Using AirPods Pro 2 with iPhone and iPad to Help Protect, Assess, and Assist Hearing

October 2024

Contents

Overview

Apple is one of the world's most popular headphone brands, and in conjunction with iPhone and iPad, AirPods have further enhanced user communication and media experiences. The Health app on iPhone and iPad is a secure home for health data and trends (Apple Inc. 2023), including Headphone Audio Levels, Environmental Sound Levels (from the Noise app on Apple Watch), and Environmental Sound Reduction. AirPods Pro functions, such as Active Noise Cancellation (ANC), are designed to reduce unwanted environmental noises. Users can also use AirPods Pro 2 to minimize loud environmental noise exposure with Loud Sound Reduction.

To advance the understanding of factors that can affect hearing health, Apple launched the Apple Hearing Study in 2019 in collaboration with the University of Michigan School of Public Health and the World Health Organization (WHO) (Neitzel et al. 2022). Contributions from over 160,000 consented adult participants in the United States have corroborated the high prevalence of hearing loss and common obstacles to assessment and intervention (Apple Hearing Study 2022).

This paper outlines the development and validation of Apple's latest hearing health features—an over-the-counter (OTC) self-administered hearing test and an OTC self-fitting hearing aid. This paper also provides an overview of other important features, such as Hearing Protection, that can support a user's hearing health.

Introduction

Hearing loss is the third most common chronic health condition, affecting nearly a quarter of Americans (Haile, et al. 2023). According to the WHO, more than 1.5 billion people globally have hearing loss (WHO 2021). Age-related hearing loss is the most common form, and it progresses gradually over the years (Lin 2024). Exposure to loud noises can contribute to the progression (WHO 2021). Regardless of the cause, hearing loss is categorized from mild to profound (Olusanya et al. 2019). While the majority of those affected have mild hearing loss, those with little to no loss—or those with unilateral hearing loss with little to no loss in their better-hearing ear—may also experience hearing problems (Haile, et al. 2023; Olusanya et al. 2019).

Unrecognized and unaddressed hearing loss may affect a person's health and wellbeing in many ways, including by impairing daily communication and affecting social, physical, and cognitive functions (WHO 2021; Blazer et al. 2016). Hearing loss in adults is associated with an increased risk of cognitive decline (Loughrey et al. 2021), dementia (Livingston et al. 2024), falls (Jiam et al. 2016), higher medical costs (Reed et al. 2019), and other adverse health outcomes (WHO 2021; Blazer et al. 2016). These comorbidities may begin to develop many years before intervention for hearing loss is pursued given that declines in hearing occur slowly, are often not noticed by the affected individual, or may be attributed to other external causes, such as someone not speaking clearly or a room being too noisy (Lin 2024).

Ensuring access to clinically validated audiological testing and hearing technologies remains an ongoing challenge in many parts of the world (WHO 2021). The predominant model of hearing care around the world relies on clinic-based hearing evaluation and hearing aid fitting, which is time intensive, often involves significant out-of-pocket costs in many regions, and is limited by the finite availability of hearing healthcare professionals (WHO 2021). Strategies that could allow for earlier access to audiological testing and affordable hearing technologies could have a substantial impact on improving population health around the world.

Hearing Test Feature

Technical and Feature Description

The Apple Hearing Test Feature (HTF) is an over-the-counter air-conduction hearing assessment application designed to accurately assess hearing thresholds and generate a scientifically validated audiogram. Developed by Apple for use with AirPods Pro 2, it's compatible with iPhone (using iOS 18.1 or later) and iPad (using iPadOS 18.1 or later) and intended for individuals 18 years of age and older.

During the hearing test, tones are played across a range of frequencies from 250Hz to 8000Hz at various intensities. Hearing tests are typically conducted in sound booths to limit the impact of background noise on test results. To adapt an in-clinic test for self-administration, a noise check helps the user find a quiet environment for test-taking, and a fit check assesses and minimizes in-ear acoustic leak for optimal ANC performance. Technical safeguards, such as calibration while AirPods are in the case, ensure that the microphones and speakers aren't obstructed and are performing at a quality level acceptable to the test. After the test, information from all tones is used to produce an audiogram.

The HTF provides a convenient way for users to regularly monitor and evaluate their hearing health. HTF results may encourage users to make lifestyle changes and seek care, either through consultation with a hearing healthcare professional or, when appropriate, with hearing technologies like the Apple Hearing Aid Feature (HAF). The HAF, described later in this paper, is a software medical device that helps users compensate for perceived mild to moderate hearing loss. The HTF is available in select regions.

Preclinical Design and Algorithm Testing

Apple has developed a probabilistic modeling approach for pure tone audiometry testing. The algorithm consists of two main components: estimating the audiogram curve using Gaussian process classification and selecting the subsequent stimuli using Bayesian active learning.

A rule-based sampling process gathers initial data points at the start of the test. Tones are played at various frequencies, starting with 1000Hz, with incremental increases and decreases in tone decibels hearing level (dBHL) based on the user's response. This approach is an active learning algorithm and therefore doesn't rely on training data at any point. Audiograms from the National Health and Nutrition Examination Survey (NHANES) and the Apple Hearing Study were used to stress test and validate the algorithm.

Traditional tone audiometry methods test one prespecified frequency at a time at prespecified dBHL steps. In contrast, Apple's approach dynamically samples over the entire frequency and dBHL ranges, where any frequency and dBHL stimuli may be presented at any time. This results in continuous audiogram estimates at all frequencies and a more efficient test.

Clinical Validation

Hearing thresholds were measured by the HTF and compared with traditional reference audiometry in a clinical validation study that supported Food and Drug Administration (FDA) listing and authorization with other global medical device authorities. Each participant provided written informed consent to participate in the protocol approved by the institutional review board (IRB). At a screening visit after consent was provided and prior to randomization, an audiologist performed a visual and otoscopic exam of each participant's ear canals to confirm there were no occluding factors to air conduction (such as obstructive earwax or an eardrum anomaly), a tympanometry assessment of eardrum mobility, audiometric bone-conduction testing to verify the absence of conductive hearing loss, and audiometric air-conduction testing for recruitment binning by hearing loss category (to ensure that an adequate number of participants were recruited within each hearing loss category).

Eligible participants underwent repeat audiometric testing comparing audiologist-administered air-conduction testing, referred to as reference hearing test (REF), with the software-based Hearing Test Feature (HTF). Participants were randomized to receive either the HTF first followed by the REF (HTF_REF) or vice versa (REF_HTF) to control for the variable of listening fatigue or practice effects after the initial test. Figure 1 summarizes the participation activities.

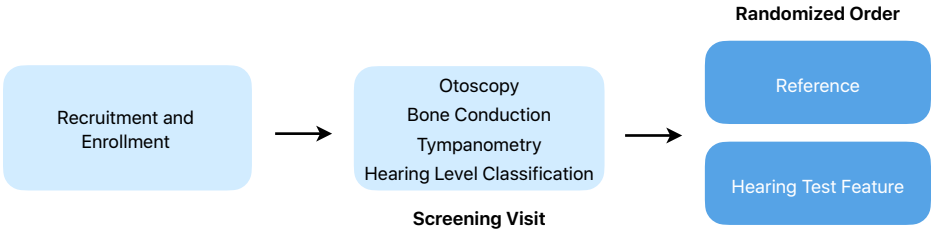


Figure 1. Hearing Test Feature Clinical Validation Study Summary

Upon completion of data collection, the HTF performance was assessed against traditional reference pure tone audiometry by comparison of the four-frequency pure tone average (4PTA: average of hearing thresholds in dBHL at 500Hz, 1000Hz, 2000Hz, and 4000Hz; primary endpoint) and eight-frequency pure tone average (8PTA: average of hearing thresholds at 250Hz, 500Hz, 1000Hz, 2000Hz, 3000Hz, 4000Hz, 6000Hz, and 8000Hz; secondary endpoint). Performance was further evaluated by appropriate classification of hearing loss according to the traditional WHO 0–4 hearing level grades (additional endpoint) (Olusanya et al. 2019) and by accuracy of the test according to the Consumer Technology Association (CTA) Standard (additional endpoint) (CTA 2023).

To assess the additional endpoints, the 4PTA in the better-hearing ear was classified as follows:

- Grade 0: ≤25 dBHL (no impairment)
- Grade 1: 26–40 dBHL (mild impairment)
- Grade 2: 41–60 dBHL (moderate impairment)
- Grade 3: 61–80 dBHL (severe impairment)
- Grade 4: ≥81 dBHL (profound impairment)

Participants

A minimum of 182 participants were targeted for enrollment to complete the clinical validation study for the primary endpoint analysis. The investigation aimed to recruit and enroll participants across a range of hearing abilities, with enrollment continuing until minimum targets for demographic factors and hearing level categories were achieved. All participants contributed audiometric data for both ears where possible, referred to as "all available ears." At the screening visit, hearing level classification binning was determined by 4PTA in the better-hearing ear, as per clinical convention. Hearing level classification for additional endpoint analysis was determined by 4PTA in the better-hearing ear by the REF.

NHANES data in adult Americans indicated a notably higher prevalence of hearing loss in individuals with male biological sex assigned at birth and those older than 60 (Goman and Lin 2016). Minimum participant recruitment targets were set for at least 33 percent of participants with female sex assigned at birth and at least 40 percent of participants younger than 60 to avoid overrepresentation of individuals with demographic characteristics strongly associated with hearing loss. Minimum enrollment requirements for hearing loss categories were 18 percent each of participants with no impairment, mild impairment, and moderate impairment and 5 percent of participants with severe to profound impairment.

A total of 202 participants were eligible, of whom 201 were randomized and completed the study. The binning characteristics of participants enrolled in the clinical validation study are presented in Table 1.

Characteristic	Validation
Number of participants	201
Age of participants (years)	58.0 (18.65)
Age < 60 years	90 (44.8%)
Age ≥ 60 years	111 (55.2%)
Sex (% female biological sex at birth)	105 (52.2%)
Hearing Level Categories	
≤25 dBHL (no impairment)	76 (37.8%)
26–40 dBHL (mild impairment)	44 (21.9%)
41–60 dBHL (moderate impairment)	60 (29.9%)
61–80 dBHL (severe impairment)	21 (10.4%)
81–85 dBHL (profound impairment)	0 (0.0%)

Data are presented as mean (SD) for continuous variables and % for categorical variables.

Table 1. Participant Characteristics in Hearing Test Feature Clinical Validation

Methods

The primary effectiveness endpoint was assessed in this study as follows:

- The median absolute difference (MAD) between REF and the HTF 4PTA using all available ears with valid data computed across the relevant frequencies

The secondary effectiveness endpoint assessed in this study was as follows:

- The MAD between REF and the HTF 8PTA using all available ears with valid data computed across the relevant frequencies

The structures of the primary and secondary endpoint hypotheses, respectively, were as follows:

H01: MAD between REF and the HTF 4PTA > performance goal (PG)
H11: MAD between REF and the HTF 4PTA ≤ PG

H02: MAD between REF and the HTF 8PTA > PG
H12: MAD between REF and the HTF 8PTA ≤ PG

PG represents prespecified performance goals of 10, 7.5, and 5 dBHL—known standard endpoints in evaluating audiometric tests—which were tested in a fixed sequence procedure to control the overall one-sided type I error at 0.025. For ease of interpretation, the results of only the most stringent performance goal criteria of 5 dBHL are presented.

The percentage of the HTF 4PTAs falling within ±5.0, ±7.5, and ±10 dBHL of the paired reference 4PTA and corresponding two-sided 95 percent bootstrap confidence intervals (CIs) were computed across all participants using all available ears.

Additionally, the REF and HTF results (dBHL) at the clinic visit are summarized descriptively (N, Mean, Std. Dev., Median, Min–Max) for each ear at each of the eight defined frequencies for 8PTA.

Results

All participants in a per-protocol (PP) analysis set (participants with available 4PTA or 8PTA from both the REF and the HTF) were used for the pivotal analyses to test the fixed sequence of primary and secondary endpoint hypotheses. No adverse events were observed related to the use of the HTF.

Primary Endpoint

The overall MAD of 4PTA between the REF and HTF was 1.81 dBHL, with a 95 percent confidence interval of (1.49, 2.30). The performance goal was met at all tested levels (10 dBHL, 7.5 dBHL, and 5 dBHL), with p-values < 0.0001 in all cases. Table 2 shows the observed performance in terms of the estimated MAD for 4PTA.

Characteristic	HTF_REF	REF_HTF	Overall
Number of paired ears	190	191	381
Number of participants	99	99	198
Estimated MAD (4PTA)*	1.56	1.98	1.81
95% bootstrap confidence interval	(1.13, 1.85)	(1.68, 2.54)	(1.49, 2.30)

* Both the adjusted and unadjusted p-values associated with the performance goal of 5 dBHL were <0.0001.

Table 2. Primary Endpoint Analysis: Median Absolute Deviation of 4PTA

Figure 2 is an X–Y plot of 4PTA values for all ears. There was a high degree of correlation between the REF and HTF 4PTA (r = 0.974).

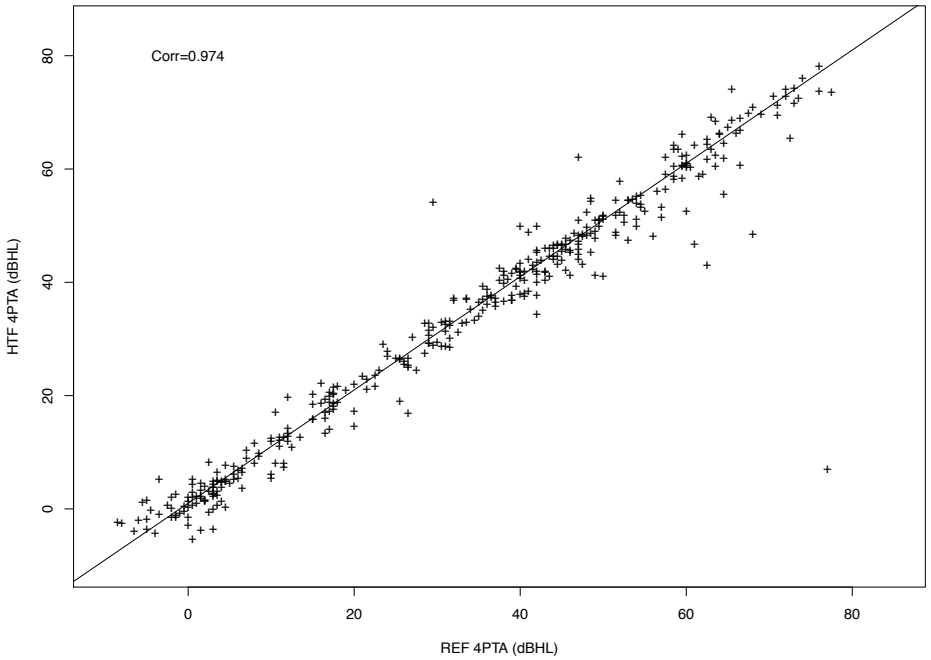


Figure 2. X–Y Plot of 4PTA Values for All Ears

The MAD of 4PTA between the REF and HTF using the better-hearing ear was 1.95 dBHL, with a 95 percent confidence interval of (1.78, 2.37). Thus, the primary endpoint performance goal was additionally met when using only the better-hearing ear for 4PTA.

Secondary Endpoint

The overall MAD of 8PTA between the REF and HTF was 1.75 dBHL, with a 95 percent confidence interval of (1.59, 1.92). The performance goal was met at all tested levels (10 dBHL, 7.5 dBHL, and 5 dBHL), with p-values < 0.0001 in all cases. Table 3 shows the observed performance in terms of the estimated MAD for 8PTA.

Characteristic	HTF_REF	REF_HTF	Overall
Number of paired ears	174	165	339
Number of participants	93	91	184
Estimated MAD (8PTA)*	1.69	1.79	1.75
95% bootstrap confidence interval	(1.39, 2.17)	(1.46, 2.08)	(1.59, 1.92)

* Both the adjusted and unadjusted p-values associated with the performance goal of 5 dBHL were <0.0001.

Table 3. Secondary Endpoint Analysis: Median Absolute Deviation of 8PTA

Figure 3 is an X–Y plot of 8PTA values for all ears. There was a high degree of correlation between the REF and HTF 8PTA (r = 0.979).

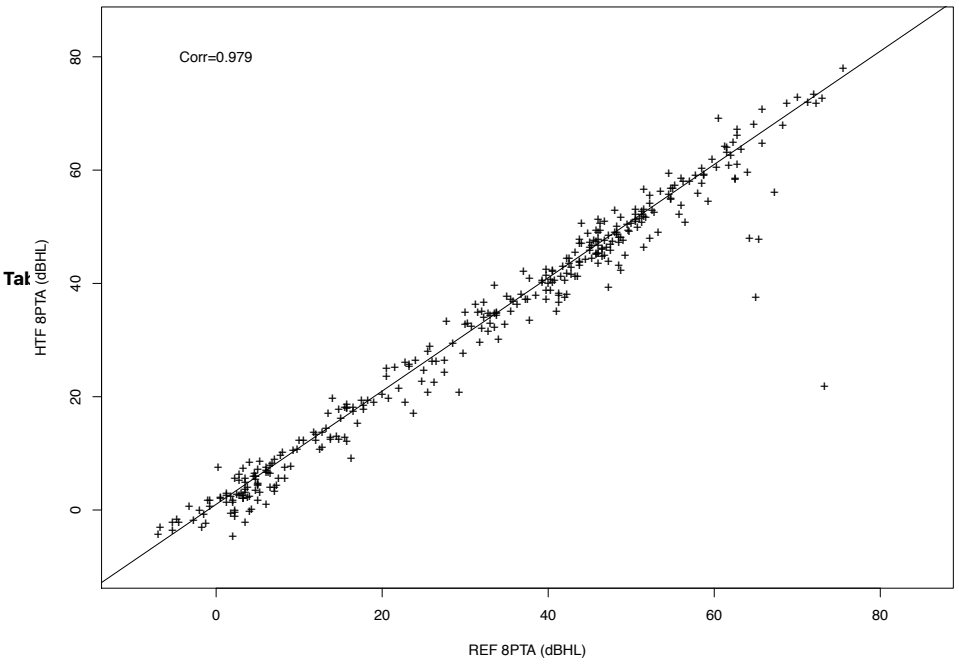


Figure 3. X–Y Plot of 8PTA Values for All Ears

A single participant in the REF_HTF group demonstrated erratic and inconsistent responses for the HTF, plotted at (65.0, 37.6) and (73.2, 21.9) for left and right ear tests, respectively. Outlier data was included in all efficacy analyses.

The MAD of 8PTA between the REF and HTF using the better-hearing ear was 1.95 dBHL, with a 95 percent confidence interval of (1.78, 2.37). Thus, as with the primary endpoint, the secondary endpoint performance goal was also met using the better-hearing ear for 8PTA.

Additional Endpoints

As presented in Table 4, the overall percent agreement between the REF and HTF in grading based on WHO classifications for 4PTA using the better-hearing ear was 86.4 percent (95 percent CI: 80.8 percent, 90.8 percent). All (100 percent) of the HTF 4PTA classifications of participants were within one category of the reference 4PTA classification.

	REF 4PTA Category					
HTF 4PTA Category	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Total
Grade 0	74	1	0	0	0	75
Grade 1	3	33	4	0	0	40
Grade 2	0	7	48	3	0	58
Grade 3	0	0	9	16	0	25
Grade 4	0	0	0	0	0	0
Total	77	41	61	19	0	198
Overall percent agreement = 171/198 = 86.4 percent (95 percent CI: 80.8 percent, 90.8 percent).						

Table 4. WHO Classification Accuracy of 4PTA Results (Using Better-Hearing Ear)

In evaluating HTF percentage accuracy, the percent of ears within ±5 dBHL of reference 4PTA overall was 88.5 percent (95 percent CI: 84.7 percent, 91.9 percent), the percent within ±7.5 dBHL of reference 4PTA overall was 95.0 percent (95 percent CI: 92.3 percent, 97.4 percent), and the percent within ±10 dBHL of reference 4PTA overall was 98.4 percent (95 percent CI: 96.6 percent, 99.7 percent).

Differences between hearing thresholds measured by the REF and HTF across all frequencies are depicted graphically in Figure 4. Overall, no substantive variability was seen in the median difference or the distribution of differences of hearing thresholds measured by the HTF compared with the REF across all measured frequencies.

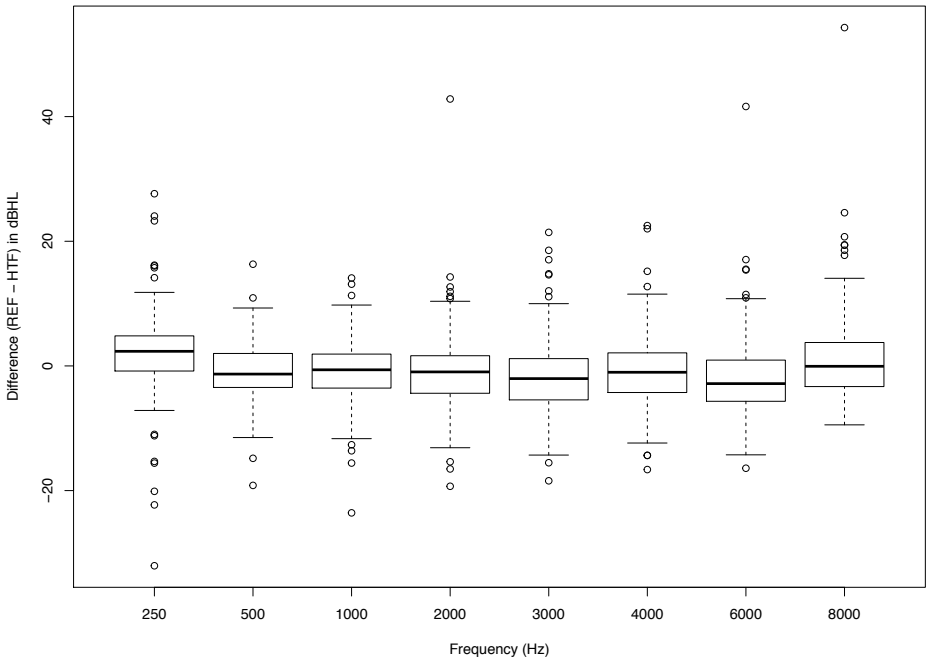


Figure 4. Difference in Frequency-Specific Hearing Thresholds Measured by the Hearing Test Feature versus Reference

Usability and Design

Hearing tests are often conducted in a clinic and administered by a healthcare professional. To adapt a traditional hearing test to a software-based, self-administered test, emphasis was placed on accounting for the user's environment, making the test intuitive, and providing easy-to-understand results to maintain test result quality and ensure that the user felt supported throughout the experience.

During onboarding, users are educated about the test and what to expect. Users watch visuals and animations to learn how to place AirPods correctly in their ears. The onboarding design was informed by usability studies that indicated the importance of providing test guidance and education to help instill confidence in the user's ability to correctly complete the test. To limit the impact of background noise, the software automatically checks that environmental noise is low enough to proceed and that the user has placed the AirPods in their ears with an appropriate fit.

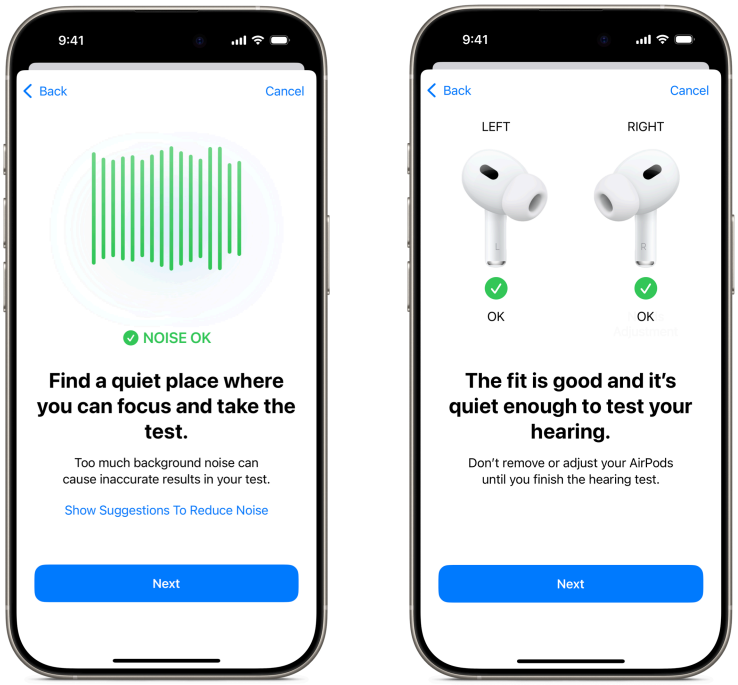


Figure 5. Hearing Test Feature Step for AirPods Pro Fit Check for Optimal Testing Conditions

For the test, users are instructed to tap the screen when they hear a tone. During the test, users see visuals that continuously move, indicating the test is in progress even if the user doesn't hear the tones being played. The visuals have been designed to avoid indicating correctness, ensuring that the results aren't biased and reassuring users that the test is proceeding as expected. In the case of loud environmental noise during the test, the test will automatically pause to limit the impact of that noise on test results, then resume if environmental noise returns to an acceptable level.

Figure 6. Hearing Test Feature User Interface Including Progress Marker

After a healthcare professional administers a hearing test, they typically explain the results to the patient and offer recommendations. Usability studies suggest that for a self-administered test, users need help interpreting test results and understanding what those results mean for them. Hence, after the test is complete, the user receives a comprehensible summary of their hearing test results, which includes a dBHL number indicating their hearing in each ear, a classification, and recommended next steps. As noted earlier, the number, often referred to as 4PTA, is an average of a user's hearing thresholds at 500Hz, 1000Hz, 2000Hz, and 4000Hz—the frequencies of sound most important for speech (ASA 2017). This number is translated to a WHO classification to help users interpret what their results mean for them.

In recognition of the importance of receiving adequate, timely care, users are provided with recommended next steps based on their result. Users with hearing loss can seamlessly set up the Apple Hearing Aid Feature, an OTC hearing aid designed for users with perceived mild to moderate hearing loss. Those with more severe degrees of hearing loss are recommended to seek care from a hearing healthcare professional

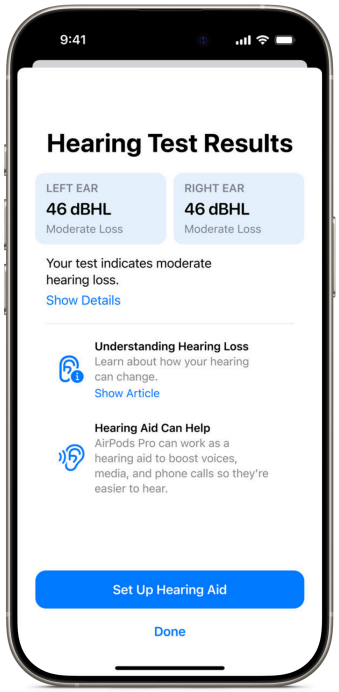


Figure 7. Hearing Test Feature Result Screen

In addition to the result summary, the user can access a detailed view of their audiogram, similar to one provided by a healthcare professional. The audiogram provides additional detail, including test results for eight frequencies and more data to help contextualize the results. Users can also see a longitudinal view of all test results in the Health app to observe changes in their hearing over time. They can export a PDF of their test results, audiogram, and longitudinal view to use in a conversation with their healthcare provider. This ability to export results is key because the HTF has been designed to augment existing healthcare offerings, empowering users to seek appropriate care earlier.

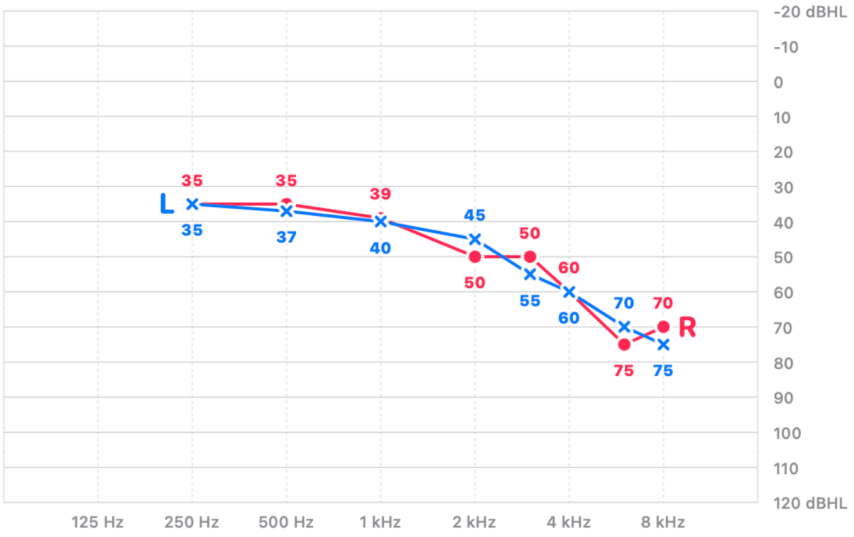
Jane Appleseed

Date of Birth: Feb 10, 1985 (Age: 39) Exported Sep 9, 2024

Hearing Test Result

× LEFT	● RIGHT
46 dBHL	46 dBHL
Moderate Loss	Moderate Loss

September 9, 2024 at 9:45 AM
Source: Apple Hearing Test



Congestion: NO
Recent Loud Noise Exposure: NO

Figure 8. Hearing Test Feature PDF to Help Facilitate Conversations with Healthcare Providers

Hearing Aid Feature

Technical and Feature Description

The Apple Hearing Aid Feature (HAF) is a self-fitting, over-the-counter air-conduction hearing aid software feature that enables users to compensate for perceived mild to moderate hearing loss. Developed by Apple for use with AirPods Pro 2, it's compatible with iPhone (using iOS 18.1 or later) and iPad (using iPadOS 18.1 or later) and intended for individuals 18 years of age and older. The Hearing Aid Feature is also supported on a compatible Mac with macOS Sequoia and later. The HAF user experience includes onboarding and informational content, facilitating the import of an audiogram from the Health app, and allowing users to customize their hearing aid settings. Users can also import an existing audiogram using their iPhone or iPad. Default settings are based on a user's audiogram from the HTF or a reliable third party, such as a hearing healthcare professional. A user can adjust the overall loudness with the Amplification control, the ratio between low- and high-pitched sounds using the Tone control, and any left-right asymmetries with the Balance control to suit their personal and situational preferences.

The hearing settings reside in the AirPods Pro firmware module, which interfaces with the microphone and audio processing to amplify sounds at specific frequencies and volumes based on the user's custom settings. The HAF, available in select regions, ensures that the maximum output signal doesn't exceed a 117 dB sound pressure level (SPL), consistent with regulations from the FDA.

Preclinical Design and Algorithm Testing

Conventional prescription hearing aids for adults commonly use the NAL-NL2 algorithm (Keidser et al. 2011), or other proprietary algorithms, for amplification. Apple's algorithm optimizes for speech intelligibility and sound quality. The HAF tuning is based on categorical loudness scaling to accommodate the reduced dynamic range for individuals with hearing loss. The goal is to improve listening comfort by providing nonlinear gain that makes speech audible to the person while maintaining comfort in listening environments with a mix of soft and loud sound events.

Consistent with typical hearing aid fitting formulas, the gain-based rules (algorithm) operate on an audiogram input. They use thresholds at 250Hz, 500Hz, 1000Hz, 2000Hz, 3000Hz, 4000Hz, 6000Hz, and 8000Hz. Thresholds at 3000Hz and 6000Hz are optional inputs, and if the input audiogram doesn't have these data points, the algorithm generates values for these frequencies by interpolating from adjacent audiogram points. Gain is provided with a bandwidth from less than 100Hz up to 10000Hz and at a resolution of 24 frequency subbands. Users can adjust settings to their preference using three controls: Amplification (±6 dB broadband gain, up to the device's maximum gain), Tone (±4.5 dB of frequency-specific gain), and Balance (6 dB interaural level difference).

Clinical Validation

Self-tuning using the HAF was tested in comparison to professional audiologist tuning in a clinical validation study that supported authorization from the FDA and other global medical device authorities. This was a separate study from the HTF validation study because the HAF is amenable to tuning from both HTF and third-party audiograms. The HAF study was designed to assess subjective improvement by participants with perceived mild to moderate hearing loss between a self-fitted HAF settings group (SF) and an industry-standard NAL-NL2 professionally fitted settings group (PF).

Each participant provided written informed consent to participate in the IRB-approved protocol. After consent was provided and participants were randomized, they were evaluated by an audiologist. This evaluation included a visual and otoscopic exam of each participant's ear canals to confirm there were no occluding factors to air conduction (such as obstructive earwax or an eardrum anomaly), a tympanometry assessment of eardrum mobility, bone-conduction testing to verify the absence of conductive hearing loss, and air-conduction testing for recruitment binning across the range of targeted hearing abilities. Participants with hearing thresholds > 60 dBHL in either ear at any frequency between 250Hz and 3000Hz or > 65 dBHL at 4000Hz were excluded due to gain requirements greater than could be delivered. Following enrollment, participants were randomized to the SF group or the PF group and proceeded through three clinic visits, as summarized in Figure 9.

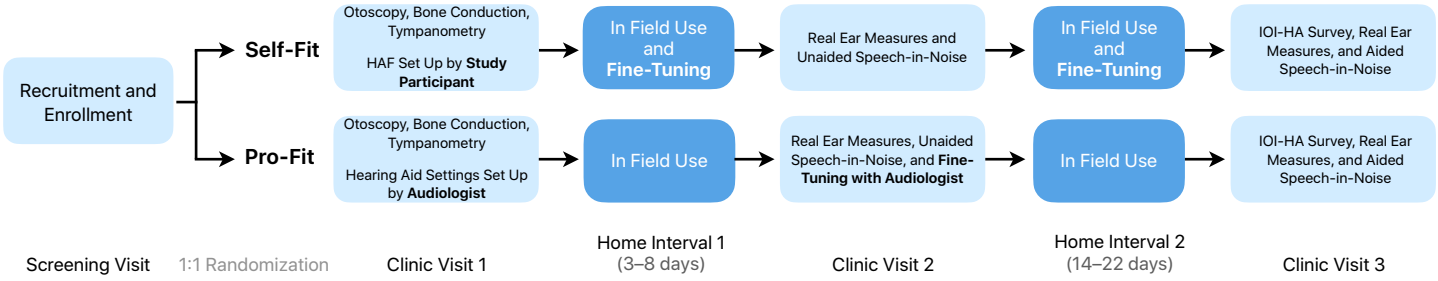


Figure 9. Hearing Aid Feature Clinical Validation Study Summary

During Clinic Visit 1, participants were provided with wireless AirPods Pro 2 and iPhone for use throughout the study. Participants in the SF group had software-assisted self-tuning of their AirPods and could further fine-tune gain to their preference using the Amplification, Tone, and Balance controls. Those in the PF group had settings selected by an audiologist who could further fine-tune as guided by participant interaction. Additionally, participants in the PF group could adjust the broadband gain to their preference during routine use (with a restriction of ±6 dB, similar to what's commonly provided by traditional hearing aids).

Participants left the clinic for Home Interval 1, wearing the AirPods in their routine environment for three to eight days. They then returned on Clinic Visit 2 to review fine-tuning settings controls independently (for SF) or to have an audiologist fine-tune the settings (for PF). This second clinic visit also included Real Ear Measures (REM) with AirPods in situ and unaided baseline Quick Speech-in-Noise (QuickSIN) testing. Participants left for Home Interval 2 for 14 to 22 days of continued AirPods use in their routine environment and were advised to adjust their software settings based on their preference. Upon returning for the final visit on Clinic Visit 3, participants completed the International Outcome Inventory for Hearing Aids (IOI-HA) survey (Cox et al. 2002), had final REM with AirPods in situ, and completed aided QuickSIN testing with their tuned AirPods in situ. Lastly, participants returned all devices and offboarded from the study.

Upon completion of data collection, HAF performance was assessed against the traditional professional fit of hearing aid settings by comparing IOI-HA survey scores for noninferiority (primary endpoint). Performance was further evaluated to verify no degradation of the speech signal relative to background noise with amplification and to assess whether gain across frequencies was comparable between groups (additional endpoints).

Participants

A minimum of 112 participants were targeted for enrollment to complete the clinical validation study for the primary endpoint analysis. The investigation aimed to recruit and enroll participants across a range of perceived mild to moderate hearing loss, with enrollment continuing until minimum targets for participants—according to demographic factors and hearing level categories—were achieved. Hearing level classification binning occurred as a function of hearing level by 4PTA in the better-hearing ear at the screening visit. Minimum participant recruitment targets were set for at least 33 percent of participants with female sex assigned at birth and at least 40 percent of participants younger than 60 to avoid overrepresentation of individuals with demographic characteristics strongly associated with hearing loss. Minimum enrollment requirements for hearing loss categories included 10 percent of participants with no impairment and perceived hearing loss as indicated by questionnaire (4PTA: 15–25 dBHL), 33 percent with mild impairment (4PTA: 26–40 dBHL), and 33 percent with moderate impairment (4PTA: 41–60 dBHL).

The binning characteristics of participants enrolled in the full analysis set (FAS) of the clinical validation study are presented in Table 5.

Characteristic	Validation
Number of participants	118
Age of participants (years)	59.2 (14.09)
Age < 60 years	50 (42.4%)
Age ≥ 60 years	68 (57.6%)
Sex (% female biological sex at birth)	69 (58.5%)
Hearing Level Categories	
15–25 dBHL (no impairment)*	27 (22.9%)
26–40 dBHL (mild impairment)	52 (44.1%)
41–60 dBHL (moderate impairment)	39 (33.1%)

* Perceived hearing loss by questionnaire.
Data is presented as mean (SD) for continuous variables and % for categorical variables.

Table 5. Participant Characteristics in Hearing Aid Feature Clinical Validation

Methods

The primary effectiveness endpoint of noninferiority between groups was assessed in this study as follows:

- IOI-HA scores: Average total IOI-HA scores were compared between Self-Fit Group and Pro-Fit Group

The IOI-HA is a seven-item validated questionnaire designed to be generally applicable in evaluating the effectiveness of hearing aid use based on user perceived benefit (Cox et al. 2002). The inventory was developed to facilitate cooperation among researchers and program evaluators in diverse settings. The survey's seven items assess daily hearing aid use, benefit, residual activity limitation, satisfaction, residual participant restrictions, impact (of hearing impairment) on others, and quality of life. Each item is scored on a Likert scale from 1 (poorest) to 5 (best) and summed to obtain a total outcome score ranging from 7 to 35, where higher scores are better.

Because the study hypothesis was a test of noninferiority, a 95 percent two-sided confidence interval based on the t-distribution for the treatment effect was computed. This interval was compared with a prespecified noninferiority margin of 3 points, which represents the minimum clinically important difference based on estimated variation from previously reported mean scores for the study instrument at the participant level (Sanchez et al. 2020; Saunders 2014). If the upper 95 percent confidence bound for the treatment effect (defined such that higher values favor the PF group) were less than or equal to 3, the null hypothesis would be rejected, and the SF treatment arm would be considered noninferior to the PF treatment arm. This is represented as follows:

$$H0: \text{Mean}(\text{IOI-HA}_{\text{Pro-Fit}}) - \text{Mean}(\text{IOI-HA}_{\text{Self-Fit}}) > 3$$
$$H1: \text{Mean}(\text{IOI-HA}_{\text{Pro-Fit}}) - \text{Mean}(\text{IOI-HA}_{\text{Self-Fit}}) \leq 3$$

$\text{Mean}(\text{IOI-HA}_{\text{Pro-Fit}})$ and $\text{Mean}(\text{IOI-HA}_{\text{Self-Fit}})$ are the average follow-up (Clinic Visit 3) IOI-HA scores for the PF and SF groups, respectively.

Additional tests were conducted to understand feature performance and confirm noninferiority to reference. QuickSIN measures the ability to hear speech in the presence of noise. Three QuickSIN test lists were administered at Clinic Visit 2 (unaided), and three different lists were administered at Clinic Visit 3 (aided). At each visit, the total test score was the sum of the correct responses across the three lists, and the average signal-to-noise ratio (SNR) loss was computed across the three lists.

REM testing measures real-time amplified sound in a participant's ear with a hearing aid in place. Available REM output values, measured in decibels of sound pressure level (dB SPL), were summarized descriptively for each ear at 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, and 8000Hz. These values were averaged across all frequencies separately for each of the two study arms (SF and PF) at each of the two clinic visits. The REM analyses were performed separately at 50, 65, and 80 dB SPL. The mean absolute difference between the SF and NAL-NL2 targets, and the PF and NAL-NL2 targets, was computed for each clinic visit at each frequency (500Hz, 1000Hz, 2000Hz, and 4000Hz) and averaged across all frequencies using all available ears. Two-sided 95 percent bootstrap confidence intervals for the mean absolute difference values were also reported.

Results

Consistent with the intention-to-treat principle, the primary endpoint analysis was performed using the FAS. A per-protocol analysis omitted one subject who didn't meet all the study protocol inclusion and exclusion criteria. All 118 randomized participants completed the study. No adverse events were observed related to the HAF.

Primary Endpoint

The mean IOI-HA total scores for the SF and PF treatment groups were 25.5 (SD = 3.03) and 26.6 (SD = 3.63), respectively, and are similar to results obtained with conventional prescription hearing aids fitted by a hearing healthcare professional (Sanchez et al. 2020; Saunders 2014). The mean difference (defined as PF – SF) between the two groups was 1.17 (SD = 3.343), with a 95 percent confidence interval of (–0.05, 2.39) for the mean difference and a p-value of 0.0036 for the noninferiority test. Thus, the null hypothesis was rejected, and the SF group was found to be noninferior to the PF group. Table 6 and Figure 10 show the observed performance in terms of the IOI-HA scores in both groups.

Characteristic	Self-Fit (N = 59)	Pro-Fit (N = 59)	Overall (N = 118)
IOI-HA total score			
N	59	59	118
Mean	25.5	26.6	26.1
Std. Dev.	3.03	3.63	3.38
Median	26.0	27.0	26.0
Min–Max	19–33	17–33	17–33
Mean difference (PF – SF)			1.17
Standard deviation of difference (pooled)			3.343
95% CI for mean difference (PF – SF)			(–0.05, 2.39)
p-value for the noninferiority test = 0.0036.			

Table 6. Primary Endpoint Analysis of IOI-HA Full Analysis Set

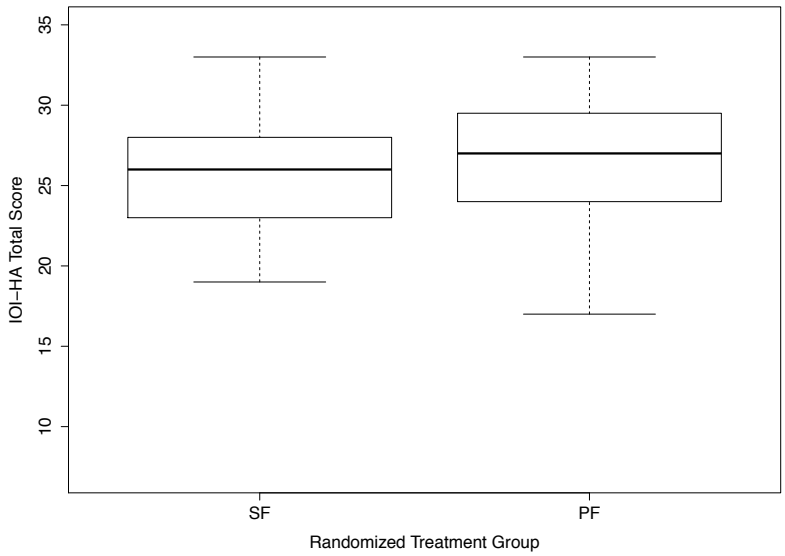


Figure 10. Boxplot of IOI-HA Scores by Treatment Group—Full Analysis Set

By-item comparison of participant responses to the IOI-HA survey identified no meaningful differences in average scores (≤0.5) between groups for each of the seven IOI-HA items. Within both groups, the average and median scores for each item were at least 3.0, with Question 6 (impact of hearing impairment on others) scoring the highest in both groups.

Additional Endpoints

Quick Speech-in-Noise (QuickSIN)

Everyday listening environments often contain both desired and undesired signals. The HAF gain formula is expected to provide similar intelligibility of speech when mixed with background noise compared with the PF formula. Analyses demonstrated that the difference in performance between the aided (Clinic Visit 3) and unaided (Clinic Visit 2) QuickSIN was comparable for both SF and PF groups, confirming that the HAF achieves speech intelligibility performance similar to professional-fit settings. Figure 11 shows a boxplot of aided QuickSIN average SNR loss by treatment group.

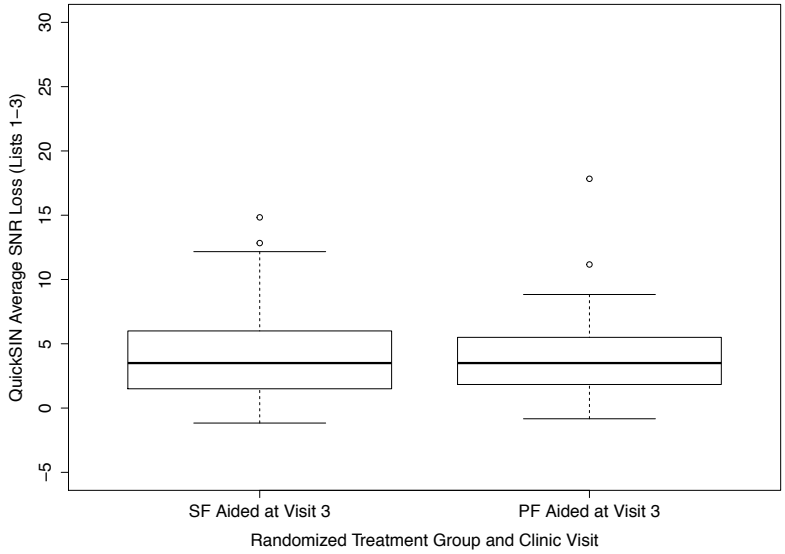


Figure 11. Boxplot of Average Aided QuickSIN Score by Treatment Group

Real Ear Measures (REM)

Some hearing aid users may initially prefer lower gain compared with professional-fit settings because those with long-standing hearing loss may need time to acclimatize to a comparably louder environment. The HAF gain formula is expected to be within a comparable range to PF, and REM analyses demonstrated no substantive differences in gain trends across the three measured conditions (50, 65, and 80 dB SPL) for both groups. Mean absolute differences were within 2 dB at the final visit, with SF group participants tending to prefer slightly less gain than PF group participants who were fitted according to NAL-NL2 targets. Boxplots for the mean difference between NAL-NL2 prescribed target gain and SF or PF gain settings, averaged across all frequencies at the final clinic visit, are shown in Figure 12.

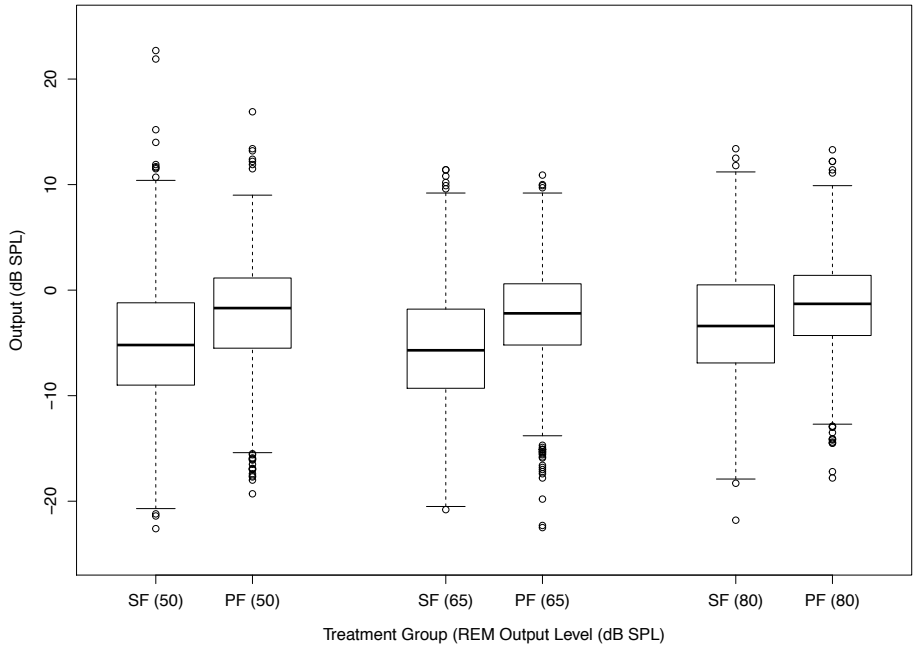


Figure 12. Mean Difference Between NAL-NL2 Target Gain and Treatment Group Gain Settings Averaged Across All Frequencies at Clinic Visit 3

In summary, the study results showed that the HAF achieved a noninferior level of performance compared with professional audiologist tuning, based on both the primary endpoint and additional analyses.

Usability and Design

Hearing aids for mild to moderate hearing loss have typically required fitting by a hearing healthcare professional. In adapting a traditional hearing aid to a self-fitting, over-the-counter hearing aid, emphasis was placed on making the setup and fine-tuning by the user easy and on par with adjustments completed by a professional to ensure that the user receives benefit. When setting up the HAF from the AirPods Pro settings, the user can easily select an audiogram. While users can seamlessly set up the HAF with the HTF results, they also have the option to use an audiogram provided by a hearing healthcare professional. Once an audiogram is selected, the software will automatically tune the hearing aid to meet the user's personalized needs. Users are then informed that they can adjust Amplification, Tone, and Balance in Control Center or in their AirPods Pro settings.

It can take time to get used to new sounds, so onboarding screens educate the user on acclimatizing to the HAF. The ability to adjust settings supports users through the acclimatization process. Users can adjust settings to fit their needs, especially as those needs change throughout the day and over time. These settings are applied to external signals such as voices and environmental sounds, as well as to media and phone calls across devices.

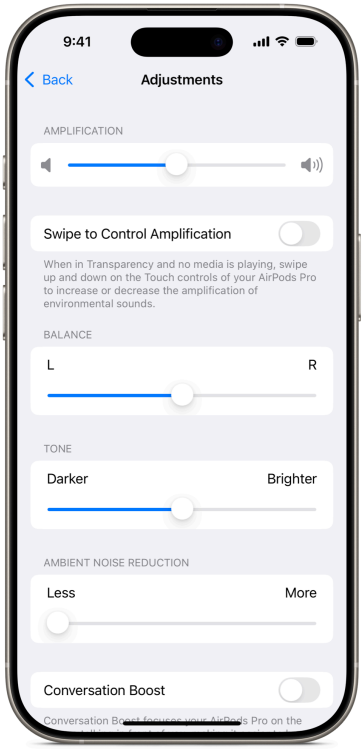


Figure 13. Hearing Aid Feature Settings Can Be Adjusted in Control Center or in AirPods Pro Settings

Discussion

Hearing loss is a common sensory deficit that, when left unaddressed, is associated with an increased risk for adverse health outcomes. Challenges in accessing clinically validated audiological testing, in part due to cost and limited availability of hearing healthcare professionals, cause individuals to be unaware of their hearing loss, further delaying access to care. Addressing hearing loss is known to improve quality of life and can help promote optimal social, cognitive, and physical functioning.

The HTF is a software-only, over-the-counter air-conduction hearing test that provides users with a self-administered assessment of hearing ability. It also provides an audiogram and hearing impairment classification based on traditional WHO grades. The HTF is designed to guide users through the assessment and communicate results in a way that benefits the user, including generating a PDF to share with healthcare providers. Audiogram results can be used in conjunction with the HAF to deliver personalized, scientifically validated hearing assistance.

The HAF is a self-fitting air-conduction hearing aid software for over-the-counter use, enabling users to benefit from hearing aid functionality to compensate for perceived mild to moderate hearing loss. While OTC hearing aids like the HAF can be set up without the need for a hearing healthcare professional, these devices don’t preclude invaluable services such as counseling, education, and treatment that a professional can provide. The availability of OTC hearing aids gives users agency over their hearing health, amplifying current care offerings as they become more engaged in their own care. The HAF can be used as part of a traditional treatment course with an audiologist or in conjunction with the Apple HTF.

The safety and effectiveness profiles of the HTF and HAF were established in separate formal clinical validation studies, which have supported device registration and authorization with global medical device authorities.

Additional Tools to Support Hearing Health

AirPods provide users with a suite of tools that can be used in conjunction with the HTF and HAF to support their hearing health.

Understanding speech in noisy environments is a common challenge for people, regardless of their hearing level. Conversation Boost focuses AirPods on the person talking in front of a user to make it easier to hear face-to-face conversations in a loud room. Users can easily turn Conversation Boost on and off from Control Center as listening conditions change, and it can also be used in conjunction with the HAF. Learn more about Conversation Boost here.

In addition to features notifying users of potentially harmful exposures, users have software-based tools to help actively protect their hearing. The Hearing Protection feature combines the latest active noise reduction algorithms on AirPods Pro 2 with functionality safeguards, useful guidance, and accredited measurement data on a compatible companion device. The feature helps protect hearing across listening modes and offers two types of protection. Leveraging Loud Sound Reduction in Transparency and Adaptive Audio modes, users receive level-dependent noise reduction with a fast reaction time, increasing sound attenuation as the environment gets louder. Transparency mode is particularly useful in loud, heterogenous environments, such as a concerts, where users may want to reduce loud noises and preserve softer sounds—all while maintaining sound quality. In Noise Cancellation mode, the noise reduction system consistently provides the maximum sound attenuation achievable, dependent on fit (up to 30 decibels NRS_A) (ASA 2007). The Hearing Protection feature is on by default and available in select regions. Learn more about the Hearing Protection feature here.

While noise can contribute to hearing loss progression, users are often unaware of their exposure to sounds that could affect their hearing. Real-time and historical feedback on headphone and environmental sound levels can help then understand this risk. Learn more about headphone audio exposure notifications here.

For users with an Apple Watch, the Noise app will periodically measure environmental sound levels throughout the day, allowing for both real-time visualization in the Noise app and historical visualization via Environmental Sound Levels in the Health app. Additionally, the Noise app will notify users if environmental sound levels could adversely affect their hearing. Learn more about the Noise app here.

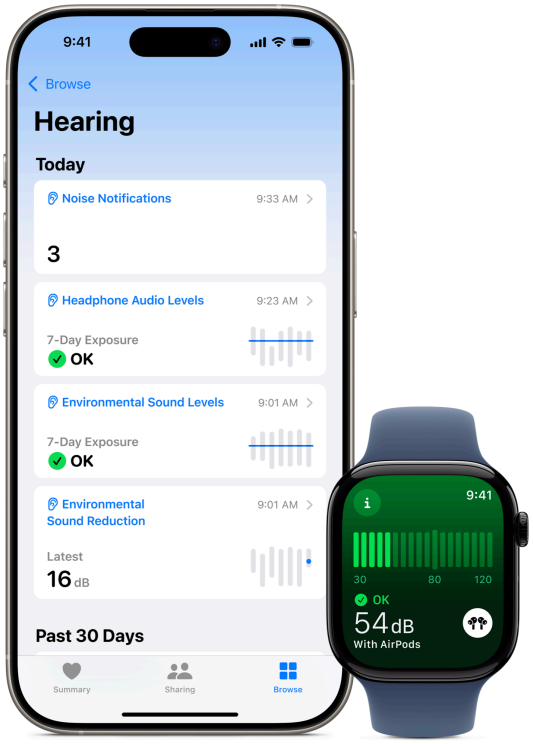


Figure 14. Hearing Categories in the Health App and Noise App on Apple Watch

To empower users with hearing loss to make the most of their Apple devices, Apple has also long supported many accessibility features, including:

- Mono Audio for users with a high degree of hearing loss in one ear
- Live Listen to stream sound from the microphone on iPhone to paired hearing aids, enhancing clarity in noisy environments
- Sound Recognition to identify environmental sounds in the immediate area
- Live Captions to generate real-time, on-device transcriptions of audio, including conversations and media

Learn more about Apple accessibility features for hearing here.

Together, these features account for the common sources of sound exposure that may exist in users' daily lives and allow them to make more informed decisions on noise exposure risks and behavior modification.

Conclusion

Hearing loss is a common and treatable sensory deficit associated with impaired quality of life and greater incident morbidity across a range of health conditions, including dementia, social isolation, and falls (WHO 2021). Compared with myopia, another treatable sensory impairment, hearing impairment often goes untreated or is treated after significant delay (Blazer et al. 2016). This concerning intervention gap is attributed in part to issues with access, stigma, and cost, despite increasing evidence of the benefits of hearing intervention on health (Blazer et al. 2016; Lin et al. 2023).

Apple products support a wide variety of features that empower users to better understand their health and provide actionable insights. The Hearing Test Feature on AirPods Pro 2 and compatible devices provides an air-conduction hearing test comparable to industry-standard clinic-based audiometry. The self-administered test offers a convenient and accessible way for users to identify potential hearing impairment with a high degree of accuracy. As described in this paper, the Hearing Test Feature can be helpful as part of assessing general hearing health.

The Hearing Aid Feature on AirPods Pro 2, also discussed here, equips users with an accessible tool to address their perceived mild to moderate hearing loss on their own, with outcomes equivalent to professional audiologist tuning. Users can customize settings on iPhone or iPad to output sound at specific frequencies and volumes through their AirPods and use this feature in conjunction with Conversation Boost to enhance speech understanding in noisy settings.

Together, the Hearing Test Feature, the Hearing Aid Feature, and other Apple hearing health features on AirPods Pro 2 and compatible iPhone and iPad devices offer a comprehensive range of tools that empower users to help protect, assess, and assist their hearing throughout their lives.

References

Apple Hearing Study. "Celebrating World Hearing Day 2022." March 3, 2022. sph-umich.shinyapps.io/world-hearing-day-2022.

Apple Inc. "Health Privacy Overview: Learn How the Health App and HealthKit Protect Your Privacy." May 2023, apple.com/ios/health/pdf/Health_Privacy_White_Paper_May_2023.pdf.

ASA (Acoustical Society of America). 2007. *Methods Of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors Are Worn*. ANSI/ASA S12.68-2007. New York, NY: ASA; reaffirmed 2020.

ASA (Acoustical Society of America). 2017. *Methods For Calculation of the Speech Intelligibility Index*. ANSI-ASA S3.5. New York, NY: ASA, approved June 6, 1997; reaffirmed July 12, 2017.

Blazer, Dan G., Sarah Domnitz, and Catharyn T. Liverman, eds. 2016. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*. National Academies of Sciences Engineering and Medicine. Washington, DC. National Academies Press. www.ncbi.nlm.nih.gov/books/NBK367633.

Cox, Robyn M., Dafydd Stephens, and Sophia E. Kramer. 2002. "Translations of the International Outcome inventory for Hearing Aids (IOI-HA)." *International Journal of Audiology* 41 (1): 3–26. pubmed.ncbi.nlm.nih.gov/12467365.

CTA (Consumer Technology Association). 2023. *Four Frequency Pure Tone Average Testing Methodology and Reporting Metrics for Consumer Facing Hearing Solutions*. ANSI/CTA-2118. Arlington, VA: CTA.

Goman, Adele M., and Frank R. Lin. 2016. "Prevalence of Hearing Loss by Severity in the United States." *American Journal of Public Health Research* 106 (10): 1820–1822. pubmed.ncbi.nlm.nih.gov/27552261.

Haile, Lydia M., Aislyn U. Orji, Kelly M. Reavis, Paul Svitil Briant, Katia M. Lucas, Fares Alahdab, et al. and GBD 2019 USA Hearing Loss Collaborators. 2024. "Hearing Loss Prevalence, Years Lived With Disability, and Hearing Aid Use in the United States From 1990 to 2019: Findings From the Global Burden of Disease Study." *Ear and Hearing* 45 (1): 257–267. pubmed.ncbi.nlm.nih.gov/37712826.

Jiam, Nicole Tin-Lok, Carol Li, and Yuri Agrawal. 2016. "Hearing Loss and Falls: A Systematic Review and Meta-Analysis." *Laryngoscope* 126 (11): 2587–2596. pubmed.ncbi.nlm.nih.gov/27010669.

Keidser, G., H. Dillon, M. Flax, T. Ching, and S. Brewer. 2011. "The NAL-NL2 Prescription Procedure." *Audiology Research* 1 (1): 88–90. pubmed.ncbi.nlm.nih.gov/26557309.

Lin, Frank R., James R. Pike, Marilyn S. Albert, Michelle Arnold, Sheila Burgard, Theresa Chisolm, et al. and ACHIEVE Collaborative Research Group. 2023. "Hearing Intervention Versus Health Education Control to Reduce Cognitive Decline in Older Adults with Hearing Loss in the USA (ACHIEVE): A Multicentre, Randomised Controlled Trial." *Lancet* 402 (10404): 786–797. pubmed.ncbi.nlm.nih.gov/37478886.

Lin, Frank R. 2024. "Age-Related Hearing Loss." *New England Journal of Medicine* 390 (16): 1505–1512. pubmed.ncbi.nlm.nih.gov/38657246.

Livingston, Gill, Jonathan Huntley, Kathy Y. Liu, Sergi G. Costafreda, Geir Selbæk, Suvarna Alladi, et al. 2024. "Dementia Prevention, Intervention, and Care: 2024 Report of the Lancet Standing Commission." *Lancet* 404 (10452): 572–628. pubmed.ncbi.nlm.nih.gov/39096926.

Loughrey, David G., Michelle E. Kelly, George A. Kelley, Sabina Brennan, and Brian A. Lawlor. 2018. "Association of Age-Related Hearing Loss with Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-Analysis." *JAMA Otolaryngology–Head & Neck Surgery* 144 (2): 115–126. pubmed.ncbi.nlm.nih.gov/29222544.

Neitzel, Richard L., Lauren Smith, Linyan Wang, Glenn Green, Jennifer Block, Michael Carchia, et al. 2022. "Toward a Better Understanding of Nonoccupational Sound Exposures and Associated Health Impacts: Methods of the Apple Hearing Study." *Journal of the Acoustical Society of America* 151 (3): 1476–1489. pubmed.ncbi.nlm.nih.gov/35364926.

Olusanya, Bolajoko O., Adrian C. Davis, and Howard J. Hoffman. 2019. "Hearing Loss Grades and the International Classification of Functioning, Disability and Health." *Bulletin of the World Health Organization* 97 (10): 725–728. pubmed.ncbi.nlm.nih.gov/31656340.

Reed, Nicholas S., Aylin Altan, Jennifer A. Deal, Charlotte Yeh, Alexander D. Kravetz, Margaret Wallhagen, et al. 2019. "Trends in Health Care Costs and Utilization Associated With Untreated Hearing Loss Over 10 Years." *JAMA Otolaryngology–Head & Neck Surgery* 145 (1): 27–34. pubmed.ncbi.nlm.nih.gov/30419131.

Sanchez, Victoria A., Michelle L. Arnold, Nicholas S. Reed, Preyanca H. Oree, Courtney R. Matthews, Ann Clock Eddins, et al. 2020. "The Hearing Intervention for the Aging and Cognitive Health Evaluation in Elders Randomized Control Trial: Manualization and Feasibility Study." *Ear and Hearing* 41 (5): 1333–1348. pubmed.ncbi.nlm.nih.gov/32251012.

Saunders, Gabrielle H. 2014. "The Performance-Perceptual Test as a Counseling Tool (PPT)." ClinicalTrials.gov. US Department of Veterans Affairs. Accessed September 12, 2024. clinicaltrials.gov/ct2/show/study/NCT00126113?term=international+outcomes+inventory+hearing+aids&draw=3&rank=12.

World Health Organization. 2021. "World Report on Hearing." www.who.int/publications/i/item/9789240020481.