Psychological Research (2025) 89:141
https://doi.org/10.1007/s00426-025-02153-w

RESEARCH

Beyond pitch: temporal processing deficits in congenital amusia

F. T. van Vugt[1,2,3,4] · L. Fornoni[1] · A. Pralus[1] · P. Albouy[1,3,4,5] · A. Caclin[1] · B. Tillmann[1,6]

Received: 3 December 2024 / Accepted: 25 June 2025 / Published online: 11 September 2025

Abstract

Congenital amusia is a neurodevelopmental disorder resulting in impaired pitch perception and memory. Here we investigated whether participants with congenital amusia have deficits in temporal processing of auditory information, in addition to pitch processing deficits. Individuals with amusia ($n=19$) and matched controls ($n=21$) were presented with sequences of five tones in which one tone was sometimes shifted in pitch or in time, and we adaptively assessed psychophysical thresholds for detecting these shifts. Pitch thresholds of the amusia group were higher (i.e., worse) than those of the control group as expected, and, crucially, time thresholds were too, although the group difference for time thresholds was smaller. Across participants, time thresholds correlated with pitch thresholds. Principal component analysis revealed that all pitch- and time-related variables (thresholds and amusia battery scores) were correlated to one component that also distinguished between amusics and controls; whereas a second component captured additional variability on the time task. Simulations suggest that prior studies had not found these time processing deficits because they had less statistical power, likely due to smaller sample sizes. The observed time processing deficit is in agreement with amusic individuals' subjective reports about their difficulties following the rhythm of the music. These data suggest that amusia deficits are not restricted to pitch, but extend to the time domain, yet with a smaller effect size, and at least when the stimuli have a clear pitch content, such as for tone sequences or music.

Significance statement

There is a long-standing debate as to whether the brain processes pitch and time information separately or with shared mechanisms. The hypothesis of separate processing is supported by reports of deficits that appear to be selective to either dimension, such as pitch processing impairments in congenital amusia. However, here we show that congenital amusics exhibit time processing deficits, albeit more subtle than the documented pitch processing deficits. This work supports a more integrated view in which pitch and time processing share some common resources.

Caclin A. and Tillmann B. contributed equally to this work.

✉ F. T. van Vugt
floris.van.vugt@umontreal.ca

B. Tillmann
barbara.tillmann@cnrs.fr

[1] Université Claude Bernard Lyon 1, INSERM, CNRS, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, Bron F-69500, France

[2] Department of Psychology, University of Montreal, Montreal, QC, Canada

[3] International Laboratory for Brain, Music and Sound Research BRAMS, Montreal, QC, Canada

[4] Centre for Research on Brain, Language and Music–CRBLM, Montreal, QC, Canada

[5] CERVO Brain Research Centre, School of Psychology, Laval University, Quebec, Canada

[6] Université Bourgogne Europe, CNRS, LEAD UMR5022, Dijon 21000, France

Introduction

Music is organized in pitch and time (e.g., McAdams, 1989; Mehr, 2024). Pitch processing has been reported to be impaired in a condition referred to as congenital amusia, a neurodevelopmental disorder affecting pitch perception and memory (Albouy et al., 2013, 2019; Ayotte et al., 2002; Peretz, 2016; Tillmann et al., 2015; Tillmann, Lévêque et al., 2016b). This phenomenon has been documented long before (Allen, 1878), but research has been accelerated with the advent of a dedicated testing battery, the Montreal Battery for the Evaluation of Amusia (MBEA) (Peretz et al., 2003; Peretz & Hyde, 2003). Peretz et al. (2003) presented short melodies to self-declared congenital amusic individuals, mostly in a same/different short-term memory paradigm, where stimuli differed either in pitch or time. Testing of a large population allowed for calculating cut-off scores (2 SD below the average), defining the threshold below which an individual's performance was considered as impaired for each of the subtests, and for the overall battery, leading to the categorization as "amusic". Using this criterion, for the pitch dimension, most amusic individuals were impaired in detecting when a tone in the target melody was replaced by a tone that was out-of-scale, violated contours, or changed the size of the used pitch intervals. For the time dimension, only a subset of the amusic participants showed impairments in detecting alterations in rhythm or discriminating between two target meter structures (Peretz et al., 2003; see also Tillmann et al., 2016a, 2023 for reviews and other data sets). The hypothesis that congenital amusia involves primarily impairments at a basic level of pitch processing, and this even outside of musical contexts, was tested in a seminal study by Hyde and Peretz (2004). Sequences of five isochronous tones of the same pitch were presented to participants, where one tone was shifted either in pitch or in time, by a varying amount (i.e., one of five changes differing in predetermined extent). Amusics showed drastic drops in detection performance for pitch differences less than a semitone, whereas they performed on a par with neurotypical, matched control individuals in terms of detecting time differences.

The hypothesis that congenital amusics' deficits are mostly restricted to the pitch domain has been challenged. Some amusic participants had difficulties tapping along to a musical piece and showed impaired rhythm during singing (Ayotte et al., 2002). Amusics have also been reported to have impaired perception of acoustic beating (Graves et al., 2023) and of amplitude modulation (Whiteford & Oxenham, 2017), which is thought to not involve pitch (Graves et al., 2023). Subjectively, some amusic individuals report experiencing deficits in temporal processing. One study reported that 73% of amusics described themselves as poor dancers, contrasting with 7% of controls (Peretz et al., 2008).

Why would amusics show deficits in some time tasks and not others? One possible explanation might be that amusics' faulty pitch processing system interferes with otherwise normal time processing. The prediction of this hypothesis would be that amusics' time processing should be impaired only when it comes to materials that have pitch content (e.g., sequences of tones with different pitch heights as opposed to non-pitched drum sounds), especially when the tones vary in pitch (as in a melody). Under this hypothesis, time processing in congenital amusia should be in the typical range with non-pitched sounds, impaired in sequences of constant-pitch stimuli, and even more strongly impaired in pitch-varying sequences, such as melodies. There is some evidence supporting that prediction. Foxton et al. (2006) presented sequences of five tones that either varied in pitch or not in a short-term memory paradigm where participants' task was to detect whether a timing change (i.e., a lengthened interval) was introduced in the second sequence. When the stimuli varied in pitch, amusic individuals' performance was significantly impaired in detecting time differences in comparison to matched control participants. Interestingly, there was a statistical trend towards group differences in a condition without pitch variation (yet still tones with pitch content), suggesting a general time processing deficit in amusia at least when sounds have a pitch content. In a duration discrimination task, Pfeuty and Peretz (2010) had participants judge whether two time intervals between tones were similar or not. Task-irrelevant pitch variations were added and contrasted with a pitch-constant baseline condition. When the tones varied in pitch even as little as a quarter of a semitone, neurotypical control participants showed impaired time discrimination performance relative to when the tones did not vary in pitch. Amusic participants' time performance was only affected for large pitch differences, but not small pitch differences, which might have not been perceived due to their increased pitch discrimination thresholds. Another study assessing both perception and action found that rhythmic performance of congenital amusic participants was impaired when using pitched stimuli (e.g., tones played with a piano timbre) relative to stimuli with less clear pitch percept, such as drums (Phillips-Silver et al., 2013).

Yet, there is also data contradicting the hypothesis that amusics' time processing is abnormal only for pitched stimuli (sounds with clearly defined pitch) or sequences with varying pitch. Recently, it was found that for pitch-varying melodies but crucially also for sequences of non-pitched (drum) sounds, beat perception and synchronization to the beat were impaired in six out of the ten tested amusic participants (Lagrois & Peretz, 2019). In sum, it remains unclear

to what extent, and under what conditions, time processing may be deficient in listeners with congenital amusia.

Turning to time processing per se, separate lines of research have studied potential deficits in time perception, referred to as beat deafness, in their own right. Time perception deficits are documented to manifest in various ways, such as difficulties finding the beat in a musical piece (beat deafness) and/or aligning one's movement to a beat (sensorimotor synchronization) (Philips-Silver et al., 2011). Individuals showing these time perception deficits often show normal pitch perception (Bégel et al., 2017; Dalla Bella & Sowiński, 2015; Sowiński & Dalla Bella, 2013; Tranchant et al., 2016). By some estimates, prevalence of beat deafness (poor synchronizers) may be around 16% of the general population (Sowiński & Dalla Bella, 2013). The finding that a much greater proportion of amusics, possibly half, have time perception deficits (e.g., Peretz et al., 2003; Tillmann et al., 2016a), therefore suggests there may be a sizable comorbidity of pitch-related amusia with temporal processing deficits, such as reflected in the reported beat-deafness.

Taken together, the picture that emerges from the existing literature is that congenital amusia is primarily a pitch processing deficit, but time processing may be impaired as well, in particular in the presence of pitch variations. According to the hypothesis that any pitch information disrupts time processing in amusic participants, time perception should be affected when using stimuli that have a clear pitch (e.g., tones played with complex harmonics), even if that pitch is the same for all stimuli. Prior work did not find such time deficits with sequences of tones having constant pitch and played by a piano timbre (Hyde & Peretz, 2004).

A limitation of the empirical data available to date on time perception deficits in amusia lies in the fact that the tool of choice to screen for amusia - the MBEA - contains rhythmic subtests as well. Since amusics are usually classified as such based on their overall MBEA score, this classification could be biased towards including individuals who have rhythmic problems to start with: it cannot be excluded that at least some individuals are designated as amusic by virtue of low MBEA scores on rhythmic, not pitch subtasks. An open question is whether individuals classified as amusic based only on the MBEA pitch subtasks (disregarding time subtasks) exhibit timing deficits or not. Although this criterion has been used in studies on amusia (Liu et al., 2010), it has not been applied to the question of time perception.

The present study tested whether individuals with congenital amusia may exhibit fine-grained timing deficits relative to controls. To be able to observe even small between-group differences, we used an adaptive thresholding procedure, which is thought to be a more sensitive measure than a constant-stimuli paradigm in some contexts (e.g., Dai, 1995), and we tested larger sample sizes than in Hyde and Peretz (2004). The same change-detection procedure in sequences of five tones was used to measure individual thresholds for changes in either pitch or time, and we assessed whether these were correlated. We repeated the comparison of time perception thresholds for amusics classified as such using pitch-only MBEA subtasks. Finally, we explored whether the objective measures of pitch and time perception were related to subjective reports of musical, and in particular temporal, processing abilities.

Methods

Participants

Twenty-one individuals with congenital amusia were recruited using advertisements distributed via posters and flyers as well as radio interviews. We further recruited 22 matched control participants without amusia for in-lab testing. Because of the low prevalence of amusia in the general population, rather than using power analysis to guide sample size, our strategy was to recruit as many amusia individuals as possible within a time frame. Participants were classified as amusic or control by comparing their scores to cutoffs (see details below). After the experiment, during quality control of the adaptive threshold procedure (see details below), we discarded data from 2 amusic and 1 control participants so that the final sample consisted of 19 amusic and 21 control participants that did not differ significantly in sex, age, dominant hand, musical training, and education level (Table 1). The study procedures were approved by the appropriate national ethics committee, and all participants provided written informed consent prior to participating. Participants reported no neurological or psychiatric disorders. All participants had normal hearing (hearing loss inferior to 30 dB at octave frequencies between 250 and 8000 Hz in both ears).

Screening (MBEA, PDT, subjective questionnaire)

Participants completed the Montreal Battery for the Evaluation of Amusia (MBEA) (Peretz et al., 2003). This battery involved melody comparison tasks (Scale, Contour, Interval, and Rhythm subtests, see Table 1), as well as meter and incidental memory subtests. In the melody comparison subtests, participants were presented with two melodies and had to judge whether they were same or different. The melodies could differ in various ways giving rise to four subtests: Scale, Contour, Interval, Rhythm. In the Meter subtest, participants heard a single melody with harmonic accompaniment and needed to respond whether it was a march (binary meter) or waltz (ternary meter). In the Incidental

Table 1 Comparison of demographic and music-related variables in the amusia and control groups

	Amusia	Control	Statistical comparison
n	19	21	-
Sex (female/male)	14/5	16/5	Fisher's Exact Test odds ratio = 0.88, $p = 1.0$
Age (years)	32.63 (15.16)	31.86 (13.11)	t[35.84] = 0.17, $p = .86$
Hand dominance (L/R)	2/17	4/17	Fisher's Exact Test odds ratio = 1.97, $p = .66$
Education (years)	15.16 (2.73)	15.71 (2.22)	t[34.73] = −0.70, $p = .49$
Musical training (years)	0.32 (0.95)	0.45 (0.84)	t[36.16] = −0.48, $p = .63$
PDT (semi-tones)	1.44 (1.35)	0.26 (0.15)	**t[18.41] = 3.80**, $p = .001$, **d = 1.26 [0.56,1.97]**
MBEA (number of correct answers, maximum score = 30 for each subtest)			
Mean (overall)	21.81 (1.75)	27.25 (1.34)	**t[33.67] = 10.95**, $p < .0001$, **d = 3.51 [2.49,4.53]**
Mean for pitch subscales (scale, contour, interval)	20.84 (2.25)	26.86 (1.88)	**t[35.28] = 9.14**, $p < .0001$, **d = 2.92 [2.00,3.84]**
Scale	21.32 (2.85)	26.67 (2.46)	**t[35.79] = −6.33**, $p < .0001$, **d = 2.02 [1.23,2.81]**
Contour	21.79 (3.07)	27.90 (1.87)	**t[29.17] = −7.52**, $p < .0001$, **d = 2.44 [1.59,3.28]**
Interval	19.42 (3.06)	26.00 (2.32)	**t[33.48] = −7.60**, $p < .0001$, **d = 2.44 [1.59,3.28]**
Rhythm	23.95 (3.84)	27.52 (2.38)	**t[29.49] = −3.50**, $p = .0015$, **d = 1.13 [0.44,1.82]**
Meter	20.00 (4.88)	26.71 (2.08)	**t[23.82] = −5.56**, $p < .0001$, **d = 1.83 [1.06,2.59]**
Memory	24.37 (3.25)	28.67 (1.20)	**t[22.38] = > −5.44**, $p < .0001$, **d = 1.79 [1.03,2.55]**
Subjective questions, selected (% yes) [95% binomial CI]			Fisher's Exact Test
n	17	13	-
"Do you notice when a musician plays a wrong note?"	17.65% [6.19, 41.03]	53.85% [29.14, 76.79]	odds ratio = 0.20, $p = .056$
"Do you notice when somebody sings out of tune?"	52.94% [30.96, 73.83]	84.62% [57.77, 95.67]	odds ratio = 0.22, $p = .12$
"I can't dance"	58.82% [36.01, 78.39]	30.77% [12.68, 57.63]	odds ratio = 3.09, $p = .16$
"I can't follow the beat of the music"	88.24% [65.66, 96.71]	7.69% [0.39, 33.31]	**odds ratio = 67.18**, $p < .0001$

Reported values are means unless otherwise indicated; standard deviations are indicated in parentheses. MBEA: Montreal Battery for the Evaluation of Amusia. PDT: Pitch Discrimination Threshold. The subjective questionnaire data was only collected for a subset of 17 amusics and 13 controls. T-test results are indicated with degrees of freedom values after Welch-Satterthwaite correction. Subjective question results are indicated as percentage yes-responses with the 95% binomial confidence interval in square brackets. Significant differences are highlighted in bold

Memory subtest, participants were presented with melodies and asked, for each, if they had heard it before in the test session. Thus, a total of six MBEA scores were obtained (scale, contour, interval, rhythm, meter, memory). To be considered amusic a participant had to obtain an average MBEA score below 23.4 (averaged across the six subtasks, out of a maximum possible score of 30) or an average score on the MBEA pitch tasks below 21.7 (averaged across scale, interval and contour; out of a maximum possible score = 30; Liu et al., 2010), or both. All participants we recruited for the amusic group satisfied that criterion except one participant with borderline scores on both measures (23.5 and 22.33); analyses reported in the remainder of the paper were repeated excluding this participant, yielding comparable results. All controls obtained scores above both cut-offs. As expected, the average scores of the amusia group differed significantly from the scores of the control group for the overall (global) MBEA battery score and also for each of the six subtests of the battery, including the rhythm and meter subtests (Table 1), with the largest effect sizes (d > 2) observed for the three pitch-related subtests. To rule out that the classification of participants based on overall MBEA scores was biased towards those with rhythmic problems, we repeated the time threshold analysis (see below) restricting our amusic group to only those who were impaired on the pitch tasks (< 21.7 averaged across scale, interval and contour tasks, following Liu et al., 2010; 5 amusic participants were discarded).

To further document the pitch perception of our sample, we had participants perform a pitch discrimination threshold task (PDT), consisting of a two-alternative forced-choice task using a two-down/one-up staircase procedure targeting 70.7% correct performance (Tillmann et al., 2009). On each trial, participants heard two pairs of 100-ms pure

tones separated by a 150-ms silent interval. Only one of the two pairs included a pitch change, and participants had to report if it was the first or the second. As previously reported (reviewed in Tillmann, Lévêque et al., 2016b), average PDT of the amusia group was significantly higher (worse) than that of the control group, even though performance overlapped between the groups (Table 1). This is consistent with previous observations that not all individuals classified as amusics according to the MBEA have pitch deficits in simple discrimination or change detection tasks (Tillmann et al., 2016a, 2023).

Finally, to document subjective experience of participants, a subset (17 amusics and 13 controls) filled out a questionnaire about their musical experience and their relationship to music. This questionnaire, based on prior work (McDonald & Stewart, 2008; Peretz et al., 2005, 2009; Sloboda et al., 2005), included more than 90 questions such as yes/no questions about activities that relate to melodic and timing perception: "Do you notice when a musician plays a wrong note?", "Do you notice when somebody sings out of tune?", "I can't dance", "I can't follow the rhythm of music" (see Table 1 for responses to these subjective questions).

MLP: adaptive threshold estimation for pitch and time

Participants' individual psychophysical thresholds were assessed for detecting changes in pitch and time using 5-tone sequences. In the standard sequences (without change), five tones were presented isochronously in a sequence, where the interval between the onsets of subsequent tones (Inter-tone-Onset Interval, IOI[1]) was 350 ms and a fundamental frequency of each tone was 1047 Hz (C6). Each tone had a duration of 100ms and was a sine wave tone created with 25ms ramping for fade-in and fade-out. In the *pitch* task, the fourth tone was on time (i.e., respecting the IOI of 350ms), but shifted upwards in pitch by a variable amount between 0 and 300 cents (300 cents = 3 semitones = 198 Hz in this case). In the *time* task, the fourth tone was delayed by a variable amount between 0 and 57% of the IOI, i.e., 200ms) relative to the other four tones, but played at the standard fundamental frequency (Ehrlé & Samson, 2005; Hyde & Peretz, 2004). This time task was defined as in prior work (van der Steen et al., 2014; van Vugt et al., 2016; van Vugt & Tillmann, 2014; Vugt & Tillmann, 2015). Implementing only one pitch direction (up) and time direction (late) was based on Hyde & Peretz (2003) reporting no effects or interactions involving the direction of pitch or time changes. In both versions of the task (pitch and time), the psychophysical threshold was determined using the adaptive Maximum Likelihood Procedure (MLP) (Deguchi et al., 2012; Green, 1993, 1995; Gu & Green, 1994; Leek, 2001; van Vugt & Tillmann, 2014). This procedure aimed to establish the amount of pitch shift or time delay, respectively, where participants responded in 50% of cases that they perceived a change.

Briefly, the MLP algorithm works as follows. Participants' probability of responding yes (change perceived) is modeled by a sigmoid psychometric function as a function of the actual stimulus change (for the pitch task, the pitch shift in cents; for the time task, the delay in ms). The equation for the psychometric curves was $p(yes) = a + (1 - a) * (1/(1 + \exp(-k*(x - m))))$, where a is the false alarm rate (the estimated proportion of yes responses when there was no change in pitch or time), k controls the slope and was fixed here at 1 $(\text{cent})^{-1}$ in the case of the pitch task, and 0.1 $(\text{ms})^{-1}$ in the case of the time task, m is the midpoint of the psychometric curve (in cents or ms) and x is the amount of pitch shift in cents for the pitch task and the amount of delay in the tone onset (in ms) for the time task. A grid of possible parameter values was created by combining a range of possible curve midpoints m (for the pitch task: 300 values linearly spaced between 0 and 300 cents; for the time task: 200 values linearly spaced in the interval between 0 and 57% IOI, i.e. 200ms) with a set of false alarm rates (a in [0, 10, 20, 30, 40%]) yielding a total of 1500 parameter combinations in the pitch task or 1000 in the time task. The likelihood corresponding to each combination of parameter values was calculated at each trial based on the participant's responses thus far using the psychometric function defined by those parameter values. The maximally likely combination was then used to determine the pitch change to be presented on the next trial (or timing delay, in the case of the timing task), except for catch trials where there was no change in pitch nor time (these trials served to test whether participants used a strategy where they always responded 'yes'). In each block, the first 12 trials contained 2 catch trials at random positions and the next 24 trials contained 4 catch trials again at random positions. No information was carried over between blocks of the same participant so as to make the blocks independent estimates of the participant's threshold. The Python implementation of this MLP procedure is available open-source at https://github.com/florisvanvugt/PythonMLP.

[1] Some authors have referred to this quantity as ITI (inter-tone-interval) but we chose the label IOI to clarify that we refer to the time between onsets of tones and not the duration of the silence between them.

Procedure

Participants completed both pitch and time tasks; the order of the two tasks was counterbalanced across participants. For each task, participants completed three blocks. Each block contained 36 trials, chosen based on recommendations (Grassi & Soranzo, 2009) that MLP requires about 24 (Leek et al., 2000) or 30 (Amitay et al., 2006) trials. Each trial consisted of a sequence of five tones. The tones and the IOI were identical except for the fourth tone, which was either shifted up in pitch or delayed in time (except for catch trials, see above). Participants were asked to respond by button press whether they perceived a change (or not). They were informed at the beginning of the block of the type of change to expect (pitch or time). The threshold task took 10–15 min including instructions, example trials, and a short familiarization phase before the first block of each type (see van Vugt & Tillmann, 2014, for details), as well as small breaks between blocks.

Data analyses

Participants' thresholds for the MLP tasks (pitch and time tasks) were determined offline as follows using previously established criteria (van Vugt & Tillmann, 2014; van der Steen et al., 2014). First, blocks with more than 30% yes responses to catch trials were discarded (as was done in prior research, e.g., Dalla Bella et al., 2017). This 30% cutoff criterion was chosen because prior methodological studies considered false alarm rates up to this percentage but not beyond (Gu & Green, 1994; Leek et al., 2000); and false alarm rates higher than 30% would leave less than 40% points differences in yes responses based on which the psychometric curves are fitted. Second, to assess whether the threshold estimate was sufficiently stable, we fit a line to the threshold estimate for the last 10 trials in each block. If the absolute value of the slope of this line exceeded 1.00 cent/trial (pitch task; cutoff based on pilot testing) or 0.34% IOI/trial (time task; cutoff based on van Vugt & Tillmann, 2014), the block was discarded. On average 2.4 (SD = 0.75) blocks remained for the pitch task and 2.6 blocks (SD = 0.70) for the time task, of which the average threshold was calculated for each participant. For three participants (two amusic and one control), no blocks remained for either pitch or time, and these participants were hence removed from the remainder of the study (see Participants section). The key analyses (see below) were repeated without this data discarding step, that is, including all three blocks for all participants, leading to essentially the same pattern of results. In particular, the key result, that amusics have lower performance in the time task, held even when no blocks were discarded.

For pitch and time thresholds separately, we ran an independent t-test to compare the groups (amusia/control). Effect sizes in terms of Cohen *d* values are reported with 95% confidence intervals. To assess whether the pitch and time task performance were related, we calculated the Spearman correlation between the thresholds indicating the degrees of freedom in parentheses. Spearman correlations were chosen over Pearson to prevent outliers from driving or masking effects. We also computed Spearman correlations between MLP pitch task threshold and PDT, which we expected to be significantly correlated. An omnibus ANOVA analysis that combines pitch and time in the same analysis is included in supplementary materials: https://osf.io/rtuq3.

We used Principal Component Analysis (PCA) to investigate the relationships between the variables using the R package *FactoMineR* (Lê et al., 2008). We constructed a matrix where the columns were the main dependent variables (MLP pitch, MLP time, PDT, as well as all the MBEA subscales separately), and the rows corresponded to the participants in the amusia and control groups. We added group (amusia/control) as a supplementary variable, meaning that it did not affect the calculation of the components, but allowed us to assess whether components differed between the groups. Scores within each column (dependent variable) were normalized. No rotation was applied. To assess whether the data was adequate for PCA, we calculated the measure of sampling adequacy according to Kaiser-Meyer-Olkin (KMO) using the *KMO* function from the *psych* package in R. Values over 0.8 are considered adequate. In the present dataset, the KMO was 0.85. The Kaiser-Guttman rule (Kaiser, 1960) was used to decide which factors were interpretable: we retained components whose eigenvalues were greater than 1, which in this case yielded two components (eigenvalues 5.30 and 1.00, respectively) together explaining 70.02% of the variance. For ease of interpretation, MLP pitch and MLP time as well as PDT scores were inverted so that higher values correspond to better performance, and log-scaled prior to PCA in order to reduce non-normality, as measured by skewness and kurtosis. A raw correlation matrix is available in the supplementary materials at: https://osf.io/uja7x.

Subjective reports (yes/no responses in the questionnaire) were compared between groups using Fisher Exact test. Where group differences were found, we tested whether these response differences were related to performance on the pitch or time MLP tasks using Mann-Whitney U tests.

Results

MLP pitch and time tasks

As expected, pitch task (MLP) thresholds were higher (i.e., worse) in the amusia group (mean = 63.6, SD = 46.6 cents) than in the control group (mean = 23.3, SD = 12.2 cents); t[38] = 3.83, $p < .001$, d = 1.24, 95% CI [0.56, 1.92] (Fig. 1A). This effect also held when removing the borderline amusic participant, t[37] = 4.02, $p < .001$, d = 1.33, 95% CI [0.62, 2.02]. To control whether this effect may be due to outliers, we also computed a non-parametric Mann-Whitney U test, which was significant, W = 352, $p < .001$, $r = .65$ [0.43,0.80]. The MLP pitch task score was significantly correlated with the PDT score (Spearman ρ[38] = 0.74, $p < .001$), which is consistent with the hypothesis that these two tasks measure a similar perceptual capacity. This correlation also held within the amusia group (Spearman ρ[17] = 0.72, $p = .0005$) but failed to reach significance within the control group (Spearman ρ[19] = 0.37, $p = .099$).

Time task (MLP) thresholds were also higher (i.e., worse) in the amusia group (mean = 12.0, SD = 8.5%IOI) than in the control group (mean = 8.0, SD = 2.7%IOI), t[38] = 2.03, $p = .049$, d = 0.66, 95% CI [0.02, 1.29] (Fig. 1B). This was also the case when removing the borderline-amusic participant, t[37] = 2.14, $p = .038$, d = 0.71, 95% CI [0.05, 1.35]. To control whether this effect may be due to the outlier observed visually (Fig. 1B; threshold > 40% IOI), we computed the same analysis removing this individual, and the group difference still held, t[37] = 2.09, $p = .043$, d = 0.69, 95% CI [0.04, 1.33]. As an additional check, we computed a non-parametric Mann-Whitney U test with all participants' data, which was also significant, W = 282, $p = .026$, $r = .35$ 95% CI [0.04, 0.61]. To rule out that this group comparison may be biased by having classified amusics based on overall MBEA scores, we repeated the analysis restricting our amusic group to only those individuals who were impaired on the pitch task (see methods; 5 amusics excluded), and found the same group difference on time task thresholds, t[33] = 2.26, $p = .03$, d = 0.80, 95% CI [0.09, 1.50]. To test whether data preprocessing steps (discarding blocks based on erroneous catch-trial responses and/or insufficient threshold stability, see Data Analysis) affected the results, we re-ran the t-test comparing time task thresholds in amusia and control groups on the entire data set (i.e., without applying any discarding), and observed the same data pattern, either just falling short of significance (t[22.2] = 2.02, $p = .056$) for the entire dataset) or significantly t[30.8] = 2.05, $p = .049$ when removing the outlier amusic participant). When we removed the amusic participants who were impaired according to the MBEA rhythm (below a cutoff score of 23 following Peretz et al., 2003; this involved six amusics), the MLP time task group difference still held, t[32] = 2.10, $p = .044$, d = 0.76, 95% CI [0.04, 1.48]. This suggests that the MLP time task may be more sensitive to time impairments than the MBEA rhythm subscale.

Pitch and time thresholds were correlated, Spearman ρ[38] = 0.42, $p = .0074$ (Fig. 1C). This correlation was still significant when removing the borderline amusic participant, ρ[37] = 0.41, $p = .0096$. This correlation was not driven by the amusic participant with a high value in both time and pitch thresholds, as the correlation was also significant

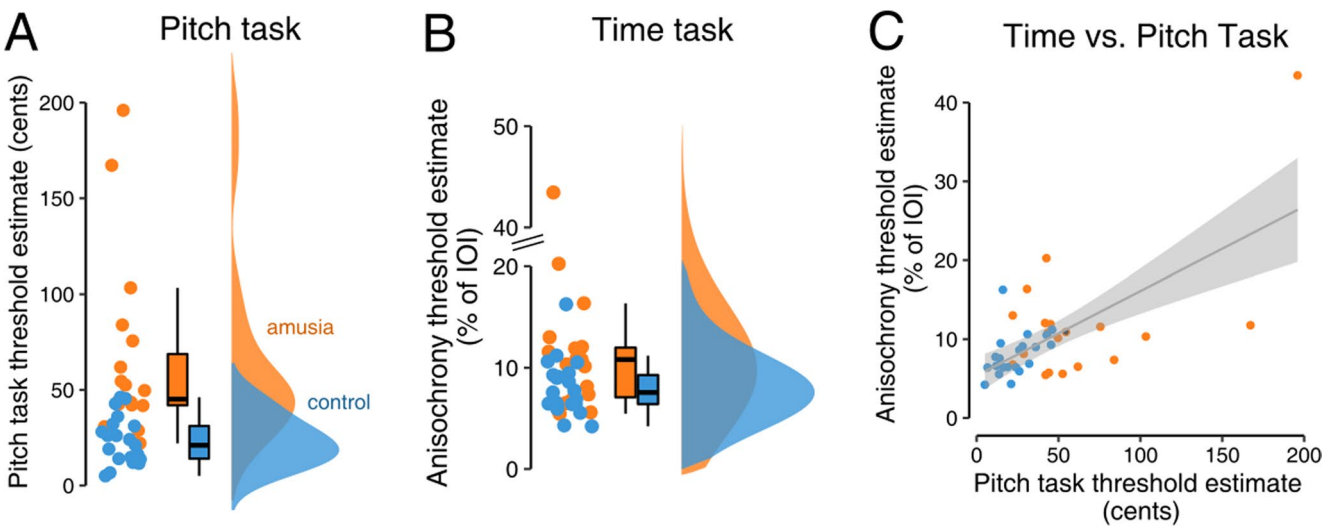

Fig. 1 The amusia group showed higher pitch thresholds (worse) than controls, as well as higher time thresholds, and although the time task effect was smaller, the thresholds of the pitch and time tasks were correlated. Note **A**: Pitch task thresholds. Here and elsewhere, dots indicate individual thresholds. Boxplots indicate group distributions as do density estimations (violin plots). **B**: Time task thresholds. **C**: Scatter plot representing the individual pitch and time thresholds. Note that the correlation was significant also without the amusic participant with high pitch and time thresholds (see main text for more details)

without this participant, ρ[37] = 0.37, p = .020. This correlation was significant within the control group, ρ[19] = 0.52, p = .015, but not within the amusia group, ρ[17] = 0.02, p = .95, even when removing the amusic outlier, ρ[16] = −0.16 p = .53. To further assess whether the overall (across groups) correlation between pitch and time thresholds was driven only by group differences, we fitted a multiple regression analysis with pitch threshold as dependent variable, and regressors time threshold and group entered simultaneously. If the correlation between pitch and time thresholds was predominantly due to group differences, we expected the time threshold regressor to not reach significance. Instead, we found that both group (t[37] = 3.05, p = .004) and time threshold (t[37] = 4.49, p < .001) were significant predictors. This pattern of findings also held when removing the borderline-amusic participant (t[36] = 4.30, p = .0001 and t[36] = 3.18, p = .003, for group and time task threshold regressors, respectively).

Effect sizes suggest that the difference between amusic and control participants' thresholds was larger in the pitch task (d = 1.33) than in the time task (d = 0.66). For the pitch task, the thresholds of 9 amusics out of 19 (47.4%) were 2 SD or more above the control mean, whereas for the time task, it was the case for only 3 amusics (16.0%). To test whether amusics differed more from controls in the pitch task than in the time task, we computed for each amusic individual the z-score of their score relative to controls (i.e., the z score for individual i was calculated from their threshold score x_i as follows, $z_i = (x_i - m_{control})/sd_{control}$ where $m_{control}$ and $sd_{control}$ are the mean and standard deviation of the control group, respectively) for both tasks separately (two values per amusic). These z-scores were significantly larger for the pitch task than the time task (t[18] = 2.58, p = .019, d = 0.53). This was also the case when removing the borderline amusic participant (t[17] = 2.57, p = .020, d = 0.54). However, this comparison of impairments between tasks may be biased by the amusic recruitment strategy using advertisements that emphasized pitch (not time) perception deficits, which is in line with prior reports (e.g., Peretz et al., 2003).

For the subjective reports, amusic individuals agreed less with "Do you notice when a musician plays a wrong note?" (18%) than controls (54%) although this difference showed only a trend towards statistical significance (p = .056) (Table 1). Across all participants, a yes-response to this question predicted a higher threshold (worse performance) for the MLP pitch task (W = 139.00, p = .046, r = .32 [−0.02, 0.58]), but not the time task (W = 102.00, p = .77, r = .05 [−0.28, 0.35]). For the questions "Do you notice when somebody sings out of tune?" or "I can't dance", amusic individuals and controls responses were not statistically different, even though the difference in percentages between the two groups was in the direction expected based on previous reports. Finally, 88% (88%) of amusic participants reported "I can't follow the rhythm of music", which was significantly greater than 8% of controls (p < .0001) (Table 1). A yes-response to this question predicted higher threshold (worse performance) for the MLP pitch task (W = 29.50, p = .001, r = -.53 [−0.68, −0.30]), but surprisingly not for the time task (Mann-Whitney W = 81.00, p = .31, r = -.17 [−0.47, 0.15]). In sum, some subjective reports confirmed differences between amusics and control individuals, and were significantly correlated with the pitch deficits, whereas no significant correlations were found with the time deficits.

Table 2 The first principal component (PC1) correlates with all variables included in the PCA, whereas the second component (PC2) correlates with the MLP time and pitch task thresholds as well as the MBEA rhythm subscale

Variable	PC1		PC2	
	r	*p*	*r*	*p*
Pitch threshold (MLP)$^+$	0.82	**<0.0001***	0.32	**0.04**
Time threshold (MLP)$^+$	0.51	**0.0007***	0.68	**<0.0001***
PDT score$^+$	0.87	**<0.0001***	0.18	0.26
MBEA scale	0.78	**<0.0001***	−0.31	0.05
MBEA contour	0.80	**<0.0001***	−0.28	0.08
MBEA interval	0.88	**<0.0001***	−0.22	0.18
MBEA rhythm	0.67	**<0.0001***	−0.34	**0.03**
MBEA meter	0.70	**<0.0001***	0.24	0.13
MBEA memory	0.79	**<0.0001***	−0.07	0.66

P-values have not been corrected for multiple comparisons (significant uncorrected ones are indicated in bold), but asterisks (*) in addition to bold font indicate those values whose significance survives FDR correction at p < .05 (Benjamini & Hochberg procedure, performed separately within each PC). $^+$ Pitch/time MLP and PDT scores were inverted so that higher scores correspond to better performance

Principal component analysis

To clarify the relationship between the various outcome measures, a Principal Component Analysis (PCA) was calculated (details in methods), which revealed two components. The first component was significantly correlated with all variables that were included (Table 2) and significantly distinguished amusics from controls (R^2 = 0.75, p < .0001) (Fig. 2). The second component was correlated with the MLP time task threshold.

Integrating present findings with prior results using simulations

The finding in the present study, that amusics differ from controls for both pitch and time tasks, appeared to differ from the closest prior study which found amusics did not differ from controls in the time task (Hyde & Peretz, 2004, although note that other studies using different stimuli and tasks did find time perception differences, e.g. Lagrois &

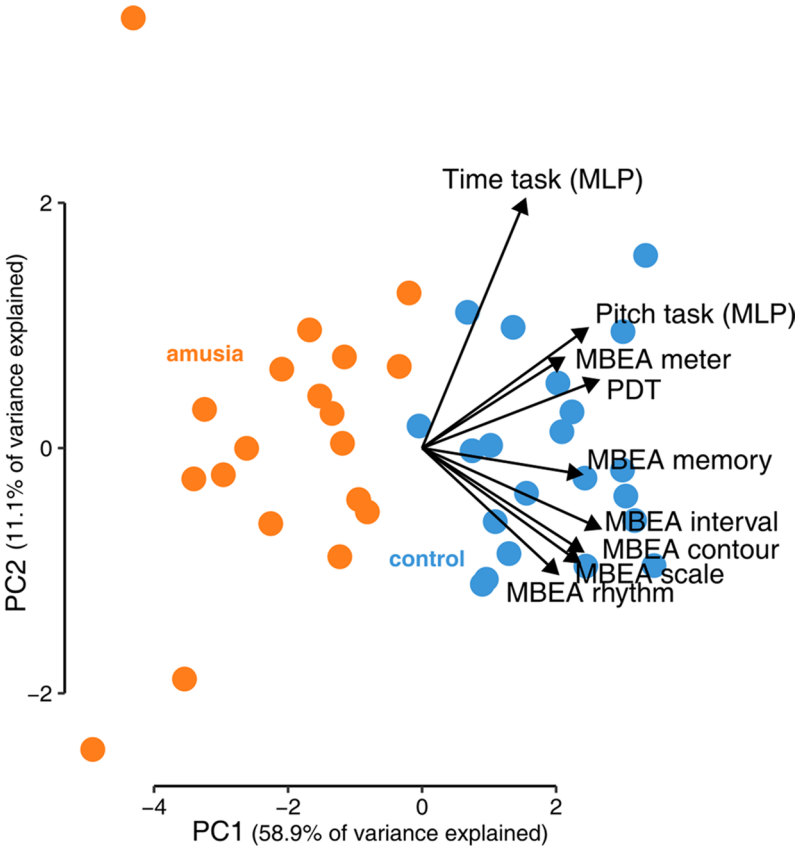

Fig. 2 Principal Component Analysis reveals that all variables are significantly related to a single component that distinguishes amusic and control individuals (PC1). A second component (PC2) captures additional variability related to temporal tasks, especially the MLP time task. Note The axes represent the first two principal components (PC1 on the horizontal and PC2 on the vertical axis). Dots represent the coordinates of individual participants, that is, their scores on the two factors. Dot colours represent the group the individual belongs to. Arrows represent the loadings of the scores of the tasks in this study (MLP pitch and time scores, PDT and MBEA). That is, the arrows indicate how for a particular task, an individual's score on this task is multiplied to yield the individual's factor loadings. Since the underlying scores are standardized, this means that the horizontal and vertical components of the arrow for each task indicates how important that task is for the component in question (PC1 and PC2, respectively). The arrows have been rescaled for clarity of presentation. In terms of interpretation, this shows that the first principal component (PC1) is associated with performance across all tasks (since all arrows have a rightward horizontal component) and separating the two groups, while the second principal component (PC2) predominantly captures variance specific to the MLP-time performance

Peretz, 2019). We performed simulations that show that both results can be explained in the case where there are real but small timing differences between amusics and controls.

Simulation procedure The methodology here differed from Hyde and Peretz (2004) in terms of sample size (20 participants in total in their study vs. 40 here) and the method of investigation (constant stimuli in their study vs. an adaptive threshold procedure here with different numbers of trials). To investigate which of these factors might explain why the present study found a different result from Hyde and Peretz (2004), we estimated the statistical power of each combination of experimental paradigm (Hyde & Peretz (2004)'s paradigm vs. the present study, hereafter MLP paradigm) and number of participants (10/10 or 19/21 amusic/control participants), thus yielding four sets of simulations. In all four cases, we simulated individual "participants" in the same way (described below) and varied only the way in which stimuli were presented to reproduce either Hyde & Peretz or MLP paradigm. A simulated participant was basically treated as a computational function f(x) that, given a particular delay (in %IOI, or, analogously, a pitch shift in cents) returns a yes ("change detected") or no ("no change detected") response with specific probabilities (see below for details on how this was achieved). These simulated participants then were submitted to either Hyde and Peretz (2004) or MLP paradigm. In either case, we could then test whether amusics differed from controls, exactly as was done in the empirical studies with real human participants. This allowed us to compute statistical power by dividing the number of simulations where significant group differences (amusics vs. controls) were found by the total number of simulations.

Simulating individual participants Each simulated participant responded to a given stimulus with yes ("change detected") or no following a given probability using the basic random number generator in Python. To calculate the probability as a function of the stimulus delay (or, analogously, pitch change), we used cumulative Gaussian psychometric response curves. Each participant had their own psychometric curve with a variable midpoint that was chosen randomly based on the data observed here: for the time task, the midpoints of amusic participants were placed at 9 (SD = 3) %IOI and those of the control participants at 7 (SD = 3) %IOI. The psychometric curve slopes were held constant at sigma = 3.5 and the false alarm rate at 5%. For the pitch task, the psychometric curves had variable midpoints at 35 cents (SD = 25) for amusics and at 12 cents

(SD = 9.6) for controls and fixed slope (sigma = 2.6) and false alarm rate (1%).

Simulating the Hyde and Peretz (2004) paradigm time task We simulated amusic and control participants, presented each with the exact same stimulus regime as used in Hyde and Peretz (2004), and used the same statistical analysis. Specifically, each of the time delays (8, 10, 12, 14, 16%IOI) was presented 36 times, yielding a total of 180 trials. A further 180 trials were included without delay. Each simulated participant thus responded to a total of 360 trials. We then reproduced the same statistical analysis performed in Hyde and Peretz (2004) as follows. For each delay, the proportion of 'yes' responses of the simulated participants was calculated, where for each simulated participant the false alarm rate was defined as the proportion of yes-responses to the 0 delay stimuli, which was subtracted from the hit rate, defined as response rates for each of the nonzero delays (hits minus false alarms). These were then submitted to an ANOVA with within-subject factor time (5 levels; note that this factor was called "temporal distance" in the original study) and between-subject factor group (amusia/control). To determine statistical power, we considered the central question whether this ANOVA yielded a main effect of group, or an interaction between group and time, or both (as any of these cases would indicate a difference between amusics and controls in the time task). To estimate statistical power, the entire procedure was repeated 1000 times, and we investigated in what proportion of these, a group effect (and/or an interaction between group and delay) was found.

Simulating the Hyde and Peretz (2004) paradigm pitch task The pitch task was simulated in a similar way and again following exactly the same procedure as used in Hyde and Peretz (2004)'s study with human participants. The stimuli presented were, as in Hyde and Peretz, 25, 50, 100, 200 and 300 cents, each presented 36 times, with 180 added trials with 0 cents pitch shift. Hit minus false alarm rates were calculated and submitted to an ANOVA with within-participant factor pitch shift (5 levels) and between-participant factor group (2 levels). Again, 1000 experiments were simulated with 10 amusics and 10 controls.

Simulating the MLP paradigm time task We used the exact same stimulus presentation regime for our simulated participants as we used for our human participants. Psychometric curves were estimated on-line trial-by-trial for the simulated participants exactly as they were for human participants and these online estimated psychometric curves were used to determine stimulus placement on the next trial. Simulated participants completed three blocks of 36 trials, each of which yielded a threshold estimate which was averaged per participant and then submitted to a t-test to compare amusics and controls. We simulated 10,000 experiments in this manner and calculated power as the proportion of times the t-test showed a significant difference between amusics and controls.

Simulation results Power in Hyde and Peretz (2004)'s study was estimated to 31.1% (corresponding to a main effect of group, an interaction between group and time, or both) (Fig. 3), meaning it was actually quite plausible (100% − 31.1% = 68.9%) that Hyde and Peretz (2004)'s study would *not* find a group difference even if it existed in reality. The estimated power for the four combinations of study paradigms and sample sizes are reported Table 3. Across the two paradigms, doubling the sample size from 20 to 40 participants yielded an approximate 20% increase in statistical power (from 30 to 50%), whereas changing the experimental paradigm yielded only a few percentage points difference in power. This result pattern suggests that of the two major differences between our study and Hyde and Peretz (2004), namely sample size and constant/adaptive stimuli, it appears to be sample size that is the most important contributor to the difference in statistical power. Note, however, that the MLP paradigm used here only requires 30% of the number of trials of the constant stimuli design of Hyde and Peretz (2004) (108 trials vs. 360 trials).

These simulations allow us to reconcile our results with those of Hyde and Peretz (2004): the fact that the prior study did not find the time effect can be explained by its lower power, while the present study did find this effect, mostly thanks to increased sample sizes.

Discussion

The present study tested amusic and control participants' pitch and time processing with a change detection task (Hyde & Peretz, 2004) using an adaptive threshold procedure. The pitch and time changes were embedded in isochronous sequences with tones of constant pitch. For both the time and the pitch task, this allows having an implementation that minimized the potentially interfering effects of the other dimension, which was not the focus of the study. The findings confirmed previous data showing that, relative to control participants, amusic individuals were less sensitive in detecting pitch changes. The key finding is that amusic individuals were also less sensitive in detecting time changes, even though these time deficits were less pronounced than their pitch deficits.

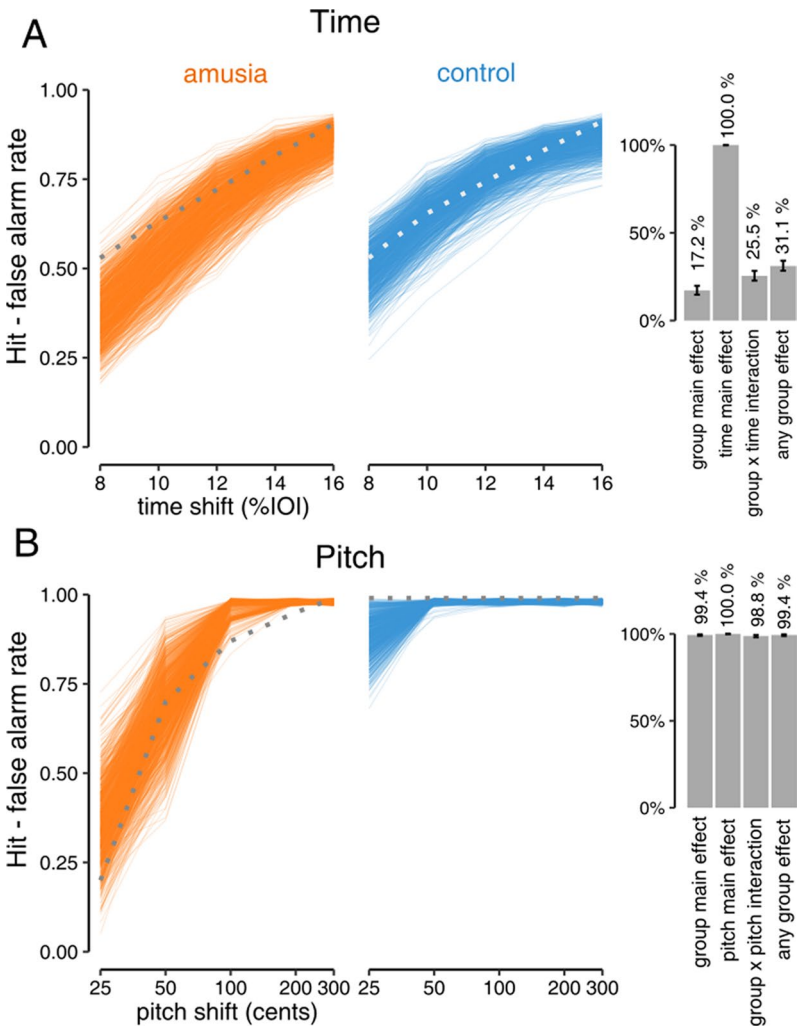

Fig. 3 Results of numerical simulations for (**A**) time task and (**B**) pitch tasks for Hyde and Peretz (2004)'s design (constant stimuli, 360 trials), with n = 20 participants (10 per group). Note Results indicate that it is plausible that prior work had not found a difference in time tasks between amusia and control individuals. We simulated possible observations for the Hyde and Peretz (2004) study using parameters in accordance with the results of the present study. The results (thin lines) align reasonably well with the Hyde and Peretz (2004) findings (their Fig. 1), whose averages are here indicated by the dashed line, both for the time (**A**) and pitch (**B**) tasks. For clarity, pitch shift stimuli x-axis are displayed on a logarithmic scale. The inlay bar plots indicate the proportion of simulations in which the group, time/pitch, or interaction effects were significant

Table 3 Summary of estimated statistical power based on simulations

Number of simulated participants (# amusics/# controls)	Power with Hyde and Peretz (2004) paradigm (constant stimuli, 360 trials) (%)	Power with MLP paradigm (adaptive, 108 trials) (%)
10/10	31.1 (30.2–36.0)	28.8 (27.9–29.7)
19/21	52.0 (48.9–55.1)	55.2 (54.2–56.1)

Estimated power in the Hyde and Peretz (2004) paradigm and MLP design for $n = 20$ or 40 participants. Note that for the Hyde and Peretz (2004) paradigm, this power aggregates findings of either a group main effect or a group x time interaction. Binomial 95% CI are indicated in parentheses

Pitch processing deficits in congenital amusia

The pitch deficits observed here with the adaptive MLP (Maximum Likelihood Procedure) pitch task were in agreement with previously reported data, including the pitch-discrimination-threshold (PDT) task data here as well as the data by Hyde and Peretz (2004) based on the constant-stimuli implementation of the change detection task. While there is overlap in performance between amusic and control groups, amusic participants needed larger pitch differences to detect a change in the sequences than did control participants. As expected based on the assumption that the thresholds derived from the pitch MLP task and the PDT (Tillmann et al., 2009) measure the same pitch discrimination ability, we observed a strong correlation between the two measures (i.e., a large effect size), and the effect sizes for the difference between amusic and control groups were similar for pitch thresholds estimated using MLP (d = 1.24) or PDT (d = 1.26). The current findings are thus in keeping with numerous studies showing impaired pitch discrimination in congenital amusia, yet with about half of the amusic participants exhibiting thresholds in the range of that of controls (see Tillmann et al., 2015, 2016a, for reviews). Note that in comparison to pitch discrimination tasks, impairments in pitch sequence (melody) memory tasks can be larger in congenital amusia, especially in challenging tasks with long sequences and/or fast stimulation rates (see Albouy et al., 2016, and Hoarau et al., 2024, for a discussion of this point).

Time processing deficits in congenital amusia

In comparison to the seminal study by Hyde and Peretz (2004), the novelty of our findings is to reveal that, relative to control participants, amusic individuals have a deficit in time perception, even for pitch-constant stimuli. This finding thus contrasts with these earlier data of Hyde and Peretz (2004) who did not find such a deficit, concluding that the amusic brain is out of tune but in-time. We here used a similar change detection task with five-tone sequences as initially used by Hyde and Peretz (2004), but now implemented with an adaptive threshold procedure (instead of constant stimuli) and pure tones (instead of piano tones). The simulations performed here revealed that the difference between these two data sets could be explained by increased statistical power in our study, which was obtained by testing a larger sample size ($N=40$ instead of $N=20$), with a limited contribution (if any) of the experimental method (adaptive vs. constant stimuli). The current findings with MLP are obtained with only 108 trials per participant vs. 360 with the constant stimuli design, thus allowing for a more efficient testing duration to reach similar power. In Hyde and Peretz (2004), data of the constant stimulus paradigm were analyzed in terms of hit minus false alarm rates and the group factor (amusia vs. control) did not reach significance for the time task. There is now increased awareness that non-significant results do not imply evidence for the absence of effects (Makin & Orban de Xivry, 2019). Our simulations suggest that this seminal study might indeed have had suboptimal statistical power to detect more subtle group differences in the time task. They further suggest that other previous studies might have also been underpowered to reveal time deficits in congenital amusia. For example, Foxton et al. (2006) tested nine amusic participants and nine control participants, allowing them to reveal impaired time processing in tone sequences that varied in pitch, but observed only a statistical trend for time processing impairment in constant-pitch sequences. While the statistical power of approximately 50–55% achieved in the present study is limited, our results nonetheless highlight the critical role of sample size in psychophysical research, and was sufficient to reveal a deficit for amusics also on the time dimension.

Amusics' deficit in detecting time irregularities in constant-pitch stimuli challenges the hypotheses that amusics' time perception is impaired only for pitch-varying stimuli and that amusia involves exclusively a pitch processing deficit (as proposed by Hyde & Peretz, 2004). The results suggest that amusics' deficit might possibly be more general, including not just pitch but also time perception deficits at least in some participants, even though more strongly expressed by deficits in the pitch dimension. Even for the pitch dimension, the observed deficit is larger when the task involves a stronger memory component. Amusic and control groups might overlap in performance for pitch perception tasks, but their performance does not overlap for pitch memory tasks (e.g., Tillmann et al., 2016b). While the present change detection task does not involve such a strong memory component as do the classically used same-different paradigms (including the MBEA), its implementation with five-tone sequences also involves a memory component. Indeed, the first three standard tones serve to establish a representation in memory of the base pitch height (for the pitch task) or of the base time interval (for the time task) relative to which the following events are evaluated. Because of their impaired pitch memory, the amusic participants might benefit less from the contextual information than do controls, rendering the change detection more difficult, whether on pitch or time dimensions. This difficulty might be further enhanced by the relatively fast pace of presentation with short tone durations. Previous research investigating pitch memory in congenital amusia have revealed that amusics suffer from abnormalities in the early stages of auditory processing, starting from the encoding stage (Albouy et al., 2016). Short tone durations (e.g., 100ms like in the present paradigm) or short inter-onset-intervals between tones are more difficult to process for amusics than for controls. For example, in a same-different short-term memory paradigm involving 4-tone sequences, controls' performance was the same for sequences implemented with short tones (100ms) followed by a 250ms silent delay or with longer tones (350ms) followed by a 350 ms delay. However, amusic participants benefited from the longer tones and the longer delay, reaching control performance, while being significantly impaired for the short tones with the shorter delay. In that earlier study, the "different" trials required a detection on the pitch dimension (a changed tone) and sequences were played isochronously. However, the early encoding steps might also affect the contextual information benefit in the present study (with a presentation rate comparable to the faster rate in Albouy et al., 2016). The contextual information trace is required to detect the change, including when this change is on the time dimension. Future studies varying stimulus presentation rates in pitch and time tasks are needed to further test this hypothesis. Furthermore, future studies should also investigate whether (and in how far) the observed time-deficit might be reduced when the pitched tones are replaced by non-pitched sounds. Ideally, time processing abilities should be directly compared in sequences of non-pitched sounds, sequences with constant-pitch sounds and pitch-varying sequences, also in line with previous investigations (e.g., Foxton et al., 2006). Recent work investigating the MMN with pitched (piano) and less-pitched (drum) sounds in congenital amusia (Quiroga-Martinez et al., 2022) suggests that amusics

are sensitive to the difference between pitched and less-pitched sounds, as are controls.

How separate are pitch and time processing?

Pitch and time thresholds were correlated across the set of participants as a whole (i.e., medium effect size): those who showed greater (i.e., worse) pitch thresholds tended to show greater time thresholds as well. This correlation between pitch and time abilities was supported by the regression analysis and by PCA, the latter revealing a common component related to all tasks. The regression analysis revealed that the correlation between pitch and time thresholds could not be explained solely by amusic-vs-control group differences, and showed that time task thresholds were a significant predictor for participants' pitch thresholds in addition to group. The PCA revealed that a large portion of the variance was captured by a component that combined pitch and time processing across different tasks (MLP pitch and time thresholds, PDT, all six MBEA subtests), and this factor allowed distinguishing amusic participants from control participants. This finding suggests that the group difference is observed across and beyond the task differences. As discussed above, this might reflect common shared mechanisms for regularity extraction and memory at play in all tasks, and that these mechanisms would be impaired in congenital amusia. The link between the pitch and time deficits in amusia is also captured by the self-report questionnaire data, showing a relationship between self-perceived rhythm difficulties and performance in the MLP pitch task.

Correlations between pitch and time processing abilities have been previously reported. Peretz et al. (2003), performance correlated significantly between the MBEA rhythm subtest and the pitch interval subtest, which can be considered as the most subtle of the three pitch-based subtests. In their study, this correlation did not extend to the meter subtest, which might be linked to its different implementation (i.e., participants judge one excerpt instead of comparing two excerpts presented in a pair like for the other subtests, thus not requiring the same amount of short-term memory). Correlations between pitch and time processing abilities have been also reported across different types of tasks, notably the scale subtest of the MBEA and beat perception and production abilities (Lagrois & Peretz, 2019). In our present study, pitch and time abilities were tested with very similar materials and using the same task instructions, revealing deficits in both tasks, but less strongly on the time dimension. The observed correlation between pitch and time processing might also be linked to findings from interference paradigms, which show that the information of the two dimensions are not processed independently. This has been shown previously for non-amusic participants with artificial material, such as tones of different pitch heights marking temporal intervals (e.g., Pfeuty & Peretz, 2010), as well as with more musical materials, whether melodies or harmonic sequences (e.g., Boltz et al., 1993; Prince et al., 2009; Lebrun-Guillaud & Tillmann, 2007).

Beyond these associations between pitch and time processing abilities, our current findings also confirm that they rely at least in part on distinct mechanisms. First of all, the group difference between amusics and controls is larger for the pitch task than the time task. Accordingly, fewer amusics (16%) exhibit time deficits at the single subject level (as defined by thresholds > 2SD away from the mean) compared to controls than for pitch (47%). Furthermore, the PCA results suggested that the second dimension mostly captured specific variance in the MLP time task. Our findings of a larger group effect in the pitch task than in the time task are in line with the data of Hyde and Peretz (2004), even though they showed a significant group effect only for pitch change detection and not for anisochrony detection (time task). Based on the results of the study by Hyde and Peretz (2004) as well as our study together with the simulations assessing the difference between paradigms and participant numbers of Hyde and Peretz (2004) and the current study, the most plausible conclusion seems to be that (at least some) amusic individuals have timing deficits, although perhaps less frequently so than pitch deficits, and such deficits would thus be less likely to be detected with relatively small sample sizes. Note however that, when time deficits are observed at the individual level in amusic participants, these deficits can be quite large and comparable in magnitude to pitch deficits (with for example here a z-score of 12.9 for the time task in the most impaired participant, which is comparable to the z-scores of 11.8 and 14.1 for the two most impaired amusic participants in the pitch task).

Regarding congenital amusia per se, it remains an intriguing and open question whether congenital amusia encompasses distinct subtypes (see also McDonald & Stewart, 2008; Tillmann et al., 2016b). One possibility would be that there exists pitch-based vs. time-based amusia (corresponding to the labels congenital amusia or tone-deafness vs. beat-based amusia or beat-deafness in the literature), with some amount of comorbidity between these two neurodevelopmental disorders. It is noteworthy that congenital amusia has recently been reported to exhibit a sizable comorbidity with dyslexia (Couvignou et al., 2019; Couvignou & Kolinsky, 2021), which could be associated to a common deficit in serial-order short-term memory affecting music (pitch) and speech (Couvignou et al., 2023), i.e., a deficit in time-based mechanisms. It thus calls for future investigations of the relationships between pitch processing and time processing in music and speech in neurodevelopmental disorders. Future work could investigate whether

amusics who are only impaired in time (according to the MBEA) show more subtle pitch deficits when measured using more sensitive discrimination tasks such as used in the present paper.

Neurophysiological data also shed light on the relationship between pitch and time processing. In congenital amusia, structural and functional anomalies have been reported mainly in right superior temporal gyrus and right inferior frontal gyrus (e.g., Hyde et al., 2011; Albouy et al., 2013, 2019). Some studies on congenital amusia have also correlated the neural correlates with participants' performance in pitch-related tasks (e.g., Hyde et al., 2006, 2007; Albouy et al., 2019), but to the best of our knowledge, this has not yet been investigated with time-related task performance (see Tillmann et al., 2023, for a recent review on neuroimaging findings in congenital amusia). However, for acquired amusia, Sihvonen et al. (2016) related lesion patterns to participants' performance in both pitch and rhythm subtests of the MBEA. Their analyses revealed links between performance in both tasks and lesion patterns in right superior temporal gyrus (including Heschl's gyrus) and the right striatum. The more general role of the basal ganglia in perceptual and cognitive sequencing as well as predictive processing has been described for rhythm and beat processing (e.g., Grahn & McAuley, 2009) and auditory language processing (e.g., Kotz et al., 2009). Together with the present findings on both pitch- and time-processing deficits, this further suggests the hypothesis that congenital amusia might be related to a more general perceptual and cognitive sequencing deficit, which expresses in particular for material that does not allow for explicit verbalization strategies (see also Tillmann et al., 2023, for further discussion). This hypothesis seems to be further supported by recent findings showing altered oscillatory dynamics in congenital amusia, bringing connections to predictive processing and active sensing (Samiee et al., 2022) as well as by observed comorbidity between dyslexia and amusia related to serial-order memory (Couvignou et al., 2023).

Acknowledgements We would like to honor the memory of the late Dr. Krista Hyde whose work laid the foundation for the current study and many key studies in this field. We thank Dr. Isabelle Peretz and Dr. Lauren Stewart for sharing their questionnaires with us.

Author contributions Conceptualization AC, BT, PA, FV; Data curation LF, AP; Investigation LF, AP; Formal Analysis FV; Methodology FV, AC, BT; Software FV; Writing - original draft FV, AC, BT; Writing - review & editing FV, AC, BT, LF, AP, PA.

Funding BT was supported by an ANER grant ("MusiC" of the Region Bourgogne–Franche Comté). 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).

Data availability The data collected as part of this study are available via the following link: https://osf.io/xsgp7/.

Declarations

Ethics approval The study procedures were approved by the appropriate national ethics committee. This study was performed in line with the principles of the Declaration of Helsinki.

Consent to participate All participants provided written informed consent prior to participating.

Consent for publication All participants signed informed consent regarding publishing their data.

Competing interests The authors declare no competing interests.

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain: A Journal of Neurology*, *136*(Pt 5), 1639–1661. https://doi.org/10.1093/brain/awt082

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Scientific Reports, 6*(1), 18861. https://doi.org/10.1038/srep18861

Albouy, P., Peretz, I., Bermudez, P., Zatorre, R. J., Tillmann, B., & Caclin, A. (2019). Specialized neural dynamics for verbal and tonal memory: FMRI evidence in congenital amusia. *Human Brain Mapping*, *40*(3), 855–867. https://doi.org/10.1002/hbm.24416

Allen, G. (1878). Note-deafness. *Mind*, *3*(10), 157–167.

Amitay, S., Irwin, A., Hawkey, D. J., Cowan, J. A., & Moore, D. R. (2006). A comparison of adaptive procedures for rapid and reliable threshold assessment and training in Naive listeners. *Journal of the Acoustical Society of America*, *119*, 1616–1625.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain: A Journal of Neurology*, *125*(Pt 2), 238–251. https://doi.org/10.1093/brain/awf028

Bégel, V., Benoit, C. E., Correa, A., Cutanda, D., Kotz, S. A., & Bella, D., S (2017). Lost in time but still moving to the beat. *Neuropsychologia*, *94*, 129–138.

Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics, 53*(6), 585–600.

Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia*, *155*, 107811.

Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology*, *36*(1–2), 1–17.

Couvignou, M., Tillmann, B., Caclin, A., & Kolinsky, R. (2023). Do developmental dyslexia and congenital amusia share underlying impairments? *Child Neuropsychology*, *29*(8), 1294–1340.

Dai, H. (1995). On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up–down methods. In The Journal of the Acoustical Society of America (Vol. 98, Issue 6, pp.

3135–3139). Acoustical Society of America (ASA). https://doi.org/10.1121/1.413802

Dalla Bella, S., & Sowiński, J. (2015). Uncovering beat deafness: Detecting rhythm disorders with synchronized finger tapping and perceptual timing tasks. *JoVE (Journal of Visualized Experiments)*, *97*, e51761.

Dalla Bella, S., Farrugia, N., Benoit, C. E., Begel, V., Verga, L., Harding, E., & Kotz, S. A. (2017). BAASTA: Battery for the assessment of auditory sensorimotor and timing abilities. *Behavior Research Methods*, *49*, 1128–1145.

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar Language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, *1455*, 75–89. https://doi.org/10.1016/j.brainres.2012.03.034

Ehrlé, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*, *58*(1), 133–147. https://doi.org/10.1016/j.bandc.2004.09.014

Foxton, J. M., Nandy, R. K., & Griffiths, T. D. (2006). Rhythm deficits in 'tone deafness'. *Brain and Cognition*, *62*(1), 24–29.

Grahn, J. A., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *Neuroimage*, *47*(4), 1894–1903. https://doi.org/10.1016/j.neuroimage.2009.04.039

Grassi, M., & Soranzo, A. (2009). MLP: A MATLAB toolbox for rapid and reliable auditory threshold Estimation. *Behavior Research Methods*, *41*, 20–28.

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Tillmann, B., & Caclin, A. (2023). Consonance perception in congenital amusia: Behavioral and brain responses to harmonicity and beating cues. *Journal of Cognitive Neuroscience*, *35*(5), 765–780. https://doi.org/10.1162/jocn_a_01973

Green, D. M. (1993). A maximum-likelihood method for estimating thresholds in a yes–no task. *The Journal of the Acoustical Society of America*, *93*(4), 2096–2105. https://doi.org/10.1121/1.406696

Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *The Journal of the Acoustical Society of America*, *97*(6), 3749–3760.

Gu, X., & Green, D. M. (1994). Further studies of a maximum-likelihood yes–no procedure. *The Journal of the Acoustical Society of America*, *96*(1), 93–101. https://doi.org/10.1121/1.410378

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*. https://doi.org/10.1016/j.neuropsychologia.2024.108960

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*(5), 356–360. https://doi.org/10.1111/j.0956-7976.2004.00683.x.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the Amusic brain: A two-site study. *Brain*, *129*(10), 2562–2570.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience*, *27*(47), 13028–13032.

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex, 21*(2), 292–299.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: A review and proposal for A model of sensory predictability in auditory Language perception. *Cortex; A Journal Devoted To the Study of the Nervous System and Behavior*, *45*(8), 982–990.

Lagrois, M. É., & Peretz, I. (2019). The co-occurrence of pitch and rhythm disorders in congenital amusia. *Cortex; A Journal Devoted To the Study of the Nervous System and Behavior*, *113*, 229–238.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*, 1–18.

Lebrun-Guillaud, G., & Tillmann, B. (2007). Influence of tone's tonal function on temporal change detection. *Perception & Psychophysics, 69*(8), 1450–1459. https://doi.org/10.3758/BF03192959

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279–1292.

Leek, M. R., Dubno, J. R., He, N. J., & Ahlstrom, J. B. (2000). Experience with a yes–no single-interval maximum-likelihood procedure. *Journal of the Acoustical Society of America*, *107*, 2674–2684.

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

Makin, T. R., & de Orban, J. J. (2019). Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *ELife*, *8*, e48175. https://doi.org/10.7554/eLife.48175

McAdams, S. (1989). Contraintes psychologiques sur les dimensions porteuses de forme en musique. In: La musique et les sciences cognitives (Eds S. McAdams et I. Deliege), Pierre Mardaga, Bruxelles, pp. 257–283.

McDonald, C., & Stewart, L. (2008). Uses and functions of music in congenital amusia. *Music Perception*, *25*(4), 345–355. https://doi.org/10.1525/mp.2008.25.4.345

Mehr, S. A. (2024). Principles of music perception. *PsyArXiv*. https://doi.org/10.31234/osf.io/bdpjc

Peretz, I. (2016). Neurobiology of congenital Amusia. *Trends in Cognitive Sciences*, *20*(11), 857–867. https://doi.org/10.1016/j.tics.2016.09.002

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

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal battery of evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75. https://doi.org/10.1196/annals.1284.006

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology*, *58*(3), 478–482.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., & Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception*, *25*(4), 331–343.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The Amusic brain: In tune, out of key, and unaware. *Brain*, *132*(5), 1277–1286.

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch—time interference in congenital amusia: Evidence from an implicit test. *Attention Perception & Psychophysics*, *72*(3), 763–774.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, *49*(5), 961–969.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., & Peretz, I. (2013). Amusic does not mean unmusical: Beat perception and synchronization ability despite pitch deafness. *Cognitive Neuropsychology*, *30*(5), 311–331.

Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception*

and Performance, 35(5), 1598–1617. https://doi.org/10.1037/a0016456

Quiroga-Martinez, D. R., Basiński, K., Nasielski, J., Tillmann, B., Brattico, E., Cholvy, F., Fornoni, L., Vuust, P., & Caclin, A. (2022). Enhanced mismatch negativity in harmonic compared with inharmonic sounds. *European Journal of Neuroscience*, *56*(5), 4583–4599. https://doi.org/10.1111/ejn.15769

Samiee, S., Vuvan, D., Florin, E., Albouy, P., Peretz, I., & Baillet, S. (2022). Cross-frequency brain network dynamics support pitch change detection. *Journal of Neuroscience*, *42*(18), 3823–3835.

Sihvonen, A. J., Ripollés, P., Leo, V., Rodríguez-Fornells, A., Soinila, S., & Särkämö, T. (2016). Neural basis of acquired amusia and its recovery after stroke. *Journal of Neuroscience*, *36*(34), 8872–8881.

Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences*, *1060*(1), 255–261.

Sowiński, J., & Dalla Bella, S. (2013). Poor synchronization to the beat May result from deficient auditory-motor mapping. *Neuropsychologia*, *51*(10), 1952–1963.

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

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital Amusias. *Handbook of Clinical Neurology*, *129*, 589–605. https://doi.org/10.1016/B978-0-444–62630–1.00033–0.

Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., & Bigand, E. (2016a). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia*, *85*, 10–18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027

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

Tillmann, B., Graves, J. E., Talamini, F., Leveque, Y., Fornoni, L., Hoarau, C., Pralus, A., Ginzburg, J., Albouy, P., & Caclin, A. (2023). Auditory cortex and beyond: Deficits in congenital amusia. *Hearing Research*, *437*, 108855. https://doi.org/10.1016/j.heares.2023.108855

Tranchant, P., Vuvan, D. T., & Peretz, I. (2016). Keeping the beat: A large sample study of bouncing and clapping to music. *PloS one, 11*(7), e0160178.

van der Steen, M. C., van Vugt, F. T., Keller, P. E., & Altenmüller, E. (2014). Basic timing abilities stay intact in patients with musician's dystonia. *PloS One*, *9*(3), e92906. https://doi.org/10.1371/journal.pone.0092906

van Vugt, F. T., & Tillmann, B. (2014). Thresholds of Auditory-Motor coupling measured with a simple task in musicians and Non-Musicians: Was the sound simultaneous to the key press?? *Plos One*, *9*(2), e87176. https://doi.org/10.1371/journal.pone.0087176

van Vugt, F. T., & Tillmann, B. (2015). Auditory feedback in error-based learning of motor regularity. *Brain Research*, *1606*, 54–67. https://doi.org/10.1016/j.brainres.2015.02.026

van Vugt, F. T., Kafczyk, T., Kuhn, W., Rollnik, J. D., Tillmann, B., & Altenmüller, E. (2016). The role of auditory feedback in music-supported stroke rehabilitation: A single-blinded randomised controlled intervention. *Restorative Neurology and Neuroscience*, *34*(2), 297–311. https://doi.org/10.3233/RNN-150588

Whiteford, K. L., & Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia*, *99*, 213–224. https://doi.org/10.1016/j.neuropsychologia.2017.03.018