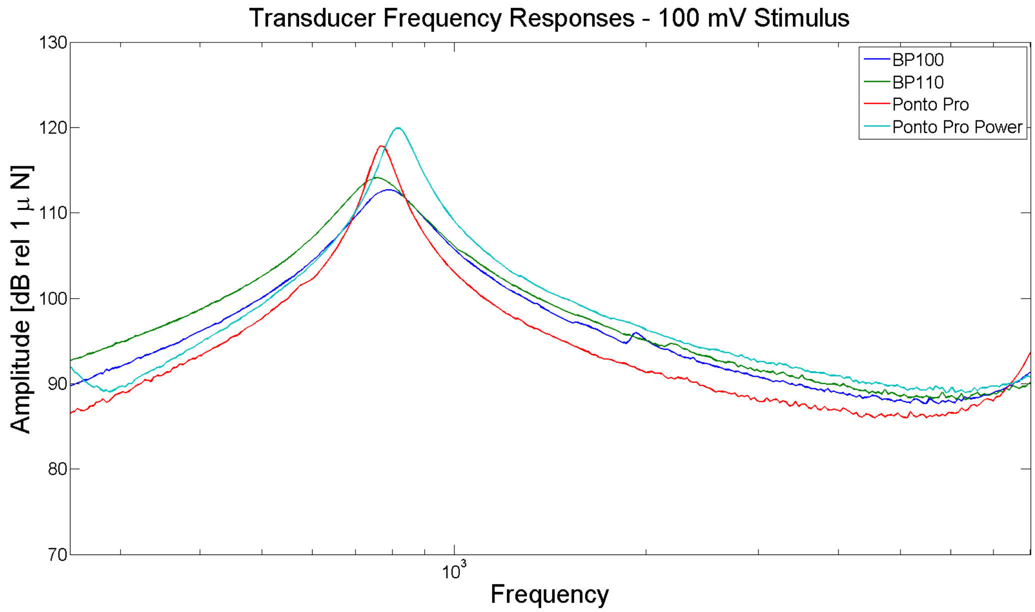


Figure 15. Frequency/Output responses of the 4 naked BAHA transducers used in the study.

Results

Normal Hearing Subjects

The 7 listeners with normal hearing were tested with one artificial BAHA user's recordings (flat conductive loss) in 2 conditions. For the first condition they rated the sound quality of the files by air conduction using standard ER-3A insert earphones. For the second condition they were tested in the same way using a B-71 audiometric bone oscillator. We were interested in whether there were any major differences between sound quality ratings measured via air conduction versus bone conduction and also if a similar pattern of results would be obtained from normal hearing listeners and those who use BAHA.

A 2 x 2 x 2 x 3 repeated measures ANOVA was run on the data from normal hearing subjects. The first independent variable was Mode of Stimulation with 2 levels (AC vs. BC); the second independent variable was Device with 2 levels (Ponto Pro vs. BP100); the third independent variable was Stimulus Level (60 vs. 75 dB SPL), and the final independent variable was Stimulus Type with 3 levels (Male Speech, Female Speech and Music). The main effect for Mode of Stimulation was approaching significance but did not reach it ($F(1, 6) = 5.33$, $p = 0.06$). This indicates that, at least for this small sample, when collapsed across all other variables, there was no main difference between Air and Bone Conduction presentations. There were main effects for Device ($F(1,6) = 104.166$, $p < 0.001$) and Level ($F(1,6) = 53.061$, $p < 0.001$), indicating that it did matter which device was being used by subjects and which level was used in the recordings of the stimuli. The main effect for stimulus was also significant ($F(2, 12) = 10.877$, $p < 0.002$). Figure 16 shows the main effect of Mode of Stimulation. Figure 17 through 19 show the significant main effects of Device, Level and Stimulus.

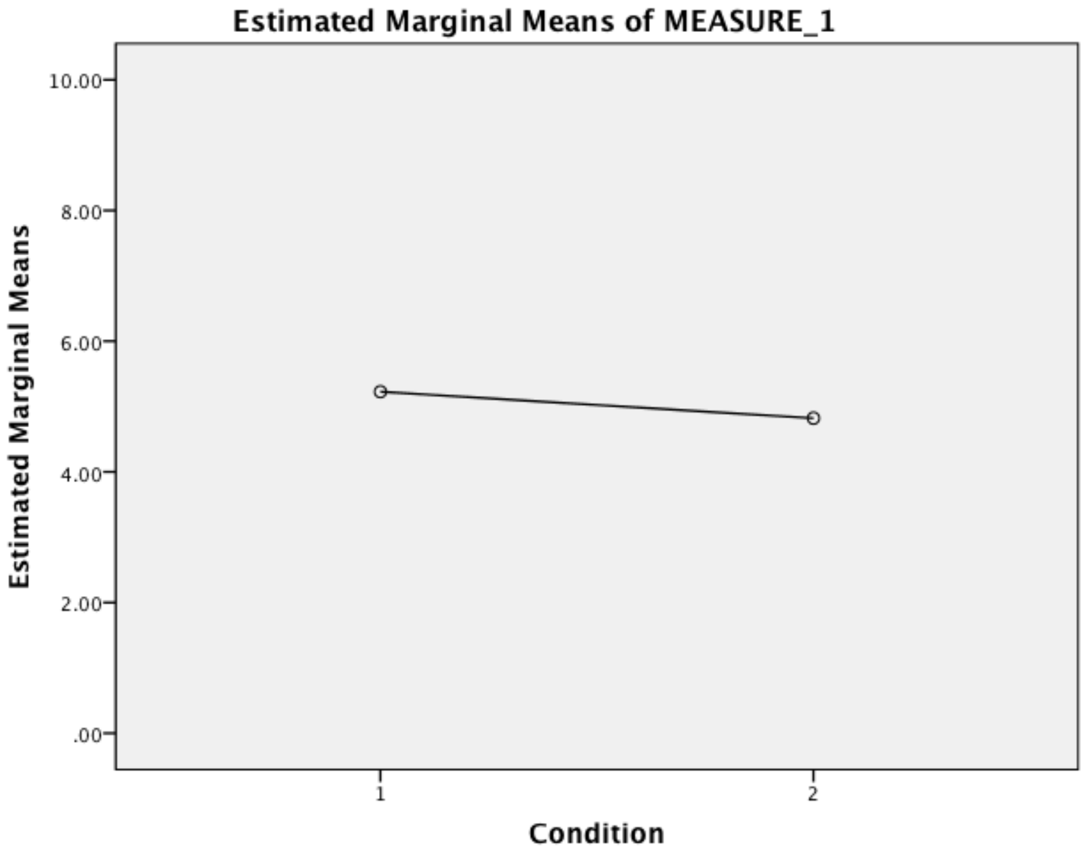


Figure 16. Main effect of mode of stimulation. Condition 1 is bone conduction and condition 2 is air conduction.

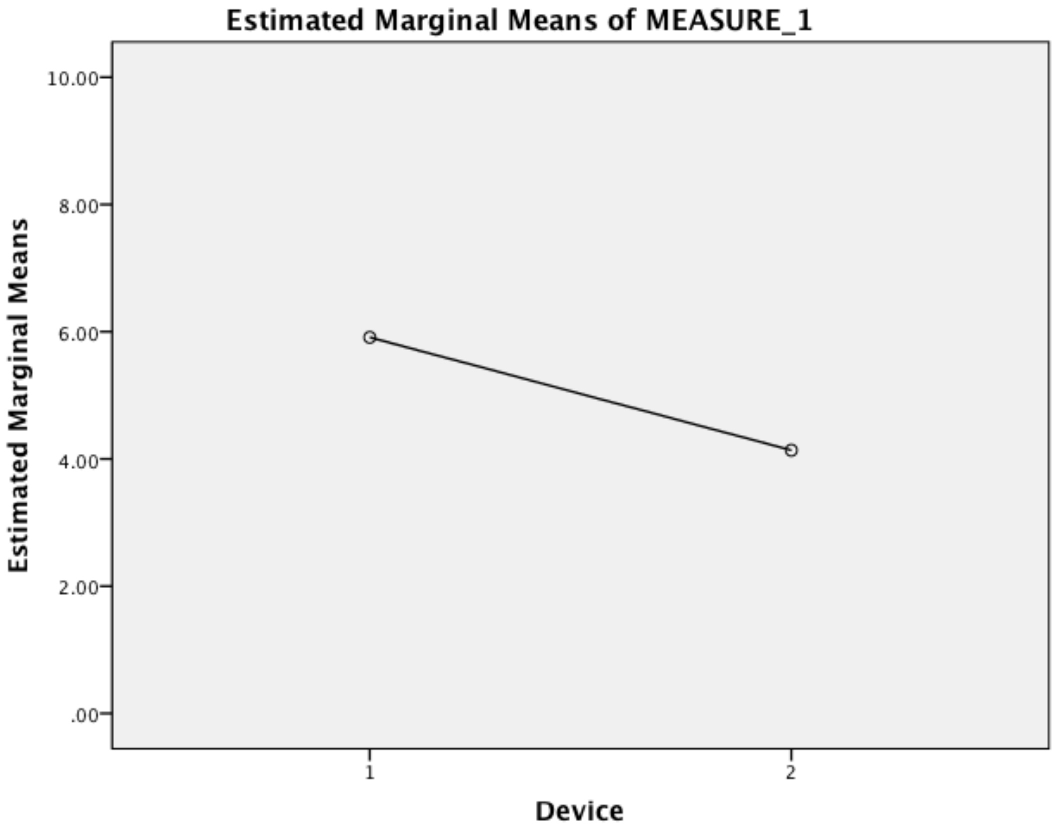


Figure 17. Main effect of device. Device 1 is the Ponto Pro; device 2 is the BP100.

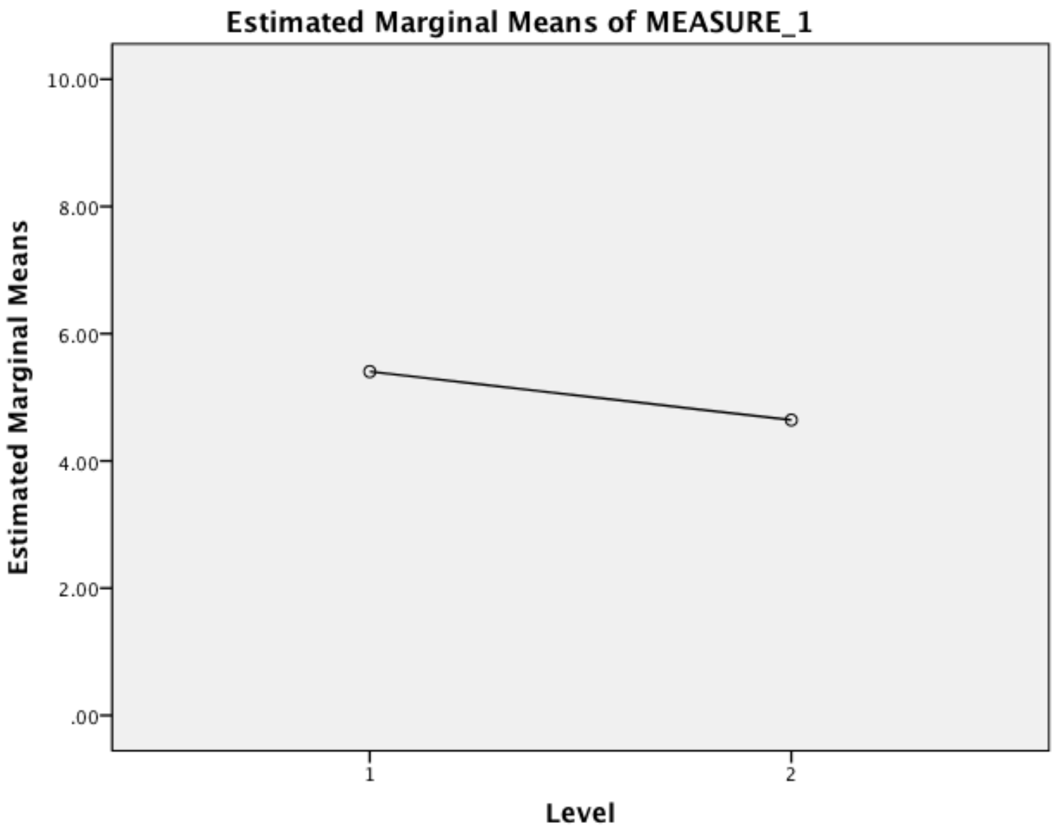


Figure 18. Main effect of level. Level 1 is 75 dB SPL input and level 2 is 60 dB SPL input.

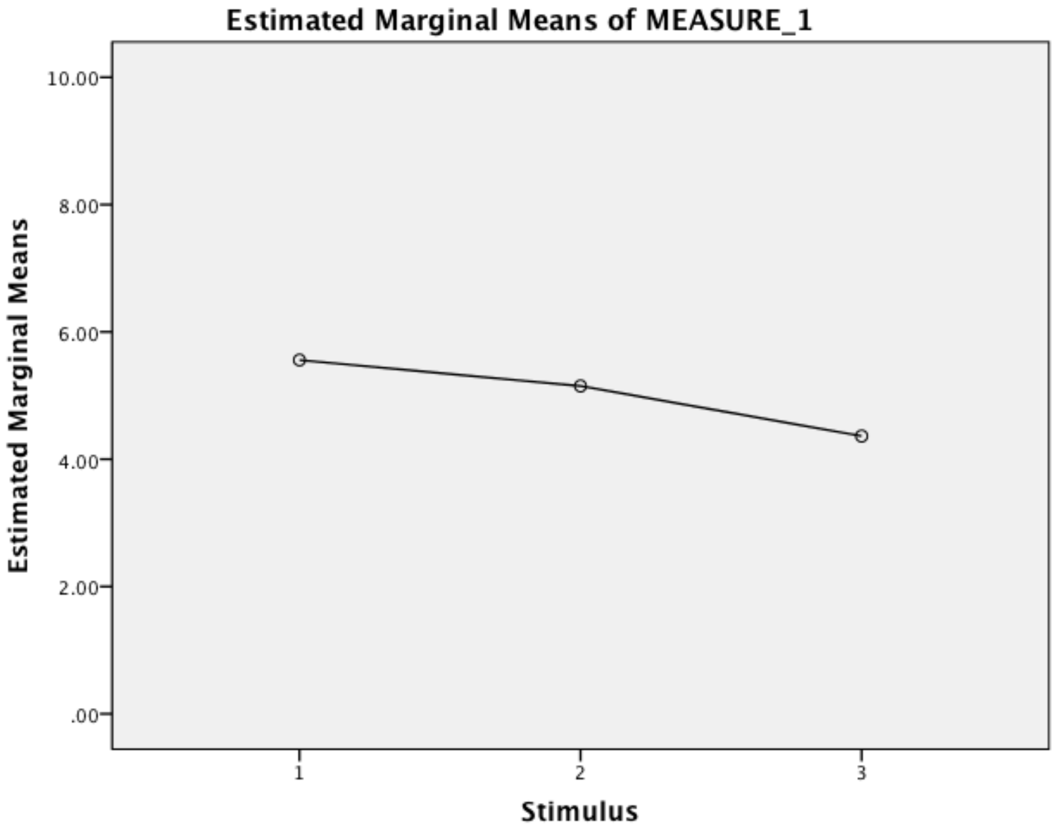


Figure 19. Main effect of stimulus type. Type 1 was male speech; type 2 was female speech and type 3 was music.

While the main effects are revealing, the more important outcome in the data was the significant 4-way interaction between Mode of Stimulation x Device x Level x Stimulus ($F(2, 12) = 7.670$, $p < 0.007$). This significant 4-way interaction indicated that differences in sound quality ratings for normal hearing listeners depended on which mode of stimulation, device, level and stimulus was under test. Figure 20 shows the overall results of the comparisons. Since we were not interested in the total post hoc comparisons from the normal hearing listeners (this was pilot exploration), we set up 20 planned comparisons. To control for the possibility of a type-1 error rate with these planned comparisons, we divided the alpha level of 0.05 by 20 to compensate for this. Thus, any comparison needed to have a p-value of less than 0.0025 in order to reach significance. For example, with 1 device, were there any conditions that lead to a significant difference between air and bone conduction sound quality ratings? The difference in scores from bone conduction to air conduction ($M = 3.893$, $SD = 1.443$) for the BP100 at 60 dB SPL input, speech female condition was significant ($t = 7.139$, $p < 0.0001$). For whatever reason, people with normal hearing rated this recording much worse by air conduction than by bone conduction. Other cases were approaching significance, but did not reach it. We also explored within a mode of stimulation (either air or bone conduction), if there were any obvious device differences for a given level and stimulus. Three significant pairings emerged. At the 60 dB SPL input level for bone conduction male speech the Ponto Pro was rated as significantly better than the BP100 (Mean difference = 4.028, $SD = 1.721$; $t = 6.191$, $p < 0.001$). The Ponto Pro was also rated better by air conduction for both the 60 dB SPL female speech (Mean difference = 4.159, $SD = 2.224$; $t = 4.903$, $p < 0.002$) and 60 dB SPL music conditions (Mean difference = 5.354, $SD = 1.193$; $t = 11.874$, $p < 0.0001$).

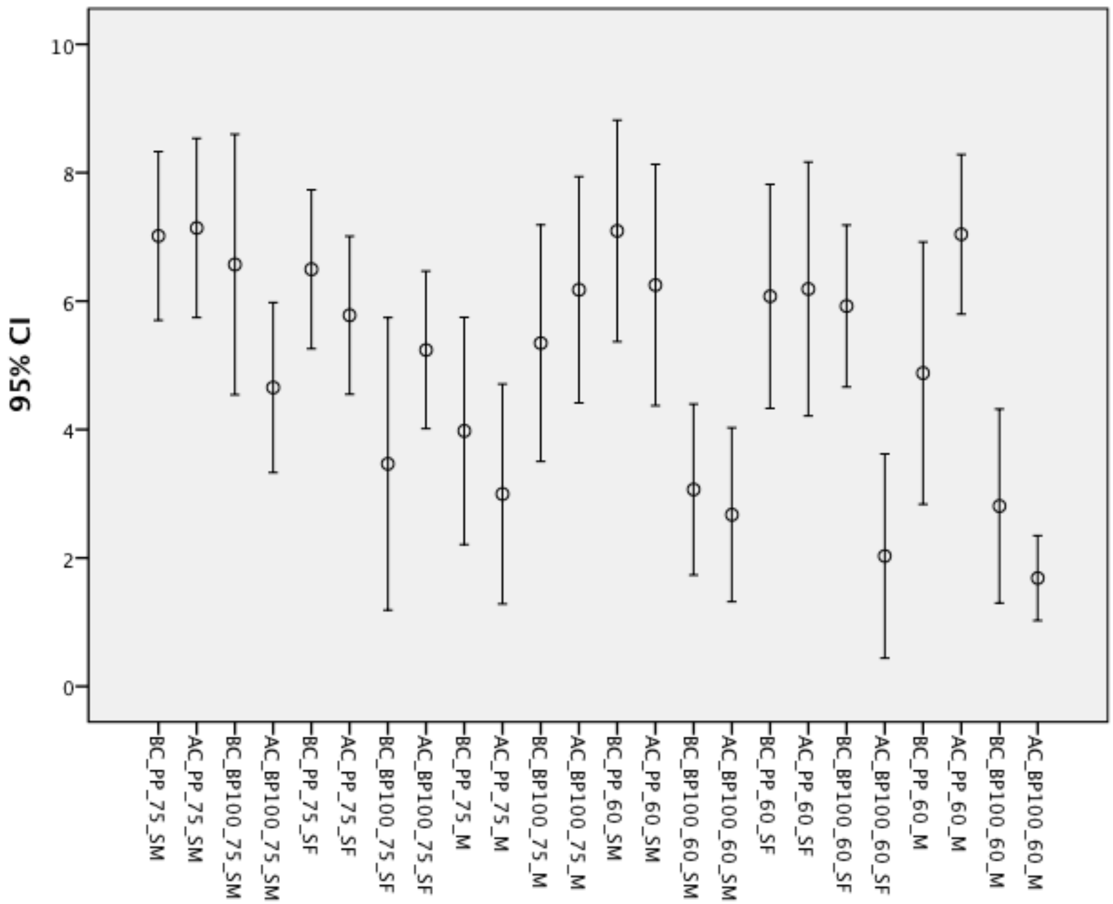


Figure 20. Overall contrasts from the 4-way significant interaction from normal hearing listeners. Error bars represent the 95% confidence interval around the mean.

Non-Power Users: Ponto Pro and BP100

The first analysis performed on the data was to determine the frequency response/audibility curves for various conditions. An example of these curves is presented in Figure 21 and 22. We have plotted the average force level thresholds for the non-power users as measured with the Ponto Pro. Additionally, we plotted the output frequency responses for the Ponto Pro and BP100 for male speech with a 60 and 75 dB SPL input level. These are the average responses prescribed by each manufacturer for the average hearing losses measured. There are some obvious differences in both the 60 and 75 dB SPL conditions.

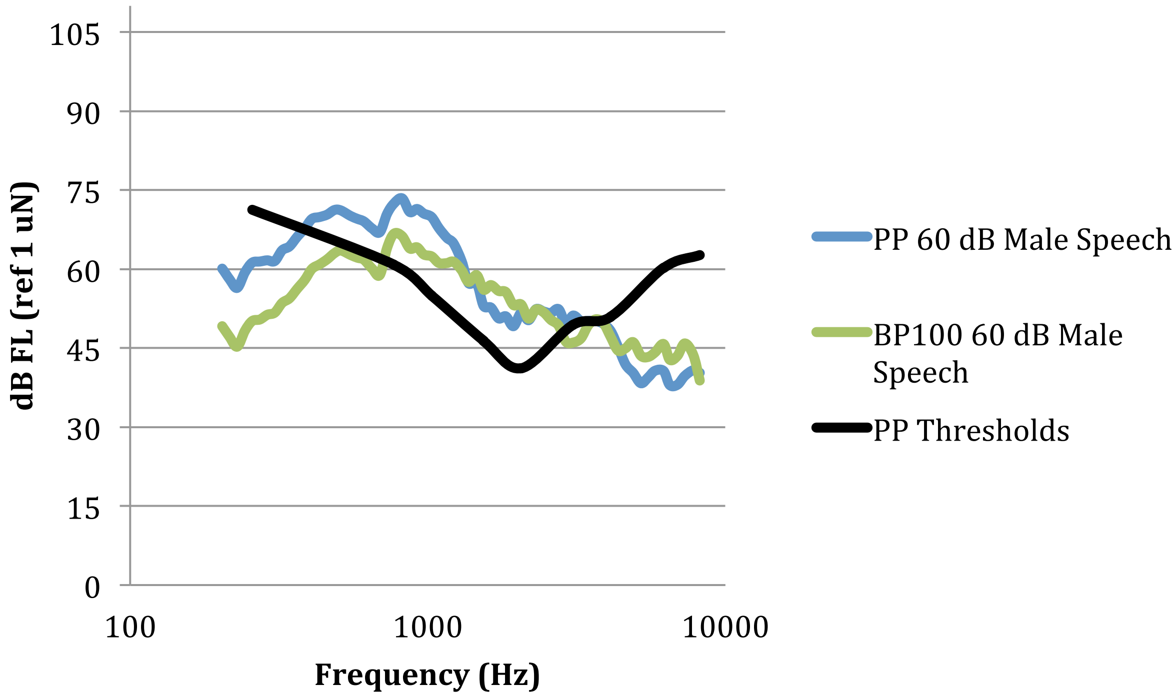


Figure 21. Frequency response/audibility curves for the 60 dB SPL male speech signals.

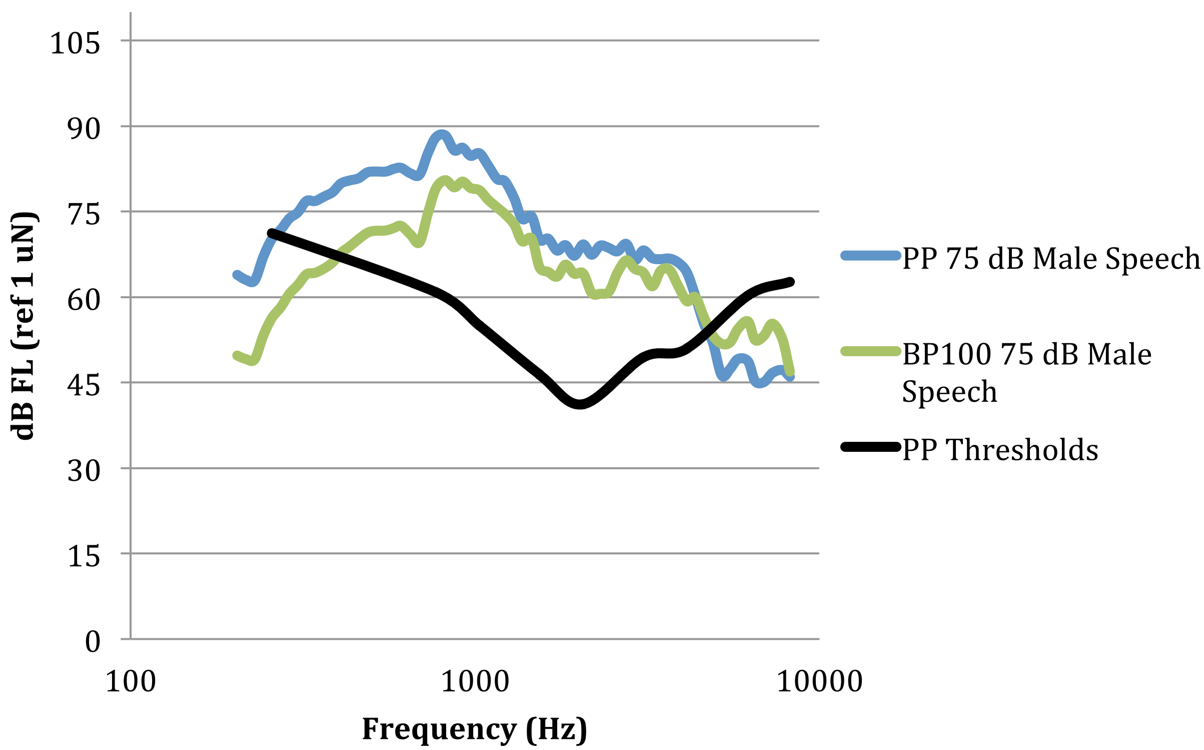


Figure 22. Frequency response/audibility curves for the 75 dB SPL male speech signals.

The second analysis performed on the data was to run a correlation matrix between conditions measured with the naked Ponto Pro and BP100 transducers. Results can be found in Table 2. The Pearson correlations that are presented were obtained by correlating the scores for the same condition (e.g., Ponto Pro, 60, speech male) with the 2 different naked transducers. The significant correlations are bolded in the table. A significant correlation was found between only two conditions.

Table 2 Correlations between scores with the different transducers (Ponto Pro and BP100) for the non Power BAHA users. Bolded rows reached significance

Condition	Pearson Correlation Between Transducers	Significance	N
Ponto Pro, 75, Speech Male	0.399	0.327	8
BP100, 75, Speech Male	0.099	0.916	8
Ponto Pro, 75, Speech Female	0.535	0.172	8
BP100, 75, Speech Female	0.679	0.064	8
Ponto Pro, 75, Music	**0.849**	**0.008**	**8**
BP100, 75, Music	0.368	0.369	8
Ponto Pro, 60, Speech Male	0.374	0.362	8
BP100, 60, Speech Male	**0.879**	**0.004**	**8**
Ponto Pro, 60, Speech Female	0.543	0.165	8
BP100, 60, Speech Female	0.676	0.066	8
Ponto Pro, 60, Music	0.481	0.228	8
BP100, 60, Music	0.658	0.076	8

Next, 2 separate 2 x 2 x 3 repeated measures ANOVAs were run on the data from the non-power Ponto Pro and BP100 users (1 for the data with the naked Ponto Pro transducer and 1 with the naked BP100 transducer). For each ANOVA, the first independent variable (IV) was

Device with 2 levels (Ponto Pro and BP100). The next IV was Level with 2 levels (60 and 75 dB SPL), and the final IV was Stimulus with 3 levels (Male Speech, Female Speech and Music).

The results of the first ANOVA run with the naked Ponto Pro transducer showed no significant main effects or interactions. Figure 23 shows the results of these conditions. There was entirely too much overlap in scores (too much variability) for any significance to emerge.

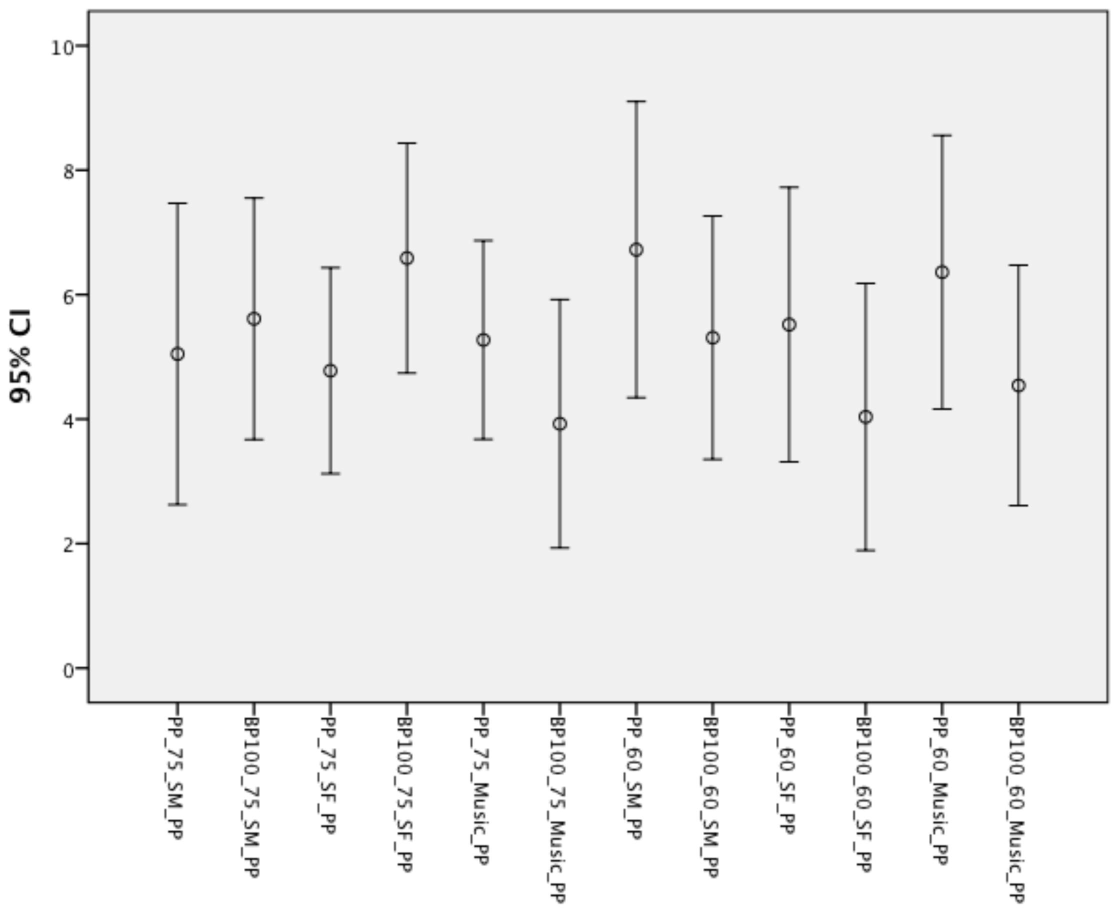


Figure 23. Sound quality results when measured using the Ponto Pro naked transducer.

Results of the second ANOVA showed the same results. None of the main effects or interactions were significant. See Figure 24. It is interesting to note that some important differences may be emerging in the data between these two different transducers. However, greater subject numbers would be required to sort out if these emerging differences were important.

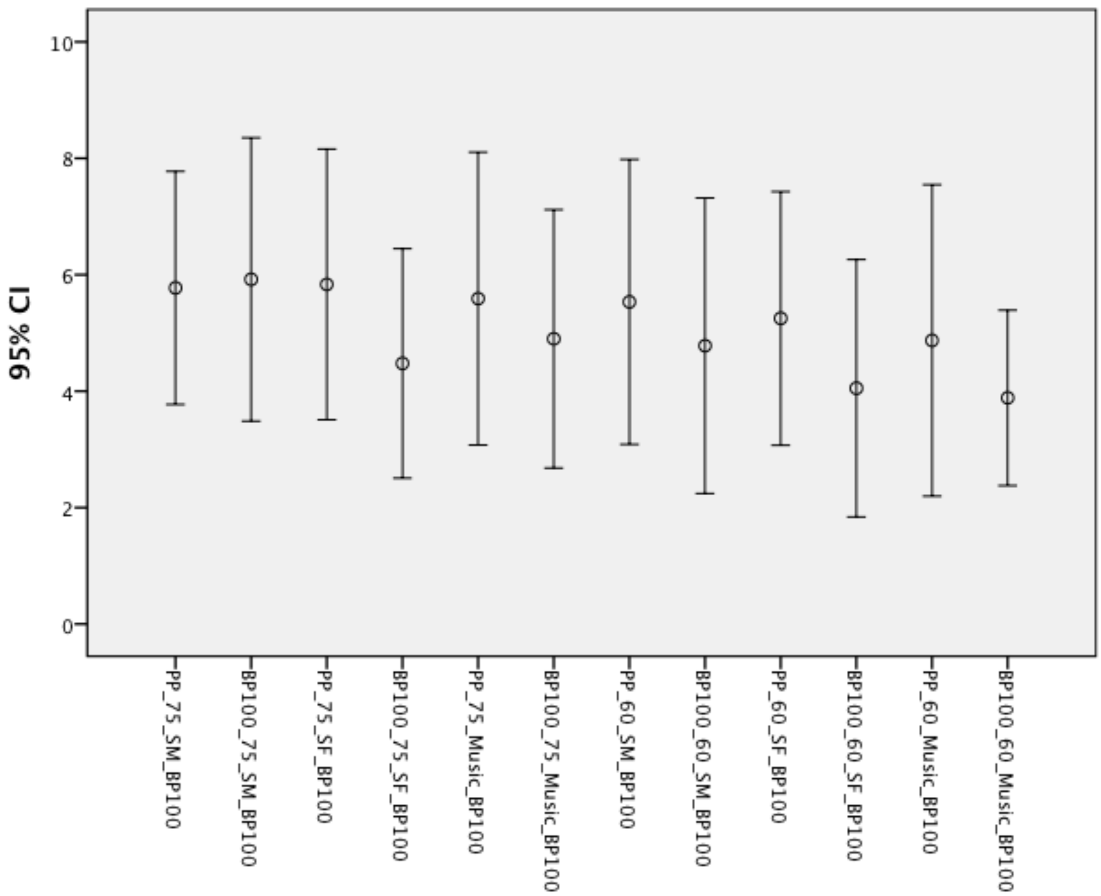


Figure 24. Sound quality results when measured using the BP100 naked transducer.

Power Users – Ponto Pro Power and BP110

As with the non-power users we began our analysis of the power users by first plotting the frequency response/audibility curves for each device. Examples are presented in Figure 24 and 25 below, showing the audibility for male speech with each device for a 60 and 75 dB SPL input.

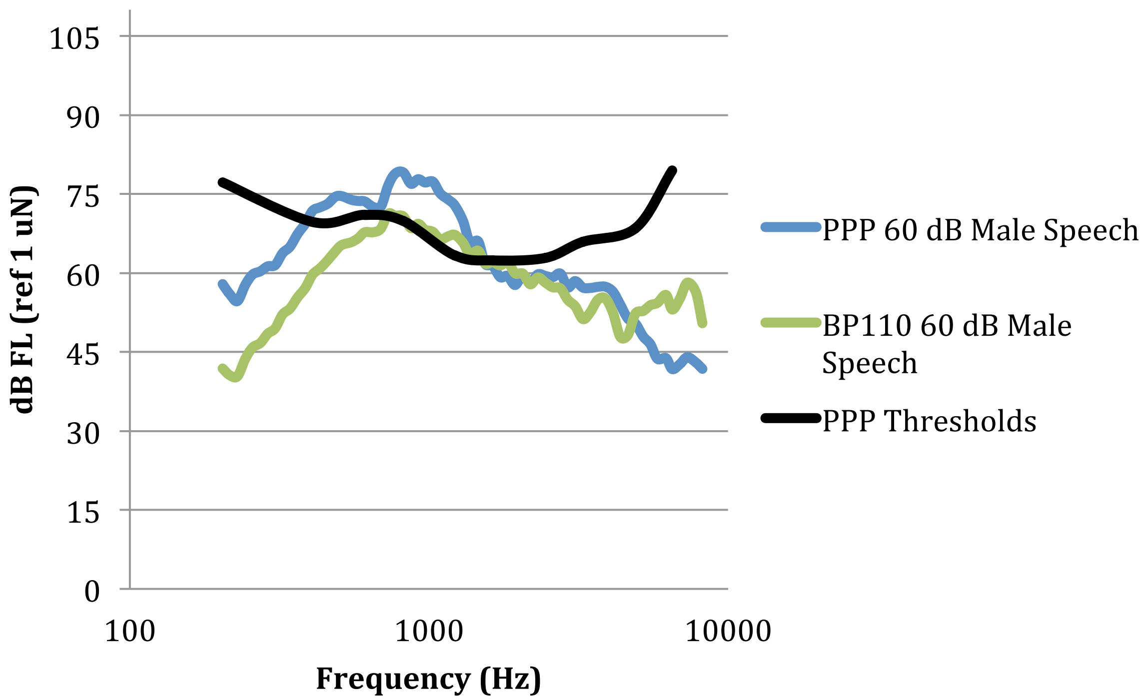


Figure 25. Frequency response/audibility curves for the 60 dB SPL male speech signal.

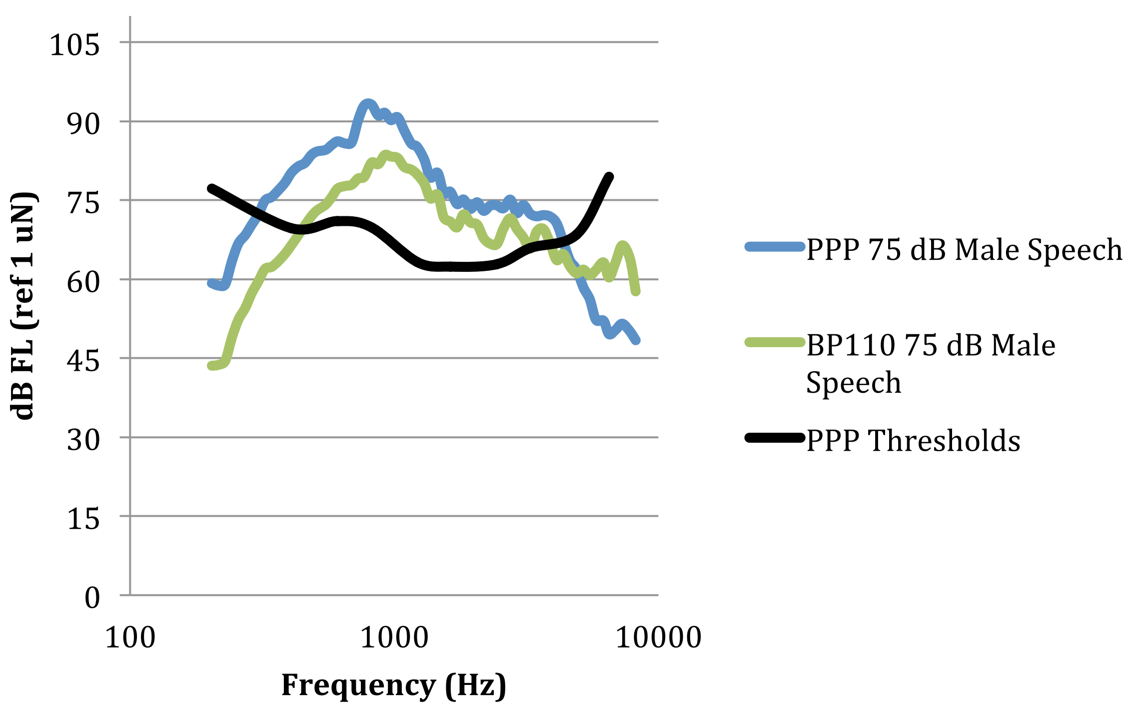


Figure 26. Frequency response/audibility curves for the 75 dB SPL male speech signals.

Next we calculated the strength of correlation between conditions measured with each of the two naked transducers (Ponto Pro Power and BP110). Better agreement was found with the power users between conditions. In the non-power group only 2 of the 12 conditions reached significance, whereas in the power group, 7 of the 12 conditions reached significance. See Table 3.

Table 3. Correlations between scores with the different transducers (Ponto Pro Power and BP110) for the Power BAHA users. Bolded rows reached significance.

Condition	Pearson Correlation Between Transducers	Significance	N
Ponto Pro Power, 75, Speech Male	0.042	0.901	11
BP110, 75, Speech Male	0.541	0.086	11
Ponto Pro Power, 75, Speech Female	**0.803**	**0.003**	**11**
BP110, 75, Speech Female	**0.729**	**0.011**	**11**
Ponto Pro Power, 75, Music	**0.881**	**0.001**	**11**
BP110, 75, Music	0.128	0.709	11
Ponto Pro Power, 60, Speech Male	0.506	0.112	11
BP110, 60, Speech Male	0.482	0.133	11
Ponto Pro Power, 60, Speech Female	**0.715**	**0.013**	**11**
BP110, 60, Speech Female	**0.971**	**0.001**	**11**
Ponto Pro Power, 60, Music	**0.637**	**0.035**	**11**
BP110, 60, Music	**0.996**	**0.001**	**11**

Again, 2 separate 2 x 2 x 3 repeated measures ANOVAs were run on the data from the Ponto Pro Power and BP110 users (1 for the data with the Ponto Pro Power transducer and 1 with the BP110 transducer). For each ANOVA, the first independent variable (IV) was Device with 2 levels (Ponto Pro Power and BP110). The next IV was Level with 2 levels (60 and 75 dB SPL), and the final IV was Stimulus with 3 levels (Male Speech, Female Speech and Music).

Results of the first ANOVA run with the naked Ponto Pro Power transducer revealed a significant main effect for Device ($F_{(1, 10)} = 13.470$, $p < 0.004$), and a significant interaction for Device x Level ($F_{(1, 10)} = 4.987$, $p < 0.050$). Figure 27 shows the main effect for device comparisons. Device 1 is the Ponto Pro Power and Device 2 is the BP110. This means that, when collapsed across all other variables, there is a small but significant advantage for the Ponto Pro Power. Figure 28 shows the interaction between Device and Level. This shows that when collapsed across all stimulus types, the ratings of the sound quality for a given device will differ depending on which level is used. Specifically people preferred the BP110 less when the signal level was low (60 dB SPL).

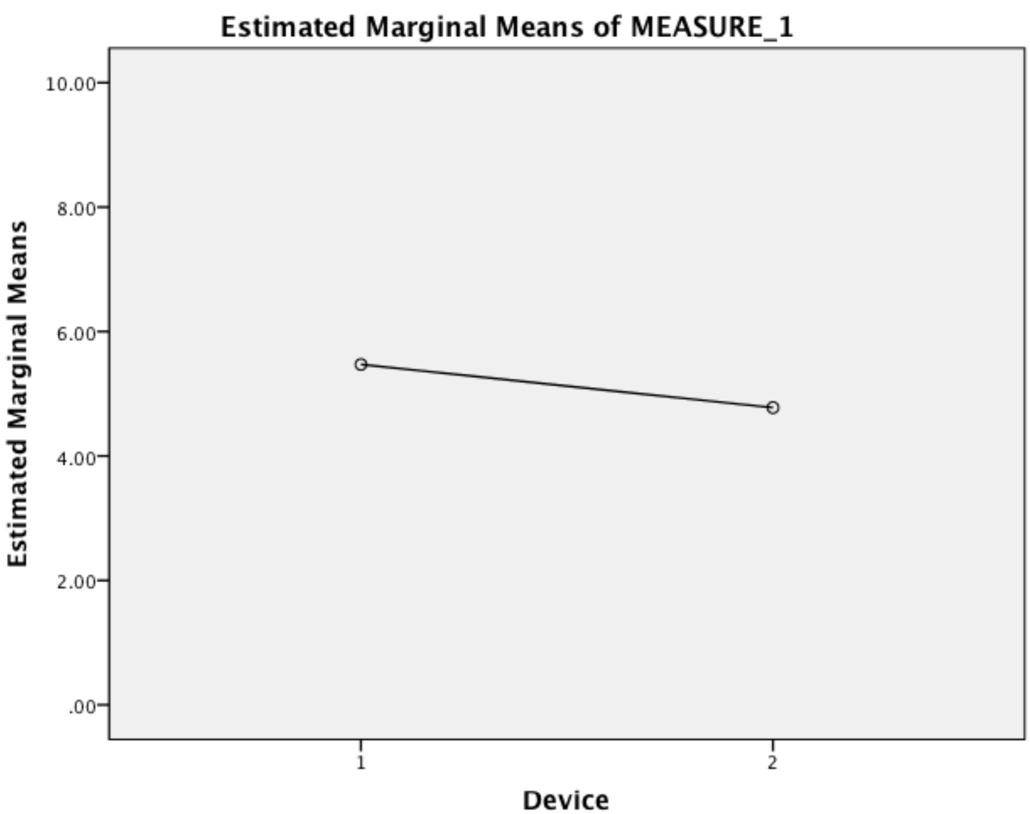


Figure 27. Main effect of Device for the Ponto Pro Power transducer condition. Device 1 is the Ponto Pro Power and Device 2 in the BP110.

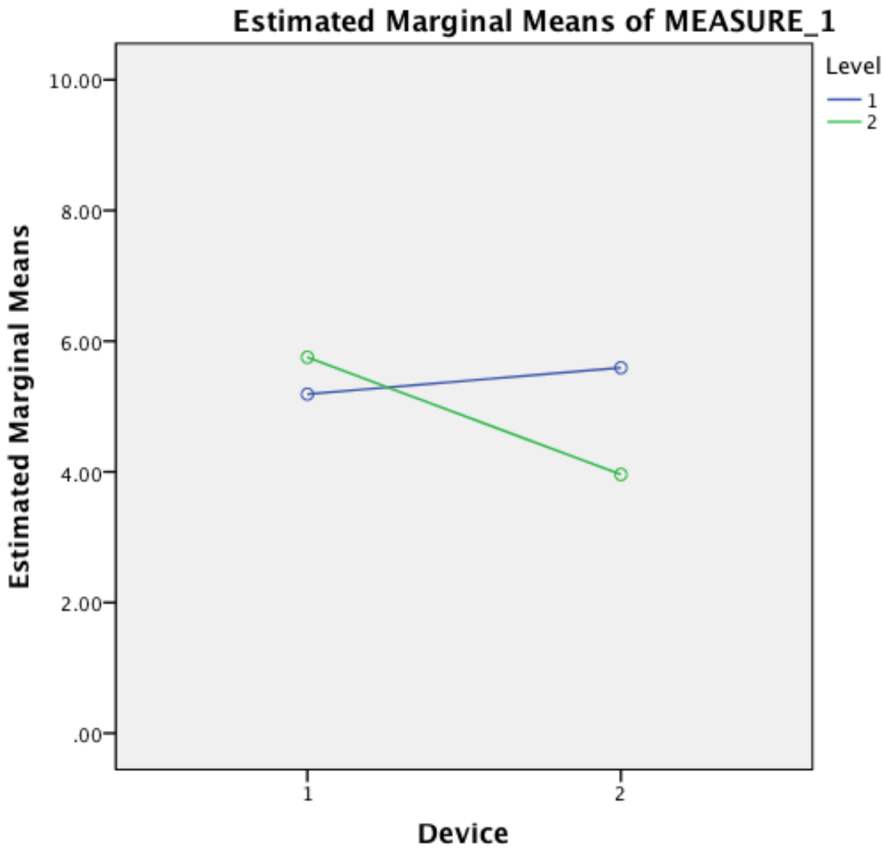


Figure 28. Interaction between Device and Level for the Ponto Pro Power transducer condition. Device 1 is the Ponto Pro Power and Device 2 is the BP110. Level 1 is 75 dB SPL input and Level 2 is the 60 dB SPL input.

Even though the 3-way interaction between Device, Level and Stimulus was not significant, we plotted the data for all conditions to determine if any trends appeared to be emerging. A trend may be emerging to indicate the power users prefer loud female speech with the BP110 to the Ponto Pro Power (see Figure 29). The opposite trend appears to be emerging with the soft female speech where it appears that people may prefer the Ponto Pro Power. Again, these are exploratory trends only.

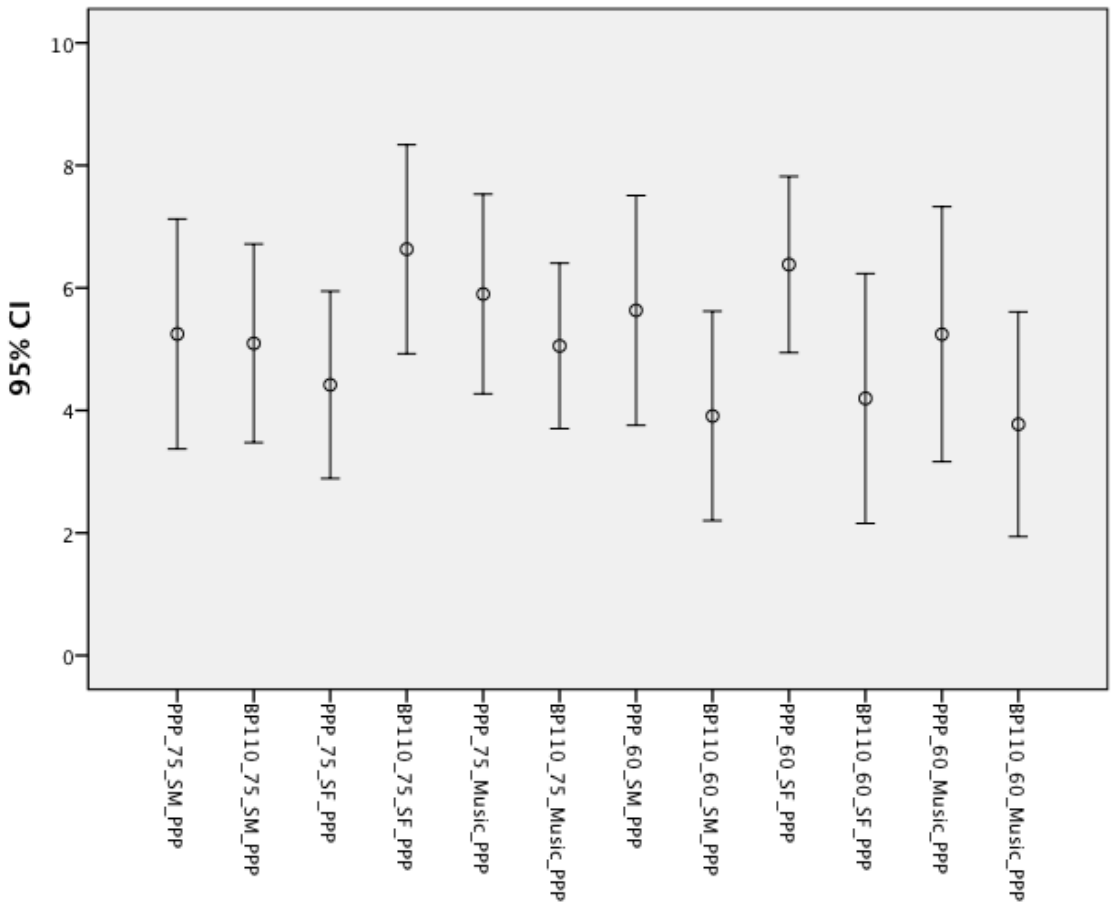


Figure 29. Means plus the 95% CI for each of the twelve conditions under test with the naked Ponto Pro Power transducer.

We ran the same ANOVA again on the power user data when they listened to the recordings using the naked BP110 transducer. We discovered similar trends as with the Ponto Pro Power Transducer. This time however, only the main effect for Device was significant. Figure 30 shows that the Ponto Pro Power was slightly preferred to the BP110.

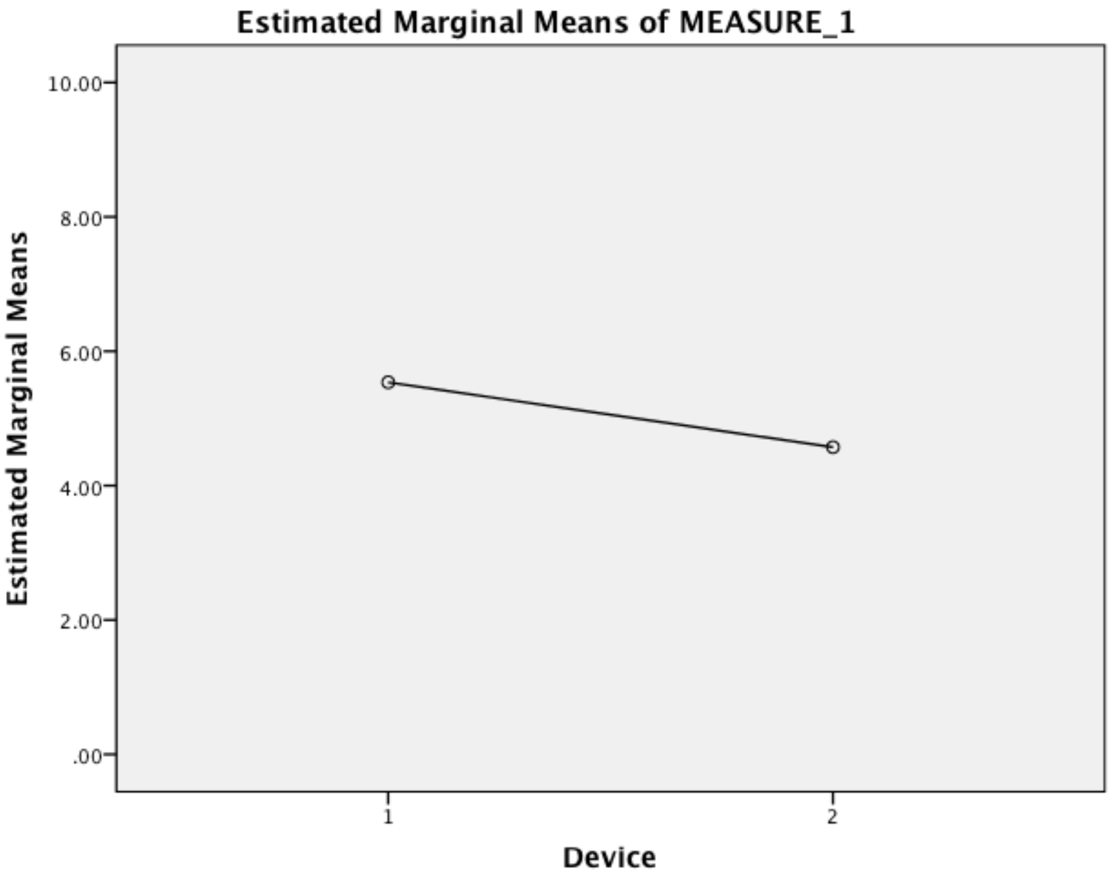


Figure 30. Main effect of Device when rated using the naked BP110 transducer. Device 1 is the Ponto Pro Power and Device 2 is the BP110.

Again, we plotted the graph with all conditions as measured with the BP110 naked transducer to see if any trends were emerging and to check the agreement with the data from the Ponto Pro Power transducer tests. See Figure 31.

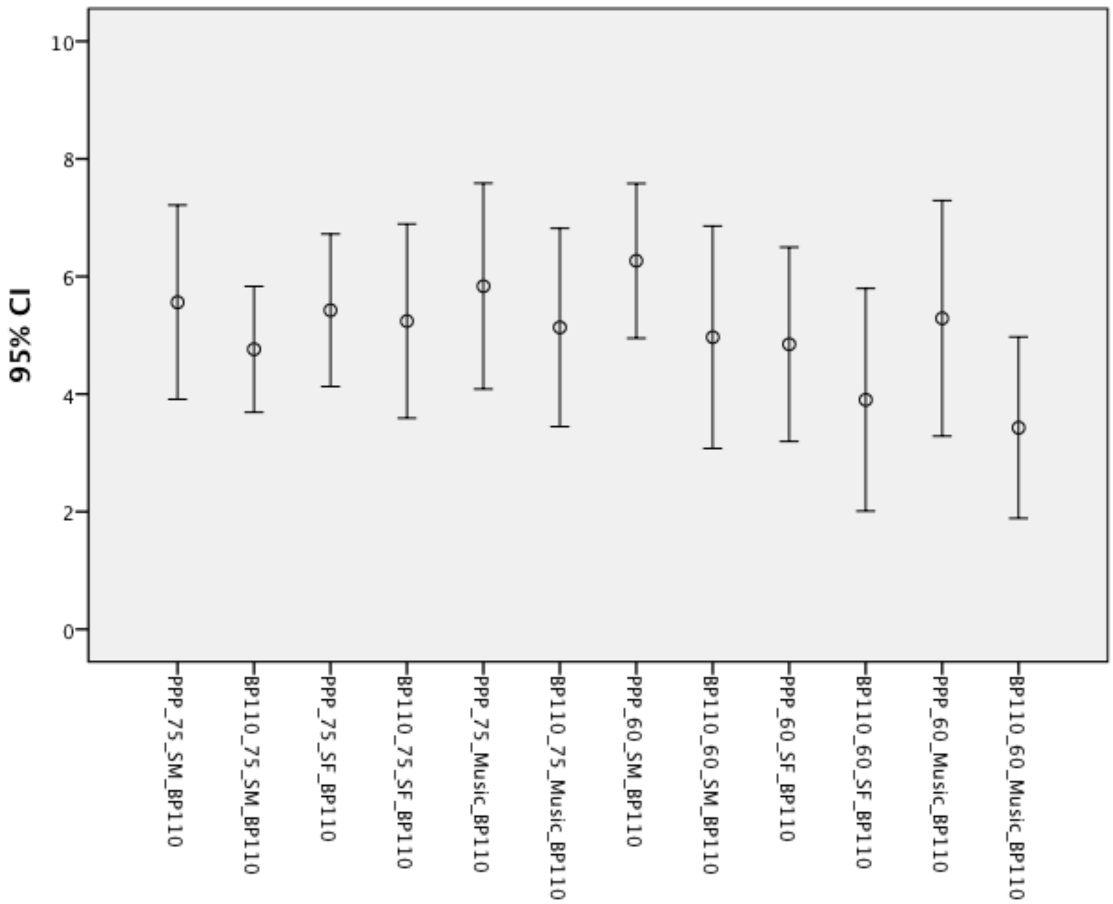


Figure 31. Means plus the 95% CI for each of the twelve conditions under test with the naked BP110 transducer.

Discussion

The primary aim of this study was to determine if significant sound quality differences existed between Oticon Medical's devices and Cochlear Corporation's devices, as these are currently the two manufacturers of BAHA devices.

Normal Hearing Subjects

Normal hearing subjects had the opportunity to rate each of the files through 2 modes of stimulation: air conduction and bone conduction. Although no main effects for mode of stimulation were found, it is worth mentioning that the data were beginning to approach significance. It is likely that with more subjects a true clinical difference would exist. This is further supported by the planned comparisons analyzing the interaction effect between the 4 variables. Although only 20 comparisons were made, the data suggest that at least one condition (BP100-60 dB SPL- female speech) resulted in normal listeners rating bone conduction superior to air conduction stimulation. Again, this result supports the idea that mode of stimulation plays a factor in how normal hearing individuals rate sound quality.

In terms of device preference, normal hearing listeners rated the Ponto Pro device superior in terms of sound quality. When comparing input levels, normal hearing individuals rated sound quality as superior when recorded at an input of 75 dB SPL. It also appears that normal hearing individuals preferred the sound quality of the male stimulus, rating it slightly superior to that of the female speaker.

The most important finding from the normal hearing ratings was the interaction effect between the mode of stimulus, device, level and type of stimulus. It appears as though the sound quality ratings differ significantly depending on the interaction of all 4 of these variables.

Although these findings are of interest, they should be interpreted with caution. It is probable that sound quality judgments could change if the normal listeners rated research devices set to various conductive hearing losses, but again for the purpose of this study only a flat conductive loss was examined. More importantly, the sample of normal listeners was small. It is likely that with both a larger sample size and analysis of various types of conductive losses, differences in the data would arise. In addition, a larger sample size would minimize the amount of variability found in the data. This data suggests that normal hearing individuals differ largely in terms of what they deem to be acceptable sound quality characteristics.

Non-power users

Data from the non-power users indicated no sound quality differences in terms of main effects or interactions. However, the results are still compelling. It appears that the non-power users rated the recordings with an extreme amount of variability regardless of device, input, or type of stimulus. Although some subjects indicated that sound quality differences existed, there was minimal consistency between participants on how the differences were rated on the scale. Ultimately, this resulted in tremendous variability in the data. For example, Figure 32 illustrates 2 different subjects' ratings of sound quality. For some conditions, these subjects appear to agree fairly closely (PP 60 dB SPL Speech Male and BP100 75 dB SPL Speech Male). However, for other conditions they differ greatly (e.g., PP 75 dB SPL Speech Female).

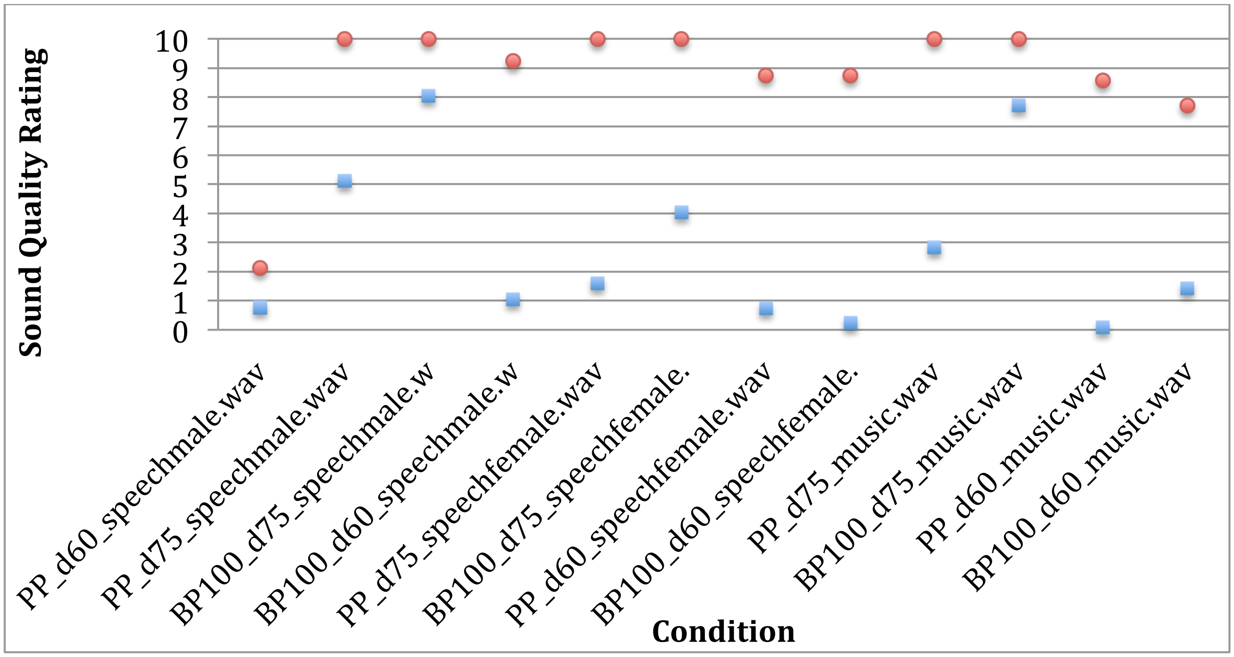


Figure 32. Differences in 2 non-power subjects sound quality ratings.

Findings from this group differed from that of normal hearing listeners. While normal hearing participants rated the Ponto Pro to have superior sound quality, the non-power BAHA users did not. It appears as though normal hearing individuals preference for sound quality differs from that of BAHA users rating the same devices. This could be due to the fact that the conductive loss experienced by the normal listeners was a flat loss. Ratings for normal hearing listeners may change if the loss was set differently.

In regards to input levels, normal hearing individuals rated recordings as having superior sound quality when recorded at an input of 75 dB SLP. Again, these findings are not consistent with what the researchers found when comparing the same devices with non-power BAHA users. It appears that normal listeners rate sound quality differently then BAHA users rating the same devices. We speculate that there may be noise floor effects (e.g., microphone noise) available to the normal hearing listeners in the 60 dB SPL conditions that may not be as obvious to the BAHA users. Anecdotally, we noted a few normal hearing subjects reported hearing more of these noise floor effects when listening by air conduction than when listening by bone conduction. The high skull impedance in the low frequencies for bone-conducted sound may explain this. It may simply be harder to hear these artifacts when listening with a bone oscillator versus headphones. More research is needed to see if this is truly the case.

In addition, it appears as though there is vast inconsistency in the correlations between the two naked transducers with only 2 out of 12 correlations being significant. This further indicates that a large amount of variability exists between individuals' subjective ratings and also within each individual's ratings. Given that the transducers have similar frequency responses, we expected that there would be greater agreement between testing with the Ponto and BP100 transducers. It is possible that the differences in scores is partially accounted for by these differences in frequency responses, however, it is more likely that subjects were too variable in their ratings of the sound quality of the devices. Greater numbers of subjects are needed.

Power Users

Results from the power participants were unique from both of the other groups and reflected a greater consistency in sound quality judgments. When examining the agreement between the correlations of the naked Ponto Pro Power and BP110, 7 out of 12 were determined to be of significance. This is substantially higher than the 2 out of 12 correlations found to be significant for non-power users.

In addition, a significant main effect for device was found regardless of which naked transducer was used, contrasting again from the non-power group. A significant interaction effect was also found for device and level when subjects rated the files using the naked Ponto

Pro Power transducer. It appeared that individuals rated the BP110 as having lower sound quality when the input level was lower (60 dB SPL).

Again, there were large amounts of variability found in the data for power users. The frequency response difference between the 2 power devices appeared to agree closely with the non-power devices. This further supports the researchers' claim that it is unlikely that the small differences in the frequency response between the non-power devices contributed substantially to the high level of variability.

It is probable that with a larger sample size not only would the variability between the subjects be reduced, but also other findings may begin to emerge. Researchers noted that trends were beginning to show for a 3-way interaction between device, level and stimulus. In addition, it appears that there may be a trend towards power users preferring loud female speech with the BP100 to the Ponto Pro Power. Lastly, the opposite appears to be emerging for softer female speech, for which it appears that people may actually prefer the Ponto Pro Power.

Conclusion

The study indicated that both normal hearing listeners and power users showed a slight preference for the Oticon Medical devices. Although no significant findings were evident with non-power users, some interesting trends were beginning to emerge and with continued research and a larger sample size it is likely these may become more prominent. Power users on the other hand appeared to have more agreement in their sound quality ratings with a significant main effect for the Ponto Pro device, regardless of which isolated transducer they were listening to the files with. One interesting finding from this study was found in terms of the audibility of signals for each user with the different manufacturer prescriptions. It may well be that the greater audibility shown in figures 21, 22, 25 and 26 is responsible for the trending towards better sound quality ratings for the Oticon devices. A further investigation is planned where we look into whether or not audibility, if properly controlled for, has a significant impact on any differences in sound quality judgments.

The large amounts of variation in the data from this study would likely be minimized with a larger sample. In addition, it appeared that experimenter differences lead to large amounts of variation. In the future, participants need to be given clearer instructions in terms of how to rate differences in sound quality. Participants would also benefit from having the chance to practice making sound quality judgments prior to the actually testing. This would familiarize them with the task and ensure that they are making judgments based solely on sound quality and not on intelligibility for example. Using a smaller sound quality rating scale may also reduce the large amounts of variability. The 0-10 scale allowed for a large range of judgments.

Lastly, it would be beneficial to examine inter-rater reliability ratings between the same isolated transducers (e.g., isolated Ponto Pro vs. isolated Ponto Pro). By only evaluating inter-rater sound quality ratings between the 2 different the isolated transducers (e.g., isolated Ponto Pro vs. isolated BP100), it is difficult to determine how much the transfer functions of each transducer are impacting the differences between ratings.

In general the research findings need to be interpreted with caution, as the sample sizes for each condition were small. Although only a pilot study, the results are promising in terms of providing a more accurate picture of which devices can offer the best advantage in terms of sound quality. This study indicates that further research needs to be conducted in the area of sound quality.

Bibliography

Bance, M., Abel, S., Papsin, B., Wade, P., Vendramini, J. (2002) A comparison of the audiometric performance of bone anchored hearing aids and air conduction hearing aids. *Otology & Neurotology*, 23, 912-919.

Bess, F.H., Humes, L.E. (2003) Audiology: The Fundementals, 4th Edition. Lippincott Williams and Wilkins

Hakansson, B. (2003) The balanced electromagnetic balanced transducer: A new bone conduction transducer. *Acoustical Society of America*, 113, 818-825

Hakansson, B., Carlsson, P. (1994) The bone anchored hearing aid: Principal design and audiometric results. *Ear, Nose & Throat Journal*, 47, 670-679.

Hodgetts, W., Chen, G., Parsa, V. (2004) Advanced measures of bone anchored hearings aids: Do they correlate with perceptual judgments. *Canadian Acoustics*, 32, 74-75.

Hodgetts, W. (2011) Thesis abstract. University of Alberta, Department of Speech Pathology and Audiology, April 2011.

Kompis, M., Krebs, M., Hausler, R. (2007) Speech understanding in quiet and in noise with the bone anchored hearing aids Baha Compact and Baha Divino. Oto-Laryngologica, 127, 829- 835.

Mylanus, E., van der Pouw, K., Snik, A., Cremers, C. (1998) Intraindividual comparison of the bone anchored hearing aid and air conduction hearing aids. *Otolaryngol head and neck surgery*, 124, 271-276.

Pearsons, K. Bennett, R. Fidell, S. (1977) Speech levels in various noise environments. Project report on contract 680 – 2466. Washington, DC: U.S. Environment Protection Agency

Snik, A., Mylanus, E., Proops, D., Wolfaardt, J., Hodgetts, W., Somers, T., Niparko, J., Wazen, J., Sterkers, O., Cremers, C., Tjellstrom, A. (2005) Consensus statements on the BAHA system: Where do we stand at present. *Annals of Otology, Rhinology & Laryngology*, 144(12), 195:1-12.