Hearing Research 477 (2026) 109639

Contents lists available at ScienceDirect

Hearing Research

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

Training of pitch contour perception: Influence of pitch processing abilities and multisensory interactions

Caliani Hoarau[a,b,*], Agathe Pralus[a], Salomé Serres-Blain[a], Lesly Fornoni[a], Patrick Bouchet[a], Barbara Tillmann[a,c], Yohana Lévêque[a], Anne Caclin[a]

[a] Université Claude Bernard Lyon 1, INSERM, CNRS, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500, Bron, France
[b] Humans Matter, Lyon, France
[c] Laboratory for Research on Learning and Development, Université de Bourgogne, LEAD-CNRS UMR5022, Dijon, France

ARTICLE INFO

Keywords:
Amusia
Pitch contour
Pitch training
Multisensory
Unisensory

ABSTRACT

Pitch contour processing is critical in music and speech processing, and is likely to be hampered in peripheral and central auditory disorders. Pitch processing is typically altered in congenital amusia, a neurodevelopmental disorder characterized by deficits of music perception and production, which are related to altered pitch processing. In this study, we examined the effects of multisensory (audio-visual, AV) versus unisensory (auditory-only, A) pitch contour training on the ability to discriminate pitch direction changes in individuals with amusia and their matched control participants. After two consecutive sessions of training (A and AV in a counterbalanced order), improvement in accuracy and reaction times (RTs) was found in both participant groups, with the largest RT improvements for the participants with worst RTs at study onset. When evaluating separately the effect of each training type (A or AV) on performance, amusic participants exhibited improvement in both accuracy and reaction times for large pitch changes after the AV training. In control participants, AV training led to faster reaction times for the small and large pitch changes. AV training gains were numerically larger than A training gains, but a direct comparison provided weak evidence for a difference in these training gains. Overall, both amusic and control participants showed improved performance after the training (whether A or AV), indicating that pitch contour processing can be trained even after a single session, with some limited evidence for more efficient training in a multisensory context. These findings offer new perspectives for long-term audio-visual pitch training in congenital amusia and for populations with other types of pitch perception deficits.

1. Introduction

Pitch perception plays a crucial role in auditory cognition, notably for music processing, speech processing, and auditory scene analysis. In music, pitch processing is necessary to recognize melodies and perceive harmonies. In the context of speech processing, pitch is one of the cues conveying prosody, and pitch contour contributes to intonation, leading to the understanding of a speaker's intention and emotion (Nooteboom, 1997; Pihan, 2006; Tang et al., 2017). Moreover, pitch is essential in auditory scene analysis. In a noisy environment, it allows listeners to segregate sounds from different sources (Oxenham, 2008). Thus, pitch processing contributes to making sense of our auditory environment and interpreting complex auditory stimuli.

Pitch perception can be impaired in peripheral and central auditory disorders. Pitch perception deficits and hearing-in-noise difficulties, which rely on pitch perception, are reported both in Cochlear Implant (CI) users (Gates and Mills, 2005; Limb and Roy, 2014; Oxenham, 2008; Pralus et al., 2021) and presbycusis patients (Lavie et al., 2014; Schneider and Pichora-Fuller, 2001). Pitch perception deficits can also be observed in acquired or neurodevelopmental central auditory deficits (Ayotte et al., 2002; Clark et al., 2015; Peretz et al., 2002). The latter is the case in congenital amusia, a neurodevelopmental disorder characterized by a deficit of music perception and production (see Ayotte et al., 2002; Foxton, Dean, et al., 2004; Peretz et al., 2002; Peretz and Hyde, 2003, for the first description of this disorder). Congenital amusia is characterized by difficulties in detecting out-of-key notes in a melody, singing in tune, and recognizing familiar melodies without lyrics. Congenital amusia cannot be explained by hearing loss, brain damage, or lack of music exposure (Peretz, 2013), and is estimated to affect about 1.5–4% of the population (Peretz et al., 2003; Peretz and Vuvan, 2017).

* Corresponding author.
E-mail address: caliani.hoarau@inserm.fr (C. Hoarau).

https://doi.org/10.1016/j.heares.2026.109639
Received 20 July 2025; Received in revised form 21 April 2026; Accepted 23 April 2026
Available online 29 April 2026

The core deficit in congenital amusia is an impairment in pitch perception and memory (see Tillmann et al., 2023, for a recent review). In particular, individuals with amusia have difficulties in pitch direction and pitch contour processing tasks, such as when they are required to determine which of two gliding tones is rising up (Foxton, Dean, et al., 2004) or the direction of discrete (non-gliding) pitch changes (Hoarau et al., 2024; Loui et al., 2008). Interestingly, this difficulty in identifying pitch direction is observed with both music and speech materials (Foxton, Dean, et al., 2004; Liu et al., 2010, 2012).

Pitch perception can be enhanced thanks to auditory training. Music training, in particular, involves extensive pitch processing and has been shown to enhance both music and speech processing. When comparing musicians and non-musicians, numerous studies revealed better pitch discrimination abilities in musician participants than in non-musician participants (Fujioka et al., 2004; Herholz and Zatorre, 2012; Kishon--Rabin et al., 2001; Magne et al., 2021; Micheyl et al., 2006; Tervaniemi et al., 2005). After pitch discrimination training, even non-musician participants improved their performance compared to untrained participants (e.g., Carcagno and Plack, 2011; Grimault et al., 2002; Micheyl et al., 2006; Whiteford and Oxenham, 2018). Beyond pitch discrimination, pitch contour processing can also be improved by specific training in non-musician participants, with for example improvements in pitch direction discrimination threshold (Liu et al., 2017) and pitch contour pattern comparison (Foxton, Brown, et al., 2004). These results suggest that both pitch discrimination and pitch contour processing can be improved with auditory training.

Auditory training also allows for improving auditory perception in peripheral and central auditory disorders, and this despite the associated pitch perception deficits. Cochlear implants (CI) users, for example, improved in contour identification performance after a pitch contour identification training (Galvin et al., 2007). In congenital amusia, studies showed that individuals could benefit from auditory training to enhance their pitch contour processing (Liu et al., 2017) and pitch discrimination abilities (Whiteford and Oxenham, 2018). Thus, it would be valuable to explore ways to optimize the training in order to maximize its effectiveness. Notably, it is well documented that multisensory integration can enhance perception (Albouy et al., 2015; McGurk and MacDonald, 1976). Visual cues, whether informative (i.e., conveying information congruent with the content of the auditory stimuli) or non-informative (i.e., conveying only temporal information congruent with the auditory stimuli), improve auditory performance in speech perception tasks (Bernstein et al., 2013; McGurk and MacDonald, 1976; Rouger et al., 2008; von Kriegstein, 2012), pitch perception tasks (Albouy et al., 2015; Pralus et al., 2021), and various non-verbal auditory tasks (Shams and Seitz, 2008; von Kriegstein, 2012). For example, judging a tone as being higher or lower is easier when visual stimuli are presented and are spatially aligned (lower or higher positions) to the auditory stimuli, compared to incongruent or auditory-only stimuli (Lidji et al., 2007; Melara and O'Brien, 1987; Pralus et al., 2021; Rusconi et al., 2006). Furthermore, it has been reported that multisensory integration becomes stronger when one of the sensory modalities is impaired (Frassinetti et al., 2005; Rouger et al., 2007). Albouy et al. (2015) observed that non-informative visual cues associated with auditory stimuli could help individuals with congenital amusia and matched control participants to improve their pitch processing with higher accuracy and shorter reaction times compared to a condition with auditory stimuli only. Capitalizing on these facilitatory audiovisual interactions, using multisensory integration in training programs could promote learning. For example, visual training associated with congruent auditory stimuli facilitates visual learning compared to visual-only training using a motion detection task (La Rocca et al., 2020; Seitz et al., 2006; Shams and Seitz, 2008) and improves visual rehabilitation (Dundon et al., 2015). Up to now, several studies investigated the benefits of audio-visual training compared to unisensory (visual or auditory-only) training on visual perception (Dundon et al., 2015; La Rocca et al., 2020; Seitz et al., 2006; Shams and Seitz, 2008) or speech perception (Bernstein et al., 2013, 2014). However, to the best of our knowledge, no studies investigated the benefits of such multisensory training on non-verbal auditory perception.

The present study investigated whether pitch contour training could improve pitch processing in congenital amusic individuals and their matched controls. We tested how participants' initial contour processing abilities could predict the training success (see for examples Wu et al., 2022; Yehezkel et al., 2016 in the visual domain) and whether audiovisual training could be more effective to improve pitch contour processing than auditory-only training, and in particular for participants with weaker performance at the start, given the increased efficiency of multisensory interactions in participants with perceptual deficits in one of the modalities. In contrast to the study of Liu et al. (2017) which used auditory training to improve pitch contour processing in congenital amusia, here we used both auditory and audiovisual training. To this aim, participants with congenital amusia and control participants were trained on a pitch direction identification task in both an auditory only (A) and audiovisual (AV) condition in randomized order, with one training session per condition. Based on previous studies (see above), we hypothesized worse pitch direction identification performance for amusic participants than control participants, and that performance of both groups would improve after training (i.e., amusic participants, despite their deficits, but also non-musician control participants), in particular after the AV training compared to the A training.

2. Materials and methods

2.1. Participants

Thirty-six non-musician participants were recruited for this study. Control participants were recruited from the general population (note that prevalence of congenital amusia is low in the general population (around 1.5–4%, Peretz et al., 2003; Peretz and Vuvan, 2017). Amusic participants were recruited through targeted announcement and a first online screening including a short questionnaire and the scale subtest of the Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al., 2003). Potential amusic participants were then invited for a full screening session in the laboratory, including the full MBEA (Peretz et al., 2003), and a Pitch Discrimination Threshold test (PDT) using a two-alternative forced-choice task (Tillmann et al., 2009). The MBEA comprises six subtests that evaluate various aspects of music perception and memory (scale, contour, interval, rhythm, meter, and memory). To be considered as amusics, participants had to obtain a global score inferior to 23 (maximum score=30) and/or a MBEA pitch score (average of the first three subtests) inferior to 22 (maximum score=30). Overall, twelve participants with congenital amusia and twenty-four controls matched for age, gender, education, and musical education participated in the study (see Table 1 for demographic information, which also includes MBEA and PDT scores for 11 control participants, collected as part of other projects in the lab).

An audiometry in silence ensured that all participants had normal hearing (no hearing loss inferior to 25dB at octave frequencies from 500 to 4000 Hz in both ears). All of them self-reported normal or corrected-to-normal vision and reported no neurological or psychiatric diseases. Participants gave their written informed consent, and were paid for their participation. As per French regulations, the study procedures were approved by a Comité de Protection des Personnes (ethics committee) SUD-EST IV (N°ID RCB: 2012-A01209–34).

2.2. Behavioral tests and training procedure

2.2.1. Stimuli

Auditory stimuli. For each trial, the stimuli were two synthetic piano tones (rendered via Cubase software, Steinberg) lasting 350 ms and

Table 1
Demographic data of participants and performance in the screening tests (mean ±standard deviations together with minimum (Min) and maximum (Max) scores within each group are reported). The mean audiometric thresholds from 250 to 4000 Hz and from 6000 to 8000 Hz across both ears was calculated (according to the recommendation of the International Bureau for Audiophonology). MBEA pitch score corresponds to the average of the scale, contour, and interval subtests in the MBEA. Groups were compared with Student *t* tests (two-sided), except for the sex-ratio and handedness where a Fisher test was used. F: female; M: male. R: right-handed; L: left-handed. MBEA: Montreal Battery of Evaluation of Amusia (Peretz et al., 2003). PDT: Pitch Discrimination Threshold (Tillmann et al., 2009). *Only 11 of the 24 control participants were tested with the MBEA and PDT.

Group	Amusics (n=12)	Controls (n=24)	p-value (group comparison)
Age (years)	29.42 (±13.12) Min: 18 Max: 57	30.21 (±11.24) Min: 19 Max: 57	0.86
Sex	8 F, 4 M	16 F, 8 M	1
Education (years)	13.92 (±2.50) Min: 12 Max: 17	15.08 (±2.00) Min: 12 Max: 20	0.18
Musical education (years)	0.08 (±0.29) Min: 0 Max: 1	0.21 (±0.59) Min: 0 Max: 2	0.40
Handedness	11R, 1L	24R, 0L	0.33
Audiometric thresholds (dB): 250 to 4000 Hz	3.29 (±4.70) Min: −1.5 Max: 11.50	2.64 (±3.37) Min: −3 Max: 9.5	0.65
6000 to 8000 Hz	3.96 (±6.72) Min: −6.25Max: 17.5	6.52 (±7.19) Min: −1.25Max: 25	0.36
MBEA global score* (maximum score=30, cut-off=23)	22.18 (±1.20) Min: 20.67 Max: 24.17	26.82 (±1.49) Min: 24.67 Max: 29.50	<0.001
MBEA pitch score* (maximum score=30, cut-off=22)	20.69 (±1.53) Min: 18.00 Max: 23.33	27.09 (±1.38) Min: 25.00 Max: 29.33	<0.001
PDT (semitones)*	1.26 (±1.20) Min: 0.33 Max: 4.42	0.28 (±0.15) Min: 0.09 Max: 0.58	0.02

presented without silent Inter-Stimulus-Interval (ISI) in between. The fundamental frequencies F0 of the tones were between 262 and 523 Hz. All tones belonged to the C major key (C3 to C4). The pitch interval between the two tones could be of one or two semitones (referred to a minor or major second in music theory, corresponding to 2nd pitch interval, 25% of trials), three or four semitones (referred to a minor or major third, corresponding to 3rd pitch interval, 25% of trials), five or six semitones (referred to a perfect or augmented fourth, corresponding to 4th pitch interval, 25% of trials), or seven semitones (referred to a perfect fifth, corresponding to 5th pitch interval, 25% of trials), either ascending or descending.

Visual stimuli. The visual stimuli were two light gray disks (with a luminance of 60cd/m2) connected by a white bar on a black screen and appearing from the left to the right of the screen (Fig. 1A). They were presented simultaneously with the onset of the tones (first disk with the first note, connecting bar and second disk with the second note) and their positions were informative of the pitch of the note (with higher positions corresponding to higher pitches).

Apparatus. A 15-inch computer screen with a refresh rate of 75 Hz was positioned 150 cm in front of the participant. The screen background color was black, with a luminance of 20 cd/m2. Auditory stimuli were presented with two loudspeakers positioned at each side of the screen at a level of 70 dB SPL A, measured at the level of the participants' ear with a Brüel & Kjaer type 2239 sonometer. The experiment took place in a quiet room. We used Presentation software (Neurobehavioral systems, Albany, CA, USA) for the presentation of stimuli and to record participants' responses.

2.2.2. *Testing blocks: pitch direction of change identification (DCI) task (audio-only)*

Participants were tested with a pitch Direction of Change Identification (DCI) task. This task involved listening to two tones and determining whether the pitch contour was "rising" (the second tone was higher) or "falling" (the second tone was lower) from the first to the second tone by clicking on the left or right mouse button. Participants could answer 100 ms after the beginning of the second tone and until 2 s after its end. The response buttons corresponding to "rising" and "falling" were counterbalanced between participants. Participants were required to answer as rapidly as possible. Inter-trial silent intervals were between 1500 and 2000 ms. No feedback was provided at the end of the trials and the testing blocks. Participants were tested with the DCI with auditory-only stimuli before and after each training session in order to evaluate the gain of the training (see Fig. 1B for the experimental design). The three testing blocks were composed of 96 trials, half ascending and the other half descending, with 24 trials of each of the four tested pitch intervals (2nd, 3rd, 4th, and 5th) and lasted 3 to 4 min each.

Before the first testing block, participants performed a familiarization block composed of twelve trials, half audio-only and half audiovisual. The familiarization block was composed of six descending and six ascending pairs of tones without feedback.

2.2.3. *Training sessions: audio-only or audiovisual DCI training*

Participants were trained with the DCI task in an auditory (A) and audiovisual (AV) condition. No feedback was provided either at the end of the trials or of the training sessions. The order of the two training sessions (A and AV) was counterbalanced across subgroups (Fig. 1B). Each training session was composed of 192 trials including 96 trials with ascending and 96 descending pairs of tones, equiprobably distributed with respect to pitch intervals (2nd, 3rd, 4th, and 5th). For the AV training, half of the trials was *auditory only* and the other half was *audiovisual*. Each training session lasted about 6 to 8 min. The total duration of the experiment was about 40 min (familiarization phase, 3 testing blocks, 2 training sessions, see Fig. 1B).

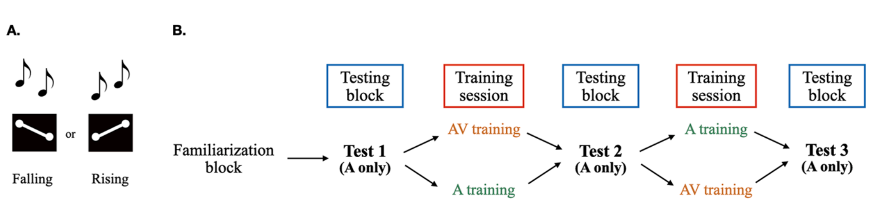

Fig. 1. Experimental design. (A) Pitch Direction of Change Identification (DCI) task. Participants had to determine whether the pitch contour is "falling" or "rising" from the first to the second tone. (B) Training procedure. The three Testing blocks are auditory only. The order of the audio-only (A) or audiovisual (AV) Training sessions was counterbalanced across participants.

2.3. Data analyses

Bayesian statistical analyses were conducted with the software JASP (JASP 0.19). Bayesian statistics allow for comparing different models including the null model and to select the one with the best evidence given the data. We report Bayes Factor (BF_{10}) as a relative measure of evidence in favor of an alternative hypothesis (the tested model) compared to the null hypothesis (the null model). Results are interpreted as follows: BF_{10} between 1 and 3 is considered as weak evidence in favor of the tested model, between 3 and 10 as positive evidence, between 10 and 100 as strong evidence and a BF_{10} higher than 100 as decisive evidence (Lee and Wagenmakers, 2014). In addition, we ran analyses of effects and reported $BF_{inclusion}$, which indicates the evidence for one effect over all of the tested models. Bayesian *t*-tests were performed as post-hoc tests for effects and interactions associated with positive (or more) evidence. For Bayesian t tests, we report effect size (δ: 0.2=small, 0.5=medium, 0.8=large) and 95% credible interval (CI) when BF_{10} was in favor of the alternative hypothesis (Faulkenberry et al., 2020). We tested varied priors, reflecting weak to strong informative prior assumption to ensure that our results were not sensitive to the choice of the prior but were driven by the data. We choose a Cauchy prior scale of 0.5 for all analyses (De Jong, 2019).

D-prime (d') and reaction times (RT) for correct responses during the testing blocks were analyzed in three steps as described below. d' was calculated using Hit and False Alarm rates with Hit corresponding to a correct answer for an upward trial and False Alarm corresponded to an incorrect answer for a downward trial. Mean RT across trials, relative to the end of the stimulus, were calculated for the correct responses. Visual inspection of the data did not reveal very long or too early RT, and no RT were removed (all RTs between 200 and 1700 ms).

2.3.1. Overall effect of training

Overall training gains for each difficulty level were computed as the d' in the third testing block minus the d' in the first testing block (i.e., the overall effect of training after completing the two training sessions, without taking into account the training condition, A or AV) and compared to zero with one-tailed one-sample Bayesian *t*-tests, separately for each group. One-tailed tests were chosen because we had a priori hypotheses about the direction of training effects. The overall effect of the training was further analyzed with a Bayesian repeated measure analysis of variance (ANOVA) on d' with Training (pre- vs. post-training sessions, i.e., first vs. third testing block) and Difficulty (pitch interval between the two sounds: 2nd, 3rd, 4th, and 5th) as within-participant factors, and Group (amusics, controls) as between-participants factor. The same analyses were performed for correct RTs.

2.3.2. Influence of initial performance level on training gains

Then the correlation between participants' initial d' (averaged across all difficulty levels) and training gains (also averaged across all difficulty levels) was analyzed using Bayesian Pearson correlation. Additionally, we assessed the correlations between initial d' or training gains and: MBEA global score, MBEA pitch score, and PDT, also with Bayesian Pearson correlations. The same analyses were performed for RTs. Testing the correlations between initial performance and training gains allowed us to determine whether the profile of participants (best or worst performers) influenced the benefits of training. Previous studies investigated the impact of participants' initial performance in visual tasks and found a correlation between initial performances and training gains (Wu et al., 2022; Yehezkel et al., 2016), with larger training gains for worst performers. Correlation analyses were first run across all participants and then separately for each participant group.

2.3.3. Difference between A and AV training

We first compared the initial d' and RTs in participants that were exposed to the audio-only or the audiovisual training as the first training session, separately for each group, using Bayesian t tests. To anticipate the results, it was observed that the initial d' and reaction times was different between control participants that were exposed to the audio-only or the audiovisual training as the first training session, with faster RTs in participants who started with the AV training (Figure S1). We thus removed from this analysis the data of 12 control participants in order to balance the two arms with respect to initial RTs (all $BF_{10} <$ 1.205 after removing these participants' data, see results section for details and Figure S1). We then analyzed the effect of A vs. AV training on amusic (n=12) and control (n=12) participants' performance (d' and RTs). Bayesian one-tailed one-sample *t*-tests against 0 were used to assess the gains of the A and AV training on participants' d' for each level of difficulty of the DCI test (A and AV training gains were the difference in performance between each post- and pre-training session). Then, we performed Bayesian repeated measures ANOVAs on training gains (separately for d' and RT), with Training type (A vs. AV) and Difficulty (pitch interval between the two sounds: 2nd, 3rd, 4th, and 5th) as within-participant factors, and Group (amusics, controls) as between-participants factor.

3. Results

A two-tailed one-sample Bayesian *t*-test was conducted to assess whether participants' performance was above chance level at the study onset. We observed strong to decisive evidence that participants performed above chance level (d'>0) for each difficulty level and group (BF_{10}>38.103) except for the 2nd pitch interval in the amusic group (BF_{10}=0.317).

3.1. Overall effect of training pitch direction of change identification

Overall effects of training on performance (d', RT) in the DCI task is illustrated in Fig. 2. See Table 2 for the details of training gains.

3.1.1. D-prime

Regarding the overall training gains (without taking into account the training condition, A or AV), in the amusic group, we found positive evidence for a difference compared to 0 of the training gain for the 2nd pitch interval (BF_{10}=4.05, δ=0.57, CI:[0.08, 1.19]), weak evidence for a difference compared to 0 of the training gain for the 3rd and 5th pitch interval (BF_{10}=2.52, δ=0.49, CI:[0.06, 1.07] and BF_{10}=1.34, δ=0.37, CI:[0.03, 1.92], respectively), and no evidence for a difference compared to 0 of the training gain for the 4th pitch interval (BF_{10}=0.69) (see Table 2 for all comparisons to 0 of the training gains). When averaging gains across difficulty levels, we found positive evidence for a difference compared to 0 of the training gains on d' (BF_{10}=4.64, δ=0.6, CI:[0.09, 1.23]). In the control group, we found positive evidence for a difference compared to 0 of the training gain for the 3rd pitch interval (BF_{10}=4.793, δ=0.44, CI:[0.08, 0.85]) and weak evidence for the 5th pitch interval (BF_{10}=1.755, δ=0.33, CI:[0.03, 0.72]) and no evidence for the other pitch intervals (all BF_{10}<0.964). When averaging gains across difficulty levels, we found positive evidence for a difference compared to 0 of the training gains (BF_{10}=3.225, δ=0.4, CI:[0.06, 0.80]).

The overall effect of the training was further analyzed with a Bayesian repeated-measure ANOVA on d' (see Fig. 2). For the Bayesian ANOVA, the best model explaining the d' included Training, Group, Difficulty, and the interaction between Group and Difficulty ($BF_{10}=1.463\times10^{31}$). This model was 1.69 times more likely than the model including the main effect of Training, Group, and Difficulty ($BF_{10}=8.634\times10^{30}$), 3.74 more likely than the model including the main effect of Training, Group, Difficulty, the interaction between Training and Group, and the interaction between Group and Difficulty ($BF_{10}=3.909\times10^{30}$) and at least 6 times more likely than any of the other tested models (all $BF_{10}<2.300\times10^{30}$). The analysis of effects across all models revealed a decisive specific effect of Difficulty ($BF_{inclusion}=\infty$) and Group ($BF_{inclusion}$=1171.690), a positive effect of

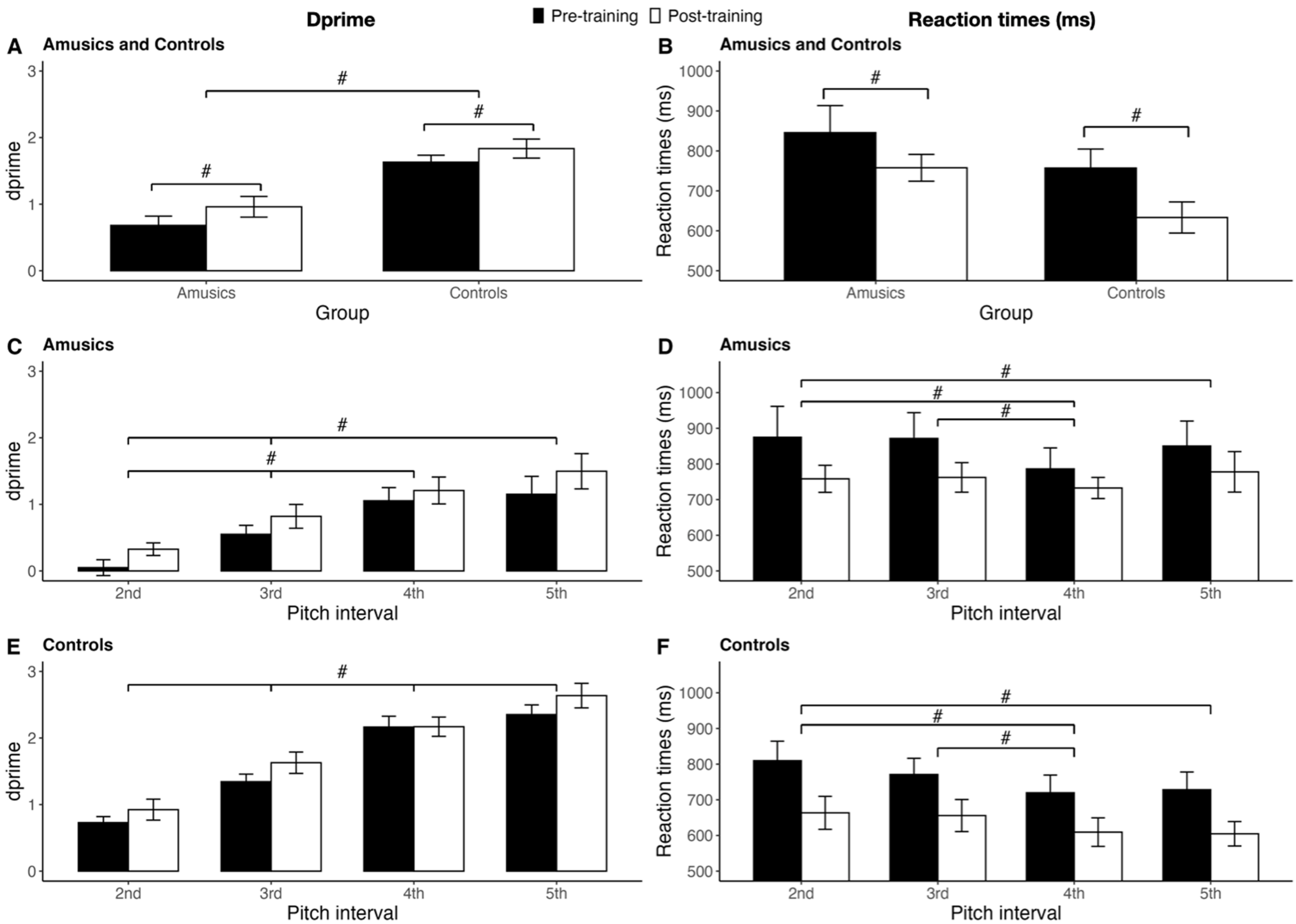

Fig. 2. Overall effect of training. (A) Pre- (black bars) and post-training (white bars) mean d' of amusic (n=12) and control (n=24) participants. (B) Pre- (black bars) and post-training (white bars) correct reaction times (in ms) of amusic and control participants. (C) Pre- (black bars) and post-training (white bars) mean d' of amusic participants for each pitch interval (2nd, 3rd, 4th, and 5th). (D) Pre- (black bars) and post-training (white bars) correct reaction times (in ms) of amusic participants for each pitch interval. (E) Pre- (black bars) and post-training (white bars) mean d' of control participants for each pitch interval. (F) Pre- (black bars) and post-training (white bars) correct reaction times (in ms) of control participants for each pitch interval. Error bars correspond to standard error of the mean. # indicates BF10>3 (A,B: training effect; C, E: difficulty effect per group, and D,F: difficulty effect).

Training ($BF_{inclusion}$=4.703) and of the interaction between Group*-Difficulty ($BF_{inclusion}$=3.540). No other specific effects were found ($BF_{inclusion}$<0.5). d' increased after training compared to pre-training (Fig. 2A). In addition, amusic participants had lower d' than control participants (Fig. 2A). For the effect of Difficulty, post-hoc analysis revealed strong to decisive evidence for a difference between smaller and larger pitch intervals: participants' d' were lower when pitch intervals were smaller compared to larger pitch intervals (for all pairwise comparisons, BF_{10}>92). Then, regarding the interaction between Group and Difficulty, in amusics (Fig. 2C), post-hoc analysis revealed no evidence for a difference between 4th and 5th pitch intervals (BF_{10}=0.671), but strong to decisive evidence for a difference between the other pitch intervals (for all pairwise comparisons, BF_{10}>16.379). In controls (Fig. 2E), we found strong to decisive evidence for a difference between all pitch intervals (for all pairwise comparisons, BF_{10}>15.589). Effect sizes based on a frequentist ANOVA are presented in Supplementary data.

3.1.2. Reaction times

Regarding overall training gains on RT for each difficulty level in the amusic group, we observed positive evidence for a difference compared to 0 of the training gain for the 3rd pitch interval (BF_{10}=7.140, δ=−0.68, CI:[−1.34, −0.13]), and weak evidence for a difference compared to 0 of the training gain for the 2nd and 5th pitch interval (BF_{10}=1.944, δ=−0.44, CI:[−1.01, −0.04] and BF_{10}=1.221, δ=−0.36, CI:[−0.89, −0.03] respectively). No evidence for a difference compared to 0 of the training gain for the 4th pitch interval was found (BF_{10}=0.824) (see Table 2 for all comparison to 0 of the training gains). When averaging RT gains across difficulty levels, we found weak evidence for a difference compared to 0 (BF_{10}=2.520, δ=−0.49, CI:[−1.07, −0.06]). In the control group, we found strong evidence for a difference compared to 0 of the training gain for the 2nd (BF_{10}=44.387, δ=−0.66, CI:[−1.12, −0.23]), 3rd (BF_{10}=77.450, δ=−0.72, CI:[−1.18, −0.27]), and 4th pitch intervals (BF_{10}=62.557, , δ=−0.70, CI:[−1.15, −0.26]) and decisive evidence for a difference compared to 0 of the training gain for the 5th pitch interval (BF_{10}=188.753, δ=−0.8, CI:[−1.28, −0.34]). When averaging RT gains across difficulty levels, we found decisive evidence for a difference compared to 0 (BF_{10}=791.933, δ=−0.93, CI: [−1.43, −0.44]).

For the Bayesian ANOVA, the best model to explain the RT data was the one including Training and Difficulty factors (BF_{10}=88456.496, see Fig. 2). This model was 1.2 times more likely than the model including the main effect of Training, Difficulty, and Group (BF_{10}=73953.619), 2.31 more likely than the model including the main effect of Training, Difficulty, Group, and the interaction between Training and Group (BF_{10}=38222.49), 3.58 times more likely than the model including the main effect of Training, Difficulty, Group, and the interaction between Difficulty and Group (BF_{10}=24685.71) and at least 7 times more likely

Table 2
Training gains for d' and reaction times (in ms) for amusic and control participants, for each difficulty level, and averaged across all difficulty levels. Bold font indicates $BF_{10}>3$ when comparing the training gain to zero (with one-tailed one-sample Bayesian *t*-test).

Pitch interval	Amusics (*n*=12) Average	2nd	3rd	4th	5th	Controls (*n*=24) Average	2nd	3rd	4th	5th
Training gains for d'	0.28 (±0.40) Min: −0.33 Max: 0.99 BF_{10}= 4.643	0.28 (±0.41) Min: −0.62 Max: 0.78 BF_{10}= 4.051	0.27 (±0.48) Min: −0.59 Max: 1.15 BF_{10}= 2.518	0.15 (±0.70) Min: −0.81 Max: 1.29 BF_{10}= 0.685	0.36 (±0.84) Min: −1.16 Max: 1.60 BF_{10}= 1.344	0.20 (±0.47) Min: −0.67 Max: 1.3 BF_{10}= 3.225	0.20 (±0.75) Min: −1.19 Max: 1.67 BF_{10}= 0.964	0.29 (±0.59) Min: −0.68 Max: 1.58 BF_{10}= 4.793	0.005 (±0.65) Min: −1.23 Max: 1.37 BF_{10}= 0.296	0.29 (±0.81) Min: −0.99 Max: 2.16 BF_{10}= 1.755
Training gains for reaction times (ms)	−88 (±155.48) Min: −316.00 Max: 132.69 BF_{10}= 2.520	−116.58 (±231.32) Min: −517.25 Max: 179.75 BF_{10}= 1.944	−109.33 (±136.85) Min: −375.75 Max: 75.25 BF_{10}= 7.140	−53.67 (±194.41) Min: −479.25 Max: 225.5 BF_{10}= 0.824	−72.44 (±187.67) Min: −449.5 Max: 145.5 BF_{10}= 1.211	−123.87 (±123.4) Min: −405.75 Max: 117.13 BF_{10}= 791.933	−146.28 (±201.29) Min: −812.75 Max: 220.75 BF_{10}= 44.387	−115.12 (±147.23) Min: −291.5 Max: 322.25 BF_{10}= 77.387	−110.37 (±145.06) Min: −372.25 Max: 184.25 BF_{10}= 62.557	−123.7 (±142.49) Min: −461.5 Max: 94.75 BF_{10}= 188.753

than any of the other tested models (all $BF_{10}<11113.82$). The analysis of effects across all models revealed a decisive specific effect of Training ($BF_{inclusion}=612.831$) and a strong effect of Difficulty ($BF_{inclusion}=27.514$). No other specific effects were found ($BF_{inclusion}<0.6$). Participants' RT were lower (better) after training than before training (Fig. 2B). For the effect of Difficulty, post-hoc tests revealed positive to decisive evidence for a difference between RTs for small pitch intervals (2nd or 3rd) compared to large pitch intervals (4th or 5th) (2nd vs. 4th, $BF_{10}=370.040$; 2nd vs. 5th, $BF_{10}=4.297$; 3rd vs. 4th, $BF_{10}=37.418$, all other $BF_{10}<0.82$, Fig. 2D,F). Participants had lower (faster) RT for the large pitch intervals compared to small pitch intervals. Effect sizes based on a frequentist ANOVA are presented in Supplementary data.

3.2. *Correlation between training gains and initial performance*

The main findings of the correlation analyses are illustrated in Fig. 3.

3.2.1. *D-prime*

Correlation analysis revealed no evidence for a correlation between participants' initial d' and the training gains ($r=-0.098$; $BF_{10}=0.243$; n=36) (Fig. 3A). In addition, positive to decisive evidence was found for correlations between the initial d' and the MBEA global score ($r=0.730$; $BF_{10}=401.757$; n=23, decisive evidence), the MBEA pitch score ($r=0.562$; $BF_{10}=10.127$; *n*=23, strong evidence), and PDT ($r=-0.535$; $BF_{10}=7.738$; *n*=23, positive evidence). Participants with better MBEA-global and MBEA-pitch scores or with lower PDT (better threshold) had higher d' at study onset (Fig. 3B, C, D). No correlations were observed between the training gains and MBEA, MBEA pitch, PDT (all $BF_{10} < 1$).

In addition, we analyzed the correlations separately by group. In amusic participants (n=12), no evidence for a correlation was observed between their initial d' and the gains of the training ($BF_{10}=0.485$). However, positive or weak evidence for a correlation between their initial d' and the MBEA-global score ($r=0.635$; $BF_{10}=3.188$; positive evidence), PDT ($r=-0.466$; $BF_{10}=1.017$; weak evidence) were found. No evidence for a correlation between their initial d' and the MBEA-pitch score was found ($BF_{10}=0.370$). In addition, we found no evidence for a correlation between the training gains and the MBEA-global, MBEA-pitch, and PDT ($BF_{10}<0.721$). In control participants (n=24), no evidence for a correlation was observed between their initial d' and the gains of the training ($BF_{10}=0.254$). The correlation analyses revealed positive evidence for a correlation between their initial d' and the PDT ($r=-0.753$; $BF_{10}=8.671$, *n*=11), but no evidence for the correlations of their initial d' with MBEA-global or MBEA-pitch score ($BF_{10}<0.683$). In addition, no evidence for a correlation between the training gains and the MBEA-global, or MBEA-pitch, or PDT was found ($BF_{10}<0.509$).

3.2.2. *Reaction times*

When analyzing RT, we found decisive evidence for a correlation between participants' initial RT and the training gains ($r=0.653$; $BF_{10}=1566.858$, n = 36). Participants with higher (slower) RT before training benefit more of the training than participants with lower (faster) RT at study onset (Fig. 3E). No correlations were observed between initial RT or the training gains on RT and MBEA global, MBEA pitch, PDT (all $BF_{10}<1$, *n*=23) (Fig. 3F, G, H).

For RT in the amusic group, decisive evidence for a correlation between the initial RT and the gain of the training on RT ($r=-0.9$; $BF_{10}=374.602$) was found. However, no evidence for a correlation was observed between their initial RT and the MBEA-global, the MBEA-pitch or the PDT ($BF_{10}<0.49$), as well as no evidence for a correlation between their training gains on RT and their MBEA-global, or MBEA-pitch, or PDT ($BF_{10}<0.38$). Then, analyzing RT in the control group, we found strong evidence for a correlation between their initial RT and the training gains on RT ($r=-0.574$; $BF_{10}=14.605$, *n*=24). Weak evidence for a correlation between controls' initial RT and their PDT was

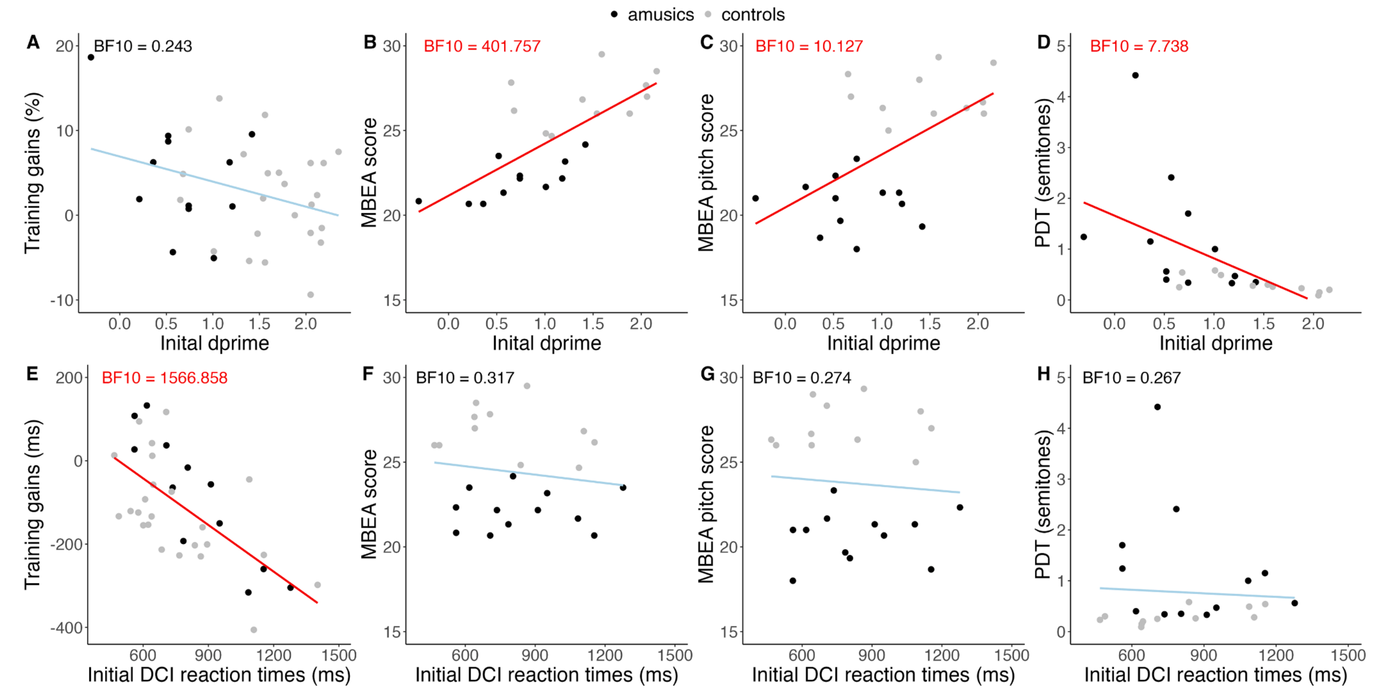

Fig. 3. Correlations between initial performance, training gains, and screening tests. Top row: d'; bottom row: correct reaction times. Correlations are computed across participants from both groups in all panels, amusic participants' data are represented with black dots (n=12) and control participants' data with gray dots (n=24 in panels A & E, n=11 in panels B, C, D, F, G, H). (A) Correlation between participants' initial mean d' and the benefits of the training (post-training d' minus pre-training d'). (B) Correlation between participants' initial mean d' and the MBEA score. (C) Correlation between participants' initial mean d' and the MBEA pitch score. (D) Correlation between participants' initial mean d' and the PDT. (E) Correlation between participants' initial reaction times and the benefits of the training (post-training response times minus pre-training reaction times). (F) Correlation between participants' initial reaction times and the MBEA score. (G) Correlation between participants' initial reaction times and the MBEA pitch score. (H) Correlation between participants' initial reaction times and the PDT. The red correlation line indicates BF10>3; the blue correlation line indicates BF10<1.

observed (r = 0.620; BF_{10}=2.343, *n*=11). However, no evidence for a correlation was observed between their initial RT and the MBEA-global, or the MBEA-pitch (BF_{10}<0.4), as well as no evidence for a correlation between their training gains on RT and their MBEA-global, or MBEA-pitch, or PDT (BF_{10}<0.5).

3.2.3. Correlation between d' and reaction times

Correlation analyses revealed no evidence for a correlation between initial d' and initial RT (BF_{10}=0.751, *n*=36) nor between training gains on d' and training gains on RT (BF_{10}=0.270, *n*=36).

Then, we analyzed the correlation between d' and reaction times separately for each group. In the amusic group, no evidence was found for a correlation between initial d' and initial RT (BF_{10}=0.492) and weak evidence of a correlation between training gains on d' and training gains on RT was found (r=0.502; BF_{10}=1.237). In addition, in the control group, weak evidence for a correlation between their initial d' and RT was found (r=−0.434; BF_{10}=2.110) and no evidence of a correlation between training gains on d' and training gains on RT was found (BF_{10}=0.268).

3.3. Direction of change identification task: A vs. AV training

3.3.1. Comparisons between the two arms at study onset

Initial d' and RTs (Testing block 1) were analyzed with Bayesian *t*-tests within each group (amusics, controls) to verify that there was no difference between participants who started with the A training and those who started with the AV training, which would preclude to compare the two training regimens, as training gains on RT depend heavily on initial RT (see 3.2 just above). We found no evidence for an effect of the arm assignment on the initial d' in the amusic group (BF_{10}=0.474). However, we found positive evidence for an effect of the arm assignment on the initial d' in the control group (BF_{10}=3.836). The initial d' was lower for control participants who performed the auditory training as the first training compared to those who performed the audiovisual training as the first training. We found no evidence for an effect of the arm assignment on initial RTs in the amusic group (BF_{10}=0.48). However, we found positive evidence for an effect of the arm assignment on RTs in the control group (BF_{10}=7.92, see Figure S1). The initial RTs were higher (i.e., worse) for control participants who performed the auditory training as the first training compared to those who performed the audiovisual training as the first training. Thus, to counteract initial RTs difference in the two arms. for the control group, we removed from the subsequent analysis the data from the six control participants with the higher RT for those who started with the A training and the six control participants with the lower RT for those who started with the AV training (Figure S1). After removing data from these participants in the control group, we found only weak evidence of an effect of the arm assignment on initial d' (BF_{10}=1.205) and no evidence of an effect of the arm assignment on RTs in the control group (BF_{10}=0.488). We thus analyzed the effect of A vs. AV training on data from twelve amusic and twelve control participants only (Fig. 4 and Table 3; Sections 3.3.2 and 3.3.3).

3.3.2. D-prime: A vs. AV training

For the training gains on the d' (see Table 3), results of the Bayesian one-sample one-tailed *t*-tests of the amusics group revealed weak evidence for A training gains difference to 0 for the 2nd pitch interval (BF_{10}=1.094, δ=0.30, CI:[0.02, 0.82]). No evidence for A training gains difference to 0 for all pitch intervals and when averaging across all levels of difficulty (all BF_{10}<0.768) was found. For the AV training, positive evidence for a difference compared to 0 of the AV training gains for the 5th pitch interval (BF_{10}=6.249, δ=0.65, CI:[0.11, 1.30]) was found. No evidence for training gains difference to 0 was found for other pitch intervals (all BF_{10}<0.695). When averaging across all levels of difficulty,

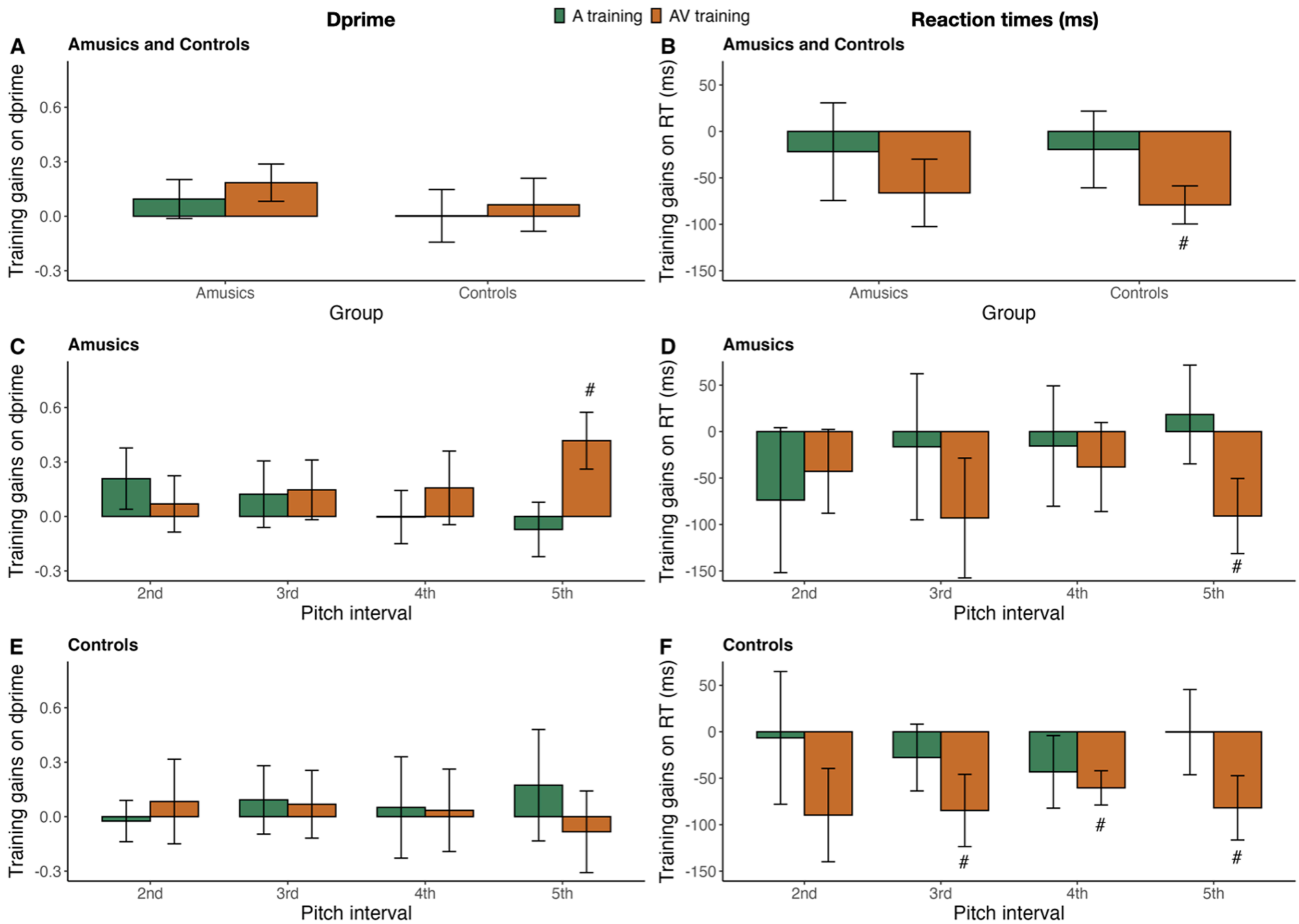

Fig. 4. Auditory and audiovisual training gains in amusic (n=12) and control (n=12) participants. Gains of the auditory (A, in green) and audiovisual (AV, in orange) trainings for each level of difficulty of the DCI test were computed as the d' in the test block just after the training session of interest minus the d' in the testing block immediately preceding the training session of interest. The same calculations were made for the reaction times. (A) Training gains on d'. (B) Training gains on reaction times (in ms). (C) Training gains on d' for the 2nd, 3rd, 4th, and 5th pitch intervals for amusic participants. (D) Training gains on reaction times (in ms) for each pitch interval for amusic participants. (E) Training gains on d' for each pitch interval for control participants. (F) Training gains on reaction times (in ms) for each pitch interval for control participants. Comparisons of training gains to 0: # indicates BF10>3 when comparing training gains to zero (one-tailed one-sample Bayesian t-tests). Error bars correspond to standard error of the mean.

we found only weak evidence for a difference compared to 0 of the AV training gain (BF_{10}=2.075, δ=0.41, CI:[0.04, 0.99]).

In the control group, results of the Bayesian one-sample one-tailed *t*-tests revealed no evidence of difference compared to 0 of the A and AV training gains for any of the pitch intervals and when averaging across all levels of difficulty (all BF_{10}<0.575).

For the Bayesian ANOVA, the model that best explains the A and AV training gains on d' was the null model, followed by the model with the Group effect (BF_{10}=0.267). None of the other tested models explained the data better than the null model (all BF_{10}<0.262). Effect sizes based on a frequentist ANOVA are presented in Supplementary data.

3.3.3. Reaction times: A vs. AV training

In the amusic group, RT analyses revealed no evidence of difference compared to 0 of the A training gains for any of the pitch intervals and when averaging across all levels of difficulty (all BF_{10}<0.815). However, we found positive evidence for a difference compared to 0 of the AV training gain for the 5th pitch interval (BF_{10}=3.611, δ=−0.55, CI:[−1.16, −0.08]) and weak evidence for a difference compared to 0 of the AV training gain for the 3rd pitch interval (BF_{10}=1.371, δ=−0.38, CI:[−0.92, −0.03]). No evidence of difference compared to 0 of the AV training gains was observed for the other pitch intervals (all BF_{10}<0.817). When averaging across all levels of difficulty, we found only weak evidence for a difference compared to 0 of the AV training gain (BF_{10}=2.136, δ=−0.46, CI:[−1.03, −0.05]). See Table 3 for the details of A and AV training gains.

In the control group, results of the Bayesian one-sample one-tailed *t*-tests revealed no evidence of difference compared to 0 of the A training gains for any of the pitch intervals and when averaging across all levels of difficulty (all BF_{10}<0.956). However, we found strong evidence of a difference compared to 0 of the AV training gain for the 4th pitch interval (BF_{10}=14.097, δ=−0.81, CI:[−1.51, −0.19]) and positive evidence of a difference compared to 0 of the AV training gain for the 3rd (BF_{10}=3.307, δ=−0.54, CI:[−1.14, −0.07]) and 5th pitch intervals (BF_{10}=4.204, δ=−0.58, CI:[−1.20, −0.08]). In addition, we found weak evidence of a difference compared to 0 of the AV training gain for the 2nd pitch interval (BF_{10}=2.038, δ=−0.45, CI:[−1.02, −0.05]). Without taking into account the levels of difficulty, the analysis revealed strong evidence for a difference compared to 0 of the AV training gain (BF_{10}=31.407, δ=−0.96, CI:[−1.71, −0.28]).

For the Bayesian ANOVA of RT, the model that best explains the data was the null model, followed by the model with the Training type effect (BF_{10}=0.530). None of the other tested models explained the data better than the null model (all BF_{10}<0.243). Effect sizes based on a frequentist ANOVA are presented in Supplementary data.

Table 3
Training gains for d' and reaction times (in ms) for amusic and control participants after the A and AV training. A: auditory training; AV: audiovisual training. Bold font indicates $BF_{10}>3$ when comparing the training gain to zero (one-tailed one-sample Bayesian t-test).

Pitch interval	Training gains (d')					Training gains (reaction times, ms)				
	Average	2nd	3rd	4th	5th	Average	2nd	3rd	4th	5th
A training (Amusics, n=12)	0.10 (±0.37) Min: −0.47 Max: 0.84 BF_{10}=0.768	0.21 (±0.59) Min: −0.81 Max: 1.40 BF_{10}=1.094	0.12 (±0.64) Min: −0.78 Max: 1.34 BF_{10}=0.630	−0.003 (±0.51) Min: −0.79 Max: 0.87 BF_{10}=0.367	−0.07 (±0.52) Min: −1.16 Max: 0.61 BF_{10}=0.279	−21.81 (±182.22) Min: −407.06 Max: 352.94 BF_{10}=0.508	−73.79 (±270.31) Min: −725.00 Max: 359.50 BF_{10}=0.815	−16.33 (±272.48) Min: −734.25 Max: 365.50 BF_{10}=0.432	−15.54 (±224.48) Min: −310.75 Max: 520.25 BF_{10}=0.443	18.44 (±184.09) Min: −246.00 Max: 405.00 BF_{10}=0.300
AV training (Amusics, n=12)	0.18 (±0.36) Min: −0.23 Max: 0.94 BF_{10}=2.075	0.07 (±0.54) Min: −0.78 Max: 0.82 BF_{10}=0.521	0.15 (±0.57) Min: −0.92 Max: 1.06 BF_{10}=0.773	0.16 (±0.70) Min: −1.22 Max: 1.01 BF_{10}=0.695	0.42 (±0.54) Min: −0.39 Max: 1.29 BF_{10}=6.249	−66.2 (±125.66) Min: −232.25 Max: 182.50 BF_{10}=2.136	−42.79 (±156.37) Min: −356.50 Max: 249.75 BF_{10}=0.817	−93.00 (±223.38) Min: −402.75 Max: 358.50 BF_{10}=1.371	−38.13 (±166.13) Min: −298.25 Max: 271.25 BF_{10}=0.706	−90.88 (±140.08) Min: −260.50 Max: 216.25 BF_{10}=3.611
A training (Controls, n=12)	0.003 (±0.50) Min: −0.78 Max: 1.15 BF_{10}=0.378	−0.02 (±0.39) Min: −0.61 Max: 0.65 BF_{10}=0.325	0.09 (±0.65) Min: −0.87 Max: 1.19 BF_{10}=0.541	0.05 (±0.97) Min: −1.16 Max: 2.27 BF_{10}=0.416	0.17 (±1.06) Min: −1.06 Max: 2.73 BF_{10}=0.575	−19.44 (±142.96) Min: −200.56 Max: 355.63 BF_{10}=0.532	−6.56 (±247.53) Min: −322.75 Max: 700.25 BF_{10}=0.398	−27.73 (±124.57) Min: −277.00 Max: 146.00 BF_{10}=0.691	−43.13 (±135.20) Min: −319.25 Max: 182.75 BF_{10}=0.956	−0.35 (±158.85) Min: −219.50 Max: 424.00 BF_{10}=0.375
AV training (Controls, n=12)	0.06 (±0.51) Min: −0.57 Max: 1.48 BF_{10}=0.515	0.08 (±0.81) Min: −0.99 Max: 1.58 BF_{10}=0.485	0.07 (±0.65) Min: −1.22 Max: 1.05 BF_{10}=0.488	0.04 (±0.79) Min: −1.16 Max: 1.16 BF_{10}=0.416	−0.08 (±0.78) Min: −1.08 Max: 1.80 BF_{10}=0.296	−79.16 (±71.07) Min: −238.50 Max: −2.38 BF_{10}=31.407	−89.69 (±174.01) Min: −479.50 Max: 165.75 BF_{10}=2.038	−84.69 (±134.67) Min: −298.25 Max: 141.50 BF_{10}=3.307	−60.38 (±63.95) Min: −169.75 Max: 7.25 BF_{10}=14.097	−81.88 (±119.90) Min: −372.00 Max: 96.75 BF_{10}=4.204

4. Discussion

The present study investigated whether pitch contour processing could be improved by training in participants with congenital amusia and matched controls, testing whether an audiovisual training regimen could be more effective than auditory-only training. As expected, based on previous research (Foxton, Dean, et al., 2004; Hoarau et al., 2024; Liu et al., 2017; Loui et al., 2008), amusic participants exhibited lower initial performance (d') than control participants when they had to distinguish whether pitch contour was rising or falling between the two tones. Analyses without taking into account the training condition (auditory-only or audiovisual) revealed positive evidence for training gains on d' in both amusic and control participants. In addition, participants improved their reaction times after the two consecutive 6-to-8-minute training sessions. Auditory training improved pitch contour processing in control participants, as expected, but also in individuals with amusia, despite their deficit in pitch processing. Furthermore, we found that improvement in reaction times depends on participants' initial reaction time. For both amusic and control groups (and also across all participants), participants with higher (slower) reaction times benefit more from the contour pitch training. Finally, limited evidence for higher benefits of multisensory audiovisual training compared to auditory-only training on pitch contour perception was found. Analyzing training gains, positive evidence for a difference compared to 0 was found only after audiovisual training, both for amusic and control participants. However, direct comparisons between gains after audio-only or audiovisual training revealed no evidence for a higher benefit of the audiovisual training compared to auditory-only training. The implications of these findings for our understanding of congenital amusia, pitch contour training, and multisensory interactions are discussed below.

4.1. Pitch contour processing in non-musicians and in congenital amusics

As expected, we observed lower performance in the pitch direction task in the amusic group compared to the non-musician control group already at study onset. Both amusics and control participants exhibit a decreased *d'* when the task difficulty increases (smaller pitch intervals) with no evidence for an interaction between the group effect and the difficulty of the task (pitch interval between a 2nd to a 5th). This result is in line with previous studies reporting pitch contour deficits in congenital amusia (Foxton, Dean, et al., 2004; Hoarau et al., 2024; Loui et al., 2008). It should be noted that the pitch contour deficit is observed in amusic participants even when the change in pitch between the two tones is greater than their pitch discrimination thresholds. Here amusic participants exhibited pitch discrimination thresholds between 0.33 and 2.41 semitones (and one amusic presented a discrimination threshold of 4.42 semitones) while group differences in the contour task (DCI) were observed also for the 3rd (3 or 4 semitones) and 5th (7 semitones) pitch intervals. These findings suggest that the pitch contour deficit would not only be due to a simple pitch discrimination impairment, but is also related to a pitch memory impairment (Tillmann et al., 2009, 2016, 2023) and/or a deficit of specific contour processing mechanisms (Stewart, 2011). Identifying pitch direction also relies on memory, as participants must memorize the first tone to compare it with the pitch of the second tone, yet timescales are different compared to other tasks (e.g., melodies over several seconds for the MBEA vs. pitch contours within <1 s here). In keeping with the hypothesis of a pitch memory and/or contour deficit as the core deficit in congenital amusia, impacting performance in a variety of tasks, performance in the direction of change identification task used here correlated with other pitch tasks across participants, namely the MBEA-global and MBEA-pitch scores, which rely heavily on memory, contour and/or interval processing abilities.

This pitch contour processing deficit was observed in previous studies for both discrete and gliding tones, with the largest impairments for discrete changes (Liu et al., 2012), as used here. In keeping with the importance of pitch contour also in speech perception, it has been observed that individuals with amusia exhibit deficits in pitch processing in speech material (Tillmann, Rusconi, et al., 2011; Tillmann et al., 2015), including speech intonation perception (Liu et al., 2010) and lexical tone discrimination (Tillmann, Burnham, et al., 2011).

4.2. Training of pitch contour

Overall, the present results suggest that pitch contour perception can be improved through auditory training in both individuals with congenital amusia and control participants. We observed positive evidence for an effect of the training on participants' *d'* and decisive evidence for reaction times. Liu et al. (2017), demonstrated that amusics improved their pitch direction identification threshold after completing 10 pitch contour training sessions, whereas a group of amusic participants not enrolled in training did not improve in pitch contour at re-test. In addition, their trained amusics improved their MBEA contour sub-scale performance compared to untrained amusics, and, interestingly, the larger improvement in pitch threshold was observed between the first and the second training. In their study, a larger effect size for the main effect of training on the pitch direction identification threshold was reported (η^2_G=0.47) compared to the present study (η^2_p=0.243). A larger effect size has also been observed by Whiteford & Oxenham (2018) (η^2_p=0.644) after 4 training sessions with psychoacoustical tasks (pitch or localization discrimination), suggesting a stronger training effect with larger numbers of training sessions. Using comparison of pitch contour sequence (pattern) training, Foxton, Brown, et al. (2004) demonstrated that non-musician participants improved their perception of pitch contour pattern after 7 training sessions. Similarly, it has been observed that non-musician participants could improve their pitch contour processing when identifying pitch direction in syllables after three sessions of pitch contour training (Wayland et al., 2010). Moreover, previous studies demonstrated that musical training enhances contour processing, as evidenced by comparisons between musicians and non-musicians (Fujioka et al., 2004; Schön et al., 2004; Wayland et al., 2010). This improvement is also observed at the cortical level: musicians present larger mismatch negativity (MMN) in response to pitch contour deviations compared to non-musician individuals (Fujioka et al., 2004). The present study showed that even one session of auditory training could be beneficial in improving pitch perception in individuals with congenital amusia and in controls. No group difference has been observed on training gains. Despite deficits in pitch contour processing, individuals with congenital amusia could benefit from pitch training (as in Liu et al., 2017; and Whiteford and Oxenham, 2018). Although control participants showed higher performance at study onset compared to participants with amusia, they were also able to further improve their pitch processing thanks to the training sessions. To go further, it could be interesting to investigate long-term pitch training with a simple direction of change identification task as used here, and see whether the training effect can generalize to other auditory abilities. It has been emphasized that trial-by-trial feedback could be relevant to improve training effect (Whiteford and Oxenham, 2018). In the present study, no feedback was provided which could have limited the benefits of the training. Note however that receiving frequent negative feedback, which would be the case for amusic participants, could diminish participants' motivation. A potential training facilitation with feedback should be further investigated.

In the present study, we found no correlation between participants' initial *d'* and the training gains on *d'*, however, we found decisive evidence for a correlation between participant's initial RTs and RT training gains. In the study of Liu et al. (2017), no correlation was observed between pre- and post-training discrimination threshold. To the best of our knowledge, no previous studies investigated the benefits of pitch training on RTs. We neither found a correlation between RTs and d' at study onset nor in training gains, suggesting different mechanisms for the training effects on these two measures or that RTs are a more sensitive measure than d' in this context. The single session of training had more effect on RTs than on d' which could reflect higher confidence of the participants in their performance leading to faster RTs. Nevertheless, we also observed some improvement on *d'*.

4.3. Unisensory vs. multisensory training

Several studies highlighted improved pitch perception thanks to visual cues (Albouy et al., 2015; Lidji et al., 2007; Pralus et al., 2021; Rusconi et al., 2006). In the present study, we investigated whether this effect of visual cues on auditory perception could be beneficial in pitch training. We expected that participants would benefit more from the informative visual stimuli which was associated with the tones to improve their pitch contour processing compared to the auditory-only training. Using multisensory stimuli allows attention to be directed using temporal cues. In addition, the congruency between the auditory and visual stimuli facilitates multisensory integration (Kim et al., 2008; La Rocca et al., 2020). The present study found only limited evidence for a larger benefit of audiovisual training compared to auditory-only training. Indeed, when comparing to zero the training gains after either a single session of audio-only training or a single session of audiovisual training, sizable evidence for training gains was only found after audiovisual training (for amusics, in both d' and RTs for 5th pitch intervals, and for controls, in RTs for 3rd, 4th, and 5th pitch intervals, note that this analysis was done on only half of the control data due to differences in initial RTs in the two arms of the study). However, the training gains did not differ between training conditions (audiovisual or auditory-only) when compared directly. Nevertheless, the effects were in the expected direction, showing a larger benefit for the audio-visual training over the auditory training, for both d' and RT (Fig. 4 and Table 3).

The only limited additional benefit of multisensory stimulation compared to unisensory stimulation could be explained by several factors. Firstly, as in any unisensory vs. multisensory training comparison, we here compare two active trainings, where we expect to have benefit for both. Secondly, we used a single training session design. Indeed, longer training with more training sessions could be more beneficial, especially with the multisensory training procedure, yet it would require to use a between-subject design to compare the two training regimens (as in Bernstein et al., 2013; Seitz et al., 2006). Seitz et al. (2006) observed that participants who underwent audio-visual motion discrimination training (n=5) showed greater performance improvement after 10 days of training compared to those who performed visual training over 10 days (n=5). In the same way, Bernstein et al. (2013) found that four days of multisensory audio-visual training led to greater auditory perceptual learning compared to auditory-only training in participants with normal hearing (AV training n=18; A training n=19). However, this benefit of the audio-visual training compared to the auditory-only training was not observed using a within-subjects training design with multiple sessions (Bernstein et al., 2013, 2014). In the present study comparing two types of training, the advantage of studying the effect of each after a single training session is that it allows us to compare them directly within participants (as participants do not reach ceiling performance after a single session). Overall, this study tested for the first time the effect of auditory-only and audiovisual pitch contour training, and some limited evidence in favor of more efficient learning in the audiovisual context has been observed. The present study is limited by its small sample size, and its design with fixed levels of frequency changes used in both groups. This leads to limited power: in particular here, the relatively small number of participants might have impacted the testing of interactions between pitch interval, training type, and group on training gains. Future studies could (1) target more specifically interval sizes where performance is likely to be improved (here the larger intervals, such as a 5th for participants with congenital amusia), ideally using an adaptive design (as in Van Vugt et al., 2025), and (2) include larger samples of participants. We could expect that increasing the sample size in future studies would provide stronger evidence, notably regarding the difference between audio-only and audio-visual training in specific ranges of performance. Nevertheless, the present study provides first evidence for the interest of pitch contour training with audiovisual stimuli in congenital amusia and typical populations, and provides new perspectives that can be further exploited in additional studies, including other populations with pitch processing difficulties (e.g., patients with cochlear implants).

4.4. Perspectives

Overall, we observed that amusic and control participants benefit from pitch training to improve their impaired pitch contour processing, as also observed in previous studies (Liu et al., 2017; Whiteford and Oxenham, 2018), even after a single session of training and whether implemented with auditory only or with both auditory and visual information. The difference between the audiovisual training over the audio-only training was limited, yet numerical differences were in the expected direction, with larger training gains after audio-visual training. Thus, it could be interesting to use multisensory training in a longer remediation program in individuals with congenital amusia, as well as for other disorders characterized by pitch processing deficit such as cochlear implant users (Hong and Turner, 2006; Oxenham, 2008; Pralus et al., 2021).

Fundings

This work was funded by a grant of the Fédération pour la Recherche sur le Cerveau (FRC) to AC and BT. BT was supported by an ANER grant of the Region Bourgogne-Franche Comté (MusiC). This work was conducted within the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060) of Université de Lyon, within the program "Investissements d'avenir" (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR).

Statements and declarations

Ethics approval: The study procedures were approved by a Comité de Protection des Personnes SUD-EST IV (N°ID RCB: 2012-A01209–34). Written informed consent was obtained from all participants.

CRediT authorship contribution statement

Caliani Hoarau: Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis. **Agathe Pralus:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Salomé Serres-Blain:** Writing – review & editing, Visualization, Investigation, Formal analysis, Data curation. **Lesly Fornoni:** Writing – review & editing, Investigation. **Patrick Bouchet:** Writing – review & editing, Conceptualization. **Barbara Tillmann:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Yohana Lévêque:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Anne Caclin:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declared no conflicts of interest.

Acknowledgments

The authors wish to thank Gwenaëlle Mariette for help with data acquisition.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.heares.2026.109639.

References

Albouy, P., Lévêque, Y., Hyde, K.L., Bouchet, P., Tillmann, B., Caclin, A., 2015. Boosting pitch encoding with audiovisual interactions in congenital amusia. Neuropsychologia 67, 111–120. https://doi.org/10.1016/j.neuropsychologia.2014.12.006.

Ayotte, J., Peretz, I., Hyde, K., 2002. A group study of adults afflicted with a music-specific disorder. Brain 125, 238–251. https://doi.org/10.1093/brain/awf028.

Bernstein, L.E., Auer, E.T., Eberhardt, S.P., Jiang, J., 2013. Auditory perceptual learning for speech perception can be enhanced by audiovisual training. Front. Neurosci. 7. https://doi.org/10.3389/fnins.2013.00034.

Bernstein, L.E., Eberhardt, S.P., Auer, E.T., 2014. Audiovisual spoken word training can promote or impede auditory-only perceptual learning: prelingually deafened adults with late-acquired cochlear implants versus normal hearing adults. Front. Psychol. 5. https://doi.org/10.3389/fpsyg.2014.00934.

Carcagno, S., Plack, C.J., 2011. Subcortical plasticity following perceptual learning in a pitch discrimination task. J. Associat. Res. Otolaryngol. 12 (1), 89–100. https://doi.org/10.1007/s10162-010-0236-1.

Clark, C.N., Golden, H.L., Warren, J.D., 2015. Acquired amusia. In: Handbook of Clinical Neurology, 129. Elsevier, pp. 607–631. https://doi.org/10.1016/B978-0-444-62630-1.00034-2.

De Jong, T., 2019. A Bayesian Approach to the Correction for Multiplicity. Center for Open Science. https://doi.org/10.31234/osf.io/s56mk.

Dundon, N.M., Bertini, C., Làdavas, E., Sabel, B.A., Gall, C., 2015. Visual rehabilitation: visual scanning, multisensory stimulation and vision restoration trainings. Front Behav. Neurosci. 9. https://doi.org/10.3389/fnbeh.2015.00192.

Foxton, J.M., Brown, A.C.B., Chambers, S., Griffiths, T.D., 2004a. Training improves acoustic pattern perception. Current Bio. 14 (4), 322–325. https://doi.org/10.1016/j.cub.2004.02.001.

Foxton, J.M., Dean, J.L., Gee, R., Peretz, I., Griffiths, T.D., 2004b. Characterization of deficits in pitch perception underlying "tone deafness. Brain 127 (4), 801–810. https://doi.org/10.1093/brain/awh105.

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., Làdavas, E., 2005. Audiovisual integration in patients with visual deficit. J. Cogn. Neurosci. 17 (9), 1442–1452. https://doi.org/10.1162/0898929054985446.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2004. Musical training enhances automatic encoding of melodic contour and interval structure. J. Cogn. Neurosci. 16 (6), 1010–1021. https://doi.org/10.1162/0898929041502706.

Galvin, J.J., Fu, Q.-J., Nogaki, G., 2007. Melodic contour identification by Cochlear implant listeners. Ear & Hearing 28 (3), 302–319. https://doi.org/10.1097/01.aud.0000261689.35445.20.

Gates, G.A., Mills, J.H., 2005. Presbycusis. The Lancet 366 (9491), 1111–1120. https://doi.org/10.1016/S0140-6736(05)67423-5.

Grimault, N., Micheyl, C., Carlyon, R.P., Collet, L., 2002. Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. Percept. Psychophys. 64 (2), 189–197. https://doi.org/10.3758/BF03195785.

Herholz, S.C., Zatorre, R.J., 2012. Musical training as a framework for brain plasticity: behavior, function, and structure. Neuron 76 (3), 486–502. https://doi.org/10.1016/j.neuron.2012.10.011.

Hoarau, C., Pralus, A., Moulin, A., Bedoin, N., Ginzburg, J., Fornoni, L., Aguera, P.-E., Tillmann, B., Caclin, A., 2024. Deficits in congenital amusia: pitch, music, speech, and beyond. Neuropsychologia 202, 108960. https://doi.org/10.1016/j.neuropsychologia.2024.108960.

Hong, R.S., Turner, C.W., 2006. Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. J. Acoust. Soc. Am. 120 (1), 360–374. https://doi.org/10.1121/1.2204450.

Kim, R.S., Seitz, A.R., Shams, L., 2008. Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS ONE 3 (1), e1532. https://doi.org/10.1371/journal.pone.0001532.

Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y., 2001. Pitch discrimination: are professional musicians better than non-musicians? J. Basic Clin. Physiol. Pharmacol. 12 (2), 125–144. https://doi.org/10.1515/JBCPP.2001.12.2.125.

La Rocca, D., Ciuciu, P., Engemann, D.-A., Van Wassenhove, V., 2020. Emergence of β and γ networks following multisensory training. NeuroImage 206, 116313. https://doi.org/10.1016/j.neuroimage.2019.116313.

Lavie, L., Banai, K., Attias, J., Karni, A., 2014. How difficult is difficult? Speech perception in noise in the elderly hearing impaired. J. Basic Clin. Physiol. Pharmacol. 25 (3), 313–316. https://doi.org/10.1515/jbcpp-2014-0025.

Lee, M.D., Wagenmakers, E.-J., 2014. Bayesian Cognitive Modeling: A Practical Course, 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759.

Lidji, P., Kolinsky, R., Lochy, A., Morais, J., 2007. Spatial associations for musical stimuli: a piano in the head? J. Experim. Psychol.: Human Percept. Perform. 33 (5), 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189.

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

Liu, F., Jiang, C., Francart, T., Chan, A.H.D., Wong, P.C.M., 2017. Perceptual Learning of Pitch Direction in Congenital Amusia. Music Perception 34 (3), 335–351. https://doi.org/10.1525/mp.2017.34.3.335.

Liu, F., Patel, A.D., Fourcin, A., Stewart, L., 2010. Intonation processing in congenital amusia: Discrimination, identification and imitation. Brain 133 (6), 1682–1693. https://doi.org/10.1093/brain/awq089.

Liu, F., Xu, Y., Patel, A.D., Francart, T., Jiang, C., 2012. Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. Brain and Cognition 79 (3), 209–215. https://doi.org/10.1016/j.bandc.2012.03.008.

Loui, P., Guenther, F.H., Mathys, C., Schlaug, G., 2008. Action–perception mismatch in tone-deafness. Current Biology 18 (8), R331–R332. https://doi.org/10.1016/j.cub.2008.02.045.

Magne, C., Schon, D., Besson, M., 2021. Musician Children Detect Pitch Violations in Both Music and Language Better than Nonmusician Children. Behavioral and Electrophysiological Approaches 18 (2), 2021.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264 (5588), 746–748.

Melara, R.D., O'Brien, T.P, 1987. Interaction Between Synesthetically Correspond. Dimensions.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219 (1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004.

Nooteboom, S., 1997. The prosody of speech: melody and rhythm. The Handbook of Phonetic Sci. 5, 640–673.

Oxenham, A.J., 2008. Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. Trends Amplif. 12 (4), 316–331. https://doi.org/10.1177/1084713808325881.

Peretz, I., 2013. The Biological Foundations of music. The Psychology of Music. Elsevier, pp. 551–564. https://doi.org/10.1016/B978-0-12-381460-9.00013-4.

Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., Jutras, B., 2002. Congenital Amusia. Neuron 33 (2), 185–191. https://doi.org/10.1016/S0896-6273(01)00580-3.

Peretz, I., Champod, A.S., Hyde, K., 2003. Varieties of musical disorders. Ann. N. Y. Acad. Sci. 999 (1), 58–75. https://doi.org/10.1196/annals.1284.006.

Peretz, I., Hyde, K.L., 2003. What is specific to music processing? Insights from congenital amusia. Trends Cogn. Sci. 7 (8), 362–367. https://doi.org/10.1016/S1364-6613(03)00150-5.

Peretz, I., Vuvan, D.T., 2017. Prevalence of congenital amusia. European J. Human Genet. 25 (5), 625–630. https://doi.org/10.1038/ejhg.2017.15.

Pihan, H., 2006. Affective and linguistic processing of speech prosody: DC potential studies. In: Progress in Brain Research, 156. Elsevier, pp. 269–284. https://doi.org/10.1016/S0079-6123(06)56014-5.

Pralus, A., Hermann, R., Cholvy, F., Aguera, P.-E., Moulin, A., Barone, P., Grimault, N., Truy, E., Tillmann, B., Caclin, A., 2021. Rapid assessment of non-verbal auditory perception in normal-hearing participants and Cochlear implant users. J. Clin. Med. 10 (10), 2093. https://doi.org/10.3390/jcm10102093.

Rouger, J., Fraysse, B., Deguine, O., Barone, P., 2008. McGurk effects in cochlear-implanted deaf subjects. Brain Res. 1188, 87–99. https://doi.org/10.1016/j.brainres.2007.10.049.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Barone, P., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. Proceed. Nat. Acad. Sci. 104 (17), 7295–7300. https://doi.org/10.1073/pnas.0609419104.

Rusconi, E., Kwan, B., Giordano, B., Umilta, C., Butterworth, B., 2006. Spatial representation of pitch height: the SMARC effect. Cognition 99 (2), 113–129. https://doi.org/10.1016/j.cognition.2005.01.004.

Schneider, B.A., Pichora-Fuller, M.K., 2001. Age-related changes in temporal processing: implications for speech perception. Semin. Hear. 22 (03), 227–240. https://doi.org/10.1055/s-2001-15628.

Schön, D., Magne, C., Besson, M., 2004. The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology 41 (3), 341–349. https://doi.org/10.1111/1469-8986.00172.x.

Seitz, A.R., Kim, R., Shams, L., 2006. Sound facilitates visual learning. Current Bio. 16 (14), 1422–1427. https://doi.org/10.1016/j.cub.2006.05.048.

Shams, L., Seitz, A.R., 2008. Benefits of multisensory learning. Trends Cogn. Sci. 12 (11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006.

Stewart, L., 2011. EPS Prize Lecture: characterizing congenital amusia. Quarterly J. Experim. Psychol. 64 (4), 625–638. https://doi.org/10.1080/17470218.2011.552730.

Tang, C., Hamilton, L.S., Chang, E.F., 2017. Intonational speech prosody encoding in the human auditory cortex. Science 357 (6353), 797–801. https://doi.org/10.1126/science.aam8577.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., Schröger, E., 2005. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp. Brain Res. 161 (1), 1–10. https://doi.org/10.1007/s00221-004-2044-5.

Tillmann, B., Albouy, P., Caclin, A., 2015. Congenital amusias. In: Handbook of Clinical Neurology, 129. Elsevier, pp. 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., Peretz, I., 2011a. Congenital amusia (or Tone-Deafness) interferes with pitch processing in tone languages. Front Psychol. 2. https://doi.org/10.3389/fpsyg.2011.00120.

Tillmann, B., Graves, J.E., Talamini, F., Lévêque, Y., Fornoni, L., Hoarau, C., Pralus, A., Ginzburg, J., Albouy, P., Caclin, A., 2023. Auditory cortex and beyond: deficits in congenital amusia. Hear. Res. 437, 108855. https://doi.org/10.1016/j.heares.2023.108855.

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., Caclin, A., 2016. Impaired short-term memory for pitch in congenital amusia. Brain Res. 1640, 251–263. https://doi.org/10.1016/j.brainres.2015.10.035.

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., Peretz, I., 2011b. Fine-grained pitch processing of music and speech in congenital amusia. J. Acoust. Soc. Am. 130 (6), 4089–4096. https://doi.org/10.1121/1.3658447.

Tillmann, B., Schulze, K., Foxton, J.M., 2009. Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. Brain Cogn. 71 (3), 259–264. https://doi.org/10.1016/j.bandc.2009.08.003.

Van Vugt, F.T., Fornoni, L., Pralus, A., Albouy, P., Caclin, A., Tillmann, B., 2025. Beyond pitch: temporal processing deficits in congenital amusia. Psychol. Res. 89 (5), 141. https://doi.org/10.1007/s00426-025-02153-w.

von Kriegstein, K., 2012. A multisensory perspective on Human auditory communication. In: Murray, M.M., Wallace, M.T. (Eds.), The Neural Bases of Multisensory Processes. CRC Press/Taylor & Francis. http://www.ncbi.nlm.nih.gov/books/NBK92846/.

Wayland, R., Herrera, E., Kaan, E., 2010. Effects of musical experience and training on pitch contour perception. J. Phon. 38 (4), 654–662. https://doi.org/10.1016/j.wocn.2010.10.001.

Whiteford, K.L., Oxenham, A.J., 2018. Learning for pitch and melody discrimination in congenital amusia. Cortex 103, 164–178. https://doi.org/10.1016/j.cortex.2018.03.012.