Audio Engineering Society

Conference Paper 7

Presented at the AES International Conference on Headphone Technology 2025 August 27-29, Espoo, Finland

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Subjective test method for evaluating ergonomics of in-ear and intra-concha headphones

Markus Vaalgamaa[1], and Laura Laaksonen[1]

[1]SoundlyFit

Correspondence should be addressed to Markus Vaalgamaa (markus@soundlyfit.io)

ABSTRACT

We present the motivation and testing methods for subjective evaluation of ergonomics of earbuds, including in-ear, intra-concha headphones, and hearing aids, by human testers. Earbuds have become increasingly popular for enjoying music, videos, and communication over recent years. However, many users and reviews report that headphones are not always comfortable, highlighting the need for effective testing methods. This paper introduces an efficient test method that requires approximately 15 minutes per device per tester, facilitating a fast and reliable way to compare different headphones. Our focus has been on practical and significant aspects of physical and functional comfort, addressing questions such as how easy it is to position the headphone in the ear, its comfort during wear, and whether it remains securely in place. The method outlined in this paper has been successfully implemented in several studies over the past years. Finally, we include some example results and lessons learned from our research to assist others in effectively applying this method.

1 Introduction

Wireless earbuds have become a popular way to enjoy music, watch movies, and make calls. According to Statista [1], headphone sales were expected to reach about 586 million units in the year 2024. A significant portion of these sales comprise true-wireless-stereo (TWS) headphones, which generated approximately 32 billion USD in revenue in 2023 and are projected to increase by about 15% annually [2]. One prominent type of TWS headphone is the in-ear earbud, which consists of a body and an ear tip as illustrated in Figure 1. The hard body primarily resides inside the concha of the ear, while the soft ear tip seals against the ear canal. Another popular related type is the intra-concha headphone, depicted in Figure 2, which does not feature a separate ear tip. One of the key requirements for successful headphones is their ergonomics and comfort.



Figure 1. In-ear headphone i.e. earbud and its parts.

Good ergonomics encompasses a variety of factors, as shown in Table 1. We selected these topics based on multiple iterations of ergonomic testing, aiming to identify the factors that users have described as most important for the ergonomics of earbuds and related devices. The priority of these factors varies based on

the user's device usage, the shape and condition of their ears, and personal preferences. For instance, listening to an audiobook in a quiet home from a stationary position results in different ergonomic priorities compared to listening to music in a noisy environment while commuting to work.



Figure 2. Example of an intra-concha headphone. Unlike in-ear models, these do not have ear tips.

Factor	Meaning
1. Comfort	Physical comfort resulting from the contact between the device and the ear.
2. Easiness to put to ear	How easy it is to insert the device into the ear?
3. Secure fit	Does the device stay securely in the ear?
4. Acoustic seal/audio quality	Does the device seal well against the ear, allowing for good and consistent audio performance?
5. Occlusion	The extent to which the user hears their own speech and body sounds louder than without the device.

Table 1. Key factors of ergonomics, related to physical perception

The importance of ergonomics is well illustrated by a large-scale online questionnaire conducted by Qualcomm in early 2023 [3]. This study spanned 7 countries and involved a total of 7,000 users. The study summarizes that "This year, comfort is the #1 purchase driver for true wireless earbuds, driven by the fact that consumers are now wearing their devices for longer periods of time than ever before."

The test method outlined in this paper stems from the ergonomics tests conducted in collaboration with Sense Labs at Force Technologies in Denmark starting at 2018. The method has undergone several iterations by the authors and their colleagues during the work at Huawei and in the current company.

In our testing, we have focused on metrics related to physical experience. For instance, the comfort metric concentrates on physical and functional comfort, while excluding other aspects such as acoustic or psychological comfort, as detailed in [4]. We selected these metrics because designers of headphone products can significantly influence physical and functional comfort during a design cycle of these devices.

Various subjective ergonomics tests for in-ear headphones have been documented in the literature, as referenced in works such as [5], [6], [7], and [8]. However, we found no paper that provides a comprehensive explanation, motivation, and practical learnings for subjective testing. Many of these papers primarily focus on ear anthropometry, which involves the dimensions and shapes of ears, and only briefly mention comfort tests to explore the relationship between comfort and ear measurements. While these studies affirm the effectiveness of their methods, many lack detailed definitions or contain multiple detailed metrics with unclear definitions, which can make practical testing inefficient.

The guideline we adhered to was to ensure that the ergonomic test method was both efficient and accurate, leading to the development of the subjective method presented in this paper. This method is relatively quick, taking approximately 15 minutes on average to test one device per user. It effectively captures the most common daily usage scenarios of headphones. Practically, this method allows us to test 3-4 devices per hour per test subject, enabling practical testing of about 20-30 people within a work week. The differences between short- and long-term comfort are further discussed in later sections of the paper.

The organization of the paper is as follows: Section 2 presents the subjective method questionnaire and provides motivation for the chosen metrics. Sections 3 and 4 discuss the methodologies for testing with real devices or 3D printed mockups and how audio quality can be effectively assessed using the latter. Section 5 describes the testing session in detail. Section 6 outlines the analysis methods used to interpret the data collected. Section 7 presents example results from ergonomic tests of 5 different earbuds. Section 8 elucidates the differences and similarities between short- and long-term subjective testing. Section 9 offers the conclusions derived from the research.

2 Subjective testing questions

When a test participant arrives for testing, he or she undergoes four phases during the test session: 1. Explanation of the testing procedure, 2. Pretesting of devices to select suitable eartips or other adjustable parts, 3. Actual ergonomic testing of one device at a time, and 4. Reevaluation of the relative ranking of devices. In this section, we will define the questions

used in phase 3, which generate a significant portion of the results. We employ a Mean Opinion Score (MOS) type scale ranging from 1 to 5 allowing participants to use 0.5 increments, which turns the scale into 9-point rating scale. This method has proven highly effective in practice, offering sufficient precision while remaining user-friendly. Previously, we have also successfully implemented a 0 to 10 step scale.

The first question pertains to the "Easiness to find a position," as shown in Table 2. This question is asked immediately after the participant has found a suitable position for the device in their ear. The experience of the user is sometimes reflected here, and first-time users may experience a considerable learning curve during testing. For example, the first device might be difficult to place in the ear, while the last device might find its position more easily. Therefore, it's important to review the responses to this question during the final, fourth phase of testing.

Easiness to find position	
Very easy	5
	4.5
Easy	4
	3.5
Fair	3
	2.5
A bit difficult	2
	1.5
Difficult	1

Table 2. Easiness to find a position

After the first question, we will ask the participant to take a roughly 5-minute walk and engage in practical testing with the device. Following this, we will inquire about perceived comfort, as detailed in Table 3. Typically, we repeat this question three times—first asking about overall comfort, and then separately about comfort in the ear canal and concha parts. The latter questions aim to determine which of the ear parts offers better or lesser comfort. We often also ask separately about which of the ear parts experienced lower comfort, as shown in Table 4. Additionally, the participant may be asked to identify the specific locations of the worst points of discomfort in the ear. These details can be extremely valuable in pinpointing the physical parts of the device that impact comfort.

Table 5 presents the question used to evaluate stability in the ear. We have learned that it is beneficial to properly introduce the scale beforehand and engage in a discussion about it. This is particularly important as the gradings between 1-3 can be difficult for participants to evaluate if they do not have prior experience with similar devices. While the scale can be taken literally to simplify testing, a more beneficial approach is to ask participants to use careful consideration when comparing stability between devices. We also suggest that participants re-evaluate the stabilities of the devices during the fourth part of the testing. This method ensures a more thorough and reflective assessment, potentially leading to more accurate and useful results.

Comfort scale	
Very comfortable	5
	4.5
Comfortable	4
	3.5
Fair	3
Relatively	2.5
uncomfortable	2
	1.5
Uncomfortable	1

Table 3. Comfort question, which can be asked about the overall experience, and separately for the ear canal/eartip and concha/body of the device.

Comfort between both ears	
Right much worse	3
Right worse	2
Right slightly worse	1
About equal	0
Left slightly worse	-1
Left worse	-2
Left much worse	-3

Table 4. Comfort between both ears

Secure fit/dropping of device	
Very firm and condident on stay	5
	4.5
Ok	4
Dit felt that it	3.5
could drop	3
did need to	2.5
touch it	2
It actually	1.5
dropped	1

Table 5. Stability in ear

Stability is followed by a question about the audibility of own speech and body sounds, as detailed in Table 6. Often, participants are not familiar with the concept of occlusion. To explain this, you can ask participants to talk while closing their ears with their hands. This demonstration helps illustrate the acoustic impact of closing the ear canals, which can often result in an overly loud and slightly bass-heavy amplification of

one's own voice and body sounds. The scale for this question is set up so that a high score means the participant feels like there is nothing blocking their ears. Note that occlusion, as such, is a question of audio domain; however, it has a clear link to the usage and positioning of earbuds in the ear, and thus occlusion impacts ergonomics.

The final question regarding overall ergonomics consolidates all prior experiences and inquiries. This question can be framed as: "How satisfied would you be with the overall ergonomics of this device if you had received or purchased it (regardless of its price)?"

We used the English version of the template with all participants and explained the terms to them as needed.

Own speech and body sounds	
No body sounds heard	5
Similar to without device	4.5
	4
Heard a bit more body sounds	3.5
	3
Heard clearly more body sounds	2.5
	2
Annoyingly loud body sounds	1.5
	1

Table 6. Occlusion - assessing the audibility of own speech and body sounds

Overall ergonomics	
Excellent	5
	4.5
Good	4
	3.5
Fair	3
	2.5
Poor	2
	1.5
Bad	1

Table 7. Overall ergonomics

3 Testing with real devices or 3D printed mockups

In the realm of ergonomics testing, a common challenge arises: the availability of reference devices already on the market versus the need to test prototype devices that are not yet functional. This leads to the question of how to effectively combine the testing of aesthetically polished, fully functional devices with those that are non-operational and may visually appear as "prototypes." Drawing from practical experience, it is crucial to approach this issue with careful consideration. Below, we will explore a few possible solutions for the dilemma.

If all devices being tested are finalized and ready, then conducting tests with the real devices is the most straightforward approach. This method has advantages, as the acoustics and ergonomics will be exactly like those of the final products. However, based on experience, this approach also presents several challenges. Some devices might look more appealing than others, or users might recognize certain brands. These factors can easily bias the opinions of users, similar to what has been observed in loudspeaker tests comparing audio quality with the speakers either visible or hidden. In tests where speakers are visible, typically a larger and more visually appealing speaker is favored, regardless of a different ranking achieved in a non-visible speaker test or what technically measured audio performance might indicate. This demonstrates the impact of visual bias in audio testing.

Another practical challenge arises in the need to connect all tested devices to a Bluetooth player for audio testing. Switching between different eartips or adjustable parts during the test can slow down the process. Alternatively, having as many devices as there are eartip sizes complicates the setup, as it requires an equivalent number of Bluetooth devices to act as the player for music and audio. Consequently, using real devices introduces significant practical challenges in the reproduction of audio.

A practical solution for the above-mentioned problems is high-quality 3D printed mockups, as depicted in Figure 3. Current 3D scanning and printing technology enables the creation of extremely realistic shapes, while for example using transparent materials it is difficult to recognize the brand of the device. The most significant advantage of this approach is the flexibility to seamlessly integrate fully functional devices with prototypes that are not yet operational. This method facilitates unbiased testing by focusing solely on the ergonomics and design aspects without brand and aesthetic influences.



Figure 3. 3D printed high quality mockups for ergonomics testing

One crucial aspect of realistic testing is adjusting the weight of the devices to match the real ones. Figure 4 illustrates a practical solution: incorporating heavier materials, such as a suitably sized metal nut, into the 3D print to increase weight, or alternatively, removing part of the 3D printed material from the inside to reduce weight.

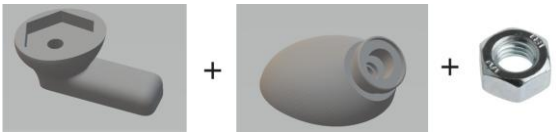


Figure 4. 3D printing in two parts allows for the adjustment of the headphone mockup's weight using a nut. The accompanying graphs also illustrate the vent designed for using a leak tester in eartip selection and/or controlling occlusion.

The use of carefully modeled 3D bodies enables the integration of real eartips or other adjustable parts that accompany the device. At a detailed level, various 3D printing techniques, along with varnishing and painting, can be applied to create a realistic fine structure for the body. This helps to achieve a realistic friction between the 3D printed parts and the ear, enhancing the authenticity of the testing experience.
The final option involves combining actual devices with 3D printed mockups, which incorporates both the benefits and challenges of each approach. However, based on practical experience, this is not recommended as it often leads to easy biases; the 3D prints can be quickly recognized, which may influence the results.
Finally, if mockups are used in testing, there remains a challenge regarding how to test audio quality. We will address this issue in the next section.

4 Fast audio quality testing during ergonomics testing with a leak tester

Audio quality is a crucial part of the earbud experience, and it impacts ergonomics since people try to optimize the placement of the earbud in the ear to maximize both audio quality and ergonomics at the same time. Testing of ergonomics without audio reproduction can be inaccurate since the depth and position of the earbud in the ear depend on the perceived audio quality. A better seal between the earbud and the ear typically leads to higher audio quality; however, this can also increase occlusion, which may, in turn, reduce the overall experience. Testing the above is more realistic with a real device; however it can also be estimated with mockups in the following way.

The hearing aid industry has developed a method for conducting rapid audio quality testing on ear-worn devices using a device known as a Leak Tester. This device is a small-scale air pump that connects to the mockup via an elastic tube, as illustrated in Figure 5. The mockup features an air vent that runs through the body to the ear canal side. The pump generates a small and safe increase in air pressure, then evaluates whether the pressure remains stable—indicating a good seal between the eartip/device and the ear—or if the pressure drops, which would suggest an air leak between the device and the ear.



Figure 5. A leak tester enables quick and repeatable measurements of the seal between the ear tip and ear canal; in this instance, the ear tip used scores 2 out of 10 on seal quality.

We have utilized the Bachmeier leak tester, although there are other devices available on the market. Analyzing the test results is straightforward. If the tester indicates a seal between 4-10, then the seal is considered very good, which means the audio reproduction quality with typical drivers in earbuds, and for instance, passive noise attenuation, would be excellent. Conversely, if the seal measures 0 or 1, then the audio quality would likely be compromised. If the seal scores between 2-3, the audio quality could be fair, but it might exhibit some sound coloration (such as a lack of low frequencies) or reduced noise attenuation performance.
One of the greatest benefits of the leak tester is its speed and repeatability. The tester delivers results in a second, and its good repeatability allows for quick evaluations of various device positions in the ear, such as normal, loose, and deeper placements. One of the drawbacks of this method is that it cannot be used for designs that do not create a seal. Additionally, the leak tester does not account for occlusion or the potential passive or active methods used in some earbuds to minimize occlusion.

To summarize Sections 3 and 4, if the focus of testing is primarily on ergonomics of earbuds rather than aesthetics or final audio quality, we recommend testing with high-quality prototypes. If 3D printed mockups are prepared as described in this section—making them visually equally appealing, using realistic weights, and enabling the use of a leak tester—this approach allows for extremely fast and repeatable testing of all key aspects of ergonomics. Additionally, the use of a leak tester facilitates seamless and rapid testing of fundamental audio quality of earbuds, even when real and prototype devices are used in combination.

5 Test session

Before the testing session starts, the devices need to prepared, the questionnaire finalized and participants with suitable age, gender and geographical distribution invited to the test. To minimize learning effects and biases, randomizing the order of the devices for each participant is recommended. Conducting the tests in a quiet room with neutral decor and lighting, situated in a peaceful area of the office, creates an ideal environment. Ensure there is a sufficiently large table to accommodate all devices, making them easily accessible to the participant.

Participants should walk for about 5-10 minutes with each device, so having easy access to a corridor, stairs, or outdoor space is preferable. Upon arrival, begin by explaining the testing procedure—providing an overview of the testing method, questions, and scales. Allow participants to ask questions in a relaxed and stress-free manner, which, although it might take 10 to 15 minutes, can significantly expedite the actual testing process.

Next, before proceeding to the actual ratings, select suitable sizes for any changeable parts, such as eartips, for each device. It is beneficial to have all devices and their respective changeable parts readily available, as shown in Figure 3, enabling participants to quickly assess different sizes and select their preference. This step may be followed by an audio quality or leak tester check, allowing the participant to evaluate the same device with another changeable part if necessary.

During this phase, carefully note the final tip size selected for each device, along with any results from the leak tester or audio quality tests. This part may take about 5 minutes per device, depending on the participant's experience and the variety of eartips or other parts to be tested.

Once all eartips have been selected, the actual ergonomic test can start, as depicted in Figure 6. The advantage of conducting tip testing beforehand is that participants are already familiar with the devices and their ergonomics, facilitating faster and smoother testing now that some experience has already been gained.



Figure 6. Practical ergonomics testing setup, with all devices and questionnaires readily accessible for a smooth and efficient testing process.

Devices are tested sequentially. The testing begins with the participant placing the device in their ear, followed by asking them about the "Easiness to find a position" as per Table 2. It is crucial to immediately gather and document any comments from the participants during this initial step. Subsequently, the participant is instructed to take a 5-minute walk, during which they are encouraged to perform various common head and body movements they deem relevant. Upon return, the remaining questions are asked, and additional comments from both the participant and the test conductor are recorded.

To conclude the testing session, it is beneficial to review the results once more, from the first device to the last. Participants are asked to quickly try the first and last devices again to verify if their experiences align with the recorded responses. If discrepancies are noted, adjustments to the responses are permissible to ensure they accurately reflect the comparative experiences across devices. Finally, determine which devices performed best and worst in each metric.

6 Analysis

Ergonomics tests are assessed similarly to any subjective evaluation of audio, video, or multimedia quality, involving typical statistical analysis. This analysis includes checking the dataset for missing or incorrectly filled data, calculating averages, counts, standard deviations, and confidence intervals. From our experience, a minimum of 30 participants is generally required to achieve statistically significant differences between devices, which is slightly higher than the often-used minimum of 20 participants in audio tests. The number of required participants can

vary depending on the experience level of the test participants and the subtlety of the differences being evaluated between devices.

In our observations, ergonomic tests tend to show larger differences between reported metrics compared to audio tests with the same number of participants. This can be attributed to the significant variation in ear sizes among individuals. A device that fits well naturally feels more comfortable; however, a mismatch in the size of the device and the person's ear can lead to discomfort and issues with the device staying in place, among other metrics. Therefore, ergonomics is largely influenced by the interaction between the sizes and shapes of the tested devices and the ears.

When compared to audio quality testing, such as listening through loudspeakers, the shape of the ear does not significantly impact the experience, and individuals with varying ear sizes tend to have more consistent experiences. Consequently, we observe a higher standard deviation in ergonomics than in audio quality tests, indicating that more participants are typically needed for ergonomic testing to achieve a similar level of statistical accuracy as in audio quality testing.

While this paper does not delve into factors that might impact or bias the testing results, it is important to note that skin sensitivity and cultural and language differences can affect how individuals express their ergonomic experiences. Similar physical contact between the ear and the device might be perceived and rated differently by different individuals in terms of comfort or other ergonomic metrics. An insightful reference on this subject is [9], which describes how comfort is experienced in nine different countries.

In addition to analyzing numerical metrics, it is beneficial to summarize textual comments, particularly to explain differences between devices and to investigate the reasons behind the numerical responses provided by participants.

7 Example results

We conducted a comprehensive ergonomics test involving 51 participants on five different earbuds from autumn 2023 to spring 2024, as detailed in Table 8. Many of these devices were purchased in spring 2023; thus, the results might be considered partially outdated by mid-2025. However, they still provide a valuable overview of ergonomic testing.

The participants hailed from 13 countries, with 45 from European backgrounds, 3 from Asia, 2 from Africa, and 1 from Central America. The average age of the participants was 36 years. We aimed for a broad age distribution, including 10 participants below 20 years old and 11 above 50 years old, with the youngest being 16 and the oldest 62 years. The gender distribution included 27 females and 24 males.

Short	Device
Ap	Airpods Pro 2
Ja	Jabra Elite 7 Pro
Jb	JBL Tour Pro
Se	Sennheiser Momentum True Wireless 3
Xx	Prototype device

Table 8. Devices evaluated in ergonomic testing

We present the results below as an example and offer guidance on key areas to focus on. It is important to note that the results and rankings of the devices might differ if similar tests were conducted in other parts of the world. However, we affirm that this test was carried out professionally and without bias, with no preference given to any of the devices.

Figure 7Figure 10 display the average results for the key metrics. In Figure 7, the top-performing devices achieve average scores slightly above 4, which corresponds to "easy" for ease of placement in the ear, "comfortable" for comfort, and "ok" for stability in the ear. For the "own body sound" metric, the highest-rated device scored a 3.5, which is described as "Did feel that it could drop." Additionally, the range — i.e., the difference between the highest and lowest average scores — is approximately 0.6 to 0.7 on these scales, representing 15-17.5% of the entire 1 to 5 scale used for these metrics.

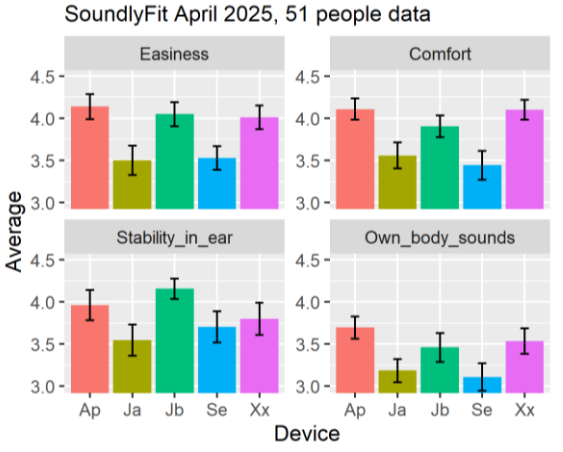


Figure 7. Average metrics with 95% confidence intervals are shown with black lines.

The 95% confidence intervals are around ±0.14 for comfort and increase to ±0.17 for stability per device, which can be considered quite suitable compared to

the differences observed in the averages between devices. However, when these confidence intervals are compared to some audio quality tests, they appear wide, despite the considerable participation of 51 individuals in the test. The statistical reason for this is the large standard deviations in each metric, which range from 0.75 to 0.8 (17-20% of the scale) per device across all these metrics. These deviations are larger than those typically seen in many audio quality tests.

The large standard deviations are attributable to the broad range of scores that participants assigned. Some participants were highly satisfied with many of the devices, often awarding scores between 4 and 5, likely because the devices were well-suited to the size of their ears. Conversely, there were participants for whom the devices did not fit well, possibly due to having ears that were smaller, larger, or differently shaped compared to the design of the devices. As a result, these participants assigned lower scores.

To illustrate the wide spread of scores, Figure 8 shows a very useful metric: the proportion of participants who gave a high score, in this case, 3 or higher. These participants can be considered "happy" or formally "satisfied" users (for each metric). For example, it is evident that for the Easiness and Comfort metrics, the best devices received scores of 3 or higher from over 90% of participants. While this is impressive, it also indicates that there are 5-10% of participants for whom these devices are not suitable.

For the other devices, the proportions of scores 3 and higher range from 70 to 80%, suggesting that 20-30% of participants encounter some challenges with these devices. This variability highlights the importance of considering diverse ear shapes and sizes in the design of such devices to accommodate a wider range of users.

The 5% to 30% rate of participants who assigned scores between 1 and 2.5 suggests that there is room for improvement in the design of these devices to better meet the needs of these individuals. A detailed study of the challenges faced by these participants could potentially lead to more effective solutions. The most obvious hypothesis is that their dissatisfaction may stem from unique aspects of their ear shapes that make the current designs less suitable. Addressing these challenges presents a significant opportunity. By refining the design to accommodate these diverse ear shapes, manufacturers can not only increase the proportion of "happy/satisfied" users but also broaden the market appeal of the devices. This improvement could lead to increased sales and positive feedback, particularly from those who typically avoid in-ear type devices due to fit issues.

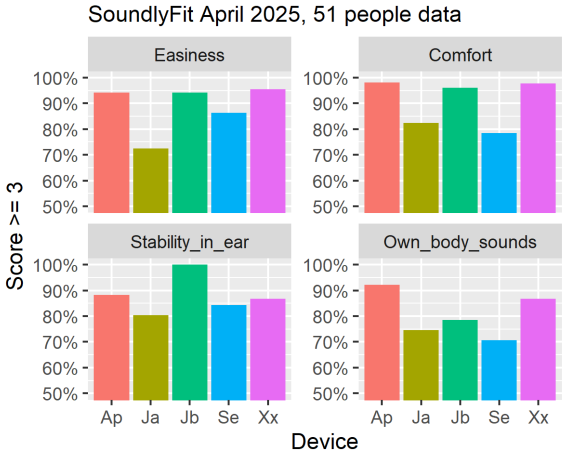


Figure 8. Proportion of participants having score 3 or higher for metrics shown in Figure 7

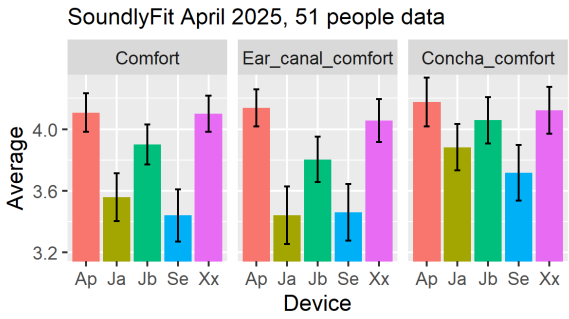


Figure 9. Comfort metrics: overall and separated for ear canal and concha. 95% confidence intervals shown with black lines

In Figure 9, we present the Comfort results, distinguishing between overall comfort and specific comfort in the ear canal and concha. The rationale for evaluating Comfort in three separate aspects is that the overall Comfort, depicted in the first plot of Figure 9, encapsulates both ear canal and concha comfort. The data reveals that ear canal comfort closely mirrors overall comfort, whereas concha comfort averages about 0.2 points higher. This significant finding suggests that ear canal comfort is the primary factor influencing overall comfort. In most cases, concha comfort is satisfactory, indicating that potential improvements could focus on the eartip or adjacent areas. This insight is invaluable for designers, mechanical engineers, or acoustic engineers working on these devices, guiding them to concentrate on the correct mechanical components.

Figure 10-Figure 11 further delve into the overall ergonomics of these devices. The highest-performing devices achieve a score just below 4, categorized as

"Good," while devices with lower scores reach just above 3, categorized as "Fair." The range of scores is 0.7, representing about 17% of the Ergonomics scale, which spans from 1 to 5. According to Figure 11, between 70% to 94% of participants are satisfied with these devices, indicating that there is potential for improvement for the remaining 6-30% of users.

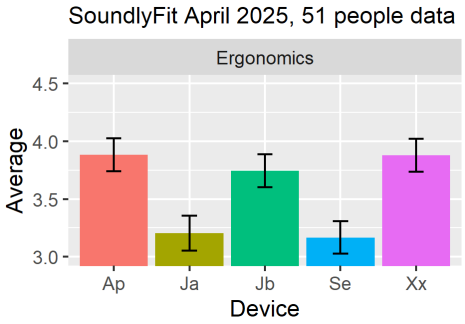


Figure 10. Overall ergonomics with 95% confidence intervals shown with black lines

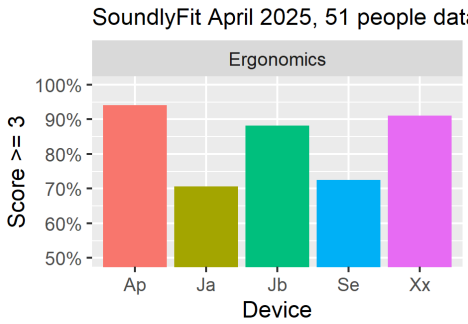


Figure 11. Proportion of participants having score 3 or higher for overall ergonomics

The similarity between the overall ergonomics and comfort scores suggests a correlation between these metrics. Participant comments further underscore that the overall ergonomics in this test were often influenced by the experience of comfort, highlighting the importance of this factor in the design and evaluation of earbud devices.

8 Short vs long-term testing and comfort

Within this paper, we propose and recommend short-term ergonomics testing, typically lasting 15 minutes per device. This raises a common question: Is this duration sufficient since people often use headphones for 30 minutes or more, and hearing protectors and hearing aids even longer?

The simple answer is that if there are problems in a short-term ergonomics test of a particular device, these issues will not disappear over a longer usage period. This situation is akin to wearing shoes that are too tight or an uncomfortable shirt—if they are not comfortable within a few minutes, testing them for an hour or more is pointless.

We have also systematically studied this difference. While most of our testing focuses on short-term tests, we have extended some studies to longer durations, such as 2 hours or an entire day. Due to the logistical and financial challenges, these extended tests typically involve a smaller group of participants, around ten. We cannot publish the results here due to confidentiality, but we can summarize the key findings.

We have observed the following trends between short-term (15 min) and long-term (1-2 hours or day) comfort testing based on our statistical analysis:

First, for about 50% of the users, the long-term comfort is lower than short-term comfort. This decline follows an exponential pattern versus time, meaning it drops a bit faster at the beginning, but the drop stabilizes over time. The level of drop depends on each metric and device but could be around 0.5 to 1.5 in comfort depending on the device when comparing 15 minutes to 2 hours usage. The main driver seems to be that if there were some minor problems in the short-term test, those problems just grow in long-term testing. In addition, in testing we have performed so far, we have seen that the ranking of the devices has been the same in short- to long-term testing.

Second, for about 50% of the test subjects, the long-term comfort remains about the same as the short-term comfort. There can be a small variation, say ±0.5 in comfort in the long term. For example, there may be 10% of all participants who feel the long-term comfort is slightly better since they adjusted to the new design or found a comfortable position for the headset in the ear. However, in a similar fashion, there can be another 10% of users for whom the comfort was slightly lower. To be more precise, we have not seen a statistically significant trend of people "learning" to tolerate a device with ergonomic problems. This is a statement that is often speculated in review meetings of ergonomic results of a short-term test. So, in the big picture, these two minor groups—one with slightly higher and another with slightly lower ergonomics—will equalize each other. Therefore, we can conclude that for about 50% of participants, the comfort was similar in short- and long-term tests. However, for another 50%, there will be a drop in comfort. The drop level depends on the device. But the key point is that if there were some problems in the short-term test, those will not vanish;

rather, the problems just amplify in long-term testing. The realization of this factor means that while the short-term test may not directly estimate accurate absolute results of the long-term ergonomic test, the short-term test can indicate the key problems of the design and allows much faster and effective testing of devices.

9 Conclusions

This paper presents an effective subjective test method to evaluate the ergonomics of earbuds, including in-ear, intra-concha headphones, and hearing aids. The testing method, which takes approximately 15 minutes per device per test participant, offers a fast and reliable way to compare headphones with one another. Over time, the method has evolved; we have optimized the number of metrics, omitting those that are less crucial. Our focus has been on practical and significant aspects of physical and functional comfort, with questions such as how easy it is to position the headphone in the ear, whether it is comfortable to wear, and if it stays in place. Conversely, we have omitted certain aspects such as acoustic or psychological comfort, as these can be assessed using other methods.

We have successfully applied this method in several studies over the past few years. We present example results from a large-scale ergonomic test involving 51 participants, which we hope will provide guidance on how to conduct practical testing and what kind of results can be expected. Additionally, we provide statistical data supporting the often-heard assertion that "one size does not fit all." This means that even if a device exhibits good ergonomics for some individuals, it does not guarantee universal appeal due to the significant variation in ear shapes and sizes. For this reason, it is advisable first to involve a sufficient number of people, preferably at least 30, in an ergonomic test. Secondly, it is important to analyze the percentage of test participants who evaluate the device as good versus bad. Particularly, a systematic study of the experiences of participants who have rated the device poorly in some metrics can be invaluable in efforts to improve the product. Such analysis can lead to insightful revelations about differences in ear shapes, age-related variations, gender differences, or even variations among different countries.

A free copy of the template is available https://www.soundlyfit.io/post/ergonomic-testing-template, or it can be requested from the authors.

10 Acknowledgements

The authors would like to thank the team at Sense Labs on Force Technologies Denmark during 2018 and 2019: Nick Zacharov, Tore Stegenborg-Andersen, Christer Volk, Thor Bundgaard Nielsen and Søren Vase Legarth. A substantial contribution to 3D scanning, ear shapes, came from Kuuloxi, Finland. A special thanks to Tommi Sipilä and Hanna Pirttinen. We also like to thank our ex-colleagues at Huawei Finland of improve the testing method: Tero Takala, Pasi Rautiainen and Markus Kultanen.

References

[1] Statista/Ahmed Sherif, "Unit shipments of headphones worldwide from 2013 to 2024 by Statista", (4th April 2024)

[2] Global market insights. "True Wireless Stereo Earbuds Market 2024-2032, (Nov 2024)

[3] Qualcomm company material. The State of Sound 2023 Report - A global analysis of audio consumer behaviors and desires. (July 2023)

[4] O. Doutres F. Sgard , J. Terroir , N.Perrin, C. Jolly , C. Gauvin , A. Negrini, "A critical review of the literature on comfort of hearing protection devices: Definition of comfort and identification of its main attributes for earplug types", International Journal of Audiology (July 2019)

[5] Haseung Song, Gee Won Shin, Yeree Yoon and Sangwoo Bahn, "The Effects of Ear Dimensions and Product Attributes on the Wearing Comfort of Wireless Earphones", Applied sciences 10 (24) (Dec 2020)

[6] S.-K. Stavrakos and S. Ahmed-Kristensen, "Assessment of anthropometric methods in headset design", International Design Conference, Dubrovnik, Croatia (May 2012)

[7] Yan Yan, Yonghong Liu & Haining Wang, "An earphone fit deviation analysis algorithm", Nature portfolio, scientific report 13:1084 (2023)

[8] F. Fu and Y. Luximon, "Comfort and fit perception based on 3D anthropometry for ear-related product design", Applied Ergonomics, Volume 100, (April 2022)

[9] P. Vink, S. Anjani, S. Udomboonyanupap, G. Torkashvand, T. Albin, S. Miguez, W. Li, C. Reuter & A. Vanacore "Differences and similarities in comfort and discomfort experience in nine countries in Asia, the Americas and Europe", Ergonomics, Volume 64, (2021) - Issue 5.