# Phonak
# Field Study News

## Audiogram and AudiogramDirect: comparison of in-clinic assessments

In-situ audiometry, such as Phonak AudiogramDirect, allows clinicians to measure hearing thresholds directly through a client's hearing aids. Retrospective analyses of 167 722 in-clinic assessments of standard audiometry and AudiogramDirect show new insights.

Author: Charlotte Vercammen / June 2020

### Key Highlights

- Results showed that the outcomes of standard and in-situ audiometry (Phonak AudiogramDirect) were highly correlated, suggesting a high level of agreement between the methods.

- The mean difference between the outcomes for pure-tone averages was < 1 dB HL – negligible from a clinical point of view.

- 95% of the differences between outcomes varied within an interval of ±15 dB HL for pure-tone averages. This range of variation can be considered within a clinically acceptable range and is, in fact, expected.

- The results are frequency-dependent, with a trend towards lower variation for middle frequencies (see Appendix 2 for frequency-specific results).

### Considerations for practice

- While the outcomes of in-situ audiometry are highly related to those obtained through standard audiometry, differences in outcomes between both are expected.

- In-situ audiometry, such as Phonak AudiogramDirect, allows clinicians to measure hearing thresholds directly through a client's hearing aids. This can be done to increase the accuracy of a fitting, or to estimate hearing thresholds when no clinical audiometer is available. An example would be when a client is unable to come into the hearing care center and the appointment is taking place remotely.

PHONAK
life is on

## Introduction

In-situ audiometry, available within Phonak Target fitting software as AudiogramDirect, allows clinicians to measure hearing thresholds directly through a client's hearing aids. With the hearing aids in place ("in-situ"; Block, 2008), the hearing aid receivers deliver sound stimuli to the ear, while the fitting software controls stimulation level (O'Brien, Keidser, Yeend, Hartley, & Dillon, 2010).

When only standard audiometric data is available, fitting formula algorithms use an estimated value of the Ear-to-Coupler Level Difference (ECLD) for a chosen hearing instrument to calculate target gain. By performing in-situ audiometry, individual deviations from those estimated values - due to residual ear canal volume, hearing aid insertion depth, acoustic coupling seal, applied venting, and individual receiver characteristics (Block, 2008) – can, in part, be considered. This will increase the accuracy of a fitting, resulting in a better match to target.

Another application of in-situ audiometry is that it allows clinicians to estimate Air Conduction (AC) hearing thresholds when no clinical audiometer is available (Kiessling et al., 2015). In this respect, it is important to know to what extent in-situ audiometric outcomes agree with or vary from standard audiometric outcomes, the gold standard for diagnosis of hearing loss in clinical practice (Roeser et al., 2007). To this end, the current retrospective study compared AC hearing thresholds obtained through Phonak AudiogramDirect and standard audiometry during the same in-clinic appointment for a large sample of 167 722 fittings. Based on previously collected data, we hypothesized AudiogramDirect outcomes to fall within a tolerance range of $\pm$10-15 dB HL when compared to hearing thresholds obtained through standard audiometry (Omisore, 2011).

## Methodology

A large cross-sectional sample of 167 722 Phonak Marvel hearing aid fitting logs was retrospectively analyzed. The data was collected through Phonak Target software of clinicians worldwide who participated in the Phonak Target Improvement Program by enabling the logging function in the fitting software.

Fitting data was included in the analyses based on the following criteria: fittings performed through Phonak Target software version 6.1 or more recent, for the Marvel platform, using the Adaptive Phonak Digital proprietary fitting formula, for clients who were 18 years of age or older, when standard audiometry and in-situ audiometry (through AudiogramDirect) were performed during the same in-clinic appointment. Simulated fitting logs (originating from the Target training module) were excluded. All personal information was removed from the fitting logs, besides age and country (for descriptive results – see Appendix 1).

Statistical analyses were performed using R software (R Core Team, 2020). Pearson correlation analyses were used to investigate the potential relationship between standard audiometric and AudiogramDirect outcomes. The Bland-Altman method was used to investigate potential differences between the two measures (Bland & Altman, 1986; Giavarina, 2015). With A, outcomes obtained by the first measure and B, outcomes obtained by the second measure, the Bland-Altman method plots the average of paired outcomes ((A+B)/2; $x$ axis) as a function of the difference of paired outcomes (A-B; $y$ axis). This allows the calculation of two parameters, the 'bias' between the two methods and the 'lines of agreement' (Bland & Altman, 1986; Giavarina, 2015).

The 'bias' represents the mean of the differences between the paired data (Bland & Altman, 1986; Giavarina, 2015). As an example, a hypothetical bias of 10 units would mean that on average, the outcomes of the second method (B) would be 10 units lower than the outcomes measured by the first method (A), i.e. on average, the second method would underestimate the outcomes compared to the first method. Ideally, the bias would be 0.

The 'lines of agreement' represent the area on the Bland-Altman plot within which 95% of the data on the differences between the two measures lie. It should then be decided whether this range of variation is within clinically acceptable limits (Bland & Altman, 1986; Giavarina, 2015).

The analyses were performed for pure-tone averages across three frequencies (500, 1000, 2000 Hz; PTA3) and four frequencies (500, 1000, 2000, 4000 Hz; PTA4) – see Results section. The analyses were also performed for individual frequencies (250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000 Hz) – see Appendix 2. For every separate analysis, missing values were pairwise removed. In other words, if a fitting log had missing values for any of the frequencies included in PTA3 or PTA4, either obtained through standard audiometry or through AudiogramDirect, the fitting log was not included in the respective analysis. The total number of datapoints included in each analysis is indicated in the Results section by $n$.

# Results

There was a strong, positive correlation between PTA3 obtained through standard audiometry and AudiogramDirect, which was statistically significant ($r$ = .90, $n$ = 167 219, $p$ < .0001, 95% CI [.90, .90]). There was also a strong, positive correlation between PTA4 obtained through standard audiometry and AudiogramDirect, which was statistically significant ($r$ = .89, 95% CI [.89, .89], $n$ = 166 796, $p$ < .0001). The scatterplots in Figure 1 visualize the results for PTA3 (upper panel) and PTA4 (lower panel).

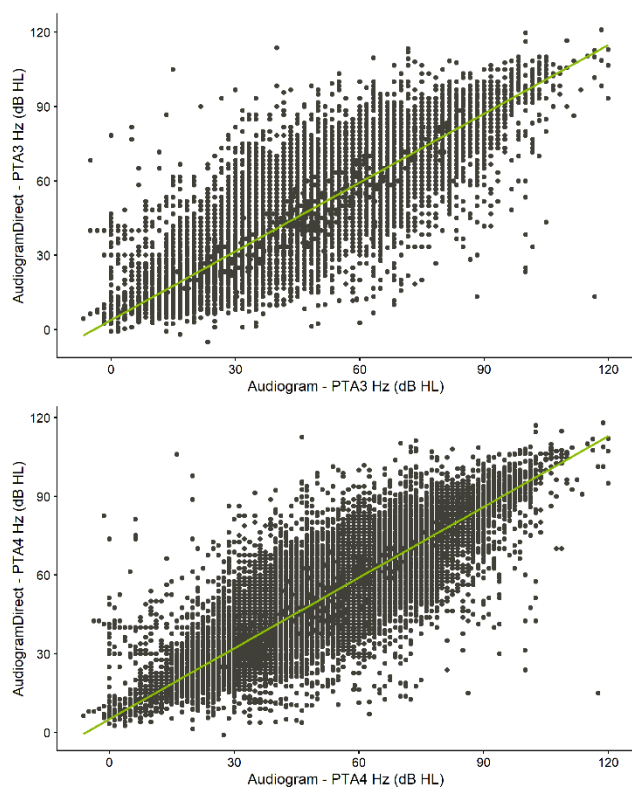

Figure 1. Scatter plots of PTA3 (dB HL; upper panel) and PTA4 (dB HL; lower panel), visualizing the outcomes obtained through standard audiometry ($x$ axis) as a function of AudiogramDirect ($y$ axis).

The Bland-Altman analyses (Bland & Altman, 1986; Giavarina, 2015) revealed a bias of –0.50 dB HL for PTA3 ($n$ = 167 219, 95% CI [–0.53, –0.46]). The lines of agreement fell between a lower limit of –14.73 dB HL (95% CI [–14.79, –14.67]) and upper limit of 13.74 dB HL (95% CI [13.68, 13.80]). The bias for PTA4 was –0.005 dB HL ($n$ = 166 796, 95% CI [–0.04, 0.03]). The lines of agreement fell between a lower limit of –13.26 dB HL (95% CI [–13.31, –13.20]) and upper limit of 13.25 dB HL (95% CI [13.19, 13.30]). The Bland-Altman plots in Figure 2 summarize the results for PTA3 (upper panel) and PTA4 (lower panel).
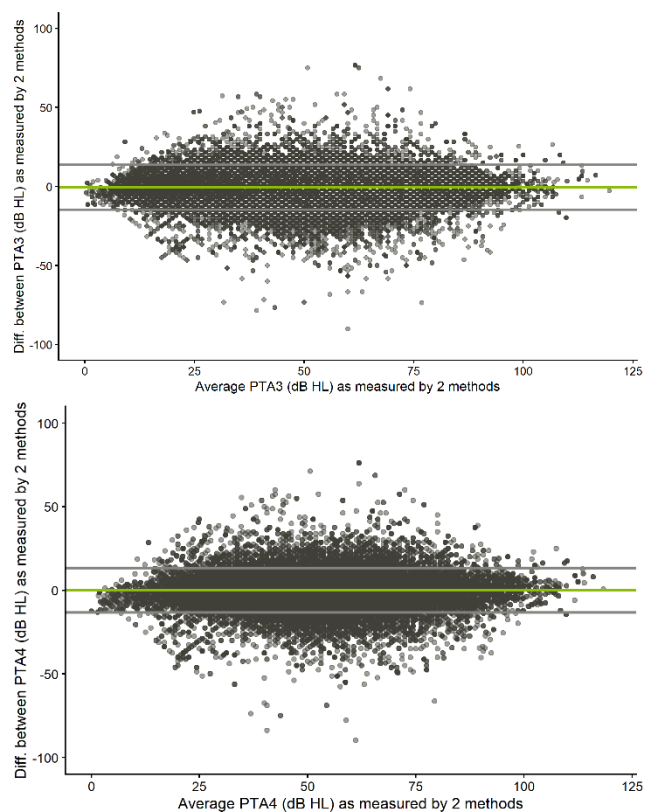


Figure 2. Bland-Altman plots for PTA3 (upper panel) and PTA4 (lower panel). With A and B representing, respectively, AC hearing thresholds obtained through standard audiometry and AudiogramDirect, the figures visualize the average of paired outcomes ((A+B)/2; $x$ axis) as a function of the difference of paired outcomes (A–B; $y$ axis). The horizontal green lines represent the bias, the horizontal gray lines the lines of agreement.

# Discussion

Results of these retrospective data analyses showed a high correlation between AC hearing thresholds obtained through standard audiometry and AudiogramDirect, both for PTA3 and PTA4. Thereby the results showed a high level of agreement between both methods. The Bland-Altman analyses showed that the bias between the two methods was small, –0.50 dB HL and –0.005 dB HL for PTA3 and PTA4, respectively. In other words, for a big sample, the mean difference between PTA3, respectively PTA4, obtained through standard audiometry and AudiogramDirect, was smaller than 1 dB HL – negligible from a clinical point of view. In addition to the bias, the lines of agreement represent the differences in PTA3, respectively PTA4, as obtained through standard audiometry and AudiogramDirect, for 95% of the data. Thereby they give an idea of the variation. For both averages, PTA3 and PTA4, the current results show that this range falls with ±15 dB HL. This range of variation is in line with previously collected data (Omisore, 2011) and is, in fact, expected. Indeed, while standard audiometry and in-situ audiometry are highly related to each other, in-situ audiometry takes into account the hearing aid insertion depth, acoustic coupling seal in the

ear canal, effects of venting, receiver characteristics, and chosen hearing aids. Standard audiometry does not. For both methods, as with all behavioral measures, variations due to perceptual imprecisions can also induce variation, as does data handling, e.g. by manually entering hearing threshold data in the fitting software.

It is important to note that all data included in these retrospective analyses was collected during in-clinic visits. Consequently, we cannot generalize the results to other, potentially less acoustically shielded, settings. It is also important to note that the extent to which in-situ audiometric outcomes agree with or vary from standard audiometric outcomes is frequency-dependent (Kiessling et al. 2015). The results reported in this paper were analyzed for pure-tone averages. The results for individual frequencies can be consulted in Appendix 2.

## Conclusion

PTA3 and PTA4 outcomes obtained through standard audiometry and AudiogramDirect are highly correlated. For both PTA3 and PTA4, the mean difference between the two audiometric methods is smaller than 1 dB HL – negligible from a clinical point of view. The differences between both methods for PTA3 and PTA4 varied within an interval of ±15 dB HL for 95% of the fittings included in these data analyses. This range of variation can be considered within a clinically acceptable range and is, in fact, expected. The variation between both methods could be explained by perceptual imprecisions during behavioral testing, and imprecisions during data handling. The variation between both methods could also be explained by methodological differences, as in-situ audiometry takes into account the hearing aid insertion depth, acoustic coupling seal in the ear canal, effects of venting, receiver characteristics, and chosen hearing aids, whereas standard audiometry does not.

## Acknowledgements

## References

Bland, M. J., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 307–310.

Block, M. (2008). Adding Precision to the Initial Hearing Aid Fitting. *The Hearing Professional, 57*(4), 7–10.

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica, 25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Kiessling, J., Leifholz, M., Unkel, S., Pons-Kühnemann, J., Jespersen, C. T., & Pedersen, J. N. (2015). A comparison of conventional and in-situ audiometry on participants with varying levels of sensorineural hearing loss. *Journal of the American Academy of Audiology, 26*(1), 68–79. https://doi.org/10.3766/jaaa.26.1.8

O'Brien, A., Keidser, G., Yeend, I., Hartley, L., & Dillon, H. (2010). Validity and reliability of in-situ air conduction thresholds measured through hearing aids coupled to closed and open instant-fit tips. *International Journal of Audiology, 49*(12), 868–876. https://doi.org/10.3109/14992027.2010.503664

Omisore, D. (2011). AudiogramDirect: In-situ hearing tests at their best. *Phonak Field Study News*, July, 1–3. Retrieved from: www.phonakpro.com/evidence. Accessed April 29th, 2020.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roeser, R. J., Valente, M., & Hosford-Dunn, H. (2007). *Audiology diagnosis*. New York, USA: Thieme Medical Publishers Incorporated
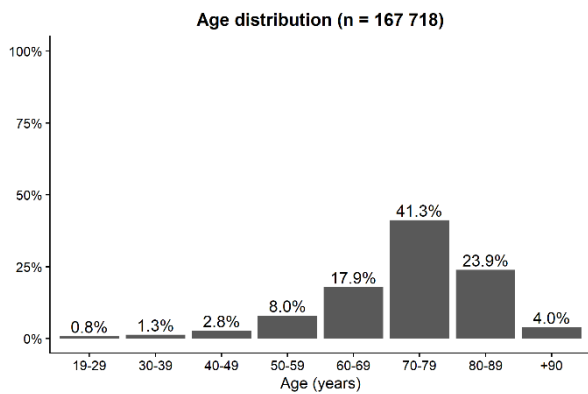
# Appendix 1



**Figure 3. Age distribution of clients whose fitting log files were included in the retrospective data analyses reported on this study. The full sample contained 167 722 data points – there were 4 missing values for Age.**
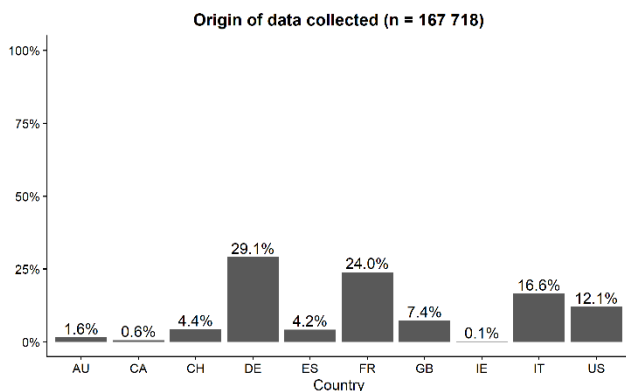


**Figure 4. Distribution of countries where the fitting logs included in the retrospective data analyses originated from. The full sample contained 167 722 data points – there were 4 missing values for Country.**

# Appendix 2

| Frequency | n | Pearson's r [95% CI] | p value |
|---|---|---|---|
| 250 Hz | 159 267 | 0.8244 [0.8228; 0.8260] | < 0.0001 |
| 500 Hz | 166 937 | 0.8651 [0.8639; 0.8663] | < 0.0001 |
| 750 Hz | 80 349 | 0.8877 [0.8862; 0.8891] | < 0.0001 |
| 1000 Hz | 166 982 | 0.8944 [0.8934; 0.8953] | < 0.0001 |
| 1500 Hz | 93 274 | 0.8962 [0.8949; 0.8975] | < 0.0001 |
| 2000 Hz | 166 979 | 0.8848 [0.8838; 0.8858] | < 0.0001 |
| 3000 Hz | 135 983 | 0.8729 [0.8716; 0.8741] | < 0.0001 |
| 4000 Hz | 167 115 | 0.8621 [0.8609; 0.8633] | < 0.0001 |
| 6000 Hz | 135 349 | 0.8408 [0.8393; 0.8424] | < 0.0001 |

**Table 1. Frequency-specific overview of Pearson correlation analyses investigating the potential relationship between AC hearing thresholds**

obtained through standard audiometry and AudiogramDirect. The table shows the number of data points included per analysis (n; column 2), Pearson correlation coefficients with 95% confidence intervals ([95% CI; column 3]), and corresponding p values (column 4).

| Frequency | n | Bias (dB HL) [95% CI] | LL (dB HL) [95% CI] | UL (dB HL) [95% CI] |
|---|---|---|---|---|
| 250 Hz | 159 748 | −3.2823 [−3.3339; −3.2308] | −23.8788 [−23.9669; −23.7907] | 17.3141 [17.2261; 17.4022] |
| 500 Hz | 167 460 | −2.4825 [−2.5284; −2.4365] | −21.2945 [−21.3730; −21.2159] | 16.3295 [16.2510; 16.4081] |
| 750 Hz | 80 588 | −0.1009 [−0.1597; −0.0422] | −16.7781 [−16.8785; −16.6777] | 16.5763 [16.4759; 16.6767] |
| 1000 Hz | 167 515 | −0.9928 [−1.0330; −0.9527] | −17.4339 [−17.5025; −17.3652] | 15.4482 [15.3796; 15.5169] |
| 1500 Hz | 93 555 | 0.4538 [0.4021; 0.5055] | −15.3580 [−15.4464; −15.2697] | 16.2657 [16.1773; 16.3540] |
| 2000 Hz | 167 433 | 1.9637 [1.9243; 2.0032] | −14.1788 [−14.2462; −14.1114] | 18.1062 [18.0388; 18.1736] |
| 3000 Hz | 136 337 | 1.3774 [1.3332; 1.4216] | −14.9329 [−15.0084; −14.8574] | 17.6877 [17.6122; 17.7632] |
| 4000 Hz | 167 116 | 1.4657 [1.4227; 1.5086] | −16.0902 [−16.1636; −16.0168] | 19.0215 [18.9482; 19.0949] |
| 6000 Hz | 135 505 | 4.2438 [4.1895; 4.2980] | −15.7174 [−15.8100; −15.6247] | 24.2049 [24.1122; 24.2975] |

**Table 2. Frequency-specific overview of Bland-Altman analyses to investigate potential differences between AC hearing thresholds obtained through standard audiometry and AudiogramDirect. The table shows the number of data points included per analysis (n; column 2), the bias between the two methods (column 3), as well as lower limits (LL; column 4) and upper limits (UP; column 5) of the lines of agreement. The last three parameters include 95% confidence intervals ([95% CI]).**

## Conclusion Appendix 2

For the individual frequencies, hearing thresholds obtained through standard audiometry and AudiogramDirect are highly correlated (all Pearson correlation coefficients > .80). The mean difference between the two audiometric methods is smaller than ±5 dB HL for all individual frequencies. The differences between both methods lies within an interval of ±24 dB HL for 95% of the data, across all frequencies. It is important to note that the results are highly frequency-dependent, with a trend towards lower variation for middle frequencies. The frequency-specific outcomes revealed in this study lie within a broader range (±24 dB HL) than the ±15 dB HL range reported on in Omisore (2011). This difference could potentially be explained by methodological differences. The Omisore (2011) study recruited participants with mild-to-moderate hearing losses, and an equally big

group of participants with moderately severe-to-profound hearing losses. Receiver and earpiece choice were controlled. The current retrospective analyses included log files linked to a wide range of hearing losses, receiver and earpiece choices – not necessarily balanced groups. The current dataset also consisted of data retroactively pulled from Phonak Target fitting software. Registration accuracy of data, such as when data is manually entered in the fitting software, could therefore not be controlled.

## Authors and Investigators

### Principal Investigator & Author

Charlotte Vercammen obtained a Master of Science in Speech Language Pathology and Audiology Sciences, and a PhD in Biomedical Sciences at the University of Leuven (KU Leuven) in Belgium. She joined Phonak HQ in 2018, where she currently works as Clinical Research Manager in the Global Audiology team. Her research interests lie in the fields of cognitive hearing science, healthy aging, and auditory neuroscience.