Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transient mirror-type pitch inversion during auditory rehabilitation in cochlear implant and normal-hearing listeners

Krzysztof Rafał Szymański
University of Białystok
Krzysztof Gawryluk
University of Białystok
Marek Brancewicz
m.brancewicz@uwb.edu.pl
University of Białystok

Article

Keywords: cochlear implants, auditory rehabilitation, pitch perception, perceptual learning, mirror-type mistakes

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9756183/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Transient mirror-type pitch inversion during auditory rehabilitation in cochlear implant and normal-hearing listeners

Krzysztof R. Szymański[1], Krzysztof Gawryluk[1] and Marek Brancewicz[1,*]

[1] University of Bialystok, Faculty of Physics, 15-245 Białystok, Poland
*m.brancewicz@uwb.edu.pl

Keywords

cochlear implants, auditory rehabilitation, pitch perception, perceptual learning, mirror-type mistakes

ABSTRACT

We analyse longitudinal auditory rehabilitation data from cochlear implant (CI) users performing melodic contour identification tasks using a remote platform. To address sparse and noisy measurements, we apply a neural multi-scale smoothing approach inspired by scale-space theory to reconstruct continuous learning trajectories. We identify a systematic error pattern – mirror-type pitch inversion (M-mistakes), defined as reversal of contour direction – that occurs more frequently than expected by chance and is present in both CI users and normal-hearing listeners. This indicates a general feature of pitch processing. We introduce a probabilistic model with three perceptual modes: correct recognition, pitch inversion, and random anticoincidence responses. The model links recognition score and M-mistake rate to underlying perceptual states. Our results show that M-mistakes form a transient but dominant mode at intermediate performance levels, revealing a non-monotonic learning process. Time-resolved trajectories demonstrate that listeners typically pass through a stage of systematic pitch inversion before achieving accurate perception. These findings suggest that such errors reflect structured intermediate representations and should be explicitly addressed in auditory training.

Introduction

Pitch perception is often considered a highly precise auditory ability, as reflected by low pitch discrimination thresholds in normal-hearing listeners. However, this precision does not directly translate to higher-level perceptual tasks such as interval or contour processing. In particular, interval discrimination thresholds are substantially higher than would be expected based on basic pitch sensitivity, suggesting that relational aspects of pitch perception involve additional complexity [1]. A striking illustration of this complexity is the tritone paradox, first described by Shepard [2] and further investigated by Deutsch [3–5], in which the perceived direction of pitch change between two tones separated by a tritone is ambiguous and depends on listener-specific factors such as language background and auditory experience. Moreover, experimental studies have revealed asymmetries in processing pitch direction: both normal-hearing and cochlear implant (CI) users may exhibit different performance for ascending and descending intervals, although the interpretation of this effect

remains unclear and may depend on experimental procedures [1]. Together, these observations indicate that pitch direction and contour processing are not trivial extensions of simple pitch encoding and may involve distinct and potentially ambiguous perceptual mechanisms.

Recent advances in remote auditory rehabilitation platforms enable continuous monitoring of perceptual skills in CI users over extended periods of time. In particular, the EARPLANTED in-home rehabilitation system described in [6] was designed to enhance pitch and timbre perception while simultaneously collecting large longitudinal datasets. The platform implements melodic contour identification tasks based on the paradigm introduced by Galvin et al. [7], which has become a standard tool for assessing pitch contour perception in CI listeners. The availability of such long-term behavioural data creates new opportunities to study the dynamics of auditory learning at an individual level.

However, real-world longitudinal measurements are typically sparse, noisy, and irregularly sampled, which limits the direct application of standard analytical methods. To address this limitation, we adopt a multi-scale analysis framework inspired by scale-space theory. In classical formulations, signals are represented across a continuum of Gaussian-smoothed scales, with the Gaussian kernel uniquely satisfying the axiomatic requirements of scale-space representations [8, 9]. More recently, these ideas have been extended using learned nonlinear operators implemented as deep neural networks, enabling adaptive smoothing tailored to empirical data distributions [10]. In the present work, we apply such neural multi-scale smoothing to reconstruct continuous trajectories of auditory rehabilitation from irregular behavioural observations.

To quantify perceptual dynamics, we analyse both recognition accuracy and structured error patterns in melodic contour identification. We identify a systematic class of mistakes—mirror-type pitch inversion (M-mistakes)—in which the direction of a melodic contour (rising vs. falling) is reversed. These errors occur more frequently than expected by chance and are observed in both CI users and normal-hearing (NH) participants, suggesting that they reflect a general property of pitch processing rather than a device-specific limitation.

To interpret these findings, we introduce a probabilistic framework that describes listener responses in terms of three perceptual modes: correct recognition, systematic pitch inversion, and random anticoincidence responses. This formulation links observable quantities—recognition score and M-mistake rate—to underlying perceptual states. The temporal evolution of these modes reveals a non-monotonic learning process in which listeners typically pass through a transient stage dominated by systematic inversion errors before achieving stable correct perception.

These results are consistent with broader research in perceptual learning, which indicates that learning proceeds through structured intermediate representations rather than simple monotonic refinement [11–13]. In auditory cognition, similar transitional effects have been reported in pitch and interval learning and are often associated with selective attention and experience-dependent plasticity [14, 15]. Music cognition studies further suggest that contour-based representations may dominate early stages of perception, preceding accurate encoding of pitch direction [16, 17], while efficient coding theories predict that relational representations of pitch can introduce directional ambiguities [18, 19].

Taken together, these considerations motivate the central hypothesis of this work: that structured error patterns such as mirror-type pitch inversion reflect intrinsic properties of auditory processing and learning dynamics. By combining large-scale longitudinal data with neural multi-scale analysis and probabilistic modelling, we aim to characterize these dynamics

and provide a unified framework for understanding both performance and error structure in auditory rehabilitation.

Methods

Basic characteristics

Nine categories of melodic contour are presented graphically to the participant (rising, flat-rising, rising-falling, flat, falling-rising, flat-falling, and falling). The task is to select the contour that best matches the perceived stimulus. The contours differ in the interval structure between successive non-flat tones, which can take values of 1, 2, or 4 semitones. This parameterization yields 25 distinct melodic contours for a given timbre (musical instrument). In addition, three different timbres are used, resulting in a total of 75 distinct melodic stimuli. Each stimulus is presented with equal probability. Further details of the experimental paradigm are provided in [6].

The underlying perceptual stimulus space consists of 25 distinct melodic contour classes. Participants, however, provide responses using a 9-category interface corresponding to coarse-grained contour labels. The probabilistic model is defined at the level of the 25-element stimulus space, while observed responses are obtained via a deterministic many-to-one mapping from the stimulus space onto the response categories.

A natural basic performance measure is the score defined as the fraction of correct responses k in n trials (total number of responses):

$$s = \frac{k}{n}. \tag{1}$$

Another important characteristic is the fraction of M-mistakes, denoted by m, defined as

$$m = \frac{k_M}{n}, \tag{2}$$

where k_M denotes the number of incorrect responses in which a rising contour is mistakenly identified as falling (or vice versa). These two measures will be further analysed, as they reflect important aspects of psychoacoustic processing. Both characteristics may be interpreted either as mean values over the rehabilitation process or as time-dependent quantities. In both cases, their uncertainty must be properly estimated. When treated as local time-dependent measures, appropriate averaging procedures must be applied.

Since the measured quantities are based on a finite number of trials, they should be treated as statistical estimates rather than exact values. In particular, the empirical success probability is defined as $p = k/n$. The uncertainty of this estimate is evaluated under the assumption of binomial statistics, reflecting the discrete nature of the data, possible asymmetry of confidence intervals, and limited sample sizes. For this reason, we apply the Clopper–Pearson method [6], which provides exact confidence intervals based on the cumulative distribution function of the binomial distribution. The cumulative distribution function is defined as:

$$F(k, n, p) \;=\; P(X \leq k | n, p), \tag{3}$$

where k is the number of observed successes, n is the number of trials, p is the success probability in a single trial. The lower and upper bounds of the confidence interval, p^L and p^U are obtained by solving

$$F(k-1,n,p^L)=\alpha_L, F(k,n,p^U)=\alpha_U, \quad (4)$$

where $\alpha_L = 0.159, \alpha_U = 0.841$ correspond to the one-sigma confidence limits (central 68% probability interval).

Data smoothing

Data smoothing was performed using a deep learning approach based on neural networks trained on fixed-length windows of the time series of length N. Each observation is represented by a five-dimensional feature vector comprising the time coordinate, measured probability, lower and upper uncertainty bounds, and a weighted local moving average, yielding an input dimensionality of $5N$. The model reconstructs a continuous function from irregular and noisy observations using a feed-forward architecture with three parallel branches whose outputs are combined linearly. Two branches contain hidden layers with nonlinear activations (softsign and hyperbolic tangent), while the third provides a direct linear mapping, improving reconstruction stability. Training was conducted on synthetically generated datasets derived from diverse analytical functions with Bernoulli-corrupted observations and stochastic augmentations, using the ADAM optimizer [20] and a mean squared error loss.

Real clinical data may contain an arbitrary number of points, which typically does not match the window size N used during training. To address this issue we apply a multi-scale sliding window approach. Separate networks are trained for several values of N ranging from 40 to 500. For a dataset containing M observations the largest network scale satisfying $N \leq M$ is selected as the main smoothing scale, the corresponding network is applied to all possible windows covering the data then the predictions from overlapping windows are averaged.

Standard sliding-window algorithms often perform poorly near the boundaries of the dataset, where fewer windows overlap. In our method this problem is mitigated using additional networks trained at smaller scales. Near the beginning and end of the time series the algorithm supplements the main scale predictions with networks corresponding to smaller window sizes. This strategy effectively performs adaptive scale selection, allowing the smoothing process to adjust to the reduced amount of available data. Such adaptive scale combinations are conceptually related to scale-space representations, where signals are analysed simultaneously at multiple levels of smoothing [8–10].

To estimate the uncertainty of the neural smoothing procedure itself, we employ a parametric bootstrap approach [21].

The neural multi-scale smoothing and adaptive scale combinations method allows us to reconstruct a continuous monitoring of auditory recovery including scores and M-mistakes over extended time periods.

Subjects

CI users were recruited through outreach efforts at partnering universities and advertisements on social media platforms. No financial compensation was provided for participation in the study. Initially, data from bilaterally implanted users were aggregated under a single account per participant. However, as the study progressed, it became advantageous to collect data separately for each implant; this approach was subsequently adopted as the standard for the majority of participants. Similarly, NH volunteers were recruited using the same dissemination channels. The volunteers did not undergo a medical examination; instead, they self-reported normal hearing in at least one ear. Each volunteer was assigned a single account. The characteristics of the participants are presented in Tab. 1.

A total of 285 282 melodic contour identifications were collected between October 24, 2022, and April 23, 2026.

Ethics statement

The study was reviewed and granted exemption from ethics approval by the Bioethics Committee of the Medical University of Bialystok (Reference Number: APK.002.586.2021; December 16, 2021). All procedures were performed in accordance with relevant guidelines and regulations and with the Declaration of Helsinki.

Users of the EARPLANTED platform provided informed consent electronically by accepting the platform's user agreement, which included information regarding data collection and the use of anonymized data for research purposes. Registration on the platform was not possible without providing such consent.

The study included two minor participants. For one participant, written informed consent was obtained from a parent or legal guardian. The second minor participant took part voluntarily using anonymous participation without collection of personally identifying information.

This manuscript contains no identifying participant information.

Tab. 1. Characteristics of the users.

	Mean Age [years]	Min. Age [years]	Max. Age [years]	Female no	Male no	Undiscl. no
CI	51.7	12.3	86.2	40	53	4
NH	32.4	15.1	74.3	62	28	0

Results

Recognition score and M-mistakes

The two main performance characteristics for cochlear implant (CI) users and normal-hearing (NH) participants are shown in Fig. 1 and Tab. 1. Horizontal lines indicate the performance level corresponding to purely random responses. Black markers represent confidence intervals, which become large in cases of poor statistical support (i.e., small numbers of trials). In some cases (bottom-right panel), participants achieve perfect recognition performance; here the M-mistake rate is zero, while the uncertainty remains finite due to the limited number of trials.

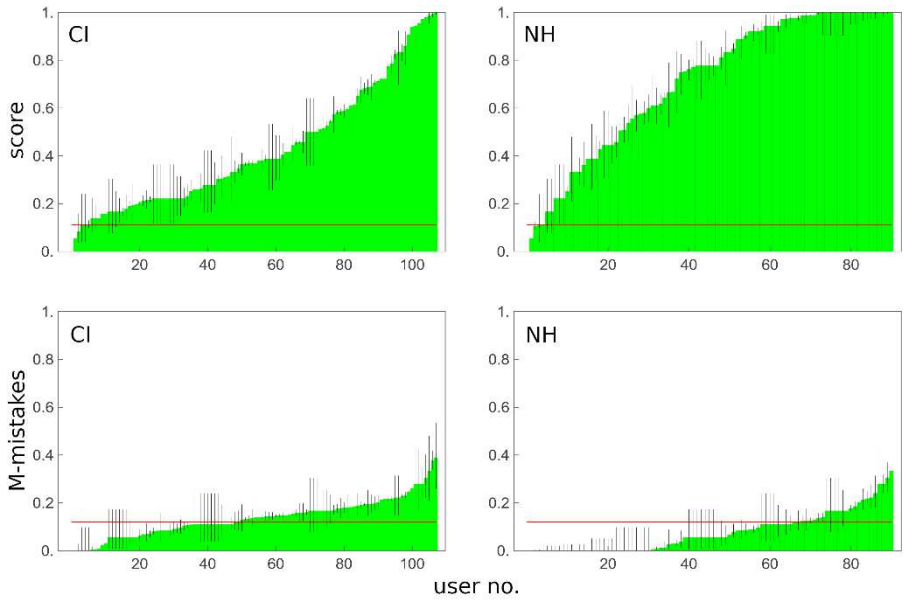


Fig. 1. Recognition score (top panels) and M-mistake rate (bottom panels) for cochlear implant (CI) users (left) and normal-hearing (NH) participants (right). Horizontal lines indicate the performance level expected for purely random responses.

Each trial requires the identification of one melodic contour from nine possible alternatives. Only one response is correct, while the remaining eight are incorrect. Among the incorrect options, one represents a mirror-type mistake (M-mistake), defined as a reversal of the pitch contour direction, such that a rising contour is perceived as falling and vice versa. Under the assumption of random responding, the probability of selecting a mirror-type response is $p_M = 0.12$, which serves as the statistical baseline expected by chance.

Across the dataset, observed M-mistake rates are frequently higher than this baseline, typically ranging from 0.1 to 0.3 and in some cases exceeding 0.4. This indicates that mirror-type responses occur more often than expected under random conditions. Consequently, these errors cannot be interpreted as purely stochastic; rather, they reflect systematic perceptual confusions in the processing of pitch contour direction.

A correlation between recognition score and M-mistake rate is shown in Fig. 2. Participants with low recognition scores generally exhibit M-mistake rates close to the random baseline (approximately 0.1–0.2). In this regime, responses are likely dominated by guessing, as the contour structure is not reliably perceived. Interestingly, some participants show M-mistake rates below the random baseline, suggesting that incorrect responses are not uniformly distributed but instead biased toward other types of mistakes.

As recognition performance increases, the M-mistake rate also tends to increase. Participants with intermediate and high scores (typically above ~0.4) often exhibit M-mistake rates between 0.2 and 0.4. This pattern suggests that while the global contour shape is correctly perceived, the direction of pitch movement is frequently misclassified, leading to systematic reversal mistakes.

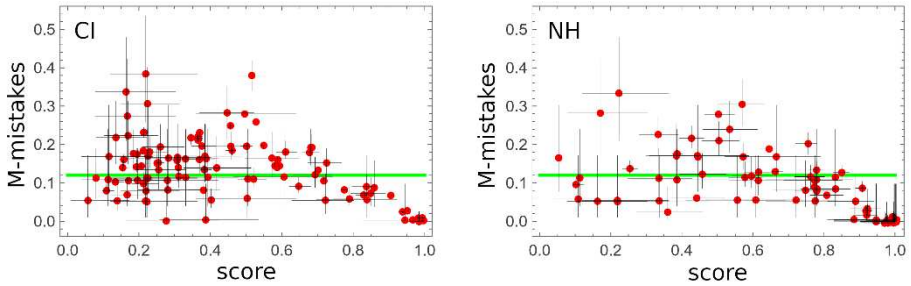


Fig. 2. Average recognition score versus M-mistake rate for cochlear implant users (left) and normal-hearing participants (right). The green horizontal line indicates the value expected from random guessing ($p_M = 0.12$).

Different perceptual modes

The relationship between the global score and the number of M-mistakes can be interpreted within a probabilistic framework in which listener responses arise from three distinct perceptual modes (Mode 1, Mode 2 Mode 3). Each mode corresponds to a characteristic processing strategy and is associated with a specific probability of occurrence.

In Mode 1, the listener accurately represents the melodic contour, leading to consistently correct responses. In Mode 2, the contour shape is correctly identified but the pitch direction is systematically inverted; as a result, all non-flat contours yield M-mistakes. In Mode 3, the listener fails to extract the contour and responds by randomly selecting from the available options, excluding the mirror contour. Consequently, the mirror response does not occur in this mode, which can be interpreted as an *anticoincidence* strategy.

Consider the set of melodic contours proposed by Galvin [7]. The system contains one flat contour and eight non-flat contours (Fig. 3). In the present platform the non-flat contours are realized with three different interval sizes, which yields in total 1+3×8=25 distinct contours. All contours are presented with equal probability. For each perceptual mode one may therefore compute the expected values of two basic characteristics: the score s, defined as the probability of a correct response (Eq. 1), and the parameter m, defined as the probability of an M-mistake (Eq. 2).

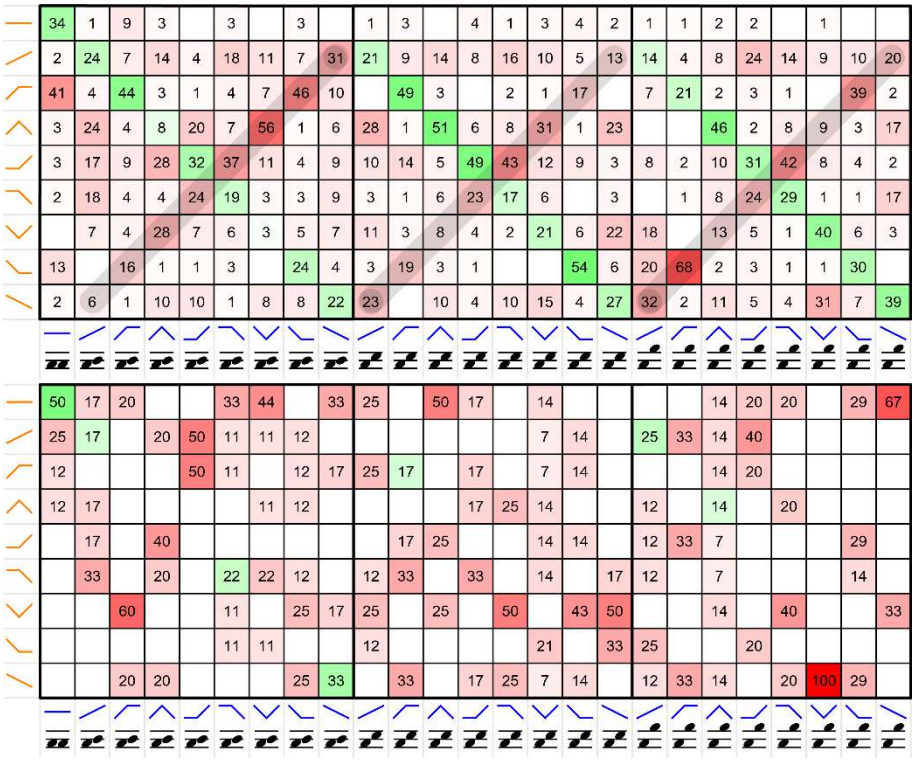


Fig. 3. Representative confusion matrices for users with IDs 212015 (top) and 311071 (bottom). Blue symbols schematically depict melodic contours, with note labels indicating interval sizes. Red symbols represent the nine response options in the Galvin test. The main diagonals, highlighted by green squares, indicate correct-response probabilities, whereas red squares denote error probabilities (all values in %). Colour intensity reflects the corresponding probabilities. Empty cells correspond to 0% and are omitted for clarity. In the top panel, thick grey lines mark antidiagonals associated with enhances probabilities of M-mistakes.

In the Mode 1 all responses are correct; thus, for 25 trials the listener produces 25 correct answers and no errors of any kind, which corresponds to $s = 1$ and $m = 0$. In the second mode only the flat contour is identified correctly, whereas for all other contours the direction is inverted, producing on average 24 M-mistakes and no other errors. Consequently $s = 1/25$ and $m = 24/25$. In the third mode the listener answers randomly under the anticoincidence constraint. Over 25 trials the expected outcome consists of three correct responses and twenty-two incorrect responses, none of which are M-mistakes. Hence $s = 3/25$ and $m = 0$. Let the probabilities of occurrence of the three perceptual modes be denoted by P, Q and R, respectively. Although other perceptual modes may exist, the present model is restricted to these three. The expected values of the observable characteristics s and m can then be expressed as weighted averages of the mode-specific predictions:

$$s = P + \frac{1}{25}Q + \frac{3}{25}R, \tag{5}$$

$$m = \frac{24}{25}Q. \tag{6}$$

Since the probabilities must satisfy the normalization condition

$$P + Q + R = 1, \tag{7}$$

the mode probabilities can be reconstructed from the measured characteristics s and m. Solving the system yields

$$P = \frac{300s + 25m - 36}{264}, \tag{8}$$

$$Q = \frac{25m}{24}, \tag{9}$$

while the remaining probability is given by

$$R = \frac{25(1 - m - s)}{22}. \tag{10}$$

The model predicts characteristic patterns in the relationship between the mean score and the mirror-type error rate. In particular, M-mistakes are expected to be rare when the listener operates predominantly in Mode 3, because the mirror contour is excluded from the random response set. As the listener begins to recognize contour shapes but still confuses pitch direction, Mode 2 becomes dominant, producing a large proportion of mirror-type errors. Finally, when full contour recognition develops, Mode 1 dominates and the overall error rate becomes very small. This interpretation is consistent with the empirical observations. Table 2 illustrates representative examples from the dataset showing typical results covering score s from low to high values. Table for all users is given in Supplementary material.

Tab. 2. Summary of key metrics for representative users, including score (s), number of M-type mistakes (m), and probabilities associated with Modes 1–3 (P, Q, R).

ID	s	m	P	Q	R
311241	0.39(2)	0.14(1)	0.32(2)	0.14(1)	0.54(2)
311174	0.457(6)	0.248(3)	0.406(7)	0.258(6)	0.336(9)
311149	0.714(4)	0.109(3)	0.685(5)	0.113(3)	0.202(5)
210001	0.904(2)	0.070(2)	0.898(2)	0.073(2)	0.029(2)

Listeners with very low scores exhibit relatively small mirror-type probabilities. For example, user 311241 has a score of 0.39 and a mirror-type rate of 0.14, implying that only about 14% of trials correspond to Mode 2, while the majority correspond to the anticoincidence random mode. At intermediate performance levels, mirror-type errors become much more prominent. User 311174, with a score of 0.46, shows a mirror-type rate of 0.25, indicating a substantial contribution of Mode 2 responses. In this regime, listeners appear to recognize the contour shape but systematically invert pitch direction. At very high performance levels, the overall error rate becomes small, and mirror-type errors represent only a small fraction of all trials. For example, user 210001 achieves a score of 0.90, corresponding to a total error rate of 0.10. Although the mirror-type rate among these errors is relatively high, the absolute probability of mirror responses and probability of Mode 2 remains small.

A correlation between the probabilities P and Q of Modes 1 and 2, respectively, for all users is shown in Fig. 4. It is expected that as P approaches 1, Q should tend to 0. This behavior is indeed clearly observed in Fig. 4. However, an unexpected feature is that Q exhibits a maximum for intermediate values oP (approximately 0.4–0.6). To quantify this effect, we performed polynomial fitting of the data using orthogonal distance regression [22], as

uncertainties are present in both variables P and Q. Since the exact functional form describing the correlation is unknown and the uncertainties are relatively large, we tested polynomial models of orders 2, 3, and 4. The resulting fits are shown as lines in Fig. 4 and consistently reveal a pronounced maximum for both implanted and non-implanted patients, independent of the polynomial order used.

Correlation between the R and P probabilities are shown in Fig. 5. As expected, for poor recognition anticoincidence mode is dominating and tend to zero when recognition tend correct one.

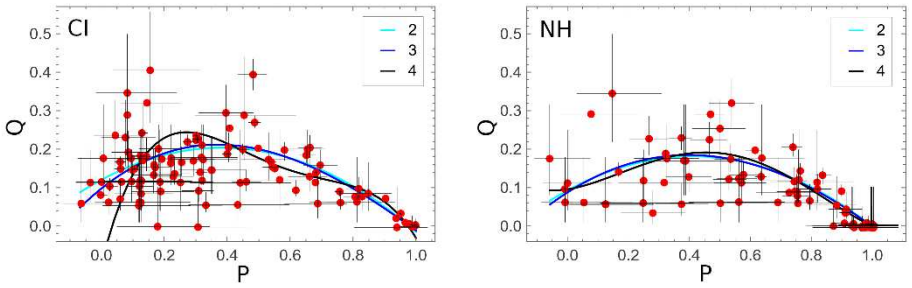


Fig. 4. Correlation between probabilities of Mode 1 (P) and Mode 2 (Q). The lines represent an orthogonal distance regression fits using an n-th order polynomial model $Q(P)$, where n is indicated in the legend.

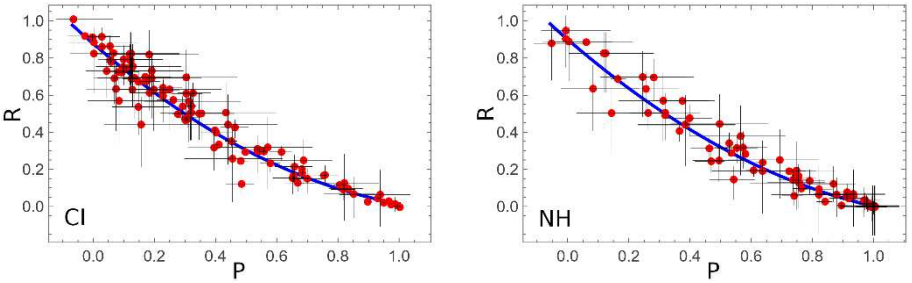


Fig. 5. Correlation between probabilities of Mode 1 (P) and Mode 3 (R). Blue line is result of orthogonal distance regression fit of the second order polynomial $R(P)$.

Dynamics of individual perceptual modes

The proposed model has so far been successfully applied to characterize average properties across all users, both implanted and normal-hearing (NH). However, a subset of users undergo particularly intensive training, making it possible to observe changes in their individual temporal perceptual modes. An example of such behaviour is shown in Fig. 6 for user ID = 311231. The data, grouped into one-hour intervals (light colours), exhibit substantial dispersion, indicating strong variability. These fluctuations are largely statistical in nature rather than physiological. To address this, a smoothing procedure based on the deep machine learning algorithm described in the Methods section was applied. The resulting smoothed data (dark colours in Fig. 6) reveal a much clearer temporal trajectory while preserving experimental uncertainty. The contrast between raw and smoothed data is particularly striking: whereas one-hour binning introduces significant noise, the neural multi-scale smoothing uncovers the underlying structure of the perceptual dynamics.

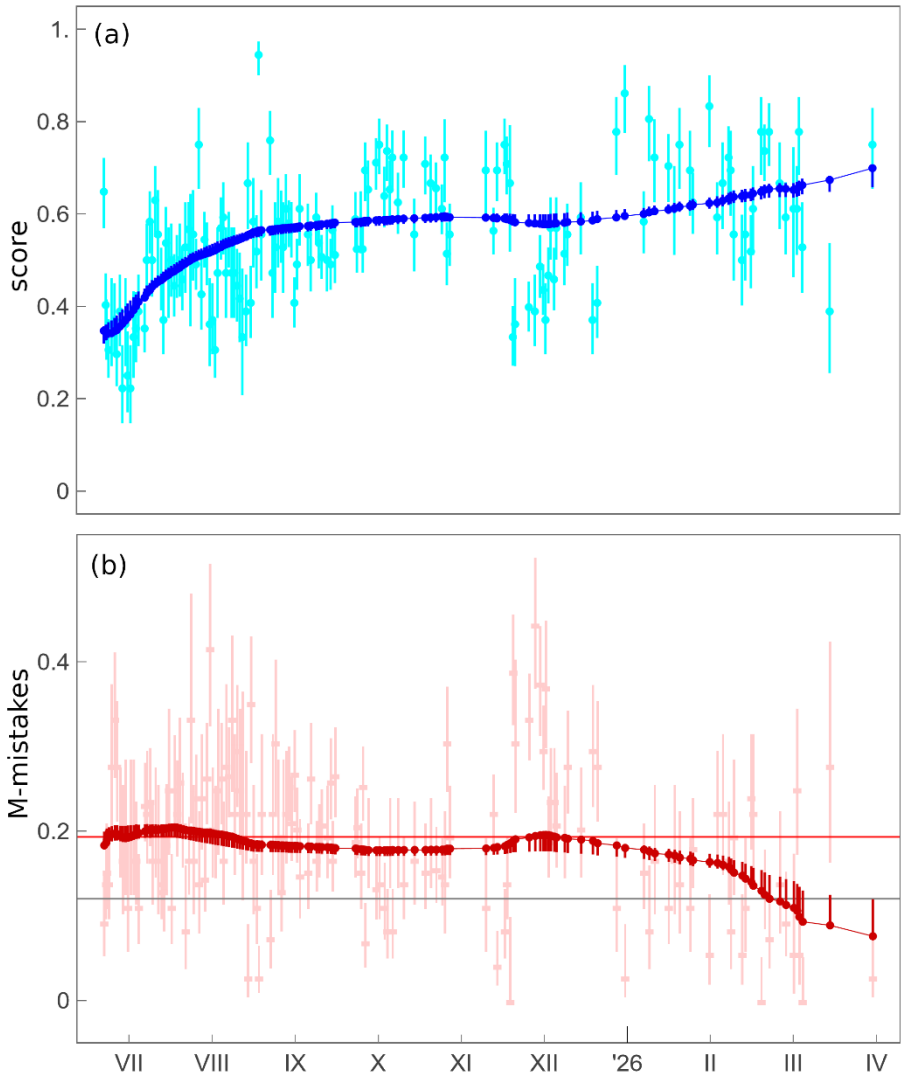


Fig. 6. (a) Time evolution of the score s for user 311231. Light blue points represent values obtained from one-hour data grouping, while the dark blue curve shows the result of neural multi-scale smoothing. (b) Time evolution of the M-mistake rate m. Light red points correspond to one-hour grouped data, and the dark red curve shows the smoothed estimate. Error bars represent experimental uncertainties. The gray horizontal line indicates the expected error level for random responses, while the red horizontal line denotes the mean M-mistake rate.

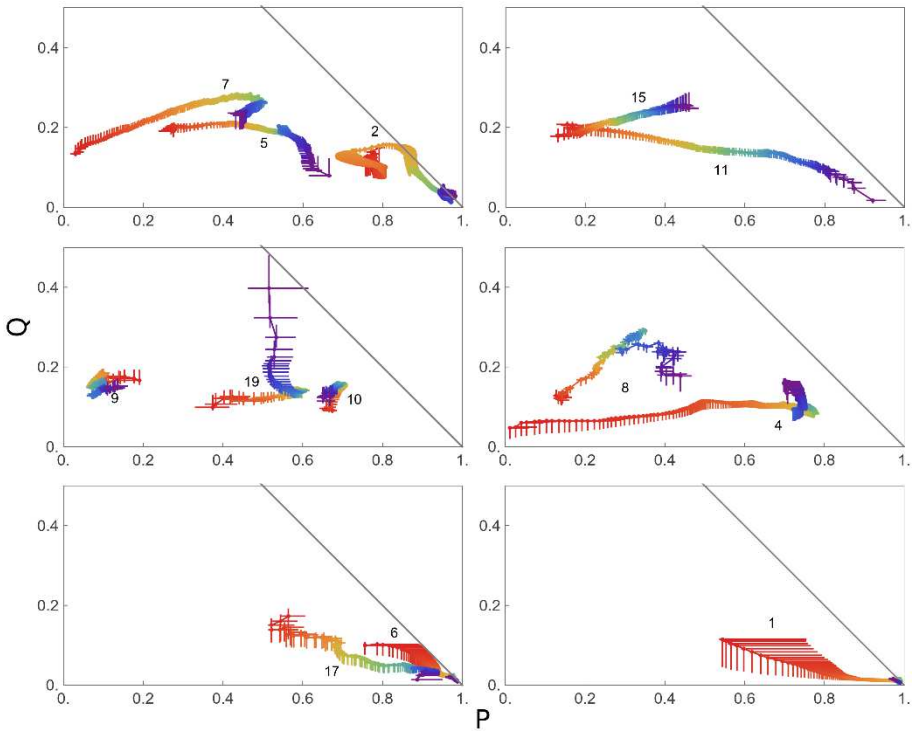


Fig. 7. Individual time-dependent trajectories in the (P, Q) plane, representing the probabilities of Mode 1 (correct recognition) and Mode 2 (mirror confusion), respectively. Each trajectory is color-coded according to time, with red indicating early stages of training and violet corresponding to later stages. Error bars of matching colors denote uncertainties in the estimated probabilities. Gray line corresponds to condition $P + Q = 1$. Numbers adjacent to the trajectories identify individual users: 1 – 311152, 2 – 210001, 4 – 311149, 5 – 311231, 6 – 2110011, 7 – 311174, 8 – 311071, 9 – 311133, 10 – 311081, 11 – 311132, 15 – 210004, 17 – 211005, 19 – 311165.

Fig. 7 provides a compact representation of these dynamics in the (P, Q) space obtained for most active users. Within experimental uncertainty, all (P, Q) points lie in the region $P + Q \leq 1$; see the gray line in Fig. 7, which delineates the boundary of the admissible region. The trajectories generally move toward higher P and lower Q, consistent with learning progression. A key feature is that many trajectories pass through a region where Q is relatively high before decreasing. This supports the interpretation that listeners often enter a transitional phase in which contour shape is recognized but pitch direction is systematically inverted. In addition, the trajectories are clearly non-linear. They follow curved paths, frequently exhibiting a pronounced local maximum in Q at intermediate values of P. This observation is consistent with the previously identified maximum of Q at moderate P (≈0.4–0.6), indicating that Mode 2 represents a distinct and prominent perceptual strategy during learning rather than a residual source of error. Finally, inter-individual variability is clearly visible. While the overall trend is consistent across users, the speed of progression and the extent of the intermediate Mode 2 dominance differ substantially. This suggests that the learning process is heterogeneous and may depend on factors such as training intensity, prior experience, or individual auditory processing characteristics.

Discussion

The present results show that auditory rehabilitation does not follow a simple monotonic trajectory toward improved performance, but rather progresses through transitions between distinct perceptual modes. At low performance levels, responses are dominated by the anticoincidence random mode (Mode 3), characterized by largely random choices that avoid mirror contours. As performance increases, a clear intermediate regime emerges in which mirror-type errors (Mode 2) become prominent. The observed maximum of Mode 2 at intermediate values of P (≈0.4–0.6) indicates that listeners often acquire contour shape representations before correctly encoding pitch direction. Thus, systematic errors appear to constitute an intrinsic stage of perceptual learning rather than mere stochastic noise.

At higher performance levels, Mode 1 (correct recognition) gradually dominates while Mode 3 diminishes. Notably, mirror-type confusions persist as the principal remaining error category, suggesting that pitch direction encoding is more demanding than contour shape recognition and may rely on partially separable perceptual mechanisms. This interpretation is consistent with previous observations showing that listeners often recognize melodic contour while misjudging pitch direction [7]. In CI users, such effects have been linked to limited spectral resolution and degraded temporal fine structure [23], although broader mechanisms of pitch perception likely also contribute [24].

A key observation is that the relationship between recognition score, mirror-type error rate, and perceptual modes is remarkably similar in cochlear implant (CI) users and normal-hearing (NH) listeners. This convergence suggests that mirror-type confusions reflect general properties of pitch contour processing rather than limitations specific to degraded auditory input [7]. The finding is consistent with evidence that interval discrimination performance can exhibit comparable characteristics in CI and NH listeners [25], supporting the view that relational pitch processing imposes common perceptual constraints across listener groups.

The transient dominance of Mode 2 provides empirical support for the hypothesis that contour shape and pitch direction are processed as partially separable dimensions. Foundational work in music cognition has shown that melodic contour constitutes a primary perceptual cue [26], often preceding accurate encoding of intervals or directional relations [16, 17]. The present results extend this framework by demonstrating that, during auditory rehabilitation, contour representations may become functionally available before pitch direction is reliably encoded. This produces an intermediate perceptual regime in which inverted responses form a coherent and dominant response pattern rather than a random collection of mistakes.

Importantly, the present study provides a dynamic perspective on auditory learning. Rather than relying exclusively on static performance measures, we track trajectories of perceptual modes over time and show that learning follows structured paths in the (P, Q) space. Many users pass through a transitional phase characterized by elevated Mode 2 probability before reaching stable Mode 1 dominance. The observed trajectories are notably non-monotonic: increases in recognition performance can temporarily coincide with increases in structured error rates. Such behavior supports the view that perceptual learning involves qualitative transitions between internal representations rather than continuous refinement alone [11–13].

This interpretation is compatible with broader theories of auditory plasticity emphasizing adaptive reorganization of perceptual categories through attention and experience-dependent learning [14, 15]. During rehabilitation, listeners may progressively

reweight perceptual cues, initially relying on contour structure while only later refining directional pitch encoding. In this context, Mode 2 may reflect the emergence of an internally consistent but incomplete perceptual mapping, in which contour relations are preserved but directional assignment remains systematically distorted.

The present findings may also be interpreted in relation to known ambiguities of pitch perception. Phenomena such as the tritone paradox demonstrate that perceived pitch direction can depend on listener-specific factors, including language background, vocal range, auditory experience, and cognitive representations [2–5]. Similarly, efficient relational pitch coding may introduce directional uncertainty even in normal-hearing listeners [18, 19]. The occurrence of mirror-type errors in both CI and NH groups therefore suggests that such confusions arise from fundamental properties of pitch processing rather than exclusively from degraded cochlear implant input.

From a practical perspective, these findings have direct implications for auditory rehabilitation. Mirror-type errors should not be interpreted solely as undesirable failures but as informative indicators of an intermediate stage of perceptual organization. Training protocols may therefore benefit from explicitly targeting pitch direction encoding rather than focusing exclusively on overall recognition accuracy. In particular, extending stimulus sets beyond standard paradigms such as Galvin-type contours [7] and introducing controlled pitch ambiguities may promote more robust and flexible perceptual representations.

Taken together, the results support a view of auditory rehabilitation as a process shaped by transitions between distinct perceptual states. Errors, especially mirror-type inversions, emerge not merely as residual noise but as meaningful markers of the current stage of perceptual organization and its underlying constraints. Incorporating this dynamic perspective into rehabilitation strategies may lead to a more mechanistic understanding of learning and ultimately to more effective auditory training.

Summary

This study shows that auditory rehabilitation follows a structured, non-monotonic progression through distinct perceptual modes rather than simple gradual improvement. M-errors emerge as a systematic and transiently dominant error pattern at intermediate performance levels in both cochlear implant users and normal-hearing listeners, indicating a general property of pitch processing.

The proposed three-mode model reveals that listeners typically learn contour shape before correctly encoding pitch direction, passing through a stage where inversion errors reflect a coherent perceptual strategy rather than noise. Using neural multi-scale smoothing, we reconstruct continuous learning trajectories that consistently pass through this intermediate regime.

These findings highlight that systematic mistakes are an integral part of perceptual learning and should be explicitly addressed in auditory training, offering a more dynamic and mechanistic understanding of rehabilitation.

References

[1] McDermott, J. H., Keebler, M. V., Micheyl, C. & Oxenham, A. J. Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *J. Acoust. Soc. Am.* **128**, 1943–1951, DOI: https://doi.org/10.1121/1.3478785 (2010).

[2] Shepard, R. N. Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* **36**, 2346–2353, DOI: https://doi.org/10.1121/1.1919362 (1964).

[3] Deutsch, D., Henthorn, T. & Dolson, M. Speech patterns heard early in life influence later perception of the tritone paradox. *Music Percept.* **21**, 357–372, DOI: https://doi.org/10.1525/mp.2004.21.3.357 (2004).

[4] Deutsch, D. The tritone paradox: An influence of language on music perception. *Music Percept.* **8**, 335–347, DOI: https://doi.org/10.2307/40285517 (1991).

[5] Deutsch, D. *The Psychology of Music* 3rd edn (Academic Press, 2013), ISBN: 9780123814609.

[6] Szymański, K. R., Gawryluk, K. & Brancewicz, M. Continuous monitoring of temporal skills during long-term in-home training by cochlear implant users. *Heliyon* **11**, e41817, DOI: https://doi.org/10.1016/j.heliyon.2025.e41817 (2025).

[7] Galvin, J. J., Fu, Q. J. & Nogaki, G. Melodic contour identification by cochlear implant listeners. *Ear Hear.* **28**, 302–319, DOI: https://doi.org/10.1097/01.aud.0000261689.35445.20 (2007).

[8] Koenderink, J. J. The structure of images. *Biol. Cybern.* **50**, 363–370 DOI: https://doi.org/10.1007/BF00336961 (1984).

[9] Lindeberg, T. Scale-Space *Theory in Computer Vision* (Kluwer Academic Publishers, 1994), DOI: https://doi.org/10.1007/978-1-4757-6465-9.

[10] Yang, Z., Zhuang, X., Lowe, M. J. & Cordes, D. A deep neural network for adaptive spatial smoothing of task fMRI data. *Front. Neuroimaging* **4**, 1554769, DOI: https://doi.org/10.3389/fnimg.2025.1554769 (2025).

[11] Goldstone, R. L. Perceptual learning. *Annu. Rev. Psychol.* **49**, 585–612, DOI: https://doi.org/10.1146/annurev.psych.49.1.585 (1998).

[12] Seitz, A. R. & Watanabe, T. A unified model for perceptual learning. *Trends Cogn. Sci.* **9**, 329–334, DOI: https://doi.org/10.1016/j.tics.2005.05.010 (2005).

[13] Ahissar, M. & Hochstein, S. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* **8**, 457–464, DOI: https://doi.org/10.1016/j.tics.2004.08.011 (2004).

[14] Francis, G. & Nusbaum, H. C. Selective attention and the acquisition of new phonetic categories. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 349–366, DOI: https://psycnet.apa.org/doi/10.1037/0096-1523.28.2.349 (2002).

[15] Siegel, J. A. & Siegel, W. Absolute identification of notes and intervals by musicians. *Percept. Psychophys.* **21**, 143–152, DOI: https://doi.org/10.3758/BF03198717 (1977).

[16] Burns, E. M. & Ward, W. D. Intervals, scales, and tuning in *The Psychology of Music* (ed. Deutsch, D.) 241–269 (Academic Press, 1982), ISBN: 9781483292731.

[17] Krumhansl, C. L. Rhythm and pitch in music cognition. *Psychol. Bull.* **126**, 159–179 (2000), ISBN: 9780122135620.

[18] McPherson, M. J. & McDermott, J. H. Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proc. Natl. Acad. Sci. USA* **117**, 32169–32180, DOI: https://doi.org/10.1073/pnas.2008956117 (2020).

[19] Moore, B. C. J. *An Introduction to the Psychology of Hearing* 6th edn (Brill, 2012), ISBN: 978-9004252424.

[20] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. Preprint at https://doi.org/10.48550/arXiv.1412.6980 (2014).

[21] Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **7**, 1–26, DOI: https://doi.org/10.1214/aos/1176344552 (1979).

[22] Fuller, W. A. *Measurement Error Models* (John Wiley & Sons, 1987), DOI: https://doi.org/10.1002/9780470316665.

[23] McDermott, H. J. Music perception with cochlear implants: A review. *Trends Amplif.* **8**, 49–82, DOI: https://doi.org/10.1177/108471380400800203 (2004).

[24] Oxenham, A. J. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification.* **12**, 316–331, DOI: https://doi.org/10.1177/1084713808325881 (2008).

[25] Luo, X., Masterson, M. E. & Wu, C. C. Melodic interval perception by normal-hearing listeners and cochlear implant users. *J. Acoust. Soc. Am.* **136**, 1831–1844, DOI: https://doi.org/10.1121/1.4894738 (2014).

[26] Krumhansl, C. L. *Cognitive Foundations of Musical Pitch* (Oxford University Press, 1990), ISBN: 9780195148367.

Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

Acknowledgements

The authors thank Prof. Piotr Zieliński, Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego, for valuable discussions.

Funding

This work was supported by the Polish Ministry of Science and Higher Education through a subsidy for maintaining the research potential of the Faculty of Physics, University of Bialystok.

Competing interests

The authors declare no competing interests.

Author contributions

K.R.S. conceived the study, performed data analysis, developed the probabilistic model, performed computational analysis and data smoothing, interpreted the results, prepared the figures, and wrote the main manuscript text.

K.G. developed the auditory rehabilitation platform, collected the data, and contributed to manuscript preparation.

M.B. contributed to the development of the auditory rehabilitation platform, contributed to data analysis, and contributed to manuscript preparation.

All authors reviewed and approved the manuscript.

Table legends

Table 1.

Characteristics of the users.

Table 2.

Summary of key metrics for representative users, including score (s), number of M-type mistakes (m), and probabilities associated with Modes 1–3 (P, Q, R).

Figure legends

Figure 1.

Recognition score (top panels) and M-mistake rate (bottom panels) for cochlear implant (CI) users (left) and normal-hearing (NH) participants (right). Horizontal lines indicate the performance level expected for purely random responses.

Figure 2.

Average recognition score versus M-mistake rate for cochlear implant users (left) and normal-hearing participants (right). The green horizontal line indicates the value expected from random guessing ($p_M = 0.12$).

Figure 3.

Representative confusion matrices for users with IDs 212015 (top) and 311071 (bottom). Blue symbols schematically depict melodic contours, with note labels indicating interval sizes. Red symbols represent the nine response options in the Galvin test. The main diagonals, highlighted by green squares, indicate correct-response probabilities, whereas red squares denote error probabilities (all values in %). Colour intensity reflects the corresponding probabilities. Empty cells correspond to 0% and are omitted for clarity. In the top panel, thick grey lines mark antidiagonals associated with enhances probabilities of M-mistakes.

Figure 4.

Correlation between probabilities of Mode 1 (P) and Mode 2 (Q). The lines represent an orthogonal distance regression fits using an n-th order polynomial model $Q(P)$, where n is indicated in the legend.

Figure 5.

Correlation between probabilities of Mode 1 (P) and Mode 3 (R). Blue line is result of orthogonal distance regression fit of the second order polynomial $R(P)$.

Figure 6.

(a) Time evolution of the score s for user 311231. Light blue points represent values obtained from one-hour data grouping, while the dark blue curve shows the result of neural multi-scale smoothing. (b) Time evolution of the M-mistake rate m. Light red points correspond to one-hour grouped data, and the dark red curve shows the smoothed estimate. Error bars represent experimental uncertainties. The gray horizontal line indicates the expected error level for random responses, while the red horizontal line denotes the mean M-mistake rate.

Figure 7.

Individual time-dependent trajectories in the (P, Q) plane, representing the probabilities of Mode 1 (correct recognition) and Mode 2 (mirror confusion), respectively. Each trajectory is color-coded according to time, with red indicating early stages of training and violet corresponding to later stages. Error bars of matching colors denote uncertainties in the estimated probabilities. Gray line corresponds to condition $P + Q = 1$. Numbers adjacent to the trajectories identify individual users: 1 – 311152, 2 – 210001, 4 – 311149, 5 – 311231, 6 – 2110011, 7 – 311174, 8 – 311071, 9 – 311133, 10 – 311081, 11 – 311132, 15 – 210004, 17 – 211005, 19 – 311165.