Gamified psychoacoustic pitch discrimination test app - a case study

Andrea Gulli¹, Enrico Piccirillo², Caterina Kihlgren², Cesare Miani^{3,4}, Enrico Muzzi³

1) HCI Lab, Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy.

2) Gruppo Otologico, Casa di Cura Piacenza, Piacenza, Italy

3) Department of Otorhinolaryngology, Academic Hospital "Santa Maria della Misericordia", Azienda Sanitaria Universitaria Friuli Centrale, P.le Santa Maria della Misericordia 15, 33100 Udine, Italy

4) University of Udine, Department of Medicine (DAME), Via Colugna 50, 33100 Udine, Italy

Abstract

Background - a pitch mismatch is considered detrimental for binaural integration and auditory performance in asymmetric hearing rehabilitated with cochlear implant. However, current psychoacoustic pitch balancing procedures are time consuming and annoying, leading to disengagement and unreliable results.

Aim - the aim of this study is to speed up the psychoacoustic pitch discrimination test maintaining high levels of reliability and engagement with a gamified mobile-based procedure appropriate for children and adults.

Methods - a preliminary case study is presented of a mobile-based application to test pitch discrimination in an adult patient affected by asymmetric hearing loss with mild sensorineural hearing loss in one ear and profound sensorineural hearing loss rehabilitated with cochlear implant in the other ear. The details of the procedure and the results of ease of use and workload questionnaires are reported.

Results - the pitch discrimination test pointed out a frequency-specific mismatch between the unaided ear and the cochlear-implanted ear. The questionnaires' scores suggest that the gamified procedure could be easily implemented in the testing protocol of cochlear implant fitting.

Conclusion - the results of the psychoacoustic pitch discrimination test could be useful to improve cochlear implant fitting. Research is ongoing to verify and validate the efficiency of the gamified pitch discrimination procedure reported in the present study.

Keywords: *Pitch, loudness, cochlear implant, asymmetric hearing, gamification*

Introduction

One broad definition of pitch is that it is a perceptual characteristic of sounds that, when they are presented at equal loudness, it enables them to be ranked along a low to high dimension (McDermott, 2011).

The ability to distinguish between pitches plays an important role in various aspects of our daily lives. For instance, being able to discern pitch is essential for perceiving music melodies (Moore, 2012), while in verbal communication (Laures and Weismer, 1999), it helps in understanding different talkers and the prosodic nuances of speech, which are particularly significant in tonal languages (Deroche et al., 2019). Additionally, pitch

is a fundamental cue for separating different sound sources, aiding in the segregation of auditory information. When access to cues related to changes in speech fundamental frequency is limited, it can negatively impact the ability to recognize speech, especially in noisy environments (Binns and Culling, 2009). Pitch, a perceptual attribute primarily determined by the frequency content of a sound (Moore, 2012), can be represented through various cues, including place pitch, which involves selecting a specific region of excitation in the cochlea (Marimuthu, Swanson and Mannell, 2016), and temporal pitch (Saenz and Langers, 2014). The mapping of

place pitch encoding relies on the established tonotopic map of the cochlea. However, this map can exhibit variability between ears and among individuals, possibly changing within the same individual over time. Furthermore, deviations in neural plasticity associated with hearing loss can lead to alterations in the tonotopic map (Koops et al., 2020).

In patients undergoing rehabilitation for hearing loss, the most significant interaural discrepancy occurs in individuals with one cochlear implant (CI) and varying degrees of acoustic hearing in the opposite ear, ranging from normal to severely impaired (Pieper et al. 2022). When acoustic hearing is supplemented with a hearing aid, it's termed as "bimodal" approach. If the individual has normal hearing, it's denoted as "single-sided deafness and CI" strategy (SSD-CI). Frequency resolution, tonotopic mismatch, and interaural mismatch exert distinct impacts on speech comprehension and spatial release from masking in simulated scenarios of bilateral CIs. Reducing interaural mismatch could be crucial for maximising binaural advantages and enhancing CI efficacy in environments with competing speech, which is a common listening situation (Xu et al., 2020).

The correct measurement of pitch discrimination may be useful for two main reasons. The first is that good results in other related measures, such as an accurate pitch ranking ability, appear to be an independent predictor of overall CI performance outcome (Kenway et al., 2015). The second is that it has been found that the pitch discrimination can be used as a reference for the fitting of the CI (Saleh, 2013).

There are also two main limitations to the clinical application of pitch discrimination tests. One is the total time required to complete the procedure, which can lead to annoyance and disengagement, and the other is that there are several variables and biases affecting this measure. Here we present a case study on a quick pitch discrimination test performed with a rigorous psychoacoustic procedure implemented in an engaging task via a mobile app.

Materials and Methods

Participant

A male, 59 years-old, affected by right ear idiopathic profound deafness (Figure 1) without comorbidities, was enrolled at the Gruppo Otologico (Casa di Cura Piacenza, Piacenza, IT) and tested 5 years after right ear CI (Cochlear Nucleus CI532 Slim Modiolar, Cochlear Limited, Sydney, AU) with an off-the-ear processor (Cochlear Nucleus Kanso, Cochlear Limited, Sydney, AU). Unaided speech reception threshold for 50% of correctly recognized bisyllabic words (SRT50) in quiet was located at 25 dB hearing level (HL), while direct input SRT50 in quiet was located at 55 dB HL. At the unaided italian Matrix Sentence Test (Hoerzentrum Oldenburg, DE), the SRT50 was 5.3 dB in the S0N0 condition and -1.8 dB in the S0N+90 (i.e. with noise on the right side) condition, while in the CI-aided condition SRT50 was 4.4 dB in the S0N0 condition and -2.7 dB in the S0N+90 condition. Prior to the experiment, electrode impedances of the implanted ear were checked.

Figure 1. Pure tone audiometry: left PTA_{500-4000Hz} *= 28.75 dB HL; right profound deafness. Full circles: right CI-aided pure tone threshold, PTA500-4000Hz = 43.75 dB HL.*

Stimuli

Acoustic stimuli were generated as exponentially decaying sinusoids, derived from the impulse responses of a resonant second-order infinite impulse response (IIR) filter (De Poli and Avanzini, 2005) with central frequency set to control the stimulus pitch and sampling frequency of 44.1 kHz. The parameter controlling timbre and duration remained constant at 0.1. During testing, audio was synthesised in real-time using Pure Data open-source software on a tablet (Apple iPad Pro 11-inch 3rd generation 256 GB, Cupertino, US) facilitated by Pure Data libpd audio synthesis library (Brinkmann et al., 2016). The stimuli were delivered monaurally through the direct input to the right CI and through a loudspeaker (Genelec 8020D, Insalmi, FI) positioned at approximately one meter lateral to the left ear. To ensure adequate loudness, the iPad's digital output was routed through a preamplifier system (Antelope Zen Go Synergy Core USB sound interface, Sofia, BG, via a USBC-3.5 mm jack adapter, and then the analog output was amplified using a RME Babyface Pro sound card, Haimhausen, DE). The system was calibrated to achieve a tolerance of 2.5 dB sound pressure level (SPL) and a peak amplitude of 103 dB SPL measured with a calibrated class 1 sound level meter (XL2 Sound Level Meter, NTi Audio, Schaan, LI).

Procedure

The pitch discrimination session included six steps:

Familiarisation, involving collaborative training trials with the experimenter.

Figure 2. *The tablet user interface running the loudness balance test (a) and the pitch discrimination test (b); the stereo line-out of the tablet is splitted into left and right channels; the left channel is connected to a soundcard and preamplifier system (c); an active loudspeaker delivers the acoustic signal to the left ear (d); the right wire is connected to the line-in jack of a wireless streamer (Cochlear MiniMic 2+, Cochlear Limited, Sydney, AU) (e); a wireless direct input to the cochlear implant processor*

stimulates the right ear while microphones are turned off (f). Continuous lines: wired connections; dashed line: wireless connection.

Determination of a degree of similarity in a scale with 5 integer steps of three different waveforms (pure tones, exponentially decaying sinusoids, and Pulse-Spreading Harmonic Complexes - PSHCs) with a 1 kHz central frequency presented with a 500 ms stimulus onset asynchrony between the two sides.

Loudness balancing, where the participant adjusted stimuli to their most comfortable loudness (MCL) and established an equal-loudness contour (ELC).

Two-interval, two-alternative forced choice (2I-2AFC) adaptive pitch discrimination testing.

A catch trial, serving as a distractor task, where the listener identified whether a pair of pure tones sounded different or the same.

Evaluation of the overall experience through brief questionnaires on task load and ease of use.

The test setup is reported in figure 2.

Loudness balancing

The participant adjusted loudness to a comfortable level, starting from 50 dB SPL in the acoustic-hearing ear. SPL was increased or decreased by 1 or 4 dB steps using four animated buttons. In this way, we estimated an equal loudness contour (ELC) for each of the three test stimuli in both ears. The first step consisted in presenting seven stimuli spaced in half-octave steps centred around 1 kHz in random order. The listener set each one at the most comfortable level. The second step consists in comparing each of these levels with the test stimulus set at the most comfortable level (MCL) presented with a 500 ms onset asynchrony. In this second step, a layout similar to the one used for MCL determination allowed sound level adjustments, with buttons enabling increments and decrements of the levels at the contralateral ear of 1 or 3 dB SPL steps. An example of the result of the ELC estimation procedure is illustrated in Figure 3.

2I-2AFC pitch discrimination

Discrimination thresholds can be measured using various methods. Adaptive procedures are usually preferred over constant stimuli as they ensure a dense sampling of the psychometric curve in the region of the threshold (Arzounian et al., 2017). That region is hard to situate prior to the experiment because of interindividual differences and training effects.

Figure 3. *Equal loudness levels set by the listener in the loudness balancing procedure.*

A common procedure, implemented here, is the two-interval two-answer forced choice (2I-2AFC). In the 2I-2AFC pitch discrimination adaptive procedure, two stimuli are presented on each trial with slightly different frequencies, and the listener must decide whether the latter has a higher pitch than the former. We utilised a weighted up-down method similar to the simple up-down rule (Kaernbach, 1991). The frequency step between the tones to be compared was adjusted to track the 75% correct point on the psychometric function. After each incorrect response, the relative frequency difference was doubled, while after a correct response, it was divided by the cubic root of 2. The initial step size was set to 10%, and it was adapted with a single adaptive track for rapid convergence over at least 16 trials, aiming to complete the experiment in approximately 30 minutes. Upon completion of 16 trials, the mean and the 95% confidence interval were computed and a catch trial was inserted. The pitch discrimination sequence stops when one of three following conditions is satisfied:

at the 16th and 32nd trials, the threshold fell in the 95% interval of confidence computed from the previous values of the same pitch discrimination sequence, hence indicating a stable value;

at the 32nd trial, the threshold has reached a value smaller than 0.3%, considered to be the just noticeable difference in frequency discrimination of a trained normal hearing individual;

the pitch discrimination sequence reached a maximum length of 48 trials (Figure 4).

Figure 4. *Temporal profiles of the three pitch discrimination sequences according to the two-interval two-answer forced choice adaptive procedure. Central frequency of the test stimulus: dashed horizontal line; correct responses: dots; incorrect answers: crosses. The*

vertical dashed lines mark the 16th and 32nd trials, where the pitch discrimination sequence is interrupted by catch trials. Horizontal dotted lines represent the upper and lower bounds of the 75% confidence interval computed using all sequence thresholds.

We tested three frequencies (750 Hz, 1 kHz, and 2 kHz) that align well with the "speech banana" and fall within the effective range of CIs (Nguyen et al., 2023). The pitch discrimination session lasted approximately 30 minutes.

Collected data were stored in the device and then transferred to a personal computer for analysis.

Graphical user interface

The pitch discrimination graphical user interface (GUI) was designed to provide an engaging experience, incorporating elements of serious game mechanics (Bhopal and Senan, 2022). The interface features two bouncing "marbles" positioned on each side of the screen, running on a track that scrolls horizontally. When the marbles bounce, corresponding stimuli are presented. A question mark and two arrows appear contextually, prompting participants to indicate whether the second stimulus is higher or lower than the first.

This visual design choice was inspired by the established space-pitch associations made by infants across cultures (Dolscheid et al., 2014). Feedback was provided through animations showing the marbles either converging or diverging based on the correctness of the response. A strawberry was awarded as a reward for correct answers, enhancing engagement, while a visible score counter tracked performance.

Quality of the experience

Sessions included assessments using a modified version of the NASA Raw Task Load Index (RTLX) and a Short User Experience Questionnaire (UEQ-S) to evaluate workload and subjective user experience (Hart & Staveland, 1988; Schrepp et al., 2017). Experimenters assisted the participant in completing the questionnaires.

Results

The participant ranked the pure tones and the PSHC as quite similar at the two ears (integer score $= 4$), and the exponentially decaying sinusoids slightly less similar (integer score = 3). The results of the pitch discrimination session are summarised in table 1 and in Figure 5.

Table 1. Results of the pitch discrimination session.

Procedure	Times (minutes)
MCL 2000 Hz	0.7
ELC (complete procedure)	2.52
Pitch discrimination 750 Hz	2.57
Pitch discrimination 1000 Hz	1.98
Pitch discrimination 2000 Hz	1 75

Table 2. Time required to complete each step of the pitch discrimination procedure.

Table 2 summarises the times taken to complete each single step of the procedure. These times represent the sum of the intervals during which the listener provided responses after the stimuli were presented. The total test duration encompasses various factors, including interstimulus intervals, time required for graphic animations and transitions between different GUIs, intermissions of catch trials, and some minutes for introducing the listener to the task. Given these factors, although the listener spent approximately 10 minutes completing the various tasks, this time was spread over a total session time of approximately 30 minutes.

Figure 6 illustrates the results of the NASA RTLX and UEQ-S questionnaires.

Figure 5. *Pitch discrimination thresholds of the three tested stimuli. The dotted line represents the identity, the triangles represent the upper and lower bounds of the 95% confidence intervals, the vertical dashed lines represent the discrimination thresholds.*

Figure 6. *NASA Raw Task Load Index (RTLX) and Short User Experience Questionnaire (UEQ-S) results. In the UEQ-S no column = 0.*

Discussion

Different methods have been used in the literature to compare pitch between electric and acoustic stimuli, including fixed stimuli and adaptive procedures (Adel et al. 2019). All of these can be potentially contaminated by non-sensory biases. Moreover, the same electrode could be matched to frequencies largely separated, suggesting that the subjects were not performing real pitch comparisons between the ears, but were judging the acoustic range of the stimulus as higher or lower than the electric stimulus. Octave confusion can also occur. In adaptive tasks the acoustic frequency presented on a given trial is based on the subject's response to the preceding trial, leading to a progressive restriction of the pitch range. Biases for these tasks are that the pitch can be strongly correlated with the starting frequency and that the pitch of the acoustic stimulus may not vary a lot between consecutive trials, which could distract the subjects from the task itself, especially if the stimuli vary across other dimensions (e.g., loudness or timbre), which may be more salient than the pitch dimension (Adel et al. 2019).

The equal loudness of pure tones is a function of frequency. Fletcher and Munson (Fletcher and Munson, 1933) determined the first equal-loudness contours (ELCs) for normal-hearing (NH) listeners, and these still form the basis for the A-weighted SPL. The stimuli used in pitch discrimination experiments are usually pure tones (Arzounian et al. 2017) or 1/3-octave band noise (Adel et al. 2019). Optimal stimuli have been designed to

simulate the broad spread of excitation, with an adjustable pulse rate to minimise intrinsic modulations after auditory filtering, known as PSHC (Hilkhuysen and Macherey 2014). In discrimination experiments with complex tones, just-noticeable differences ranging from less than one semitone to several semitones were found (Luo et al. 2014). Another study reported large individual variability in pitch-ranking thresholds, ranging from 1 to 8 semitones (Kang et al. 2009). Studies using real musical instruments have estimated pitch discrimination thresholds for CI listeners of around 80%-90% (Brockmeier et al. 2011). In all of these studies, the variance across participants was considerable, with some participants only being able to discriminate pitch with a frequency difference slightly less than 100% (one octave) and best individual thresholds around 3% (0.5 semitones). The best CI users performance is similar to pitch discrimination thresholds with musical instruments for normal-hearing listeners (Brockmeier et al. 2011). The thresholds found in this study align with the ones found in literature (Luo et al., 2014); for 750 Hz and 1000 Hz, the thresholds set to slightly less than a semitone, while it is more than doubled for 2000 Hz. The 2000 Hz threshold is not directly correlating with the pure-tone audiometry presented in Figure 1, nevertheless it is still surprisingly small compared to other results involving SSD-CI listeners. However, we observe that the stimulus with a central frequency of 2000 Hz took the longest time for the MCL determination, which could reflect some uncertainty. There are similarities among the three tested stimuli's waveforms, pure-tones and PSHCs set to 4 out of 5, and the exponentially decaying sinusoid set to 3 out of 5, indicating that the inter-aural unbalancing is limited. The PSHC is expected to be the highest ranking (Adel et al., 2019), but it resulted that the pure-tones were found similar, more than the decaying sinusoids. Nevertheless, it must be noted that the similarity question was probably too vague, because not focusing on a particular dimension (i.e., level or pitch), and it was probably not completely understood.

There are several limitations of the clinical application of pitch discrimination tests. One is the total time span of the procedure, easily reaching several hours (Arzounian et al., 2017), far beyond the attention span of an adult and even more of a paediatric patient, leading to fatigue, annoyance, and task disengagement with inconclusive or misleading results. The other is that it is a highly variable measurement, both at an inter-individual level for a normal-hearing population (Kang et al., 2009) as well as for a hearing-impaired one (Brockmeier et al., 2011), and for several types of bias interfering with the measure (Stakhovskaya et al., 2017).

Gamification refers to the application of game elements developed in the entertainment industry to a context that is not traditionally a game (Lester, 2017). Play behaviour in games can be intrinsically motivating, and can be used in a wide variety of contexts including deafness (Cano et al., 2015). Recently, the application of gamification to auditory training has led to improvements in speech perception thresholds in different competing noise conditions (Bologna, et al., 2023). Although the results are promising, attention must be paid to not increase distractions

and introduce contradictory multimodal elements.

The results of the NASA RTLX and UEQ-S indicate a high mental demand (Figure 6). The physical load is low, as expected from the passive and reflective kind of task. The temporal load is higher, probably due to the scrolling of the screen and the fact that there was no chance for repeating the stimuli, because only one listening trial was allowed per pair. The participant rated his performance as very good, while the effort and frustration were limited. These results could be probably linked to the fact that the participant was a professional audio engineer and he could have been accustomed to paying attention to small differences in auditory cues. The UEQ-S outcomes align with the results of NASA RTLX. The test experience was not perceived as particularly complicated, nor exciting, or interesting. It has been considered conventional, but leading edge as well, indicating that it was an activity familiar to the listener, but probably presented in a new way. A valuable result is that the participant found the app efficient and supportive.

Conclusion

A mobile application has been developed with the aim to improve psychoacoustic pitch matching procedures between acoustic and electric hearing. The results of the pitch discrimination task align with the literature, suggesting that the mobile app can be an efficient tool for a quick, reliable, self-administered test that can provide key elements for improving the fitting of CIs. Research is ongoing to verify and validate the results of this preliminary case study.

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. Front Neurosci. 13, 1119.
- Arzounian D., de Kerangal M., de Cheveigné A. (2017) A sliding two-alternative forced-choice paradigm for pitch discrimination. J Acoust Soc Am. 142, 167.
- Bhopal P. K. R. S., Senan N. (2022) Hearing screening test mobile games application for kids. Applied information tech computer sci. 3, 33-46.
- Binns C., Culling J. F. (2017) The role of fundamental frequency contours in the perception of speech against interfering speech. J Acoustical Soc Am. 122, 1765-76.
- Bologna W. J., Carrillo A. A., Clamage D. S., et al. (2023) Effects of gamification on assessment of spatial release from masking. Am J Audiol. 32, 210-19.
- Brinkmann P., Wilcox D., Kirshboim T., Eakin R., Alexander R. (2016) Libpd: past, present, and future of embedding pure data, New York: Proceedings of the 5th Pure Data Convention.
- Brockmeier S.J., Fitzgerald D., Searle O., et al. (2011) The MuSIC perception test: a novel battery for testing music perception of cochlear implant users. Cochlear Implants Int. 12, 10-20.
- Cano S. P., Penenory V. M., Collazos C., Fardoun H., Alghazzawi D. (2015) Training with Phonak: Serious game as support in auditory-verbal therapy for children with cochlear implants. Proceedings of the 3rd Workshop on ICTs for improving patients rehabilitation research techniques, Lisbon: Association for Computing Machinery.
- De Poli G., Avanzini F. (2005) Algorithms for sound and music computing, Milano: Università di MIlano.
- Deroche M. L. D., Lu H. P., Kulkarni A. M., et al. (2019) A tonal-language benefit for pitch in normallyhearing and cochlear-implanted children. Sci Rep. 9, 109.
- Dolscheid S., Hunnius S., Casasanto D., Majid A. (2014) Prelinguistic infants are sensitive to spacepitch associations found across cultures. Psychol Sci. 25, 1256-61.
- Fletcher H., Munson W.A. (1933) Loudness, its definition, measurement and calculation. Bell System Tech J. 12, 377-430.
- Hart S., Staveland L. (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock P., Meshkati N., eds. Human Mental Workload, North-Holland: Elsevier, 139-83.
- Hilkhuysen G., Macherey O. (2014) Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering. J Acoust Soc Am. 136, 1281.
- Kaernbach C. (1991) Simple adaptive testing with the weighted up-down method. Percept Psychophys. 49, 227-9.
- Kang R., Nimmons G.L., Drennan W., et al. (2009) Development and validation of the University of Washington Clinical Assessment of Music Perception test. Ear Hear. 30, 411-8.
- Kenway B., Tam Y. C., Vanat Z., et al. (2015) Pitch discrimination: an independent factor in cochlear implant performance outcomes. Otol Neurotol. 36, 1472-9.
- 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. J Neurosci. 40, 3178-185.
- Laures J. S., Weismer G. (1999) The effects of a flattened fundamental frequency on intelligibility at the sentence level. J Speech Lang Hearing Res. 42, 1148-56.
- Lester S.I. (2017) The effect of gamification on audiology awareness among young adults, Louisiana Tech University, Ann Arbor: ProQuest.
- Luo X., Masterson M.E., Wu C.C. (2014) Melodic interval perception by normal-hearing listeners and cochlear implant users. J Acoust Soc Am. 136, 1831-44.
- Marimuthu V., Swanson B. A., Mannell R. (2016) Cochlear implant rate pitch and melody perception as a function of place and number of electrodes. Trends Hear. 20, 2331216516643085.
- McDermott, H. (2011) Benefits of combined acoustic and electric hearing for music and pitch perception. Seminars in Hearing. 32, 103-14.
- Moore, B.C.J. (2012) An introduction to the psychology of hearing 6th ed.: Brill.
- Pieper S. H., Bahmer A. (2019) Rate pitch discrimination in cochlear implant users with the use of double pulses and different interpulse intervals. Cochlear Implants Int. 20, 312-23.
- Saenz M., Langers D. R. (2014) Tonotopic mapping of human auditory cortex. Hear Res. 307, 42-52.
- Saleh S.M. (2013) The efficacy of fitting cochlear implants based on pitch perception, London: University College London, The UCL Ear Institute.
- Schrepp M., Hinderks A., Thomaschewski J. (2017) Design and evaluation of a short version of the user experience questionnaire (UEQ-S). Int J Interactive Multimedia Art Int. 4, 103-8.
- Stakhovskaya O., Bernstein J.G., Goupell M. (2017) Non-sensory biases in a pitch-discrimination task for bilateral and single-sided deafness cochlear-implant listeners. J Acoust Soc Am. 141, S3815.
- Xu K., Willis S., Gopen Q., Fu Q. J. (2020) Effects of spectral resolution and frequency mismatch on speech understanding and spatial release from masking in simulated bilateral cochlear implants. Ear Hear. 41, 1362-71.