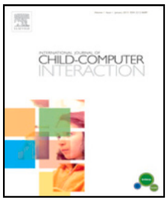
International Journal of Child-Computer Interaction 48 (2026) 100823

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci



Research paper

Pitch training for children with cochlear implants: Negotiating clinical validity through mobile game-based design

Andrea Gulli [a,*], Michele Geronazzo [a], Federico Fontana [b], Hanna Järveläinen [c], Enrico Muzzi [d], Roberta Rebesco [e], Eva Orzan [e]

[a] *Department of Engineering and Management, Università di Padova, Stradella S. Nicola, 3, Vicenza, 36100, VI, Italy*
[b] *Department of Mathematics, Computer Science and Physics, Università di Udine, Via delle Scienze 206, Udine, 33100, UD, Italy*
[c] *Institute for Computer Music and Sound Technology, Zürcher Hochschule der Künste, Pfingstweidstrasse 96, Zürich, 8031, Canton of Zürich, Switzerland*
[d] *Department of Otorhinolaryngology, Academic Hospital "Santa Maria della Misericordia", Azienda Sanitaria Universitaria Friuli Centrale, Piazzale Santa Maria della Misericordia 15, Udine, 33100, UD, Italy*
[e] *Otorhinolaryngology and Audiology, IRCCS materno infantile "Burlo Garofolo", Via dell'Istria 65, Trieste, 34137, TS, Italy*

ARTICLE INFO

Keywords:
Cochlear implant
Auditory training
Adaptive psychophysics
Personalization
Home-based rehabilitation
Single-sided deafness
Child-computer interaction
Participatory design

ABSTRACT

Auditory pitch discrimination plays an essential role in speech communication and music perception, yet reliable training and assessment remain challenging for pediatric cochlear implant users. Cochlear implants (CIs) are mainly intended to mitigate severe-to-profound hearing loss of cochlear origin. More than 1 million cochlear implants have been implanted worldwide. However, cochlear implant users still face pitch perception challenges due to an unmatched electrode-nerve interface and the lack of effective training protocols. We developed two mobile applications for children with single-sided deafness and one CI (SSD-CI): (i) a home-based pitch identification training game and (ii) a therapist-supervised adaptive pitch discrimination assessment tool. The system integrates adaptive loudness calibration (Gaussian Process Regression), a a two-interval two-alternative forced-choice (2I–2AFC) psychoacoustic procedure, and game-based interaction to balance clinical measurement validity with developmental fit and engagement. Pre-training thresholds were consistent with the literature, confirming assessment reliability. Following training, 72.2% of SSD-CI participants showed threshold reductions, with the median improving from 9.12 to 4.8 semitones. Questionnaire data indicated high engagement and manageable workload. Beyond empirical outcomes, this work demonstrates how clinically constrained psychophysical protocols can be translated into ecologically deployable child-centered interactive environments through a codesign process operating under non-negotiable clinical constraints. These findings support personalized and adaptive home rehabilitation models and suggest an approach for clinically grounded Child-Computer Interaction interventions.

1. Introduction

The ability to discriminate auditory pitch is crucial during verbal communication (Laures & Weismer, 1999) and music listening (Moore, 2013); prosody, especially in tonal languages (Deroche et al., 2019), and melody are essential for recognizing speakers and musical instruments, part of the human ability to identify and segregate sound sources (Wang et al., 2010). Reduced access to a sound's fundamental frequency, the parameter mainly responsible for pitch definition, seriously degrades speech-in-noise recognition (Binns & Culling, 2007).

Although cochlear implant (CI) users often have good word recognition ability in quiet conditions, they struggle to detect differences in pitch and timbre (Limb & Roy, 2014). As a result, CI users may hear sounds clearly but may still struggle to determine whether one sound is higher or lower in pitch than another (Zeng et al., 2014).

Studies testing CI listeners' pitch perception are characterized by large variability. Some CI users perform similarly to normal-hearing (NH) listeners, while others require much larger differences between sounds to detect a change in pitch (Brockmeier et al., 2011; Bruns et al., 2016; Kang et al., 2009; Luo et al., 2014). Moreover, perceptual maps vary over time (Saenz & Langers, 2014) due to neural plasticity linked to hearing loss (Koops et al., 2020).

---

[*] Corresponding author.

E-mail addresses: andrea.gulli@unipd.it (A. Gulli), michele.geronazzo@unipd.it (M. Geronazzo), federico.fontana@uniud.it (F. Fontana), hanna.jarvelainen@zhdk.ch (H. Järveläinen), enrico.muzzi@burlo.trieste.it (E. Muzzi), roberta.rebesco@burlo.trieste.it (R. Rebesco), eva.orzan@burlo.trieste.it (E. Orzan).

https://doi.org/10.1016/j.ijcci.2026.100823
Received 10 March 2025; Received in revised form 20 April 2026; Accepted 24 April 2026

These limitations motivate targeted auditory training approaches aimed at refining electric pitch representations and reducing functional variability across listeners.

Training approaches may be particularly effective in children, whose neuroplasticity allows perceptual abilities to adapt more readily over time. Within CI populations, the presented work focuses on patients rehabilitated for hearing loss with a CI in one ear and normal hearing in the contralateral ear, hereafter referred to as "single-sided deafness and CI" (SSD-CI). This population provides a unique opportunity: children can directly compare acoustic pitch (normal ear) and electric pitch (CI ear) during everyday listening. It is reported that SSD-CI listeners often make limited use of the implanted ear, given the dominance of the normal-hearing ear in everyday listening (Han et al., 2021).

Inspired by analogous findings in vision research (e.g., training both eyes in amblyopia (Murphy et al., 2015)), we explore whether coordinated use of both ears may help the normal-hearing ear guide the implanted ear. To our knowledge, this approach has not been explored in auditory training for SSD-CI children (Gordon et al., 2015).

This approach could be more beneficial for individuals whose neuroplasticity is more developed, i.e., mainly in younger population. Accordingly, we hypothesize that binaural training can enhance electric pitch discrimination in SSD-CI children, and that such improvement can be quantified through a clinically grounded and psychophysically valid assessment protocol embedded within a child-centered interactive system.

Clinical evaluation of pitch perception relies on psychophysical procedures whose duration, cognitive load, and rigid protocols often conflict with children's attention spans and interaction abilities. These procedures are often long and demanding, making them difficult to use with children. Rather than simplifying the protocol or merely adding gamified elements to existing tests, this work approaches system design as a negotiation between clinical constraints and child-centered interaction. Psychophysical validity remains fixed, while interaction design, task pacing, and deployment within the children's *ecology of use*, as discussed in CCI literature (Hourcade, 2022), are adapted to SSD-CI children's capabilities.

According to the CHI'25 systematic review of CCI for deaf and hard-of-hearing (DHH) children (42 papers spanning 2000–March 2024), the corpus is small; in the most recent five-year window (2020–March 2024) there are approximately 10 papers, with very little on auditory rehabilitation, and none targeting monaural pitch training for SSD-CI children (Zhao et al., 2025). Within this landscape and based on our literature search, we did not find prior CCI-codesigned studies targeting monaural pitch training in SSD-CI children; earlier works address CI children with bilateral loss or broader auditory games without a monaural pitch focus (Çetinkaya et al., 2024; Zhou et al., 2012) and do not report a documented participatory design process. This highlights a gap between clinically grounded auditory training and child-centered interactive design. Drawing from Hourcade's ten pillars of CCI (Hourcade, 2022) and from participatory design practices in disability contexts (Frauenberger et al., 2016; Spiel et al., 2017), the project used these frameworks to identify the main sites where negotiation with clinical constraints was possible. Interaction design, workload pacing, and real-world use conditions emerged as these sites, while psychophysical assessment requirements remained fixed.

The study implemented participation through repeated testing cycles in which children acted primarily as informants and testers (Druin, 2002), while clinicians defined the psychophysical requirements that could not be altered without compromising validity. Design decisions were systematically documented as responses to observed workload limits, attention patterns, and clinical measurement requirements. These documented statements provide a basis for illustrating how participants, procedures, and applications interacted in negotiating the final design (Spiel et al., 2017). Concrete examples of these pilot-session observations (e.g., hesitation during calibration or increasing response time across trials) correspond to specific design responses such as interface simplification or adaptive calibration procedures. This structured documentation responds to concerns in CCI literature about limited reporting of decision provenance and assessment rigor in work with disabled children (Zhao et al., 2025).

We therefore developed two cross-platform mobile applications for iOS and Android: a training game enabling home-based pitch identification practice, and a companion application for therapist-supervised pitch discrimination assessment. Both applications were codesigned with clinicians and therapists, tested with children during development sessions, and introduced to families through an onboarding process supporting home deployment.

Together, the two applications formed a unified system linking home-based training with therapist-supervised assessment.

Beyond interface design, the study also addresses methodological concerns in CCI research regarding the limited use of explicit learning assessment and the rarity of mixed-method pre–post evaluation designs (Baykal et al., 2023). By combining clinically interpretable psychophysical measures with participatory testing and real-world deployment, the study shows how child-centered interaction design can be integrated with rigorous evaluation of training outcomes.

Conventional pitch discrimination experiments can last more than one hour and require cognitively demanding loudness calibration procedures, creating a clear mismatch with children's attention span and interaction capacities (Brockmeier et al., 2011; Gockel et al., 2020). At the same time, the discrimination paradigm itself cannot be simplified without compromising psychophysical validity. Designing the system therefore required identifying where adaptation was possible within these constraints. To reduce workload and session duration, we adopted a recently proposed adaptive testing method to shorten assessment sessions (Gulli et al., 2023). In parallel, the use of ecologically valid auditory stimuli and gamified interaction structures supported engagement while preserving the integrity of the psychophysical task.

Personalization was intentionally limited. In assessment, adaptation operated through individualized loudness calibration and adaptive trial sequencing to ensure reliable and efficient measurement; in training, children could explore pitch relations through direct manipulation in line with CCI principles (Hourcade, 2022). In line with the view of Kucirkova (Kucirkova, 2019), our approach maintains individualized settings required for valid psychophysics while preserving opportunities for child-driven interaction, addressing concerns that fully automatic personalization may reduce agency.

The specific contributions of this work are as follows:

1. a validated test procedure for auditory pitch discrimination assessment based on a mobile game app;
2. an auditory pitch training program with proven effectiveness, based on a game app playable at home;
3. a method for accelerating auditory pitch discrimination assessments;
4. a clinically grounded participatory design lifecycle detailing how psychoacoustic validity, personalization, and child-centered interaction were negotiated within clinical constraints.

We found significant results for a group of SSD-CI children. These results provide ground for discussing the perception of electric and acoustic pitch based on auditory cues that NH individuals can interpret naturally and directly.

The remainder of the paper is structured as follows: Section 2 positions the work within auditory and CCI rehabilitation literature as well as participatory codesign models; Section 3 describes the participatory design lifecycle and the stakeholder interactions shaping the final system; Section 4 presents the stabilized experimental design; Sections 5 and 6 report procedure and data analysis; Section 7 presents results; Section 8 discusses methodological and CCI implications; and Section 9 addresses future work and concludes.

2. Related work

2.1. Auditory training and rehabilitation for hearing impairments

Auditory training has been widely proposed as a support for auditory rehabilitation in hearing-impaired populations. Reviews of the field report evidence that structured listening exercises can improve perceptual abilities, while also noting substantial variability in training paradigms, outcome measures, and reported benefits (Stropahl et al., 2019). Evidence focusing specifically on CI users indicates that training may improve auditory skills, but systematic reviews also highlight methodological heterogeneity and uneven experimental rigor across studies (Cambridge et al., 2022). Targeted interventions have demonstrated improvements in specific perceptual abilities relevant for cochlear implant listening, including pitch perception under competing place cues and melodic contour identification (Galvin et al., 2007; Vandali et al., 2015). At the same time, broader analyses of auditory-training research emphasize recurring limitations in experimental design, controls, and outcome definition, which complicate comparisons across interventions and the generalization of results (Henshaw & Ferguson, 2015).

Recent work further emphasizes the importance of ecological validity and personalization: training should target everyday listening conditions (Gaver, 1993; Keidser et al., 2020), account for individual CI characteristics (Shin & Park, 2023), and consider sample heterogeneity (Stropahl et al., 2019). These recommendations highlight a persistent tension between experimental control, individualized treatment, and ecological deployment.

When targeting children with CIs, additional design constraints arise. Hearing loss and delayed implantation may affect literacy, language, and behavior (Calderon & Greenberg, 2012), increasing susceptibility to frustration and disengagement during demanding perceptual tasks. Game-based and structured digital interventions have therefore been proposed to foster auditory memory, attention, and expression abilities (Xiang et al., 2024). Participatory design studies show that collaboration with deaf children may require approaches sensitive to visual attention patterns and communication practices during design activities (Potter et al., 2014). Complementary work proposing evidence-based guidelines for games for deaf children recommends minimizing unnecessary visual distractions and structuring interactions to support comprehension and task progression (Melonio & Gennari, 2013).

Moreover, home-based and caregiver-supported deployment is increasingly considered desirable to extend training beyond the clinic and into familiar environments (Kim et al., 2021). To our knowledge, prior work has not simultaneously integrated adaptive psychometric assessment, ecological home deployment, and child-centered participatory design within a validated protocol for CI-related auditory training.

2.2. Personalized and adaptive interactions for children

Personalized and adaptive interaction has been explored across multiple domains of child-focused technologies. In intervention science, adaptive treatments formalize personalization as rule-based adjustments of intervention intensity based on individual trajectories over time (Almirall & Chronis-Tuscano, 2016). Similarly, research on digital learning environments and serious games frequently employs mechanisms such as dynamic difficulty adjustment and individualized pacing to sustain engagement while learners progressively approach a target competence (Plass et al., 2015).

CCI research has applied related principles in the design of technologies for children with hearing impairments. Early systems such as ARTUR explored computer-assisted speech training through articulation feedback and corrective guidance (Bälter et al., 2005). Iversen and colleagues introduced embodied interaction to support language learning in children with CIs by coupling movement and auditory feedback within collaborative environments (Iversen et al., 2007). Other systems employed vibrotactile feedback to facilitate rhythm perception, demonstrating how multimodal interaction can expand access to auditory experiences for deaf children (Petry et al., 2018). More recent work, such as the BEARS virtual-reality training suite, further explores adaptive complexity, varied scenarios, and reward structures to personalize spatial-hearing training for young CI users (Vickers et al., 2021). Across these examples, child-centered adaptation is pursued mainly through multimodal feedback, embodied or game-like interaction, and engagement-oriented design, whereas direct evidence of training effectiveness is limited and uneven.

These interaction strategies are consistent with broader arguments in CCI that embodied and perceptually grounded interfaces can externalize cognitive relations and reduce the effort required to understand abstract tasks (Antle, 2013). In child-centered auditory systems, such externalization is particularly relevant because perceptual distinctions that are clinically meaningful may otherwise remain difficult to explain or sustain across repeated trials.

Spatial or physical representations can simplify task performance by making relations directly perceivable rather than relying only on verbal explanation or symbolic instruction. Prior research shows that pitch is associated with vertical spatial position from early childhood (Cuturi et al., 2019) and that such mappings influence perceptual judgments and response speed (Rusconi et al., 2006). Participatory design work also suggests that children’s engagement in repeated tasks is often sustained through recognizable feedback and progression cues embedded within interactive systems (Iversen et al., 2013).

Together, these observations suggest that interaction design may play a dual role in pediatric auditory systems: supporting engagement while making demanding perceptual or psychophysical tasks more executable for children.

While these systems demonstrate how adaptive and personalized interaction can enhance engagement and accessibility for DHH children, recent reviews highlight limitations in how their outcomes are evaluated. In particular, studies often rely on small exploratory deployments, provide limited reporting of experimental conditions, and rarely integrate mixed-method research designs or longitudinal outcome measures (Zhao et al., 2025). Similar methodological concerns have been noted in other CCI domains, where studies frequently lack explicit evaluation frameworks for measuring learning outcomes (Baykal et al., 2023). These observations suggest that adaptive interaction strategies should be complemented with evaluation protocols capable of producing interpretable measures of learning or perceptual change.

2.3. Participatory design models

Participatory design is a well-established approach within CCI, emphasizing the involvement of children in the development of technologies intended for them. Canonical formulations distinguish several roles that children may occupy in the design process — users, testers, informants, or design partners — each corresponding to different levels of influence over design decisions (Druin, 2002). Building on these distinctions, many CCI studies adopt iterative codesign practices in which prototypes are progressively refined through cycles of observation, testing, and redesign involving children and adult stakeholders. Such approaches have been widely applied in work with developmentally diverse children, including autism-focused participatory design frameworks such as IDEAS and related methods (Benton et al., 2012).

Participatory design in healthcare and rehabilitation often involves patients, caregivers, and clinicians in the development of therapeutic technologies and intervention protocols (Massey et al., 2024). In auditory rehabilitation, such approaches have been used to refine interaction scenarios and training environments for CI users (Vickers et al., 2021). However, participatory roles may need to be adapted to the constraints of the domain. For example, YoungDeafDesign shows that communication barriers and asymmetries between hearing designers and very young Deaf children can limit full design-partner participation, leading instead to contributions closer to tester or informant roles (Korte, 2022).

In the present context, these constraints become especially visible in psychophysical assessment. In auditory experiments, stimulus presentation rules, response paradigms, and calibration procedures must remain fixed to preserve measurement validity. As a result, children typically participate through the roles of testers or informants during iterative evaluations, while clinicians and researchers define the experimental parameters (Druin, 2002). In the CCI literature on empowerment, this form of participation is often described as functional empowerment, where children influence usability, engagement, and deployment conditions even when the core task structure remains unchanged (Mechelen et al., 2021).

Participatory processes nevertheless generate forms of mutual learning between designers and participants, revealing perceptual constraints, strategies, and contextual practices that inform subsequent design decisions (Qi & Yu, 2025). In rehabilitation contexts such as cochlear implantation, this exchange can be particularly salient: while children engage with training and assessment activities to develop listening abilities, researchers simultaneously gain insight into how these users perceive, interpret, and interact with auditory stimuli.

Recent analyses of participatory practices in HCI further reveal that the processes through which stakeholder observations translate into concrete design decisions are often insufficiently documented. Although iterative codesign is widely reported, the traceability between stakeholder input, emerging design tensions, and resulting implementation choices is rarely made explicit (Qi & Yu, 2025). This limitation makes it difficult to reconstruct how participatory contributions shape technological artifacts and methodological procedures.

Analytical perspectives such as Actor–Network Theory (ANT) (Latour, 2005) and Critical Discourse Analysis (CDA) (Maier & Jäger, 2016) offer conceptual tools for analyzing design processes as evolving networks of actors, artifacts, and constraints. These approaches have previously been applied in participatory design research with autistic children to analyze how interactions among participants, researchers, and technological artifacts shape design trajectories over time (Frauenberger et al., 2016; Spiel et al., 2017). Within such perspectives, methodological requirements can also be treated as actors within the design network, while the statements produced by different stakeholders provide insight into how design decisions emerge through negotiation.

The following section adopts this perspective to reconstruct the participatory lifecycle underlying the present work, describing how interactions among children, clinicians, researchers, and experimental constraints shaped the development of the resulting training and assessment ecosystem.

3. From clinical protocol to child-centered ecosystem: A participatory design lifecycle

This section documents the participatory design process through which clinical pitch assessment and an at-home training activity were developed into a coherent rehabilitation ecosystem. The process unfolded over nearly three years (November 2020–September 2023) and involved an interdisciplinary team that expanded as new needs emerged during development and piloting.

For clarity, we describe this lifecycle across three interacting spaces: the clinical protocol from which the initial requirements emerged, the technical implementation through which these requirements were translated into applications and procedures, and the domestic context in which training had to become feasible and meaningful for children and families.

We summarize the trajectory in three phases:

Phase I Destabilization of the initial clinical protocol through early observations and feasibility concerns.

Phase II Methodological realignment to ensure a reliable assessment procedure that remained practicable for children.

Phase III Ecological and regulatory integration into workflows compatible with home use and institutional requirements.

Following Druin's participation framework (Druin, 2002), children in this project primarily participated as users, testers, and informants within the assessment activities, where the core measurement structure could not be redesigned without compromising reliability. Their participation focused on interaction framing, pacing, comprehensibility, and workload. In parallel, children and families influenced the broader ecology of use (Hourcade, 2022), including the transition of training from clinic to home and the use of familiar mobile devices. Clinicians and therapists acted as key informants and codesign contributors by defining clinical requirements, ensuring feasibility with CI constraints, and supporting child-appropriate implementation.

The following subsections present the human actors involved, the three lifecycle phases, the socio-technical network through which the system evolved, the documented actor statements collected across the process, and the main negotiations through which the final training—-assessment ecosystem was stabilized.

3.1. Human actors and roles

The HCI research group comprised one principal researcher and two senior-level researchers. The principal researcher coordinated the project, implemented the mobile applications, and maintained documentation of design iterations and pilot observations. The senior HCI scientists provided methodological and technical supervision, including guidance on CCI design, digital signal processing, and experimental validation. As the need for a reliable psychophysical assessment procedure became evident, a psychoacoustics expert joined the collaboration to advise on discrimination testing procedures and stimulus design.

The clinical team included an otolaryngologist with more than ten years of experience working with CI children, an industrial engineer specializing in CI coding strategies, an audiometrist, three hospital speech therapists, and a hospital caregiver. Clinical stakeholders defined the research objectives, ensured compatibility with CI hardware and processing constraints, and evaluated comfort, workflow alignment, and age-appropriate interaction characteristics.

A rehabilitative speech therapist with more than ten years of experience with CI children (distinct from the hospital speech therapists) acted as a codesign contributor during later stages of the project. She facilitated family recruitment, organized collaborative sessions with children, and helped refine workload calibration and onboarding procedures. A GDPR specialist (Professor of Legal Informatics) served as ethics advisor, contributing to the design of consent documentation and compliant data management procedures.

Two participant groups contributed to the development process. An initial alpha-testing cohort included four SSD-CI children (1 female, 3 males; $\mu = 13.00 \pm 3.08$ years) and five NH children (5 males; $\mu = 12.80 \pm 3.92$ years). These sessions were conducted in the hospital environment, and the SSD-CI participants were recruited through clinicians among current patients. Behavioral observations from these sessions informed early interface adjustments and interaction refinements.

A second cohort participated in the final codesign refinement and experimental evaluation. This group included seven SSD-CI participants (2 females, 5 males; $\mu = 14.57 \pm 6.21$ years; CI experience 11 months–15 years) and eight NH participants (3 females, 5 males; $\mu = 13.88 \pm 3.59$ years). SSD-CI participants in this phase were recruited by the rehabilitative speech therapist among current and former patients, while NH participants were recruited through families known to the interdisciplinary team.

For transparency, we distinguish three categories of contributors that shape the digital environment (i.e., the mobile apps):

(i) codesign informants, including the HCI research group, the clinical staff, the rehabilitative speech therapist, and the GDPR advisor;
(ii) alpha testers who contributed to prototype refinement; and
(iii) participants in the final experimental evaluation. All participants completed the study. One SSD-CI participant and one NH participant were excluded from analysis due to procedural errors (incomplete training; reversed headphone channels).

3.2. Lifecycle phases across clinical, technical, and domestic contexts

The lifecycle unfolded through the three phases outlined above: clinical destabilization, methodological realignment, and ecological–regulatory integration. Early work focused on a training-oriented prototype, but discussions with clinical stakeholders progressively reframed the project around the need for a clinically valid assessment procedure. The overall timeline of these activities is summarized in the Gantt diagram provided in the Supplementary Materials.

Phase I — Destabilization of the clinical protocol

Early engagements took place in the hospital environment and focused on observing existing audiological practices and discussing their limitations for child-centered deployment. Direct observation of the clinical pitch detection procedure revealed both its diagnostic relevance and its practical limitations for children, including long session durations, loudness–pitch confounds, and limitations imposed by CI processing strategies. These observations highlighted a tension between clinically established procedures and children's interaction capacities, revealing the need to rethink how pitch perception could be assessed in a child-appropriate format.

This phase, comprising eight exploratory sessions including clinical observations, early prototype exploration, and multidisciplinary discussions with clinicians and engineers, exposed the need for a psychophysical assessment procedure that could produce reliable thresholds while remaining feasible for children. In particular, clinicians emphasized that pitch discrimination should not be confounded with loudness differences, since sound level can strongly influence pitch perception (Zheng & Brette, 2017). Ensuring perceptual loudness consistency therefore became a central requirement for the assessment task.

Phase II — Methodological realignment under psychoacoustic constraints

Phase II involved twelve iterative design and expert-consultation rounds through which the project shifted from a training-oriented prototype toward a system capable of both training and reliable pitch assessment. Consultation with a psychoacoustics expert supported the transition from detection-based testing to an adaptive two-interval two-alternative forced-choice (2I–2AFC) weighted up-down discrimination procedure (Kaernbach, 1991), suitable for threshold estimation.

Several methodological decisions emerged from these discussions. Loudness calibration was introduced to prevent stimulus comparisons from being interpreted as loudness judgments. Reaction-time logging and structured interaction records were added to support analysis of task performance. Three reference stimulus frequencies were also selected within the range effectively transmitted by CI processors and relevant for speech perception.

In parallel, the training application was developed through expert codesign involving HCI researchers and clinicians, informed by observations emerging from pilot testing of the assessment application. Its interaction design translated pitch relations into exploratory pitch identification activities supported by visual and gestural metaphors intended to make the task more interpretable for children.

By contrast, the assessment application was iteratively piloted with children during this phase in order to evaluate the feasibility of the psychophysical discrimination procedure. Its interaction design used a simplified spatial encoding and constrained response format intended to clarify the task without altering its psychophysical structure. Despite these differences, both applications relied on the same acoustic stimulus generation, progress indicators, and data logging procedures used to monitor task performance.

Phase III — Ecological and regulatory integration

Once the assessment procedure and interaction metaphors reached a stable configuration, iterative testing with children focused on feasibility and usability. Twelve pilot sessions were conducted during this phase:

- two home-based sessions with children
- four hospital-based sessions with children completing the full assessment protocol
- six hospital sessions involving clinicians and the rehabilitative therapist who evaluated feasibility, workload, and procedure configuration prior to the final experimental deployment

The sessions followed a semi-structured format comprising introduction, demonstration, full assessment, fatigue monitoring, and post-session discussion. Fieldnotes, interaction logs, and selected video recordings documented children's behavior and interaction difficulties. Children were also allowed to pause, repeat, or discontinue parts of the procedure if fatigue emerged; in one SSD-CI case, the session was interrupted after loudness balancing due to discomfort.

These sessions revealed several adjustments required for the assessment procedure. Interface layouts were simplified and loudness calibration procedures were shortened through data-driven estimation methods. Iterative observations also informed adjustments to task parameters such as the speed of feedback animations and the maximum number of discrimination trials. Observations from these sessions also informed refinements to the training application, including the introduction of short distractor activities intended to reduce fatigue during extended interaction.

Parallel discussions with the rehabilitative speech therapist refined recruitment criteria and workload expectations, while a family Q&A session clarified task difficulty, interpretation of results, and home deployment conditions. In parallel, consultations with a GDPR specialist formalized consent procedures, collaboration agreements, and a local-only data management plan. These steps culminated in IRB approval and transition to the final experimental deployment.

Across the lifecycle, the training application underwent three major interface revisions and approximately ten minor refinements, while the assessment application underwent six major revisions and roughly twenty minor adjustments, as documented in the changelog provided in the Supplementary Materials.

Together, these developments progressively reconfigured the relationships among clinical actors, technological artifacts, and interaction contexts. The following section represents this evolution through a series of actor-network snapshots.

3.3. The socio-technical network

To complement the phase-based account, we represent the design process through three actor-network snapshots corresponding to key moments in the lifecycle. Rather than repeating the chronology, these figures illustrate how the main actors, artifacts, and constraints progressively became connected through translation processes. Across the three snapshots, the network evolves from a largely uncoordinated configuration to a protocol-centered arrangement and finally to an expanded ecosystem that includes domestic and institutional actors.

The initial snapshot (Fig. 1) shows four largely separate domains: SSD-CI children, clinicians, HCI researchers, and mobile technologies. Children are represented through hearing status and developmental profile; clinicians through clinical practice and expertise; HCI researchers through design practice, research methods, and a research agenda oriented toward home-based digital training; and mobile technologies through interaction affordances, deployment potential, and embedded instrumentation. A first translation is visible at the boundary between children and mobile technologies: everyday mobile-device use

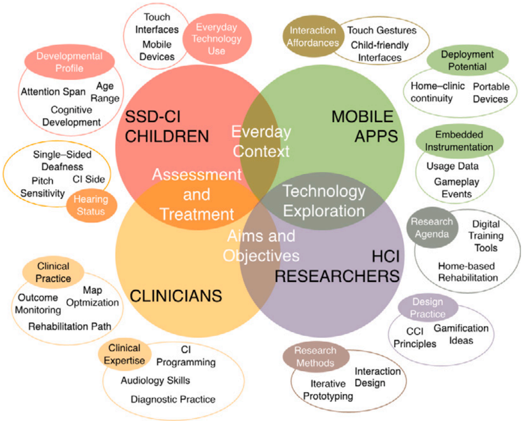


Fig. 1. Initial configuration prior to protocol destabilization.

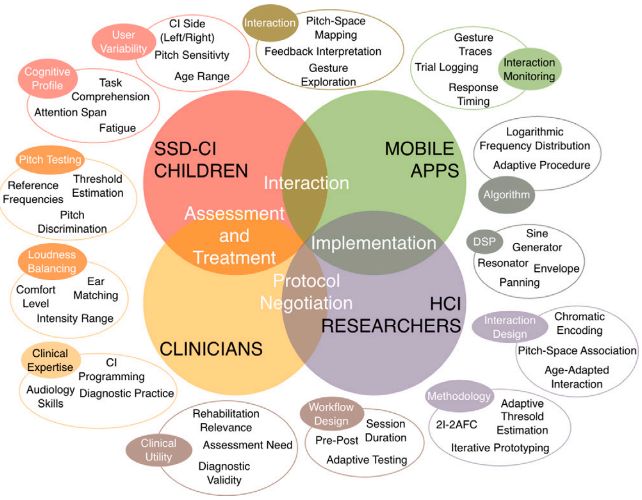


Fig. 2. Configuration following methodological realignment around protocol negotiation.

and touch interaction made mobile applications plausible as child-oriented interfaces. A second translation appears between HCI researchers and mobile technologies: the interest in home-based rehabilitation and digital training tools oriented technological exploration toward portable devices and embedded data capture. At this stage, however, these relations remained only partially connected to clinical assessment practice.

The second snapshot (Fig. 2) corresponds to the methodological realignment achieved during Phase I and Phase II. Here the network becomes reorganized around *protocol negotiation*, which acts as the central mediator linking clinical utility, psychophysical validity, interaction design, and application implementation. One clear translation concerns clinical requirements and loudness balancing: the demand for diagnostically valid pitch testing activated application- and method-related actors such as adaptive procedures, stimulus control, and interaction monitoring. A second translation concerns children's profiles and variability: attention span, fatigue, CI side, and age range triggered responses in workflow and interaction design, including adaptive testing, session-duration control, vertical pitch mapping, and age-adapted interaction. In this configuration, children, clinicians, researchers, and mobile applications are no longer parallel domains; they become connected through the emerging psychophysical protocol and its technical implementation.

Fig. 2 also reveals a third translation within the technical layer of the network. The methodological choice of a 2I–2AFC discrimination paradigm required specific technical counterparts within the mobile applications, represented by actors such as the algorithm, digital signal processing (DSP), and interaction monitoring. In other words, the psychophysical structure was translated into concrete application behaviors including adaptive procedures, controlled stimulus generation, response monitoring, and trial logging. At this stage the distinction between the two applications also becomes visible. The training application emphasizes exploratory interaction and pitch-space navigation, while the assessment application implements a constrained discrimination task. Despite these differences, both applications rely on the same underlying psychophysical protocol (stimulus, adaptive procedures), and shared data-collection methods, which ensure clinically interpretable measurements.

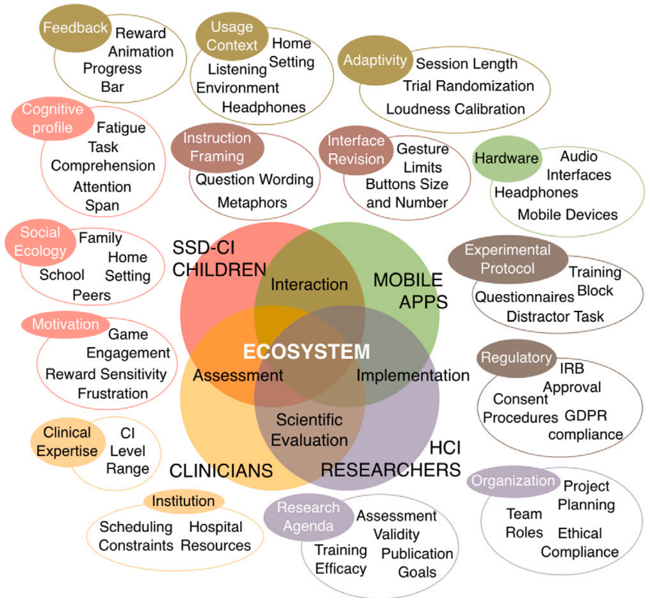


Fig. 3. Ecosystem configuration after ecological and regulatory integration.

The third snapshot (Fig. 3) represents the ecosystem configuration reached during Phase III. Relative to Fig. 2, the network expands beyond protocol construction to include actors associated with deployment, governance, and everyday use. Social actors (families, school and home settings), technical actors (hardware and listening environments), regulatory actors (ethical approval, consent procedures, GDPR compliance), and organizational actors (team roles and research agenda) enter the network alongside motivational and interactional elements such as feedback mechanisms, adaptivity, instruction framing, and interface revisions. A fourth translation is visible here: the move toward home deployment and experimental approval activated requirements that were not reducible to psychophysical validity alone. Domestic deployment introduced both technical constraints and social opportunities. Hardware and listening environments — such as headphones, mobile devices, and home acoustic conditions — imposed practical limits on stimulus delivery and experimental control. At the same time, the social ecology surrounding children, including family and peers, opened opportunities for motivational strategies such as score-based feedback and playful competition. Institutional approval, in turn, activated regulatory and organizational actors that reshaped the operational form of the applications and the experimental protocol.

Taken together, the three figures show that the final system did not result from the linear implementation of a predefined design. Instead, the actor network was progressively reconfigured as different actors acquired the capacity to orient others: children's profiles activated workflow and interaction responses, clinical requirements activated methodological and technical decisions, and deployment and regulatory actors expanded the network into a broader ecosystem. The figures

Table 1
Artifacts used to identify and document actor statements across lifecycle phases.

Data type	Data collected	Phase	Actors
Text	Research notebooks	I–III	HCI researchers, Clinicians
	Email exchanges	I–II	HCI researchers, Psychoacoustic expert
	Consent form drafts	III	GDPR expert, HCI researchers
	Q&A session notes	III	Families, HCI researchers
Sketch	Prototype drawings	I–II	HCI researchers
Video	Pilot testing recordings	III	Children
Logs	Behavioral interaction data	II–III	Applications, Clinicians, Children

therefore provide a compact account of how the project moved from an initially loose constellation of actors to a coordinated training—assessment ecosystem whose dynamics are examined in more detail through stakeholder statements in the following section.

3.4. Documenting actors' statements

Beyond the structural representation of the actor network, we documented the concrete interactions through which design decisions emerged during the lifecycle. Our analysis examined heterogeneous artifacts produced across the three phases described above, including research notebooks, prototype sketches, email exchanges, pilot-session recordings, interaction logs, and regulatory documents (Table 1). These materials provided a record of interactions, observations, and discussions occurring throughout the lifecycle.

We conducted a qualitative content analysis following established procedures (Holsti, 1969). A *statement* was defined as a discrete unit of meaning expressing either (i) a clinical or methodological constraint, or (ii) an observable child behavior with implications for design. Examples include fatigue markers, response-time drift, clarification requests, or gaze aversion during tasks. Initial segmentation was performed by the first author across the collected materials and subsequently reviewed in interdisciplinary meetings involving two HCI researchers and the rehabilitative speech therapist.

Statements were examined in relation to the design episodes in which they appeared. For example, repeated gaze aversion and increasing response times during loudness balancing were documented in interaction logs and pilot-session notes as indicators of possible cognitive overload. Similarly, recurrent clarification requests during assessment sessions (e.g., “What does this do?”) were recorded in fieldnotes and video observations during task introduction. These statements were associated with specific interaction phases of the assessment procedure and documented alongside the corresponding artifacts and actors involved.

Table 2 summarizes representative statements and observations documented across the lifecycle. Additional statements emerged during pilot-session observations in Phase III. For example, reports of fatigue (e.g., “My hand is hurting”) and visible tapping strain were recorded during loudness calibration tasks. Interaction logs also documented increasing response times across repeated trials. During exploratory sessions involving NH children, competitive exchanges and score comparison behaviors were observed during repeated game interactions. These observations were documented as part of the interaction traces collected during pilot testing.

Some statements originated from clinical and methodological discussions rather than direct observation. Clinical consultations during Phase II documented the need to separate training and assessment modules and to structure the workflow around loudness calibration followed by pitch discrimination. Psychoacoustic consultation documented the preference for discrimination paradigms over detection-based testing. Observations regarding the complexity of continuous zoom interactions informed the lower age boundary adopted for the training activity.

Table 2
Examples of stakeholder statements and observations documented across the lifecycle.

Actor	Statement	Source
Clinicians	Assessment required for clinical utility Continuous zoom too complex for <8 years	Research notebooks Research notebooks
Psychoacoustic expert	Use discrimination, not detection paradigm Adopt 2I–2AFC for methodological rigor	Email exchange Research notebooks
SSD-CI child	Increased RT across trials Visible hesitation during loudness calibration	System logs Pilot testing notes
HCI researchers	Separate training and assessment modules Implement adaptive GPR to shorten balancing	Research notebooks Research notebooks
Families	Clarify procedure before home participation	Q&A session notes
Speech therapist	Reduce cognitive load in interface	Research notebooks
GDPR specialist	Prohibit cloud storage for sensitive data	GDPR draft
Mobile apps	Repeated zoom adjustments during pad selection	System logs

Across phases, children's hesitation, response-time drift, humming along with tones, and visible fatigue were documented during pilot sessions. Family feedback contributed to the development of onboarding explanations, speech therapist recommendations supported interface simplification, and regulatory consultations required local-only data storage procedures.

Taken together, these documented statements provide a transparent account of how clinical constraints, child interaction traces, and interdisciplinary discussions informed the evolution of the system. They also provide the empirical basis for the negotiation processes analyzed in the following section.

3.5. Negotiating a rehabilitation ecosystem

The participatory lifecycle described above documented a series of tensions between clinical validity, child-centered interaction, and ecological deployment. These tensions did not concern isolated interface details only; rather, they concerned how far the procedure could be adapted without compromising the psychophysical requirements of the assessment. The final system emerged through repeated negotiations among clinicians, researchers, children, and technological artifacts, each of which introduced constraints or opportunities that redirected subsequent design choices.

The psychophysical paradigm in pitch perception

Early discussions with clinicians established that the procedure had to produce clinically interpretable thresholds while remaining feasible for children. Consultation with a psychoacoustics expert therefore shifted the design away from initial detection-oriented ideas toward a two-interval two-alternative forced-choice (2I–2AFC) discrimination procedure using a weighted up–down adaptive rule. Adaptive procedures are usually preferred because they ensure dense sampling of the psychometric curve near the threshold region (Arzounian et al., 2017).

This choice made it possible to retain a methodologically grounded threshold estimation procedure while keeping the number of trials manageable. Alternative paradigms were also considered and discarded. In particular, MUSHRA-style multi-stimulus comparison procedures were rejected because they require the simultaneous evaluation of several stimuli — typically more than five (International Telecommunication Union, 2015) — which was considered cognitively demanding for children and difficult to reconcile with the attentional limits observed during pilot sessions. Non-adaptive stimulus grids were also discarded because they would have required substantially more trials without model-based interpolation.

Loudness calibration

Clinical discussions emphasized that pitch judgments should not be confounded by differences in perceived loudness, since sound level can influence pitch perception (Zheng & Brette, 2017). This made loudness balancing a necessary precondition for the discrimination task. However, pilot observations during manual calibration documented increasing response times, hesitation, and gaze aversion, suggesting that the procedure was too demanding in its initial form. The introduction of an adaptive Gaussian Process Regression (GPR) procedure therefore did not replace the clinical requirement for loudness balancing; rather, it reconfigured how that requirement could be satisfied within a child-appropriate session duration. In this sense, the negotiation concerned not whether loudness balancing should occur, but how it could be implemented without exhausting participants before the assessment itself.

Relationship between training and assessment

Early iterations showed that a single application could not adequately support both exploratory practice and controlled psychophysical measurement. The two activities required different interaction logics: training benefited from open-ended exploration, repetition, and explicit motivational scaffolds, whereas assessment required constrained responses and reduced interactional interference. The resulting solution was to separate the system into two coordinated applications: a training game oriented toward exploratory pitch identification and a therapist-supervised assessment application implementing the discrimination protocol. This division preserved methodological control in assessment while allowing greater flexibility in the interaction design of training.

Role of reward and competition

Observations made during pilot sessions of the assessment application revealed that repeated trials could quickly reduce attention when no visible progression or comparison was present. During these sessions, several NH children spontaneously compared their results and commented on each other's performance. These behaviors suggested that comparative scoring could function as a motivational element during repeated interaction. Although these observations emerged during the evaluation of the assessment task, they informed the design of the training application, where repetition is expected to occur over longer periods at home. As a result, the training game progressively incorporated clearer reward dynamics, including visible score accumulation and a top-score system. The intention was to make repeated pitch identification practice desirable enough to sustain voluntary engagement outside the clinic.

The assessment application required a different balance. In many psychophysical experiments, feedback is avoided because it may influence response strategies during threshold estimation. However, pilot sessions showed that when small collectible elements were introduced into the interface, children displayed noticeably stronger concentration and emotional engagement with the task. For this reason, a minimal reward mechanism was incorporated into the assessment interaction. The interface adopted a visual progression metaphor involving marbles and collectible strawberries, which provided a sense of movement and accomplishment during repeated trials. The detailed interaction structure is described in Section 4.

Ecological deployment

Home-based training required compatibility with children's own devices and with everyday listening conditions, while institutional review required compliant procedures for consent, recruitment, and data storage. In parallel, ecological considerations also affected content choices: realistic but controlled damped-sine stimuli were retained because they preserved experimental control while remaining interpretable as elements of a plausible sound ecology. Mixed-method components, including onboarding discussions and post-session questionnaires, were added to document not only performance but also workload, engagement, and family interpretation of the system. Taken together, these negotiations transformed an initially technical assessment problem into a deployable training—assessment ecosystem linking home practice, therapist-supervised evaluation, and caregiver-supported participation.

4. Experimental design

Each participant completed a pre-training assessment, a series of home-based training sessions, and a post-training assessment. The protocol described below corresponds to the stabilized configuration resulting from the participatory design process detailed in Section 3. It integrates a psychophysically grounded pitch discrimination assessment with an interactive training activity designed for children, preserving methodological requirements for reliable threshold estimation while structuring interaction to remain feasible for pediatric participants. Fig. 4 details the experiment along each phase.

4.1. Pre- and post-training assessments

The pre- and post-assessments were identical and included five steps: task illustration, loudness calibration, adaptive pitch discrimination, an optional distractor task, and post-task experience reporting.

4.1.1. Illustration of the tasks

The therapist introduced the pre- and post-training tasks and supervised participants throughout the assessment. Her role was to explain the procedure to children and parents, reduce misunderstandings, and support family compliance with the full protocol (Rotfleisch & Martindale, 2021). Space–pitch mappings and visual anchors were introduced at the start to support abstract auditory judgments. Visual metaphors were used throughout to keep actions rapid, incremental, and reversible.

The therapist adapted task explanations to participants' developmental stage. All participants were introduced to the assessments through a space–pitch association, mapping ascending pitch to upward movement and descending pitch to downward movement, consistent with established cross-modal correspondences from early childhood (Dolscheid et al., 2014). For younger children, explanations drew on familiar therapeutic activities and were delivered under closer supervision in familiar rooms to prevent overload (Martini & Schindler, 2004). Older participants required less scaffolding and could complete the assessments more autonomously, with iconographic or musical examples used when useful to clarify the task.

4.1.2. Loudness balancing task

Once the tasks were illustrated, participants adjusted each stimulus to their most comfortable loudness (MCL). They then matched sounds across ears so they were perceived as equally loud for each reference stimulus. This step was necessary to reduce the risk that the subsequent pitch discrimination task would be interpreted as a loudness comparison.

To keep cognitive load child-appropriate, we adopted an active-learning variant of pairwise comparison driven by GPR with explicit noise modeling. At each trial, children matched the loudness of two exponentially damped sinusoids having different fundamental frequency by adjusting the comparison tone using three buttons. Two increase

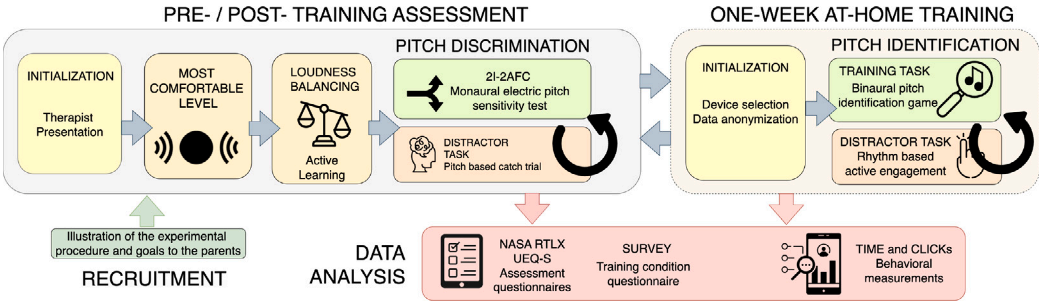


Fig. 4. Details of the experimental protocol.

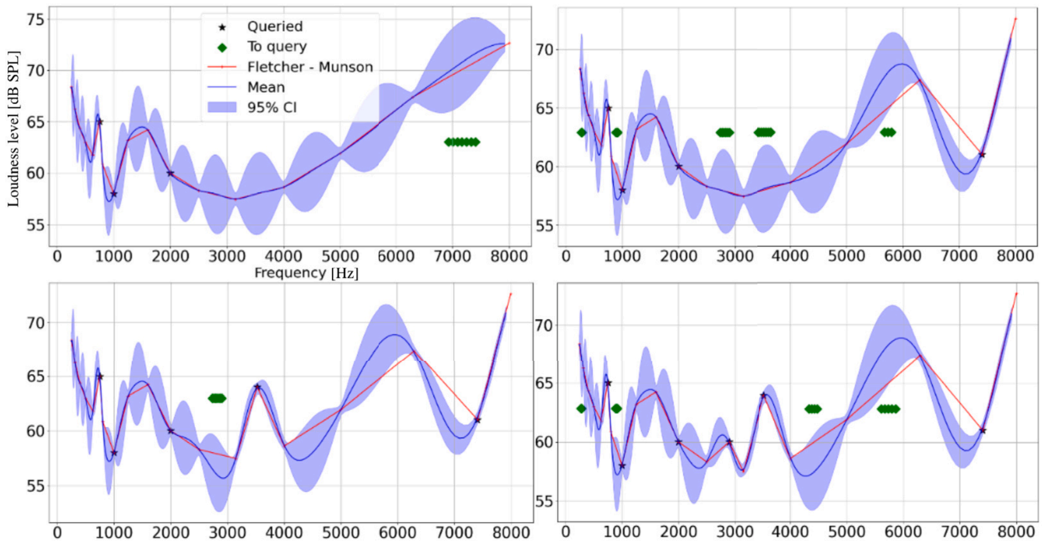


Fig. 5. First four iterations of the loudness-matching estimation procedure (from top left clockwise).

controls (+1 dB and +4 dB) were positioned in the upper part of the interface, and one decrease control (–1 dB) in the lower part, following the louder/up and softer/down spatial association (see Fig. 8 - left in Section 5.1) (Puigcerver et al., 2019).

An active learner queried the most informative comparisons so that the individual loudness-matching profile could be estimated from no more than seven responses. This upper bound reflected therapists' prior knowledge of attention span and ensured that the procedure remained within a limited child-appropriate number of adjustments while preserving adaptivity and personalization. After seven queries, the estimated profile was used to set the comparison levels for the subsequent pitch discrimination task.

Detailed mathematical formulation is reported in Appendix A; here we summarize the operational steps of the procedure. A Gaussian-process regression model (Rasmussen & Williams, 2019) iteratively estimated this loudness-matching profile over the tested frequency range. At each iteration, the system selected an informative comparison, the child adjusted the sound level, and the model updated its estimate. Compared with earlier procedures requiring many more adjustments (Schlittenlacher & Moore, 2020), our implementation limited interactive loudness matching to seven iterations per stimulus (three stimuli) in order to balance calibration accuracy with session duration. Interactive loudness adjustments were limited to seven iterations per stimulus (three stimuli) to balance calibration accuracy with session duration. Pilot sessions indicated that this configuration allowed children to complete the task comfortably while maintaining stable level estimates. Four iterations of the contour estimation process are represented in Fig. 5.

4.1.3. *Adaptive pitch discrimination task*

Pitch discrimination thresholds can be measured using various methods. In our procedure, a reference stimulus presented at the aided ear was followed, after a 500 ms delay, by a stimulus presented at the contralateral (unaided) ear with frequency and level variable according to the previously computed loudness-matching profile. The reference was presented at the participant's MCL and kept constant throughout the whole sequence. The task was to sort the two stimuli in increasing order based on their perceived pitch. We used an adaptive 2I–2AFC procedure (Kaernbach, 1991): after each response, the task became harder when the answer was correct and easier when the answer was incorrect. When no difference was perceived, participants were instructed to guess (Green et al., 1966). The frequency distance between the two sounds was adjusted automatically to estimate the point at which the child could reliably detect a pitch difference. On each sequence, the initial step size was set to 10% and adapted across 16 trials. This value was empirically determined for the session to last approximately 30 min. After 16 trials the mean and the confidence interval at 95% were computed. If the last frequency step was found within that interval, then the resulting threshold was considered sufficiently stable, the pitch discrimination sequence terminated, and the participant started with the next reference frequency; otherwise, another sequence of 16 trials was repeated for the same reference with a distractor task inserted between blocks. Block length and stopping criteria were determined during pilot sessions to balance participants' attention span with reliable psychometric convergence while limiting total session duration. An example of testing a reference tone through two sequences is shown in Fig. 6.

Three reference frequencies were tested: 750 Hz, 1 kHz, and 2 kHz. These frequencies were chosen because they fall within a range relevant

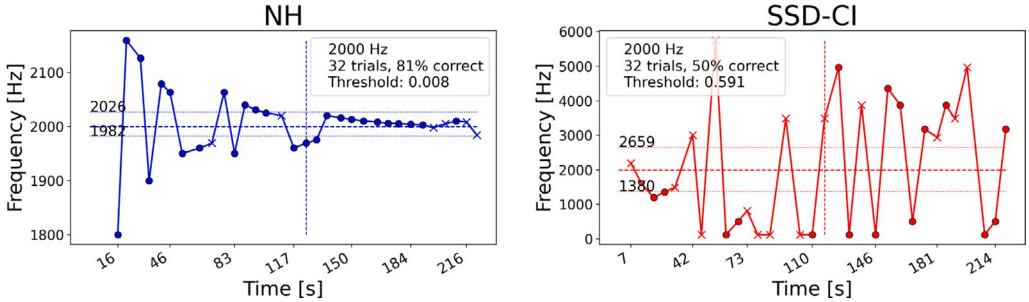


Fig. 6. Example pitch discrimination runs for the 2000 Hz reference frequency. The left panel shows an NH participant and the right panel an SSD-CI participant. Blue denotes NH listeners and red denotes SSD-CI listeners. Circles indicate correct responses and crosses incorrect responses. The dashed horizontal line corresponds to the reference frequency, and the dashed vertical line marks the beginning of the second sequence after the insertion of a distractor task. The dotted horizontal lines represent the confidence interval at 95% computed for all the answers until the second-to-last.

to everyday speech and are supported by common CI devices (Nguyen, 2023; Wagner et al., 2021).

4.1.4. *Distractor task*

If the discrimination threshold did not reach a stable value after 16 trials, a catch trial was presented before the next sequence of trials around the same reference frequency, to give participants a break by completing a different task (Adel et al., 2019) involving both visual and auditory elements. A static scene replaced the scrolling screen. Two pure tones were presented at the two ears simultaneously. Afterward, participants had to answer, “Do you hear one or two different sounds?”. Participants indicated whether they heard the same sound in both ears or two different sounds, and no feedback was provided. Such tones had the same frequency as the most recently discriminated pair, for their prospective use in a further analysis involving comparisons between audiological (typically pure tones) and ecological stimuli.

4.1.5. *Quality of the experience*

To conclude the assessment, participants answered a modified version of the NASA task load index (TLX) (Hart, 2006) called the NASA Raw TLX (RTLX) (Hart & Staveland, 1988), rating their subjectively perceived workload; furthermore they answered a short version of the User Experience Questionnaire (UEQ-S) (Schrepp et al., 2017), rating the subjective impression of the user experience. The modified version was used because of its reduced length. These questionnaires were included to document perceived workload and user experience before and after training, complementing behavioral measures with subjective reports of task difficulty and usability.

4.2. *Training at home*

Home training consisted of repeated pitch identification trials performed through a mobile game running on a tablet. As in the assessment, a distractor task was inserted between identification blocks to relieve attentional fatigue. Based on pilot observations that children disengaged more easily after repeated steady-pitch trials, the distractor activity was designed to require active engagement (Agrawal & Bech, 2023) through movement and action. Inspired by a blend of Whac-a-Mole (Garcia et al., 2018) and Guitar Hero (Yuan & Folmer, 2008), participants had to keep pace with a rhythmic auditory pattern by tapping drum-kit icons on the touchscreen.

After each session, participants completed a brief questionnaire by selecting from a menu (see Table 3) providing three options about the reproduction device, room size, volume level, as well as their mood during the game. These items documented the home conditions under which training occurred and supported later interpretation of differences in training behavior across participants.

Table 3
Menu options in the training session questionnaire.

Device	Room	Level	Mood
Headphones	Small	High	Calm
Tablet	Medium	Medium	Bored
Other	Big	Low	Upset

4.3. *Stimuli*

Standard clinical pitch discrimination tests often employ pure tones (Arzounian et al., 2017; Wagner et al., 2021) or 1/3-octave band noise (Adel et al., 2019), however both such sounds rarely are part of an everyday listening ecology. In our study, exponentially decaying sinusoids were selected for both testing and training because they provide a clear and controllable pitch while remaining closer to plausible real-world acoustic events than standard laboratory tones. This choice was informed by prior use of such sounds in sonic interaction design (Rocchesso & Fontana, 2003), speech transient modeling (Hermus et al., 2005), and impact-sound synthesis for material interactions (Aramaki et al., 2011; Chatziioannou & van Walstijn, 2015). It therefore supported our goal of assessing and training pitch perception with stimuli that remained controlled but were still compatible with everyday listening contexts.

We obtained these stimuli as impulse responses of a digital damped oscillator which allows for accurate control at runtime of the center frequency F_0 and decay time τ (Poli & Avanzini, 2005). Such sinusoids were synthesized in real time using the sound programming software Pure Data[1] via the dedicated library libpd[2] (Brinkmann et al., 2016). The waveform and spectrogram of a stimulus with $F_0 = 1000$ and $\tau = 0.1$ can be seen in Fig. 7(a).

Sample reproductions of a Roland TR-808 drum (Werner et al., 2014) were prepared for the distractor task occurring while training at home (Agrawal & Bech, 2023). A subset of these samples was selected to maximize differences in spectral content and temporal envelope, so that the distractor task remained perceptually distinct and engaging. To check that these differences were also meaningful under cochlear-implant listening conditions, the candidate sounds were additionally inspected after processing with a classic CI simulator (McDermott et al., 1992). One sample and the spectrogram resulting from CI-simulated processing is shown in Fig. 7(b).

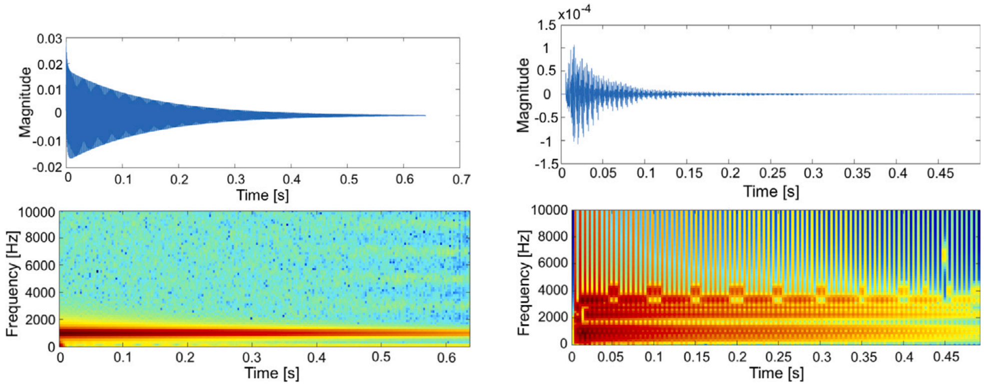
4.4. *Mobile apps and setup*

Two mobile apps were developed for both iOS and Android tablets, for cross-platform use during training and assessment, respectively.

---

[1] https://puredata.info.

[2] https://github.com/libpd/libpd.



(a) Waveform and spectrogram of a damped sinusoid with $F_0 = 1000$ and $\tau = 0.1$.

(b) Waveform and spectrogram after CI-simulated processing of a synthesized TR-808 cowbell sample.

Fig. 7. Auditory stimuli waveforms and spectra.

Both apps relied on libraries compatible also with older tablets, such as those owned by the participants. We used Swift and Objective-C provided by iOS SDK, and Kotlin and Java provided by Android SDK. We adopted the SpriteKit[3] framework for the graphics and animation rendering on XCode 15.2. The Android version of the apps were developed using AndroidX libraries for modern Android components and Jetpack features on Android Studio Flamingo 2022.2.1 Beta 4. In both applications, Pure Data was in charge of providing runtime sound synthesis at 48 kHz sampling rate. A Pure Data patch was integrated in both apps via libpd, which runs on Android as well as iOS: on Android, libpd interacted with the audio system through the AudioTrack[4] and AudioManager[5] APIs, native threads, and the app lifecycle; on iOS, libpd interacted with Core Audio via AVAudioSession.[6] We preliminarily checked that no notable audio latency was present while running the apps on an Apple iPad Pro 11-inch 3rd generation 256 GB and on a Samsung Galaxy Tab A8 64 GB. Occasional latencies around 50 ms have been reported on such setups.[7] However, delays on the order of tens of milliseconds did not significantly decrease engagement even on games where audio contributes to closed-loop interactions that are essential for the game mechanics (Halbhuber et al., 2022). Finally, the active learning algorithms for loudness balancing were implemented through custom C++ functions using the Eigen[8] and Boost[9] libraries, and then integrated within the assessment app.

The audio during the pre- and post-training test assessments was sent from the tablet to an Antelope Zen Go Synergy Core USB sound interface via a USB-C 3.5 mm jack adapter. The analog output was then amplified by an RME Babyface Pro sound card. Finally, stimuli were delivered via Sennheiser HD 600 open-back headphones with ±2.5 dB SPL tolerance and 104 dB peak SPL. This peak level accommodated our participants' diverse audiometric thresholds before loudness balancing.

At home, no specific reproduction device was prescribed for training. Participants instead reported how stimuli were reproduced while completing the post-session questionnaire, allowing home listening conditions to be documented rather than constrained.

5. Procedure

5.1. Pre- and post-training assessments

After task illustration, participants first completed the loudness balancing task. The SPL in the aided ear was initially set to 50 dB, then adjusted by listeners to a comfortable level. Participants controlled loudness by repeatedly tapping three animated virtual buttons on the GUI (see Fig. 8 - left), which respectively increased the SPL by 1 or 4 dB steps or decreased it by 1 dB steps. This GUI followed the spatial–loudness mappings described in Section 4.1.2.

During the initial explanation, the experimenter explicitly encouraged participants to overshoot and undershoot before settling on a preferred level, in order to familiarize them with the response mechanism. The loudness balancing phase typically lasted about 10 min, in line with recent computerized efficient procedures (Serpanos & Gravel, 2004) and substantially shorter than more traditional fitting protocols (Cox et al., 1997).

Participants then performed the adaptive pitch discrimination task, with the possibility of engaging in the distractor task. A screenshot of the GUI is shown in Fig. 8 - center. Two bouncing spherical "marbles" are positioned on both sides of the screen. The marbles run on a track, and the screen scrolls horizontally without noise. When they bounce, corresponding stimuli are presented. The runaway marble bounces when the test stimulus is delivered. After it jumps, it rolls out of the screen. At that point, a question mark and two arrows pointing up and down appear. Once listeners were familiar with the upward/downward space-pitch association introduced at the beginning, they made their decision based on the following question: "Where did the marble go, up or down?". This interaction relied on the well-documented association between pitch and vertical spatial position, whereby higher pitches are commonly perceived as "up" and lower pitches as "down" (Cuturi et al., 2019). Feedback was provided with an animation that showed the two marbles traveling the same track and the chaser approaching the runaway when the answer was correct, and two different tracks and more distant marbles otherwise. To increase engagement, a score counter kept track of the number of strawberries each participant won whenever a correct answer was given (Nicholson, 2015). Regarding the distractor task, participants tapped an icon: a single umbrella for the same sound, or pair of flippers for different sounds.

Finally, the experimenter and participants' parents discussed the results together and helped complete the questionnaires. While filling the questionnaires, most young children were helped by the experimenter, one parent, or the speech therapist. Data were collected and saved on the device, anonymized, and transferred to the experimenter's laptop, consistent with the GDPR-compliant local data management strategy outlined in Section 3. An example of a child performing the assessment can be seen in Fig. 8 - right.

---

[3] https://developer.apple.com/spritekit.

[4] https://developer.android.com/reference/android/media/AudioTrack.

[5] https://developer.android.com/reference/android/media/AudioManager.

[6] https://developer.apple.com/documentation/avfaudio/avaudiosession.

[7] https://puredata.info/docs/AndroidLatency, https://musicalogic.wordpress.com/2016/06/05/real-time-audio-programming-in-ios-comparing-the-play-through-latencies-of-libpd-and-the-amazing-audio-engine.

[8] https://eigen.tuxfamily.org.

[9] https://www.boost.org.

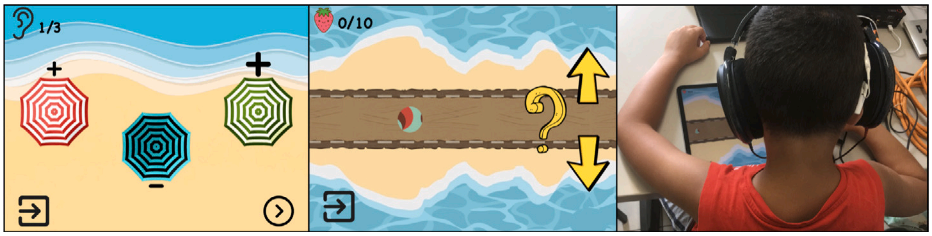


Fig. 8. Pitch discrimination app: (left) loudness balancing GUI; (center) pitch discrimination GUI; (right) SSD-CI child performing the assessment.

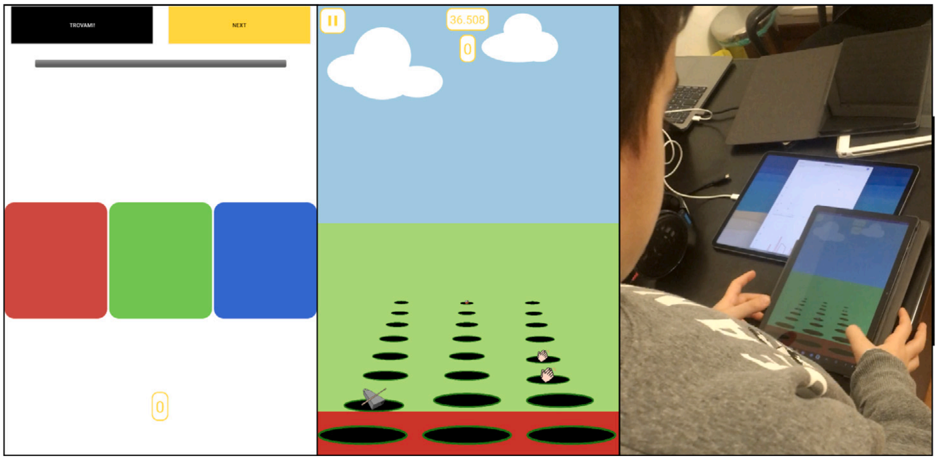


Fig. 9. Training pitch identification app: (left) pitch identification GUI; (center) distractor task GUI; (right) NH child performing the training.

5.2. Training at home

Training was conducted at home, in a quiet room chosen by the participants and their families. The volume level was individually set by each participant. Each game session proposed four pitch identification trials interspersed with three distractor tasks. The training at home consisted of at least one game every two days on the dedicated app across one week. The training was completed when at least three game results were collected. Each participant was instructed to play for at least seven days and twelve days at most.

The GUI, depicted in Fig. 9 - left, presented a rectilinear grid made of N pads, each reproducing a stimulus with a specific frequency when tapped, similar to a musical software keyboard. The frequencies of the respective stimuli always ranged from 500 Hz (leftmost pad) to 4 kHz (rightmost pad): on a grid partitioned into N pads, the $n+1$th pad from the left reproduced a damped sinusoid with center frequency $F_0 = 500{\times}2^{3n/(N-1)}$ Hz, $n = 0, \dots, N-1$. To keep a coherent audio-visual semantics, the grid was color-coded as a rainbow, with red representing the lowest center frequency and purple the highest.

The task consisted of listening to a reference stimulus that was randomly selected from the N possible damped sinusoids, and freely tapping the pads until identifying the one reproducing the same stimulus. During the task, the reference stimulus could be replayed at any time via a dedicated black-colored button. The grid could also be horizontally zoomed in and out by using a “pinch-to-zoom” gesture over the touchscreen to spawn denser pads for finer discrimination—a native gesture that supported exploration, agency, and direct manipulation (Hourcade, 2022). Zooming in duplicated and, if pinching further, tripled each pad along the grid, with the effect of getting the frequencies reproduced by adjacent pads progressively closer to each other.

Each trial proposed five such tasks with increasing difficulty, with the respective grids initially (i.e., before zooming) containing 3 to 7 pads. Participants earned 10 points for each correct identification, with the highest score saved and displayed on the homepage.

The distractor task occurred three times during the game session, each time gradually increasing the tempo starting from 70 until 100 beats per minute. Each well-synchronized tap earned one point, contributing to the overall score. The corresponding GUI is shown in Fig. 9 - center. Finally, participants answered the questionnaire shown in Table 3.

6. Data analysis

The assessment and training apps recorded all the interactions in a log file. For both pre- and post-training assessments, the app recorded each MCL, the interpolated contours for the contralateral ear, and the frequency discrimination thresholds. Thresholds were collected as percentages and mapped onto a logarithmic scale.

Data normality was checked using the Shapiro–Wilk test (Gupta et al., 2019) and the homogeneity of variances using the Levene test (Brown & Forsythe, 1974). If the data exhibited a normal distribution and homoscedasticity was met, we proceeded with a t-test; otherwise, we opted for the Wilcoxon rank sum test. When comparing thresholds from the two populations, the Mann–Whitney U rank test was used. We calculated the effect size and statistical power for each test and corrected the p-values for multiple comparisons using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995).

From the pre-training thresholds, we computed the edge frequencies of each discrimination interval around the respective reference stimulus. The number of stimuli to be identified during the training with a center frequency falling within that interval was defined as the “training intensity”. By summing up training intensities across all games, we quantified the different amounts of training each participant took for each reference stimulus. By this definition, the greater one participant’s discrimination ability, the smaller the training intensities. An example involving one NH and one SSD-CI listener is shown in Fig. 10.

We fitted a Generalized Linear Model to evaluate the contribution of the training to pitch discrimination, by examining post-training thresholds as a function of pre-training thresholds. We compared a sequence

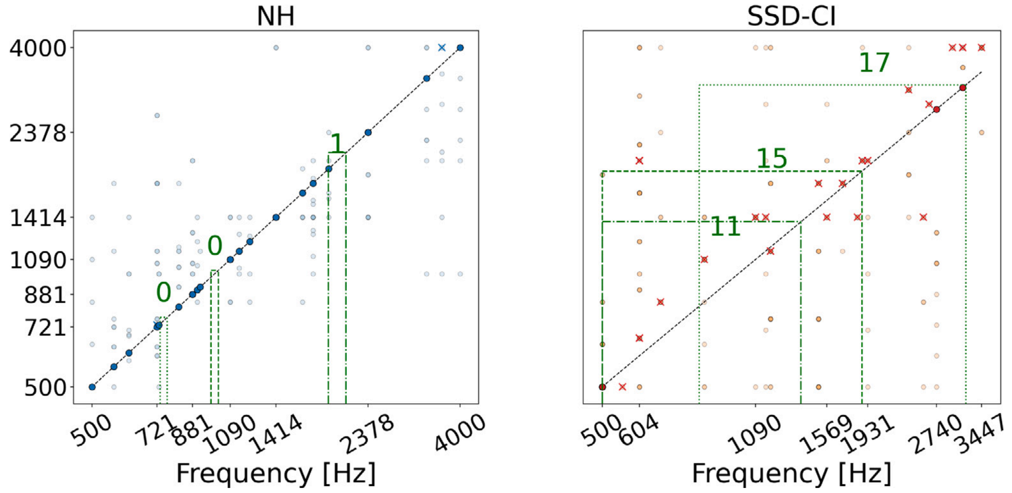


Fig. 10. Results of two training sessions from (left) one NH and (right) one SSD-CI listener. Colored markers represent the stimuli played during the training (blue for NH, red for SSD-CI), while lighter markers indicate the frequencies of the acoustic stimuli explored during the session. Circles indicate correct responses and crosses incorrect responses. The dashed diagonal line represents the locus of correct responses, where the identified frequency matches the presented stimulus frequency. Green dashed rectangles illustrate the pre-training discrimination thresholds for the three reference frequencies, namely 750 Hz, 1 kHz, and 2 kHz. The numbers in green indicate the three training intensities.

of nested models of increasing complexity (M_0: baseline ANCOVA with pre-training threshold only; M_1: with a Group term added; M_2: with a Train term added; M_3: with an interaction term Group × Train added) to evaluate the terms that were necessary to capture the effects of training and group membership. The Group × Train term was included in the complete model to test whether the training benefits differed between the SSD-CI and the NH listeners:

$$\text{Post} = \beta_0 + \beta_1 \times \text{Pre} + \beta_2 \times \text{Group} + \beta_3 \times \text{Train} + \beta_4 \times \text{Group} \times \text{Train}. \qquad (1)$$

We adopted a Gaussian distribution family with an identity link function and, after shifting the post-training thresholds to positive values, a Gamma family with a log link function. To assess adequacy and avoid overfitting, we examined residuals (Shapiro–Wilk) (McCullagh & Nelder, 2019), calculated Cook's distances to detect influential cases (Cook, 1977), and validated the models using two subject-independent strategies: an 80/20 holdout split and Leave-One-Subject-Out (LOSO) cross-validation. We also estimated robust standard errors (HC3 (Long & Ervin, 2000) and cluster-robust by subject (Huang & Li, 2021)), fitted Huber robust regressions (Huber, 1964), and compared results with mixed-effects models including random intercepts by subject.

The time spent on the apps and the total number of taps were collected from both groups, and then statistically compared by considering the pre- and post-assessments and training data. These data could provide insights into the accessibility of the apps, especially during unsupervised training at home.

The results of the NASA Raw TLX and UEQ-S questionnaires were compared item by item. Moreover, an aggregate score from the NASA Raw TLX was derived to estimate the overall perceived workload. The results of the home questionnaires are instead reported with no further analysis.

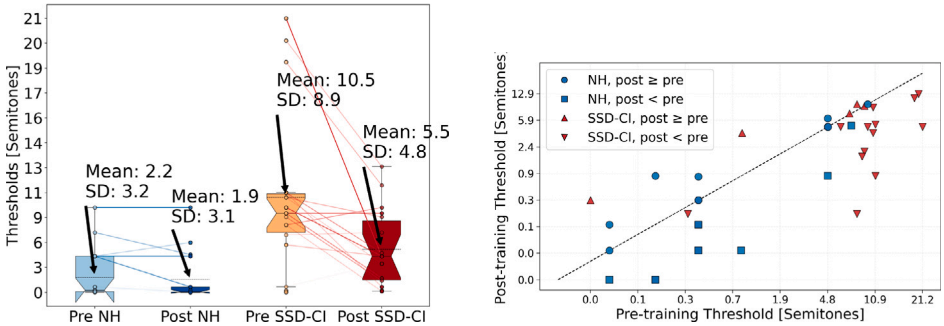
7. Results

A visual inspection of the boxplots in Fig. 11(a) reveals that the thresholds measured after the training decreased for both groups. The SSD-CI listeners' thresholds decreased from a mean of more than ten semitones ($\mu \approx 10.5$, $\sigma \approx 8.9$, median ≈ 9.1) to less than six ($\mu \approx 5.5$, $\sigma \approx 4.8$, median$\approx$ 4.8), corresponding to a reduction of almost half an octave. The NH control group's thresholds decreased from slightly more than two semitones ($\mu \approx 2.15$, $\sigma \approx 3.21$, median$\approx$ 0.76) to slightly less than two semitones ($\mu \approx 1.86$, $\sigma \approx 3.10$, median$\approx$ 0.13). At the individual level, all SSD-CI and NH participants showed improvement in at least one stimulus, with 50.0% and 57.1%, respectively, improving across all stimuli, and 33.3% and 28.6% improving in exactly two.

The pre-training thresholds of the NH participants were not normally distributed ($W = 0.87$, $p = 0.011$), whereas after the training they became normally distributed ($W = 0.92$, $p = 0.071$), furthermore the homoscedasticity was met ($W = 1.24$, $p = 0.27$). The Wilcoxon signed-rank test found a statistically significant difference between the two thresholds ($W = 20.0$, $p = 0.023$), with effect 0.91, and power 0.82. The pre- and post-training thresholds of the SSD-CI participants were not normally distributed ($W = 0.72$, $p = 1.46 \times 10^{-4}$ and $W = 0.87$, $p = 0.021$, respectively), and the homoscedasticity was met ($W = 0.021$, $p = 0.87$). The Wilcoxon signed-rank test found a significant difference between the two thresholds ($W = 29.0$, $p = 0.025$), with effect 0.83 and power 0.68.

When comparing thresholds between groups, we found a statistical difference between the pre-training thresholds (both not normally distributed, NH: $W = 0.87$, $p = 0.011$, SSD-CI: $W = 0.72$, $p = 1.46 \times 10^{-4}$, Levene $W = 0.46$, $p = 0.50$, Mann–Whitney $U = 51.5$, corrected $p = 2.09 \times 10^{-4}$, Cohen's effect size $U = 7.07$, power 1.0), between the NH pre-training thresholds and the SSD-CI post-training thresholds (NH: $W = 0.87$, $p = 0.011$, SSD-CI: $W = 0.87$, $p = 0.021$, Levene $W = 0.42$, $p = 0.52$, Mann–Whitney $U = 89.0$, corrected $p = 4.73 \times 10^{-3}$, Cohen's effect size $U = 5.14$, power 1.0), between the NH post-training thresholds and the SSD-CI pre-training thresholds (NH: $W = 0.92$, $p = 0.071$, SSD-CI: $W = 0.72$, $p = 1.46 \times 10^{-4}$, Levene $W = 0.42$, $p = 0.52$, Mann–Whitney $U = 41.5$, corrected $p = 1.34 \times 10^{-4}$, Cohen's effect size $U = 7.59$, power 1.0), and finally between the post-training thresholds (NH: $W = 0.92$, $p = 0.071$, SSD-CI: $W = 0.87$, $p = 0.021$, Levene $W = 3.04$, $p = 0.09$, Mann–Whitney $U = 73.0$, corrected $p = 1.48 \times 10^{-3}$, Cohen's effect size $U = 5.97$, power 1.0).

The Generalized Linear Models were fitted with log-likelihood minimization through the iteratively re-weighted least squares method. We checked the normality assumption on the residuals with the Shapiro–Wilk test, and in both cases it was met (Gaussian: $W = 0.97$, $p = 0.53$; Gamma: $W = 0.98$, $p = 0.78$). We first considered a sequence of nested models: M_0 (baseline ANCOVA with pre-training threshold only), M_1 (+Group), M_2 (+Train), and M_3 (+Group × Train). Model comparison showed that adding Group or Train alone (M_1, M_2) did not improve predictive performance: for M_0–M_2, hold-out RMSE was 1.16–1.23, MAE 1.02–1.05, and predictive R^2 was 0.66 for M_0, 0.62



(a) Boxplots of pre- and post-training thresholds across groups. The opacity of the red lines is proportional to the training intensity.

(b) Individual post-training thresholds as a function of the pre-training thresholds plotted on a log–log scale, with values expressing percentages of a semitone.

Fig. 11. Pre- and post-training thresholds.

for M_1, and 0.62 for M_2. LOSO cross-validation yielded negative mean predictive R^2 values (M_0: –2.96; M_1: –4.32; M_2: –4.38). The complete model including the interaction term (M_3) achieved similar hold-out accuracy (predictive $R^2 = 0.58$) but a better in-sample fit (AIC = 102.5 vs. 103.9–106.3 for simpler models). Thus, M_3 provided the best in-sample description of the data and is reported below. LOSO cross-validation confirmed that prediction errors were comparable across models (RMSE means 1.26–1.39, MAE means 1.15–1.30), with all models yielding negative predictive R^2 values (M_0: –2.96; M_1: –4.32; M_2: –4.38; M_3: –3.86). The complete model (M_3) did not reduce errors compared to simpler models but provided the lowest AIC (102.5 vs. 103.9–106.3), and uniquely captured the significant Group × Train interaction. Influence diagnostics highlighted three influential observations, two belonging to the SSD-CI group, and one to the NH group. Drop-one analyses showed that removing one of the two SSD-CI listeners reduced the Group × Train coefficient from 2.62 ($p = 0.001$) to 1.42 ($p = 0.27$), making it non-significant. In contrast, removing the others preserved a strong and significant interaction (e.g., with the other SSD-CI removed $\beta_4 = 2.99$, $p = 0.0006$, and with the NH removed $\beta_4 = 2.42$, $p = 0.003$). Using cluster-robust standard errors (clustered by subject), the pre-training threshold term maintained significance ($\beta = 1.25$, $p = 0.003$) and the Group × Train interaction remained positive and highly significant ($\beta = 2.62$, $p < 0.001$). Similarly, Huber robust regression confirmed the role of pre-training threshold ($\beta_1 = 1.33$, $p < 0.001$) and retained a positive but attenuated Group × Train interaction ($\beta_4 = 2.55$, $p = 0.035$). Group and Train main effects were not significant in either robustness check.

For the Generalized Linear Model with the Gamma family, convergence under a tolerance of 10^{-8} was reached in fifteen iterations, with a log-likelihood of –70.58, deviation of 4.43, and Pearson's chi-squared of 3.79. Cragg & Uhler's Pseudo R-squared suggests that the model explained approximately 77% of the variability observed in the post-training thresholds. Only the intercept ($\beta_0 = 1.68$, $p = 0.000$) and pre-training threshold ($\beta_1 = 0.30$, $p = 0.004$) were statistically significant predictors, whereas the group, training, and Group × Train interaction terms were not (all $p > 0.06$). Hold-out validation showed RMSE = 1.12, MAE = 0.80, and predictive $R^2 = 0.68$, which numerically outperformed the Gaussian model. However, LOSO cross-validation again revealed poor generalization, with RMSE = 1.32 ± 0.67, MAE = 1.16 ± 0.64, and negative mean predictive R^2 (–3.76 ± 5.84). These results are summarized in Table 4.

Table 4
Comparison of models. Gaussian GLM (identity link) and Gamma GLM (log link) refer to the complete specification M_3 of the formula (1). Cluster-robust estimates are based on Gaussian GLM with subject-level clustering. Huber RLM is a robust regression with Huber weights. In all models, Pre is consistently significant; the Group × Train interaction is positive and significant in Gaussian, cluster-robust, and Huber regressions, but attenuated in Gamma GLM.

Term	Gaussian GLM		Gamma GLM		Cluster-robust		Huber RLM	
	Value	p	Value	p	Value	p	Value	p
β_0	**5.53**	0.000	**1.68**	0.000	**5.53**	0.000	**5.63**	0.000
β_1	**1.25**	0.003	**0.30**	0.004	**1.25**	0.003	**1.33**	0.000
β_2	0.62	0.383	0.10	0.635	0.62	0.383	0.44	0.562
β_3	−0.04	0.908	−0.04	0.629	−0.04	0.908	−0.13	0.706
β_4	**2.62**	0.000	0.61	0.064	**2.62**	0.000	**2.55**	0.035

Mixed-effects models with random intercepts by subject were also estimated to account for the repeated-measures structure of the data. However, estimation proved unstable: the standard optimizer (lbfgs) failed due to singular covariance matrices, and only Powell optimization converged. With a simpler specification without interaction, the pre-training threshold was significant ($\beta_1 = 1.72$, $p < 0.001$) while Group and Train were not, and the random intercept variance was moderate ($\sigma^2 = 0.50$). Including the Group × Train interaction improved the fixed-effect description, with Pre remaining significant ($\beta_1 = 1.36$, $p = 0.001$), while the interaction did not ($\beta_4 = 2.15$, $p = 0.093$). The main effects of the group ($\beta_2 = 0.38$, $p = 0.647$) and train ($\beta_3 = -0.11$, $p = 0.752$) were not significant. Thus, the mixed models converged only under restricted optimization, reproduced the main fixed effect findings of the GLMs, and confirmed the instability of the Group × Train interaction given the small number of subjects and high variability between subjects.

The mean time spent on each app and the mean number of taps across participants and across training sessions are shown in Fig. 12 for both the NH and SSD-CI groups.

We found a statistically significant difference for the NH control group between the interaction times in the pre- and post-training assessments, however, they did not follow a normal distribution ($W = 0.69$, $p = 0.003$, and $W = 0.65$, $p = 0.001$, respectively). Homoscedasticity was met ($W = 0.0$, $p = 0.44$), and the Wilcoxon signed-rank test found a statistically significant difference ($W = 0.0$, $p = 0.016$, effect size 1.0). For the NH control group, the number of taps in the pre- and post-training assessments did follow a normal distribution ($W = 0.92$, $p = 0.45$, and $W = 0.90$, $p = 0.34$, respectively). Homoscedasticity was met ($W = 0.0$, $p = 0.98$), and the t-test found a statistically significant difference ($t = 4.0$, $p = 0.007$, effect size 0.47). Nevertheless, after recomputing the p-values with the False Discovery Rate (FDR) to correct for multiple comparisons, we found that only the number of taps was significantly different (corrected $p = 0.027$).

The pre- and post-training NASA RTLX and UEQ-S responses are shown in Fig. 13 for both the NH and SSD-CI groups. The means and medians across individuals did not differ significantly between the two groups in either questionnaire, nor in the overall perceived

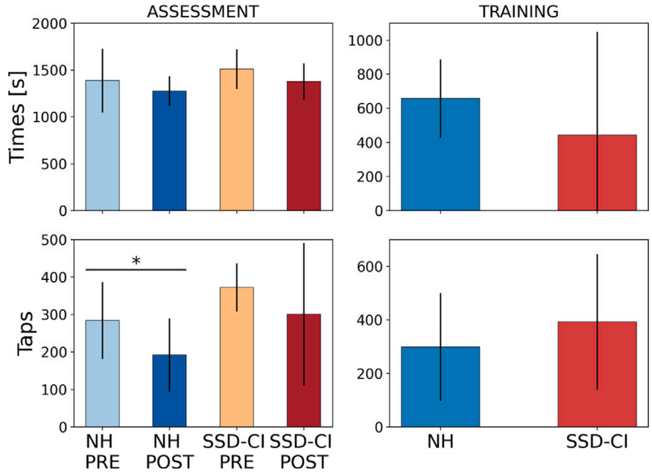


Fig. 12. Mean interaction times (above) and number of taps (below) across individuals (left) and across training sessions (right) for the NH and SSD-CI groups. Asterisks and bars indicate a significant difference (*: p <.05, **: p <.01, ***: p <.001 at post-hoc test).

workload measured with the NASA RTLX. The NASA RTLX in particular signals considerable mental demand and effort; however the UEQ-S questionnaire indicates that the task was perceived as clear and not overly complicated and the activity in general to be exciting and interesting.

Anecdotally, we report in Fig. 14 the results of the questionnaires filled at the end of each training session. 78 sessions were collected, and in 11 of them the questionnaires were not completed. The most frequent training setting was with the tablet as the reproduction device (66.67%), in a medium-sized room (43.59%), with a high volume level (52.56%), and a calm mood (75.64%). By partitioning the data between groups, we observe that 25% of the training games were played by SSD-CI listeners using headphones, compared with 10% for NH participants using the same reproduction devices. Another relevant difference is that almost three-quarters of the training game sessions were played by SSD-CI listeners at a high volume level, while NH listeners generally preferred a more uniformly distributed volume level.

8. General discussion

The system presented in this work gradually took shape from negotiations between three domains that rarely intersect directly: psychophysical assessment, child-computer interaction design, and ecological hearing conditions. Therefore, the development process did not simply implement a predefined clinical protocol but progressively translated methodological constraints into interactional forms that children could reliably engage with. Examining episodes in which this translation succeeded or failed reveals how the final training—assessment configuration stabilized across these domains.

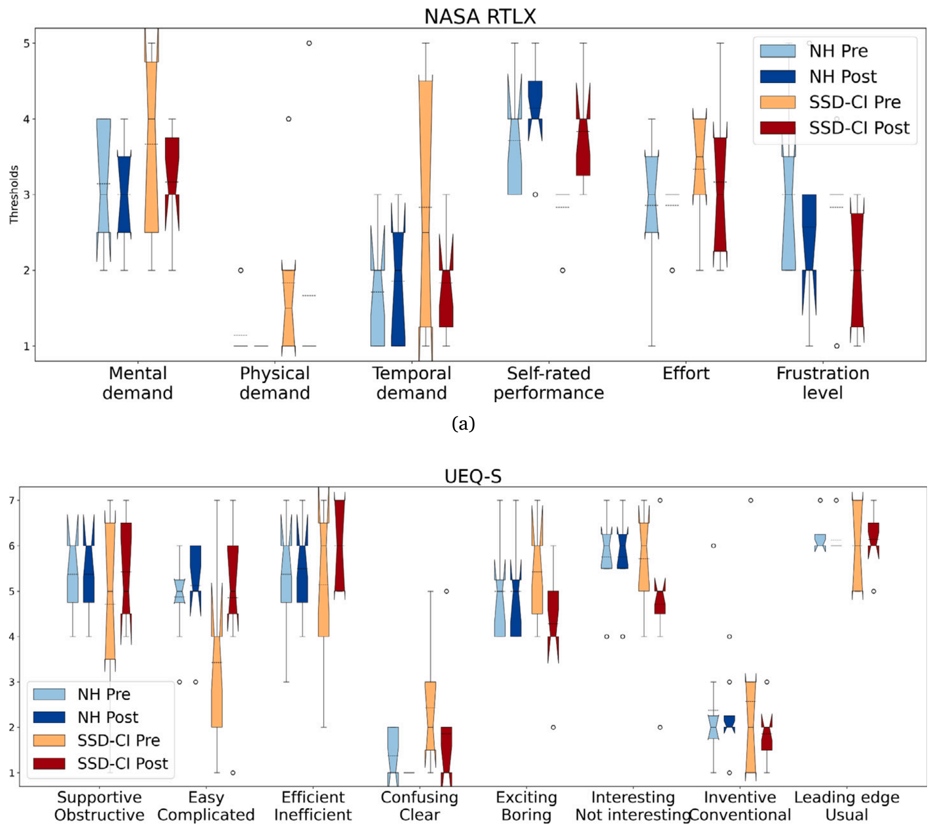
One point of tension concerned the clinical validity of the training activity. Pitch perception research offers multiple experimental approaches, including identification and discrimination paradigms. Regardless of the task structure, however, equal-loudness matching remained necessary to prevent intensity cues from influencing pitch judgments. Early pilot sessions implemented a manual loudness-balancing procedure in which users adjusted stimulus levels themselves. While methodologically correct, this approach led to hesitation and fatigue during sessions with children. The clinical constraint itself could not be removed, yet its manual implementation proved difficult to sustain for pediatric participants: conventional psychoacoustic experiments may last around one hour or longer (Brockmeier et al., 2011; Litovsky et al., 2006), whereas sustained attention in school-age children during cognitively demanding tasks typically ranges between approximately 20 and 40 min depending on age (Mahone & Schneider, 2012; Simon et al., 2023). The introduction of a data-driven active learning procedure for loudness contour estimation enabled calibration with a limited number of pairwise comparisons while preserving equal-loudness conditions. In this case, stabilization occurred not by removing the psychophysical constraint but by redistributing its enactment: calibration remained grounded in listeners' judgments, while the active learning procedure progressively estimated the loudness contour and reduced the manual adjustment required.

Other negotiations concerned the translation of psychophysical structure into visual interaction metaphors. The discrimination procedure required the sequential presentation of two acoustic stimuli under symmetrical visual conditions so that spatial cues would not influence pitch judgments. Early design explorations attempted to embed this structure within more dynamic interfaces inspired by arcade games. These prototypes proved incompatible with the neutrality required by the experimental task: visual motion patterns and asymmetrical layouts risked introducing perceptual biases (Marks et al., 2003; Spence, 2011). Therefore, the final marble-based side-scrolling interaction represented a compromise between the requirements of psychophysical rigor and the need to maintain engagement. Moving objects provided temporal structure and reward-based feedback through collectible items and score increments, while the visual arrangement preserved symmetry so that the discrimination task remained unbiased. In this sense, the interface did not merely "gamify" the experiment but translated a controlled listening procedure into an interactive format capable of maintaining engagement across repeated trials, consistent with evidence that game-based interaction structures can sustain user motivation and participation in structured tasks (Deterding et al., 2011; Hamari et al., 2014).

A similar translation occurred in the design of the training activities. The rainbow keyboard interface drew simultaneously from musical keyboards and spectrum visualizations—representations familiar to clinicians and audio researchers. Although these metaphors were not initially codesigned with children, their usability emerged gradually through iterative testing sessions in which clinicians and therapists evaluated whether the mapping remained understandable for younger participants. During early assessment sessions, several children spontaneously hummed tones while performing discrimination trials. Such behavior suggested that participants were internally rehearsing pitch relations, indicating that the interface's visual and temporal organization supported task comprehension. At the same time, hesitation and reflective response patterns revealed that the discrimination procedure required sustained cognitive effort. These observations led to the introduction of complementary training activities incorporating more active interaction structures, including the rhythmic distractor game. In this instance, the design trajectory was shaped less by theoretical preference than by the practical need to counterbalance the cognitive demands of the assessment task.

Not all design trajectories converged toward the final system. One example concerned the possibility of delivering stimuli directly through the CI audio input. From a signal processing perspective, this option would have provided tighter experimental control over the signals reaching the implant. However, direct input conflicted with the project's objective of maintaining ecological continuity between clinical assessment and everyday listening conditions. It would have introduced additional hardware dependencies and listening configurations that children do not typically encounter outside laboratory environments. Although technically feasible, the option was therefore abandoned in favor of acoustic playback, which preserved the everyday auditory ecology of SSD-CI listeners. This decision illustrates how technically sound directions may nevertheless conflict with broader deployment objectives.

Negotiation also occurred at the level of listening environments. Clinical CI testing often employs open-field loudspeaker presentation rather than headphones. During home training sessions, however, some



(a)

(b)

Fig. 13. NASA RTLX (above) and UEQ-S (below) pre- and post-training for the NH and SSD-CI groups.

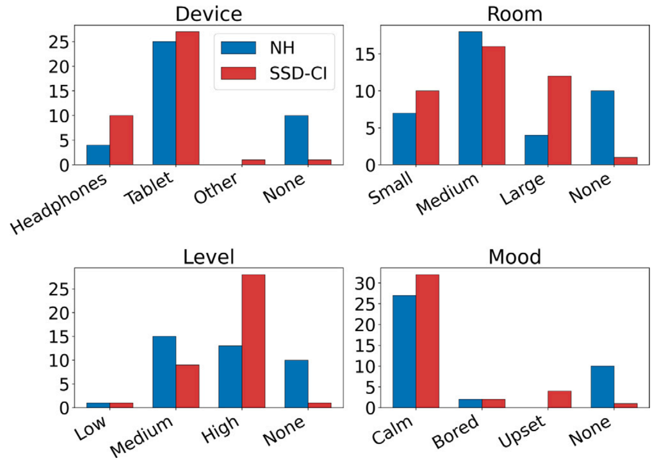


Fig. 14. Results of the questionnaires prompted at the end of each training game for each group.

participants spontaneously chose to use headphones (see Fig. 14). This behavior was unexpected but revealed how children interact with auditory stimulation in everyday digital contexts, exposing a dimension of ecological variability that could only emerge outside the clinic. Rather than eliminating this variability, the training application accommodated it while the assessment procedure maintained stricter presentation constraints to preserve measurement reliability.

Another challenge concerned the perceptual gap between designers and CI users. CI hearing differs substantially from normal auditory perception, making it difficult for researchers to anticipate how acoustic stimuli will be represented through the implant. Two complementary strategies were used to reduce this gap. A CI simulator (McDermott et al., 1992) informed stimulus selection in the distractor task, ensuring that they could be distinguished after simulated CI processing. At the same time, behavioral traces observed during pilot sessions provided indirect insight into children's perceptual engagement.

Pre-training electric pitch discrimination thresholds presented a variance of three-quarters of an octave (Brockmeier et al., 2011) and a mean slightly below one octave. These results reveal the reliability of our pitch discrimination app when compared to thresholds found in the literature (Brockmeier et al., 2011; Bruns et al., 2016). The training halved the SSD-CI threshold mean and reduced the variance to a little less than five semitones, both results being valuable. Similarly, the NH listeners' thresholds fall within values already reported in the literature, i.e., between one and two semitones using pure tones as stimuli (Zarate et al., 2012). We note that no participant was attending a musical education program at the time of the test; just one NH listener had prior musical training. Nevertheless, there was a significant improvement in the control group as well.

SSD-CI participants showed higher pre- and post-training thresholds than NH listeners, confirming their difficulty in discriminating pitch. This discrepancy between the two groups persisted also after the training, highlighting the continuous challenge of the former group to deal with the task. Nevertheless, the auditory training significantly reduced this gap; we observed a 15.56% reduction in the effect size of the difference between NH listeners' pre-training thresholds and SSD-CI participants' thresholds following the training. The plot in Fig. 11(b) shows that most observations are below the diagonal, indicating a group-level decreasing trend. Importantly, individual responses in SSD-CI were heterogeneous, with a few marked worsenings offsetting many smaller improvements. While the descriptive analyses and boxplots suggested a general post-training reduction in thresholds, the regression analyses provide a more nuanced view.

Regression analyses further clarified these patterns. The Generalized Linear Models highlight the relationship between pre-training thresholds and post-training thresholds. As displayed in Fig. 11(b), the post-training threshold is an increasing function of the pre-training threshold. Half of the NH listeners' thresholds and almost three-quarters of the SSD-CI listeners' thresholds decreased. In the regression framework, pre-training ability emerged as the most consistent predictor of post-training thresholds, while main effects of training and group were not significant. The Group × Train interaction was significant in the Gaussian models, but robustness checks and sensitivity analyses showed that this effect was fragile and strongly influenced by a single SSD-CI participant. Instead, the results highlight marked heterogeneity in SSD-CI responses, with some children improving markedly and others less clearly. The Gamma model offered comparable fit statistics in-sample but failed to generalize across subjects, suggesting caution in interpreting training efficacy. Together, these findings support the use of personalized auditory training approaches aligned with principles of precision rehabilitation. In the present system personalization was introduced selectively: automated adaptation was applied where it improved measurement efficiency — such as individualized loudness calibration and adaptive trial sequencing — while training interactions preserved opportunities for exploratory manipulation by children (Hourcade, 2022). This balance allowed the system to maintain psychophysical validity while preserving interaction freedom during training activities. For instance, recent work in CI research emphasizes the shift toward individualized diagnostics and programming to overcome the limitations of one-size-fits-all approaches (e.g., the precision medicine approach to CI care (Kim & Choi, 2022)). Moreover, in the case of single-sided deafness with CI, tailored frequency mappings and adaptive training regimens have been explicitly recommended to improve consistency and user satisfaction (Lin et al., 2025). The notion of an "auditory digital twin" further underscores how individualized computational models could help us track and predict each user's unique auditory trajectory (Geronazzo & Serafin, 2023). Although many CI outcome studies rely on group-level statistics, the broader literature on hearing devices and HCI/CCI methods suggests that mixed-methods designs and participatory personalization strategies can help translate group trends into robust, individual gains (e.g., by capturing qualitative insights, tailoring to idiosyncratic user goals, and dynamically adjusting training) (Baykal et al., 2023).

The above considerations warrant further analysis of the spectral characteristics of the stimuli after simulated CI post-processing (McDermott et al., 1992). As the example in Fig. 15 shows, two distinct peaks belonging to stimuli with different center frequencies are turned by the CI processing into "comb" components matching identical frequency bins however with varying magnitude differences across the same bins. The main effort during electric pitch decoding, hence, would become that of discriminating such differences and mapping them into a definite monaural pitch cue. This consideration pairs with previous findings showing that CI listeners rely mostly on intensity cues both during speech comprehension (Peng et al., 2012) and acoustic source localization (Gulli et al., 2024). Training binaural pitch perception may therefore enrich the SSD-CI listeners' auditory memory with a more robust imagery for pitch, linking electric to acoustic cues through ordered elements that may be identified with the combs and their individual magnitude.

The results from questionnaires suggest that gamification was beneficial for the training. Scores on the NASA RTLX suggested moderate mental demand and manageable effort, while UEQ-S responses indicated positive evaluations of both stimulation and usability. Informal observations during the sessions further revealed strong engagement behaviors: we noticed children compared scores across rounds, expressed frustration after incorrect responses, and celebrated successful trials. No participant reported perceiving the training as burdensome.

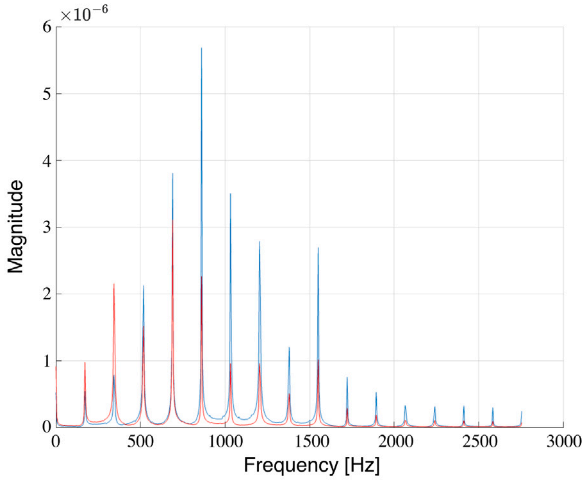


Fig. 15. Spectral magnitude of two stimuli with center frequencies respectively equal to 400 Hz and 800 Hz, after simulated CI post-processing (red and blue lines, respectively).

These reactions suggest that motivational structures embedded in the interaction enabled children to complete repeated psychophysical trials without perceiving the activity as burdensome.

Participatory processes nevertheless generated forms of mutual learning between designers and participants. In rehabilitation contexts such as cochlear implantation, this exchange can be particularly significant: training activities expose children to perceptual challenges, while iterative design enables researchers to adapt tasks and interfaces accordingly. The resulting ecosystem therefore functioned not only as a rehabilitation intervention but also as an investigative instrument through which perceptual understanding and interaction design co-evolved, echoing research-through-design approaches to sonic interaction design (Privitera & Geronazzo, 2024).

Following this design process, the training results provide evidence that the stabilized configuration achieved its primary objective: enabling children to perform repeated pitch discrimination trials without abandoning psychophysical validity. Despite these improvements, SSD-CI participants continued to exhibit higher thresholds than NH listeners, confirming the persistent perceptual challenges associated with electric pitch perception.

8.1. Limitations

The primary limitation of the current study lies in the low number of SSD-CI participants. However, even achieving this sample size was a challenge due to the fairly limited and progressively decreasing availability of SSD-CI individuals for experiments, also caused by the growing number of bilateral CI implantations in children (Nassiri et al., 2023). Moreover, the audiometric data of each participant is not reported, as is the hearing pathology. These data could have been used to partition the SSD-CI group and check if the pitch discrimination enhancement is linked to a specific hearing disability. Other hearing impairments may lead to a personalized training procedure.

A second limitation concerns the interpretability of individual differences. Training exposure was not captured with sufficient granularity, which may have obscured part of the variability in individual responsiveness. More generally, because the intervention combined several elements — adaptive calibration, repeated discrimination assessment, ecological home deployment, and game-based interaction — the present study cannot isolate the specific contribution of each factor to the observed improvements. In particular, it remains unclear how much of the outcome was associated with the training activity itself, with repeated task exposure, or with differences in how participants engaged with the home-based setup.

An in-depth study is needed to evaluate the active learning loudness balancing procedure by comparing its speed, effectiveness, and reliability to alternative clinical procedures. Psychophysical constraints also represent a limitation: we compromised between accuracy and session time by limiting the adaptive procedure to seven iterations per stimulus, a choice that may have increased measurement error in children. It remains to be verified whether longer adaptive sequences could be used in pediatric settings without exceeding children's typical attention spans, which often place practical constraints on psychophysical testing (Witton et al., 2017). More in general, the advantages and limits of data-driven methods in audiology require further investigation.

Regarding the sonic and graphical design, synchronized multimodal elements could have been improved semantically. Although the acoustic stimulus had ecological validity and no participant reported issues with it, it is unlikely that a marble jumping on a sand track would produce such a sharp impulsive sound. A different design choice (e.g., an iced track in a winter scenario) might have provided a more semantically congruent experience.

9. Conclusion and future directions

This study presented a training—assessment ecosystem for pitch discrimination in children with single-sided deafness and a cochlear implant, showing how clinically grounded psychophysical procedures can be translated into child-centered interactive systems deployable in everyday environments.

The two mobile applications offer a user-friendly and effective means of assessing and improving pitch perception through intuitive game mechanics, Gaussian processes for rapid estimation of equal-loudness contours, and adaptive psychometric techniques. Experimentation with NH participants and children with a single-sided cochlear implant demonstrated the feasibility of the approach and provided initial evidence that such systems can support both clinically interpretable assessment and repeated auditory training in pediatric populations. The measured thresholds align with established literature, suggesting the potential for the testing app to deliver accurate assessments. Post-training measurements indicate reductions in discrimination thresholds in both groups, although individual responses were heterogeneous, reflecting the variability typical of cochlear implant outcomes.

A key insight of this work is that game-based interaction structures can make demanding psychophysical procedures feasible for children without altering the methodological validity of the assessment. Reward mechanisms, exploratory interaction, and multimodal representations were deliberately designed not simply to increase engagement, but to sustain repeated perceptual trials while preserving the fixed stimulus presentation and response paradigms required for reliable threshold estimation. In this sense, gamification functioned as an enabling interface layer supporting rigorous auditory experimentation in pediatric contexts.

Future endeavors will prioritize a dedicated examination of the loudness balancing procedure and may consider the use of additional stimuli, such as the pulse-spreading harmonic complex (Mesnildrey et al., 2016), in the pitch discrimination test tailored for CI users. In particular, further work is needed to evaluate the speed, reliability, and clinical comparability of the active learning loudness calibration procedure relative to established audiological protocols. Furthermore, integrating our experimental approach with pre- and post-training audiometric assessments using speech stimuli would offer a more comprehensive evaluation of the efficacy and utility of the developed applications (Fu et al., 2004). From a CCI perspective, future work should also emphasize greater personalization of training trajectories to account for the marked inter-individual variability of pediatric CI outcomes, and adopt mixed-methods approaches that integrate clinical metrics with qualitative insights from children, families, and therapists. This dual strategy would allow us to go beyond group-level results and better capture the ecological validity and lived experiences of young CI users.

A detailed understanding of how improvements in pitch discrimination transfer to effective communication skills is essential. A speech-in-noise test campaign (Zaltz et al., 2020) following the training may shed light on the broader benefits of improved pitch discrimination. Moreover, observing whether participants continue to use the training application after the experimental period could provide insight into medium-term retention of perceptual improvements and the role of continued training in maintaining stable performance levels.

Finally, future work will address the varying needs and experiences of different user groups, as suggested by differences in subjective workload and user experience observed among participants. Further research is therefore required to refine both the auditory training paradigm and its interaction design, considering users' feedback and the broader complexities of auditory perception with cochlear implants.

CRediT authorship contribution statement

Andrea Gulli: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Michele Geronazzo:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Federico Fontana:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Hanna Järveläinen:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis. **Enrico Muzzi:** Validation, Supervision. **Roberta Rebesco:** Supervision, Methodology. **Eva Orzan:** Validation, Supervision, Resources.

Ethical approval

The test protocol was approved on August 31, 2023 by the Institutional Review Board of the Department of Mathematics, Computer Science, and Physics at the University of Udine (Ref. 2023-014-IRB-DMIF-Gulli). . All participants were recruited at the multidisciplinary rehabilitation and training center for children and adults Studio Riabilitare Insieme (Maerne (VE), Italy). Written informed consent was obtained for each participant. In case they were minors, it was obtained from their parents.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors used ChatGPT (GPT-5, OpenAI) to support the preparation of this manuscript. Specifically, it was employed to improve the clarity and fluency of text, to assist in drafting responses to reviewers, and to check consistency across revisions. The AI was not used for data analysis, experimental design, or generation of scientific results. All content was critically reviewed, edited, and approved by the authors, who take full responsibility for the manuscript.

Funding

This work was supported by the NextGenerationEU PNRR M4-C2.1.1 PRIN 2022 project “S-TWIN” (Grant No. 2022F9FWZ8_003, CUP G53D23002840006), and by the Italian Ministry of Health, Rome, Italy, in collaboration with the Institute for Maternal and Child Health IRCCS “Burlo Garofolo” (Trieste, Italy) (RC 42/22, CUP C93C25006270001).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Andrea Gulli, Roberta Rebesco, Michele Geronazzo, and Federico Fontana report research funding from the European Union and the Italian Ministry of Health. The authors declare no other competing interests.

Appendix A. Fast loudness balancing procedure

The regression function estimating the loudness contour is modeled by a multivariate Gaussian distribution:

$$p(\mathbf{f}|\mathbf{X}) = \mathcal{N}\left(\mathbf{f}\,|\,\mu(\mathbf{X}), K(\mathbf{x},\mathbf{x}')\right), \tag{A.1}$$

where $\mathbf{X} = (\mathbf{x}_1,\dots,\mathbf{x}_N)$ represents the observed data points, $\mathbf{f} = (\mathrm{f}(\mathbf{x}_1),\dots,\mathrm{f}(\mathbf{x}_N))$ the function values, μ the mean function, and $K(\mathbf{x},\mathbf{x}')$ the kernel, which is a positive definite function. From observed data points and an estimated mean function $\mathbf{f}$, a new estimation is made over the points $\mathbf{X}_*$: $\mathbf{f}_* = \mathrm{f}(\mathbf{X}_*)$. The joint distribution of $\mathbf{f}$ and $\mathbf{f}_*$ is expressed as:

$$\begin{bmatrix}\mathbf{f}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mu(\mathbf{X})\\ \mu(\mathbf{X}_*)\end{bmatrix}, \begin{bmatrix}K(\mathbf{X},\mathbf{X}) & K(\mathbf{X},\mathbf{X}_*)\\ K(\mathbf{X}_*,\mathbf{X}) & K(\mathbf{X}_*,\mathbf{X}_*)\end{bmatrix}\right), \tag{A.2}$$

whose mean is assumed to be an array of zeros. From the joint probability distribution $p\left(\mathbf{f},\mathbf{f}_*|\mathbf{X},\mathbf{X}_*\right)$ expressed by the previous equation, the conditional distribution can be analytically computed (Rasmussen & Williams, 2019):

$$p\left(\mathbf{f}_*|\mathbf{f},\mathbf{X},\mathbf{X}_*\right) \sim \mathcal{N}\left(\mathbf{K}_*^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{f}, \mathbf{K}_{**} - \mathbf{K}_*^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{K}_*\right), \tag{A.3}$$

with $\mathbf{K} = K(\mathbf{X},\mathbf{X})$, $\mathbf{K}_* = K(\mathbf{X},\mathbf{X}_*)$, and $\mathbf{K}_{**} = K(\mathbf{X}_*,\mathbf{X}_*)$. Taking into account that the observations could be noisy, the joint distribution of the observed values and the function values at new testing points is:

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\mathbf{K}+\sigma_n^2\mathbf{I} & \mathbf{K}_*\\ \mathbf{K}_*^{\mathrm{T}} & \mathbf{K}_{**}\end{bmatrix}\right), \tag{A.4}$$

where σ_n^2 represents the variance of additive independent and identically distributed (i.i.d.) Gaussian noise. The predictive equations for the Gaussian regression are:

$$p\left(\mathbf{f}_*|\mathbf{X},\mathbf{y},\mathbf{X}_*\right) \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, \mathrm{cov}(\mathbf{f}_*)\right), \tag{A.5}$$

where

$$\bar{\mathbf{f}}_* = \mathbf{K}_*^{\mathrm{T}}[\mathbf{K}+\sigma_n^2\mathbf{I}]^{-1}\mathbf{y}, \tag{A.6}$$

$$\mathrm{cov}(\mathbf{f}_*) = \mathbf{K}_{**} - \mathbf{K}_*^{\mathrm{T}}[\mathbf{K}+\sigma_n^2\mathbf{I}]^{-1}\mathbf{K}_*. \tag{A.7}$$

The Gaussian process employed a constant mean function (μ) and a Matérn covariance function with a ν parameter. The Matérn kernel belongs to a flexible class that generalizes the Radial Basis Function kernel. It is characterized by two hyperparameters, the length scale l and the smoothness parameter ν, and it is defined as:

$$K(\mathbf{x},\mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|\mathbf{x}-\mathbf{x}'|}{l}\right)^{\nu} J_\nu\left(\frac{\sqrt{2\nu}|\mathbf{x}-\mathbf{x}'|}{l}\right), \tag{A.8}$$

where J_ν denotes the modified Bessel function of the second kind. The hyperparameters μ and ν are set respectively to zero and 2.5, while the length scale is determined through likelihood maximization at each iterative step. σ_n^2 is set to 0.005. The predictions and the log marginal likelihood to be maximized are computed with the following algorithm.

Algorithm 1 Predictions and log marginal likelihood for Gaussian process regression.

```
Input: $\mathbf{X},\mathbf{y}$ (observations), $K$ (covariance function), $\sigma_n^2$ (noise level), $\mathbf{X}_*$ (new data point)
    $L = \mathrm{cholesky}(\mathbf{K}+\sigma_n^2\mathbf{I})$
    $\alpha = L^{\mathrm{T}}\backslash(L\backslash\mathbf{y})$
    $\bar{\mathbf{f}}_* = \mathbf{K}_*^{\mathrm{T}}\alpha$
    $\mathbf{v} = L\backslash\mathbf{K}_*$
    $\mathbb{V}[\bar{\mathbf{f}}_*] = K(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{v}^{\mathrm{T}}\mathbf{v}$
    $\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\mathbf{K}+\sigma_n^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log\det(\mathbf{K}+\sigma_n^2\mathbf{I}) - \frac{n}{2}\log 2\pi$
    $\bar{\mathbf{f}}_*$ (mean), $\mathbb{V}[\bar{\mathbf{f}}_*]$ (variance), $\log p(\mathbf{y}|\mathbf{X})$ (log marginal likelihood)
```

Algorithm 2 Fast estimation of equal-loudness contours

```
Input: stimulus at MCL = $\{x_T, y_T\}$
    Initialize: $\{x_0, y_0\} = \{x_T, y_T\}$
    $1 \le i \le itermax$ Find the min RMSE Fletcher–Munson: $\mathbf{X}_{FM}, \mathbf{y}_{FM}$
    Define observed data: $\mathbf{X},\mathbf{y} = \{\mathbf{X}_{FM}, \mathbf{y}_{FM}\} \bigcup_{j=0}^{i-1}\{x_j, y_j\}$
    Algorithm 1
    LBFGS Minimize negative log marginal likelihood
    Algorithm 1 with optimal hyperparameter $l$
    Compute mutual information MU from $\mathbb{V}[\bar{\mathbf{f}}_*]$ and $\log p(\mathbf{y}|\mathbf{X})$
    Random selection from $\ge 90\%$ MU points: $x_i$
    Present the two stimuli
    Get the adjusted level $\{x_i, y_i\}$
    Compute the mean $\bar{\mathbf{f}}_*$  $\bar{\mathbf{f}}_*$
```

Appendix B. Selection and participation

Participants in this study were children and young individuals with single-sided deafness and a cochlear implant (SSD-CI), as well as a control group of normal-hearing (NH) young listeners. Seven SSD-CI participants (2 female, 5 male, mean age = 14.57 ± 6.21 years) and eight NH participants (3 female, 5 male, mean age = 13.88 ± 3.59 years) were recruited from a multidisciplinary rehabilitation and training center. All SSD-CI participants had at least 11 months of experience using their cochlear implant.

For recruitment, a video presentation explaining the study's scope, procedures, and goals was shown to potential participants and their families. Informed consent was then obtained from all participants or their legal guardians in the case of minors. Ethical approval was granted by the Institutional Review Board of the Department of Mathematics, Computer Science, and Physics at the University of Udine. The signed informed consent forms are securely stored and will be preserved for a period of five years, as required by ethical guidelines.

Throughout the study, strict data privacy measures were maintained. All data collected during both the supervised assessments and the home-based training sessions were stored exclusively on local devices, with no transmission to external servers or cloud storage. Participants were informed that their data would remain confidential and used solely for research purposes.

During the study, participants engaged in pitch discrimination assessments and home-based training using mobile apps designed to be engaging and non-intrusive. Therapists and parents were involved in facilitating the process, particularly for younger children, ensuring a supportive and ethical research environment. Ethical considerations, including minimizing cognitive load and maintaining a positive user experience, were prioritized throughout the study.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijcci.2026.100823.

Data availability

Data will be made available on request.

References

Adel, Y., Nagel, S., Weissgerber, T., Baumann, U., & Macherey, O. (2019). Pitch matching in cochlear implant users with single-sided deafness: Effects of electrode position and acoustic stimulus type. *Frontiers in Neuroscience*, *13*, 1119. http://dx.doi.org/10.3389/fnins.2019.01119.

Agrawal, S. R., & Bech, S. (2023). Immersion in audiovisual experiences. *Sonic interactions in virtual environments* (pp. 319–351). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-031-04021-4_11,

Almirall, D., & Chronis-Tuscano, A. (2016). Adaptive interventions in child and adolescent mental health. *Journal of Clinical Child & Adolescent Psychology*, *45*, 383–395. http://dx.doi.org/10.1080/15374416.2016.1152555, arXiv:https://doi.org/10.1080/15374416.2016.1152555, pMID: 27310565.

Antle, A. N. (2013). Research opportunities: Embodied child-computer interaction. *International Journal of Child-Computer Interaction*, *1*, 30–36. http://dx.doi.org/10.1016/j.ijcci.2012.08.001, URL https://www.sciencedirect.com/science/article/pii/S2212868912000025.

Aramaki, M., Besson, M., Kronland-Martinet, R., & Ystad, S. (2011). Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*, 301–314. http://dx.doi.org/10.1109/tasl.2010.2047755.

Arzounian, D., de Kerangal, M., & de Cheveigné, A. (2017). A sliding two-alternative forced-choice paradigm for pitch discrimination. *Journal of the Acoustical Society of America*, *142*, 167–172. http://dx.doi.org/10.1121/1.4992030.

Bälter, O., Engwall, O., Öster, A.-M., & Kjellström, H. (2005). Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction. In *Proceedings of the 7th international ACM SIGACCESS conference on computers and accessibility* (pp. 36–43). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1090785.1090795.

Baykal, G. E., Eriksson, E., & Torgersson, O. (2023). Assessment of learning in child-computer interaction research: A semi-systematic literature review. *International Journal of Child-Computer Interaction*, *36*, Article 100578. http://dx.doi.org/10.1016/j.ijcci.2023.100578, URL https://www.sciencedirect.com/science/article/pii/S2212868923000156.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *57*, 289–300. http://dx.doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Benton, L., Johnson, H., Ashwin, E., Brosnan, M., & Grawemeyer, B. (2012). Developing IDEAS: Supporting children with autism within a participatory design team. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2599–2608). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2207676.2208650.

Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *Journal of the Acoustical Society of America*, *122*, 1765–1776. http://dx.doi.org/10.1121/1.2751394.

Brinkmann, P., Wilcox, D., Kirshboim, T., Eakin, R., & Alexander, R. (2016). Libpd: Past, present, and future of embedding pure data. URL http://www.danomatika.com/publications/libpd_pdcon_16.pdf, Presented at the 5th Pure Data Convention (Pd˜Con).

Brockmeier, S. J., Fitzgerald, D., Searle, O., Fitzgerald, H., Grasmeder, M., Hilbig, S., Vermiere, K., Peterreins, M., Heydner, S., & Arnold, W. (2011). The MuSIC perception test: A novel battery for testing music perception of cochlear implant users. *Cochlear Implants International*, *12*, 10–20. http://dx.doi.org/10.1179/146701010x12677899497236.

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*, 364–367. http://dx.doi.org/10.1080/01621459.1974.10482955.

Bruns, L., Mürbe, D., & Hahne, A. (2016). Understanding music with cochlear implants. *Scientific Reports*, *6*, 32026. http://dx.doi.org/10.1038/srep32026.

Calderon, R., & Greenberg, M. T. (2012). *Social and emotional development of deaf children: Family, school, and program effects* (pp. 188–199). UK: Oxford University Press, http://dx.doi.org/10.1093/oxfordhb/9780199750986.013.0014.

Cambridge, G., Taylor, T., Arnott, W., & Wilson, W. J. (2022). Auditory training for adults with cochlear implants: A systematic review. *International Journal of Audiology*, *61*, 896–904. http://dx.doi.org/10.1080/14992027.2021.2014075.

Çetinkaya, M. Meral, Konukseven, Ö., & İralı, A. E. (2024). World of sounds (Seslerin Dünyası): A mobile auditory training game for children with Cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *179*, Article 111908. http://dx.doi.org/10.1016/j.ijporl.2024.111908, URL https://www.sciencedirect.com/science/article/pii/S0165587624000624.

Chatziioannou, V., & van Walstijn, M. (2015). Energy conserving schemes for the simulation of musical instrument contact dynamics. *Journal of Sound and Vibration*, *339*, 262–279. http://dx.doi.org/10.1016/j.jsv.2014.11.017.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*, 15–18. http://dx.doi.org/10.1080/00401706.1977.10489493.

Cox, R. M., Alexander, G. C., Taylor, I. M., & Gray, G. A. (1997). The contour test of loudness perception. *Ear and Hearing*, *18*, 388–400. http://dx.doi.org/10.1097/00003446-199710000-00004.

Cuturi, L. F., Tonelli, A., Cappagli, G., & Gori, M. (2019). Coarse to fine audio-visual size correspondences develop during primary school age. *Frontiers in Psychology*, *10 - 2019*, http://dx.doi.org/10.3389/fpsyg.2019.02068, URL https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.02068.

Deroche, M. L. D., Lu, H.-P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S.-C., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2019). A Tonal-Language benefit for pitch in normally-hearing and Cochlear-Implanted children. *Scientific Reports*, *9*, 109. http://dx.doi.org/10.1038/s41598-018-36393-1.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "Gamification". In *Proceedings of the 15th international academic mindTrek conference: envisioning future media environments* (pp. 9–15). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2181037.2181040.

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, *25*, 1256–1261. http://dx.doi.org/10.1177/0956797614528521.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology*, *21*, 1–25. http://dx.doi.org/10.1080/01449290110108659, arXiv:https://doi.org/10.1080/01449290110108659.

Frauenberger, C., Makhaeva, J., & Spiel, K. (2016). Designing smart objects with autistic children: Four design exposès. *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 130–139). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2858036.2858050, URL https://doi.org/10.1145/2858036.2858050.

Fu, Q.-J., Galvin, J., Wang, X., & Nogaki, G. (2004). Effects of auditory training on adult cochlear implant patients: A preliminary report. *Cochlear Implants International*, *5*, 84–90. http://dx.doi.org/10.1179/cim.2004.5.Supplement-1.84, arXiv:https://doi.org/10.1179/cim.2004.5.Supplement-1.84, pMID: 18792249.

Galvin, J. J., Fu, Q.-J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing*, *28*, 302–319. http://dx.doi.org/10.1097/01.aud.0000261689.35445.20.

Garcia, B. E. R., Crocomo, M. K., & Andrade, K. O. (2018). Dynamic difficulty adjustment in a whac-a-mole like game. *2018 17th Brazilian symposium on computer games and digital entertainment (sBGames)* (pp. 439–447). Rio Grande (RS), Brazil: IEEE, http://dx.doi.org/10.1109/sbgames.2018.00020.

Gaver, W. W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, *5*, 1–29. http://dx.doi.org/10.1207/s15326969eco0501_1.

Geronazzo, M., & Serafin, S. (2023). Sonic interactions in virtual environments: The egocentric audio perspective of the digital twin. In M. Geronazzo, & S. Serafin (Eds.), *Sonic interactions in virtual environments* (pp. 3–45). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-031-04021-4_1.

Gockel, H. E., Moore, B. C. J., & Carlyon, R. P. (2020). Pitch perception at very high frequencies: On psychometric functions and integration of frequency information. *Journal of the Acoustical Society of America*, *148*, 3322–3333. http://dx.doi.org/10.1121/10.0002668.

Gordon, K., Henkin, Y., & Kral, A. (2015). Asymmetric hearing during development: The aural preference syndrome and treatment options. *Pediatrics*, *136*, 141–153. http://dx.doi.org/10.1542/peds.2014-3520.

Green, D. M., Swets, J. A., et al. (1966). *Signal detection theory and psychophysics: vol. 1*, Wiley New York.

Gulli, A., Fontana, F., Aruffo, A., Orzan, E., & Muzzi, E. (2024). Previous binaural experience supports compensatory strategies in hearing-impaired children's auditory horizontal localization. *PLoS One*, *19*, 1–25. http://dx.doi.org/10.1371/journal.pone.0312073.

Gulli, A., Fontana, F., Serafin, S., Geronazzo, M., et al. (2023). An active learning procedure for the interaural time difference discrimination threshold. In *Proceedings of the international conference on digital audio effects* (pp. 273–280). Copenhagen, Denmark: DAFx, DAFx Aalborg University Copenhagen, URL https://www.dafx.de/paper-archive/2023/DAFx23_paper_39.pdf.

Gupta, A., Mishra, P., Pandey, C., Singh, U., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, *22*, 67. http://dx.doi.org/10.4103/aca.aca_157_18.

Halbhuber, D., Köhler, A., Schmidbauer, M., Wiese, J., & Henze, N. (2022). The effects of auditory latency on experienced first-person shooter players. In *Proceedings of mensch und computer 2022* (pp. 286–296). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3543758.3543760, URL https://doi.org/10.1145/3543758.3543760.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification work? – A literature review of empirical studies on gamification. In *2014 47th hawaii international conference on system sciences* (pp. 3025–3034). http://dx.doi.org/10.1109/HICSS.2014.377.

Han, J.-H., Lee, J., & Lee, H.-J. (2021). Ear-Specific hemispheric asymmetry in unilateral deafness revealed by auditory cortical activity. *Frontiers in Neuroscience*, *15*, Article 698718. http://dx.doi.org/10.3389/fnins.2021.698718.

Hart, S. G. (2006). Nasa-Task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 904–908. http://dx.doi.org/10.1177/154193120605000909.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Advances in psychology*: *Vol. 52*, *Human mental workload* (pp. 139–183). Amsterdam: North-Holland, http://dx.doi.org/10.1016/S0166-4115(08)62386-9, URL https://www.sciencedirect.com/science/article/pii/S0166411508623869.

Henshaw, H., & Ferguson, M. (2015). How does auditory training work? Joined-Up thinking and listening. *Seminars in Hearing*, *36*, 237–249. http://dx.doi.org/10.1055/s-0035-1564456.

Hermus, K., Verhelst, W., Lemmerling, P., Wambacq, P., & Huffel, S. Van (2005). Perceptual audio modeling with exponentially damped sinusoids. *Signal Processing*, *85*, 163–176. http://dx.doi.org/10.1016/j.sigpro.2004.09.010.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading. MA: Addison-Wesley (Content Analysis).

Hourcade, J. P. (2022). *Self-published*, *Child-Computer Interaction* (2nd ed.). URL http://homepage.cs.uiowa.edu/~hourcade/book/index.php. (Accessed 11 September 2025).

Huang, F., & Li, X. (2021). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods*, *54*, http://dx.doi.org/10.3758/s13428-021-01627-0.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*, 73–101. http://dx.doi.org/10.1214/aoms/1177703732.

International Telecommunication Union (2015). *Method for the subjective assessment of intermediate quality level of audio systems (MUSHRA)*: *Recommendation ITU-R BS.1534-3*, International Telecommunication Union, Radiocommunication Sector (ITU-R, URL https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en.

Iversen, O. S., Dindler, C., & Hansen, E. I. K. (2013). Understanding teenagers' motivation in participatory design. *International Journal of Child-Computer Interaction*, *1*, 82–87. http://dx.doi.org/10.1016/j.ijcci.2014.02.002, URL https://www.sciencedirect.com/science/article/pii/S221286891400004X.

Iversen, O. S., Kortbek, K. J., Nielsen, K. R., & Aagaard, L. (2007). Stepstone: an interactive floor application for hearing impaired children with a cochlear implant. In *Proceedings of the 6th international conference on interaction design and children* (pp. 117–124). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1297277.1297301.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*, 227–229. http://dx.doi.org/10.3758/bf03214307.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J. H., Worman, T., Yueh, B., & Rubinstein, J. (2009). Development and validation of the university of washington clinical assessment of music perception test. *Ear and Hearing*, *30*, 411–418. http://dx.doi.org/10.1097/aud.0b013e3181a61bc0.

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., & Smeds, K. (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear and Hearing*, *41*(Suppl. 1), 5S–19S. http://dx.doi.org/10.1097/AUD.0000000000000944, URL https://journals.lww.com/ear-hearing/fulltext/2020/11001/the_quest_for_ecological_validity_in_hearing.2.aspx.

Kim, Y., & Choi, B. Y. (2022). Precision medicine approach to cochlear implantation. *Clinical and Experimental Otorhinolaryngology*, *15*, 299–309. http://dx.doi.org/10.21053/ceo.2022.01382, URL http://e-ceo.org/journal/view.php?number=844.

Kim, J. H., Lee, S. H., & Cho, E. Y. (2021). A systemic review of the auditory training program for hearing impairment. *Audiology and Speech Research*, *17*, 134–146. http://dx.doi.org/10.21848/asr.200090.

Koops, E. A., Renken, R. J., Lanting, C. P., & van Dijk, P. (2020). Cortical tonotopic map changes in humans are larger in hearing loss than in additional tinnitus. *The Journal of Neuroscience*, *40*, 3178–3185. http://dx.doi.org/10.1523/jneurosci.2083-19.2020.

Korte, J. (2022). YoungDeafDesign: Participatory design with Young deaf children. *International Journal of Child-Computer Interaction*, *34*, Article 100542. http://dx.doi.org/10.1016/j.ijcci.2022.100542, URL https://www.sciencedirect.com/science/article/pii/S2212868922000605.

Kucirkova, N. (2019). Children's agency by design: Design parameters for personalization in story-making apps. *International Journal of Child-Computer Interaction*, *21*, 112–120. http://dx.doi.org/10.1016/j.ijcci.2019.06.003, URL https://www.sciencedirect.com/science/article/pii/S2212868918300503.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Laures, J. S., & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research*, *42*, 1148–1156. http://dx.doi.org/10.1044/jslhr.4205.1148.

Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research*, *308*, 13–26. http://dx.doi.org/10.1016/j.heares.2013.04.009.

Lin, E. Y., Younan, S. M., Barrett, K. C., & Jiam, N. T. (2025). Personalizing cochlear implant care in single-sided deafness: A distinct paradigm from bilateral hearing loss. *Journal of Personalized Medicine*, *15*, http://dx.doi.org/10.3390/jpm15090439, URL https://www.mdpi.com/2075-4426/15/9/439.

Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., & Lake, J. (2006). Bilateral Cochlear implants in children: Localization acuity measured with minimum audible angle. *Ear and Hearing*, *27*, 43–59. http://dx.doi.org/10.1097/01.aud.0000194515.28023.4b, URL https://journals.lww.com/ear-hearing/fulltext/2006/02000/Bilateral_Cochlear_Implants_in_Children_.00004.aspx.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*, 217–224. http://dx.doi.org/10.1080/00031305.2000.10474549, URL https://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474549.

Luo, X., Masterson, M. E., & Wu, C.-C. (2014). Melodic interval perception by normal-hearing listeners and cochlear implant users. *Journal of the Acoustical Society of America*, *136*, 1831–1844. http://dx.doi.org/10.1121/1.4894738.

Mahone, E., & Schneider, H. (2012). Assessment of attention in preschoolers. *Neuropsychology Review*, *22*, 361. http://dx.doi.org/10.1007/s11065-012-9217-y.

Maier, F., & Jäger, S. (2016). Analysing discourses and dispositives: A foucauldian approach to theory and methodology. (pp. 109–136). http://dx.doi.org/10.4135/9781036235192.n5.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-Modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, *50*, 125–145. http://dx.doi.org/10.1016/S0167-8760(03)00129-6, URL https://www.sciencedirect.com/science/article/pii/S0167876003001296.

Martini, A., & Schindler, O. (2004). *La sordità prelinguale*. Torino: Omega.

Massey, J., Tsianakas, V., Gordon, A., Sadler, N., & Robert, G. (2024). Co-Designing complex therapy interventions with parents as partners in the care of children with cerebral palsy: An experience-based Co-design study in England. *Research in Developmental Disabilities*, *151*, Article 104793. http://dx.doi.org/10.1016/j.ridd.2024.104793, URL https://www.sciencedirect.com/science/article/pii/S0891422224001252.

McCullagh, P., & Nelder, J. (2019). *Generalized linear models*. New York: Routledge, http://dx.doi.org/10.1201/9780203753736.

McDermott, H. J., McKay, C. M., & Vandali, A. E. (1992). A new portable sound processor for the university of Melbourne/Nucleus limited multielectrode Cochlear implant. *Journal of the Acoustical Society of America*, *91*, 3367–3371. http://dx.doi.org/10.1121/1.402826.

Mechelen, M. Van, Musaeus, L. Have, Iversen, O. S., Dindler, C., & Hjorth, A. (2021). A systematic review of empowerment in child-computer interaction research. In *Proceedings of the 20th annual ACM interaction design and children conference* (pp. 119–130). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3459990.3460701.

Melonio, A., & Gennari, R. (2013). *How to design games for deaf children: Evidence-based guidelines* (pp. 83–92). Germany: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-00554-6_11.

Mesnildrey, Q., Hilkhuysen, G., & Macherey, O. (2016). Pulse-Spreading harmonic complex as an alternative carrier for vocoder simulations of Cochlear implants. *Journal of the Acoustical Society of America*, *139*, 986–991. http://dx.doi.org/10.1121/1.4941451.

Moore, B. (2013). *An introduction to the psychology of hearing*. Leiden, Netherlands: Brill.

Murphy, K. M., Roumeliotis, G., Williams, K., Beston, B. R., & Jones, D. G. (2015). Binocular visual training to promote recovery from monocular deprivation. *Journal of Vision*, *15*, 2. http://dx.doi.org/10.1167/15.1.2.

Nassiri, A. M., Marinelli, J. P., Lohse, C. M., & Carlson, M. L. (2023). Age and incidence of cochlear implantation in the pediatric population with congenital bilateral profound hearing loss. *Otology & Neurotology*, *44*, e492–e496. http://dx.doi.org/10.1097/mao.0000000000003932.

Nguyen, N. (2023). Bionic hearing: When is it time to get a cochlear implant? *Acoustics Today*, *19*, 37. http://dx.doi.org/10.1121/at.2023.19.1.37.

Nicholson, S. (2015). *A RECIPE for meaningful gamification* (pp. 1–20). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-10208-5_1.

Peng, S.-C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with Cochlear implants. *Trends in Amplification*, *16*, 67–82. http://dx.doi.org/10.1177/1084713812451159.

Petry, B., Illandara, T., Elvitigala, D. S., & Nanayakkara, S. (2018). Supporting rhythm activities of deaf children using Music-Sensory-Substitution systems. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–10). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3173574.3174060.

Plass, J., Homer, B., & Kinzer, C. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*, 258–283. http://dx.doi.org/10.1080/00461520.2015.1122533.

Poli, G. De, & Avanzini, F. (2005). Algorithms for sound and music computing. URL http://smc.dei.unipd.it/education/algo4smc_ch2.pdf, Teaching material for the course Informatica Musicale (Sound and Music Computing). Chapters are available under the CreativeCommons Attribution-NonCommercial-ShareAlikes, Department of Information Engineering, University of Padova.

Potter, L. E., Korte, J., & Nielsen, S. (2014). Design with the deaf: Do deaf children need their own approach when designing technology? *Proceedings of the 2014 conference on interaction design and children* (pp. 249–252). USA: Association for Computing Machinery (ACM), http://dx.doi.org/10.1145/2593968.2610464.

Privitera, A. G., & Geronazzo, M. (2024). Designing sonic interactions in intelligent reality with egocentric audio technologies. In *The routledge handbook of sound design* (p. 33). Focal Press.

Puigcerver, L., Cuadrado, S. Rodriguez, Tapia, V. Gomez, & Ordoño, J. Navarra (2019). Vertical mapping of auditory loudness: Loud is high, but quiet is not always low. *Psicologica*, *40*(2), 85–104. http://dx.doi.org/10.2478/psicolj-2019-0006, URL https://hdl.handle.net/2445/164937, 2019.

Qi, X., & Yu, J. (2025). Participatory design in human-computer interaction: Cases, characteristics, and lessons. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1–26). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3706598.3713436.

Rasmussen, C. E., & Williams, C. K. I. (2019). *Adaptive computation and machine learning series, Gaussian processes for machine learning*. London, England: MIT Press.

Rocchesso, D., & Fontana, F. (Eds.), (2003). *The sounding object*. Florence, Italy: Edizioni di Mondo Estremo, URL https://air.uniud.it/bitstream/11390/677477/1/vmain.pdf.

Rotfleisch, S., & Martindale, M. (2021). *Listening and spoken language therapy for children with hearing loss: A practical auditory-based guide*. San Diego, CA: Plural Publishing.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*, 113–129. http://dx.doi.org/10.1016/j.cognition.2005.01.004, URL https://www.sciencedirect.com/science/article/pii/S0010027705000260.

Saenz, M., & Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, *307*, 42–52. http://dx.doi.org/10.1016/j.heares.2013.07.016.

Schlittenlacher, J., & Moore, B. C. J. (2020). Fast estimation of Equal-Loudness contours using Bayesian active learning and direct scaling. *Acoustical Science and Technology*, *41*, 358–360. http://dx.doi.org/10.1250/ast.41.358.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*, 103. http://dx.doi.org/10.9781/ijimai.2017.09.001.

Serpanos, Y. C., & Gravel, J. S. (2004). Revisiting loudness measures in children using a computer method of cross-modality matching (CMM). *Journal of the American Academy of Audiology*, *15*, 486–497. http://dx.doi.org/10.3766/jaaa.15.7.3.

Shin, K. H., & Park, M. K. (2023). Personalized approach to auditory rehabilitation. *Journal of Clinical Otolaryngology–Head and Neck Surgery*, *34*, 103–110. http://dx.doi.org/10.35420/jcohns.2023.34.4.103.

Simon, A. J., Gallen, C. L., Ziegler, D. A., Mishra, J., Marco, E. J., Anguera, J. A., & Gazzaley, A. (2023). Quantifying attention span across the lifespan. *Frontiers in Cognition*, *2 - 2023*, http://dx.doi.org/10.3389/fcogn.2023.1207428, URL https://www.frontiersin.org/journals/cognition/articles/10.3389/fcogn.2023.1207428.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971–995. http://dx.doi.org/10.3758/s13414-010-0073-7.

Spiel, K., Frauenberger, C., & Fitzpatrick, G. (2017). Experiences of autistic children with technologies. *International Journal of Child-Computer Interaction*, *11*, 50–61. http://dx.doi.org/10.1016/j.ijcci.2016.10.007, URL https://www.sciencedirect.com/science/article/pii/S2212868916300174, designing with and for Children with Special Needs.

Stropahl, M., Besser, J., & Launer, S. (2019). Auditory training supports auditory rehabilitation: A State-of-the-Art review. *Ear and Hearing*, *41*, 697–704. http://dx.doi.org/10.1097/aud.0000000000000806.

Vandali, A., Sly, D., Cowan, R., & van Hoesel, R. (2015). Training of cochlear implant users to improve pitch perception in the presence of competing place cues. *Ear and Hearing*, *36*, e1–e13. http://dx.doi.org/10.1097/aud.0000000000000109.

Vickers, D., Salorio-Corbetto, M., Driver, S., Rocca, C., Levtov, Y., Sum, K., Parmar, B., Dritsakis, G., Flores, J. Albanell, Jiang, D., Mahon, M., Early, F., Zalk, N. Van, & Picinali, L. (2021). Involving children and teenagers with bilateral cochlear implants in the design of the BEARS (both EARS) virtual reality training suite improves personalization. *Frontiers in Digital Health*, *3 - 2021*, http://dx.doi.org/10.3389/fdgth.2021.759723, URL https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2021.759723.

Wagner, L., Altindal, R., Plontke, S. K., & Rahne, T. (2021). Pure tone discrimination with cochlear implants and filter-band ppread. *Scientific Reports*, *11*, 20236. http://dx.doi.org/10.1038/s41598-021-99799-4.

Wang, W., Zhou, N., & Xu, L. (2010). Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology*, *50*, 270–278. http://dx.doi.org/10.3109/14992027.2010.542490.

Werner, K. J., Abel, J. S., & III, J. O. Smith (2014). A physically-informed, circuit-bendable, digital model of the Roland TR-808 bass drum circuit. In *DAFx* (pp. 159–166). Erlangen: International Audio Laboratories Erlangen, URL https://www.dafx.de/paper-archive/2014/dafx14_kurt_james_werner_a_physically_informed,_ci.pdf.

Witton, C., Talcott, J. B., & Henning, G. B. (2017). Psychophysical measurements in children: Challenges, pitfalls, and considerations. *PeerJ*, *5*, Article e3231. http://dx.doi.org/10.7717/peerj.3231.

Xiang, Y., Zhang, Z., Chang, D., & Tu, L. (2024). The impact of gamified auditory–Verbal training for hearing-challenged children at intermediate and advanced rehabilitation stages. *Games for Health Journal*, null. http://dx.doi.org/10.1089/g4h.2023.0213, pMID: 38808466.

Yuan, B., & Folmer, E. (2008). Blind hero: Enabling guitar hero for the visually impaired. *Proceedings of the 10th international ACM SIGACCESS conference on computers and accessibility* (pp. 169–176). USA: Association for Computing Machinery (ACM), http://dx.doi.org/10.1145/1414471.1414503.

Zaltz, Y., Bugannim, Y., Zechoval, D., Kishon-Rabin, L., & Perez, R. (2020). Listening in Noise Remains a Significant Challenge for Cochlear Implant Users: Evidence from Early Deafened and Those with Progressive Hearing Loss Compared to Peers with Normal Hearing. *Journal of Clinical Medicine*, *9*, 1381. http://dx.doi.org/10.3390/jcm9051381, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7290476/.

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *Journal of the Acoustical Society of America*, *132*, 984–993. http://dx.doi.org/10.1121/1.4733535.

Zeng, F.-G., Tang, Q., & Lu, T. (2014). Abnormal pitch perception produced by cochlear implant stimulation. *PLoS One*, *9*, Article e88662. http://dx.doi.org/10.1371/journal.pone.0088662.

Zhao, J., Neto, I., Pires, A. C., Tomé-Pires, C., & Nicolau, H. (2025). Digital technologies for deaf and hard of hearing children: A systematic review, critical reflections, and future research directions. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1–19). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3706598.3714241.

Zheng, Y., & Brette, R. (2017). On the relation between pitch and level. *Hearing Research*, *348*, 63–69. http://dx.doi.org/10.1016/j.heares.2017.02.014.

Zhou, Y., Sim, K. C., Tan, P., & Wang, Y. (2012). MOGAT: Mobile games with auditory training for children with Cochlear implants. In *Proceedings of the 20th ACM international conference on multimedia* (pp. 429–438). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2393347.2393409.