Canadian Audiologist
The Official Publication of the Canadian Academy of Audiology

A Case for Reducing Judgement Noise in Our Bone Conduction Fittings

Published September 7th, 2022

Alex Gascon, R.Aud, PhD Student
Bill Hodgetts, R.Aud, PhD

Alex Gascon, R.Aud, PhD student
Bill Hodgetts, R.Aud, PhD

Glossary:

BCD: Bone-Conduction hearing Device.

dB DL: Decibel Dial-Level. These refer to *in-situ* thresholds, using the BCD as a transducer and the BCD's programming software to generate the tones.

Noise: Unwanted variability in judgments or measurements (Kahneman et al., 2021).

Percutaneous BCD: BCD worn on a skin-penetrating abutment, screwed to a titanium implant osseointegrated in the skull.

Transcutaneous BCD: In this article, we use this term for BCD worn on a soft, elastic headband (e.g., Oticon Medical Ponto or Cochlear BAHA) or an adhesive (e.g., MED-EL Adhear), passive transcutaneous with a magnetic attachment (e.g., Cochlear BAHA Attract) and active transcutaneous (e.g., MED-EL BoneBridge). Even though these BCD use very different methods to couple to the skull, we chose to group them here as audiologists face similar challenges when it comes to the verification of each of these devices.

Intro

Several advancements in the verification of bone-conduction amplification were made during the past 10 years. BCD became adjustable with manufacturer programming software, similar to digital air-conduction hearing aids. Skull-simulators connected to hearing aid analyzers were developed for clinical use, allowing clinicians to easily measure the Output Force Level of percutaneous bone-conduction hearing devices (BCD) in their day-to-day practice. DSL-BCD v1.1, a prescriptive formula for adult percutaneous BCD users was released (Hodgetts & Scollie, 2017). This framework allowed verifying percutaneous BCD similarly to air-conduction hearing aids verified with a coupler in a hearing aid analyzer.

Given that these advancements are relatively new, the extent to which the field of bone-conduction

amplification adopted this verification framework is uncertain. A survey completed in 2015 revealed that 14% of a sample of audiologists working with pediatric BCD users had access to a skull-simulator (Gordey & Bagatto, 2020). While many recent studies used skull-simulator or other measurements of the BCD Output Force Levels (e.g., Ghoncheh et al., 2022; Lunner et al., 2016), several investigations solely rely on validation measurements of BCD. Typically, validation measurements consist of questionnaires, aided thresholds obtained with stimuli presented in the soundfield or aided speech perception scores. As such, the landscape of verification practices surrounding BCD seems heterogeneous, both in clinical and research settings.

In this article, we examine how the wide range of verification practices shapes the field of bone-conduction amplification. My PhD supervisor, Dr. Bill Hodgetts, recently discussed this topic at a conference, pointing out that current verification practices in bone-conduction amplification contain various amounts of biases and *noise*, an observation inspired by Daniel Kahneman's *Noise: A Flaw in Human Judgment* (2021). To further explore this topic, we present a brief overview of common verification and validation practices in bone-conduction amplification with the lens of *Noise* (2021). *Noise* here does not refer to background noise; it refers to unwanted variability in measurements or judgements. First, we conceptualize *noise* in the context of the verification of BCD. Then, we discuss how measurements and judgements in bone-conduction amplification relate to verification practices. Finally, we provide key recommendations to reduce *noise* in bone-conduction amplification verification.

Noise in bone conduction amplification

Noise is present when a measurement tool, individual or group of individuals give different answers to the same question, and when this variability in answers is unwanted (Kahneman et al., 2021). In audiology, norms, calibrations, and clinical practice guidelines all aim to reduce this unwanted variability. There is noise inherent to the measurements tools that we use; some of it is negligible (e.g., +/- 5 dB test-retest variability of pure-tone audiometry), and some is larger (e.g., variability in behavioural hearing thresholds of a pediatric patient not engaged in the task). In addition to this, noise is expected to be present in audiology because the practice of audiology requires making professional judgments. Kahneman et al. (2021) specify that a matter of professional judgment is when there is some uncertainty about the answer, and when qualified professionals might disagree on what is the best answer. Several of our decisions in audiology are of this nature, especially in complex cases or when clinical guidelines are lacking.

In bone-conduction amplification, an example of a matter of professional judgement could be: which BCD fitting is most likely to result in the optimal hearing outcomes? Depending on the audiologist, the verification or validation tools, and other variables deemed relevant by the clinician, the answer to this could vary. Figure 1 shows the default Output Force Level in dB FL (decibel Force Level) of 3 different BCDs programmed with the same *in-situ* bone-conduction hearing thresholds. The outputs of the BCDs were measured with a Verifit Skull-Simulator and Audioscan Verifit 2. It can be noted that the different prescriptive algorithms for the same hearing thresholds result in different default outputs in dB FL, especially in the low frequencies. Given this, it is unlikely that these three fittings are equal; at least one should provide, to some extent, better hearing outcomes than the others. I assume that most audiologists would choose b) as the optimal fitting, given its good match to the DSL-BCD v1.1. targets. However, there would likely be a lot more variability in the judgments of different audiologists if the DSL-BCD v1.1. prescriptive targets were not displayed. And even if some *noise* might remain when the targets are displayed (i.e., even with Figure 1 presented as is, some audiologists might disagree on the optimal

fitting), the audiologists would use the skull-simulator objective measurement as the starting point of their judgement processes. Because skull-simulator measurements are performed in conditions that limit the potential for *noise* (e.g., inside the test box of a hearing aid analyzer, using calibrated stimuli and consistent positioning of the BCD during the measurements, without patient test-retest variability), they are an appropriate starting point for the judgement process. This starting point is critical in reducing *noise*, no matter what the "best" fitting might be.

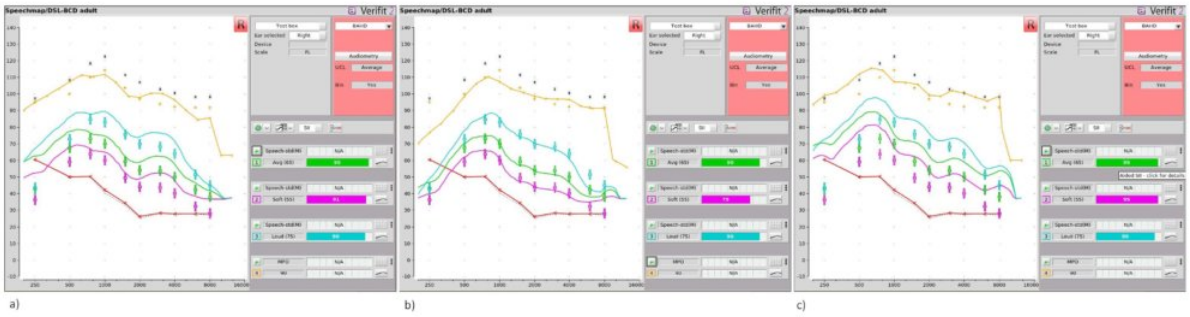


Figure 1. Output Force Level measurements for three different BCD measured with an Audioscan Skull-Simulator connected to a Verifit 2. The three devices were programmed at the default settings of their respective programming software, using the same in-situ thresholds (0 dB DL from 250 to 8000 Hz). The red line represents the corresponding dB FL (Force Level) for the in-situ thresholds (0 dB DL). Inputs were the "Standard" speech tests at 55, 65 and 75 dB SPL and a sweep at 90 dB SPL presented in the Verifit 2 test box.

Let's ask the same question, but this time, using different measurements to answer it. In this second example, the same three BCD are compared but without skull-simulator measurements. Instead, aided soundfield thresholds obtained with warble tones and aided speech reception threshold (SRT) are used to make the judgement about the optimal fitting. Figure 2 shows the thresholds obtained with one of my colleagues wearing the BCD on a soft headband and earplugs to simulate a bilateral conductive hearing loss. First, Figure 2 reveals that the aided thresholds are very similar for all devices (+/- 10 dB) and so are the SRTs. This is expected; this lack of sensitivity is a consequence of the *noise* and biases inherent to aided thresholds measurement in soundfield; as discussed by Hawkins (2004), soundfield aided thresholds testing is subject to calibration issues, test-retest reliability larger than the +/- 5 dB of diagnostic testing, and there is uncertainty about how hearing devices respond to pure-tone stimuli. Biases refer to systematic error in the measurement tool (e.g., using a soundfield speaker calibrated for 45° at 0° azimuth), while *noise* refers to unsystematic errors in the measurement (e.g., test-retest reliability, patient moving their head during testing, how the processor responds to pure-tones stimuli). Note that using recorded Ling Six (Glista et al., 2014) instead of pure tones would have reduced *noise* in this example.

Given that the three devices seem similar in Figure 2, should we conclude that none of them provide a better fit? It might be tempting to do so if we would not have the skull-simulator measurements in Figure 1. The key issue is that if the judgement process begins with *noisy* measurements, it is likely that the *noise* in the judgements from a group of different audiologists will be larger; in other words, it is reasonable to expect much larger variability in audiologists' judgements in the soundfield example vs. the skull-simulator example. This example suggests that *noise* present in the measurements may lead to more *noise* in the judgements of a group of audiologists.

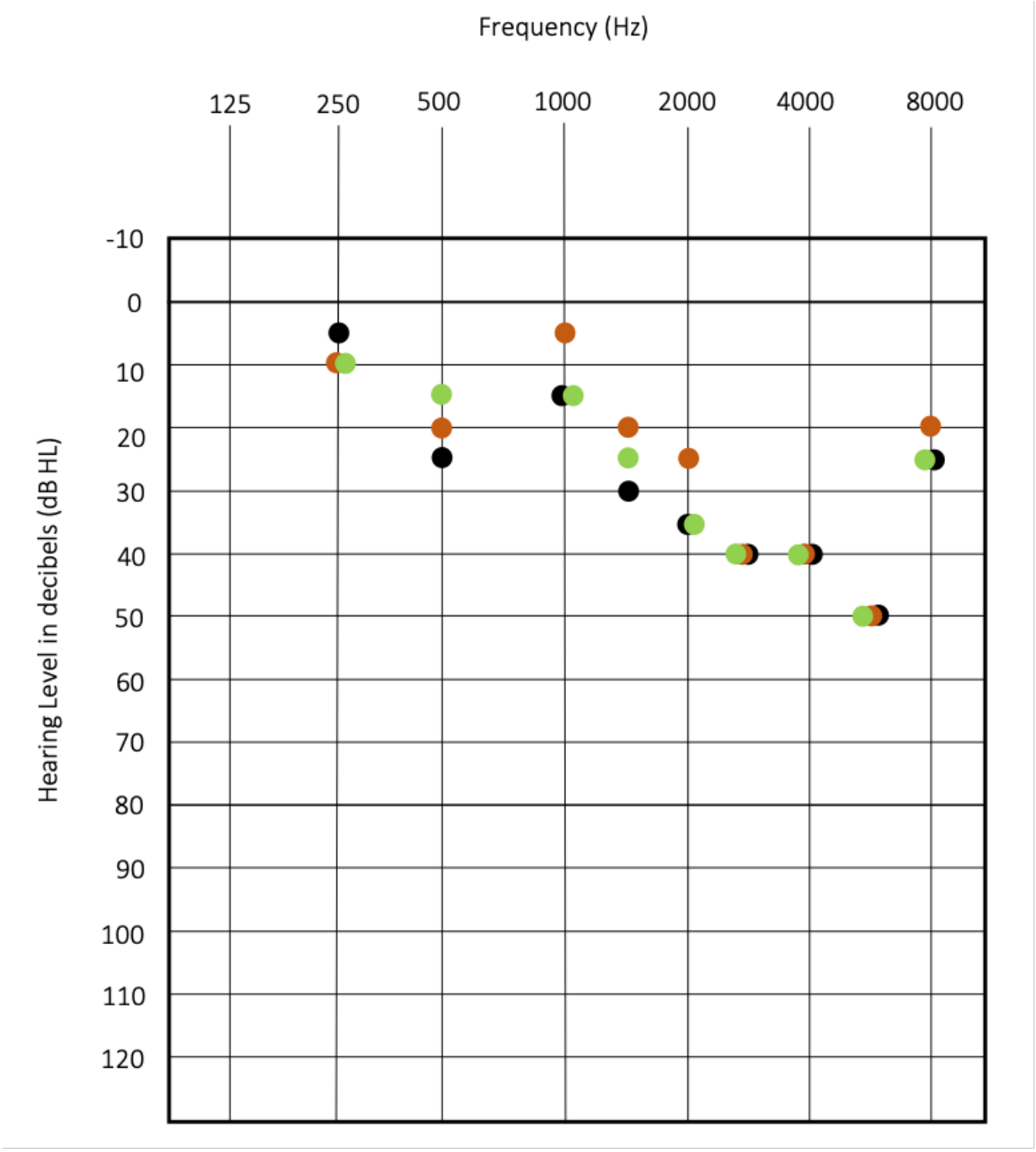


Figure 2. Aided threshold obtained in the soundfield using the three different BCD from Figure 1. My colleague wore the BCD on a soft headband and had foam earplugs in both ears. Warbles tones were presented on the right side, with a speaker at 45°.

= BCD a) SRT = 20 dB

= BCD b) SRT = 25 dB

= BCD c) SRT = 20 dB

Your verification tool should shape your judgement process - but is it?

These two examples reviewed known advantages of verifying percutaneous BCD with a skull-simulator connected to a hearing aid analyzer using concepts discussed in Kahneman's *Noise*

(2021). While variability in the judgments of audiologists might remain even when using a skull-simulator, the overall *noise* in judgements should be reduced compared to the soundfield measurements.

However, transcutaneous BCD cannot yet be verified with a hearing aid analyzer. A skin-microphone designed for the verification of transcutaneous BCD has been developed but is not yet available for clinical use (Hodgetts et al., 2018). Therefore, the scenario discussed in the second example is similar to the type of questions that arise when working with transcutaneous BCD. Many studies on transcutaneous BCD make claims using aided thresholds obtained in the soundfield. Similar to the second example, the challenges arise during the process of making judgements about the output of transcutaneous BCD based on these measurements; no matter how detailed the judgement process is or how many data points are used, the result of the judgement will remain only as good as the *noisy* aided thresholds obtained in the soundfield. The consequences of making a judgment based on these *noisy* measurements depend on the type of research or clinical question; if one is simply comparing aided vs. unaided thresholds, it is expected that the aided thresholds should be better than unaided and how much *noise* is in your measurements may not matter that much. However, if one is using aided thresholds obtained in the soundfield to compare different BCD or types of bone-conduction hearing implants to decide if one system is better than the other, the consequences of relying on *noisy* measurements to make a judgement can be more problematic.

The shape of bone-conduction verification to come

Currently, there is some level of *noise* in the verification practices in bone-conduction amplification verification. As a whole, this results in *system noise*; the unwanted variability in the verification of BCD is additive and, to some extent, impacts our field and its future directions. Some of this *noise* is due to limitations in verification tools (e.g., no objective verification for transcutaneous BCD), clinicians not using objective verification tools, and the nature of the professional judgements that constitute the practice of audiology. In this article, we proposed that objective verification tools should be a key part of the judgement processes in bone-conduction amplification verification and that using these tools can contribute to reducing *noise* in this field. The topics discussed in this article are not new, and the examples I presented could be analyzed in many other ways (e.g., sensitivity of the measurements, minimal clinically important difference), but the idea that there is *noise*, or unwanted variability in the verification practices of BCD has not yet been explored with this lens to my knowledge. *Noise* in medicine, Law and several types of professional endevours has been discussed by Kahneman et al. (2021) and comparing how the *noise* in bone-conduction amplification compares to other fields is beyond the scope of this blog article. Nonetheless, striving to reduce *system noise* in bone-conduction amplification verification is critical; you can contribute to this right now by (1) using a skull-simulator coupled to a hearing aid analyzer to verify percutaneous BCD and (2) interpret with caution aided-thresholds, whether in the clinic or when critically appraising research articles.

Key points:

- Choosing a verification tool that has an intrinsically low potential for error and biases will reduce noise more than getting skilled with a measurement technique that has a high risk of error.
- Using objective verification such as a skull-simulator connected to a hearing aid analyzer does not eliminate *noise*. But it makes it visible, and it is the first step to reducing it in our field. Skull-

simulator measurements should be standard of practice for the verification of percutaneous BCDs, in clinical and research contexts.
- The *noise* and biases inherent to certain measurements should impact how clinical and research findings in bone-conduction amplification are appraised. This is especially the case for aided soundfield thresholds.

References

Ghoncheh, M., Stenfelt, S., Maas, P., Salcher, R., Prenzler, N., Raufer, S., & Maier, H. (2022). Output performance of the novel active transcutaneous bone conduction implant Sentio at different stimulation sites. *Hearing Research*, *421*, 108369. https://doi.org.login.ezproxy.library.ualberta.ca/10.1016/j.heares.2021.108369

Glista, D., Scollie, S., Moodie, S., & Easwar, V. (2014). The ling 6(HL) test: Typical pediatric performance data and clinical use evaluation. *Journal of the American Academy of Audiology*, *25*(10), 1008–1021. https://doi.org/10.3766/jaaa.25.10.9

Gordey, D., & Bagatto, M. (2020). Fitting bone conduction hearing devices to children: audiological practices and challenges. *International Journal of Audiology*. https://doi.org/10.1080/14992027.2020.1814970

Hawkins, D. B. (2004). Limitations and Uses of the Aided Audiogram. In *Seminars in Hearing* (Vol. 25, Issue 1, pp. 51–62). https://doi.org/10.1055/s-2004-823047

Hodgetts, W. E., & Scollie, S. D. (2017). DSL prescriptive targets for bone conduction devices: adaptation and comparison to clinical fittings. *International Journal of Audiology*, *56*(7), 521–530. https://doi.org/10.1080/14992027.2017.1302605

Hodgetts, W. E., Scott, D., Maas, P., & Westover, L. (2018). Development of a novel bone conduction verification tool using a surface microphone: Validation with percutaneous bone conduction users. *Ear and Hearing*, *39*(6), 1157–1164. https://doi.org/10.1097/AUD.0000000000000572

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgment* (First edition). Little, Brown Spark.

Lunner, T., Rudner, M., Rosenbom, T., Ågren, J., & Ng, E. H. N. (2016). Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. *Ear and Hearing*, *37*(Wilson 2003), 145S-154S. https://doi.org/10.1097/AUD.0000000000000294