Can Melodic Contour Help Encoding Luminance Sequences? A Crossmodal Short-Term Memory Study

Francesca Talamini, Julia Vigl, Dominik Wille, Marcel Zentner
University of Innsbruck, Innsbruck, Austria

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

Barbara Tillmann
Université Bourgogne Europe, CNRS, LEAD UMR5022, 21000, Dijon, France

Contour, the "shape" of a melody defined by the direction of pitch changes, is a key feature for melody perception and memory. Beyond pitch, contour can also be conveyed through other auditory and visual dimensions, such as loudness, spatial position, or luminance. Previous findings have shown that musically trained individuals better memorize contour sequences implemented without pitch variation, by mentally representing them as melodies (Talamini et al., 2022). In this study, we tested nonmusicians ($N = 40$) to investigate whether short-term memory for visual contour sequences could be facilitated when preceded by a contour-matched melody. In each trial, participants compared two luminance sequences (S1, S2) and judged whether they were the same or different. Half of the trials were preceded by a melody matching the contour of S1. Music and pitch perception skills were also assessed. Results revealed that when the luminance sequences were preceded by a melody, performance was higher than without a preceding melody, but only for the most difficult sequences. Moreover, music and pitch perception skills were moderately correlated with performance in the luminance short-term memory task. These findings offer new insights into crossmodal contour perception and memory, opening new possibilities for investigating the underlying cognitive mechanisms of contour perception.

Received: August 26, 2024, accepted November 27, 2025.

Key words: contour, crossmodal correspondences, music, short-term memory, priming

In music, contour is defined as the pattern describing the direction of pitch changes in a melody (i.e., the patterns of ups and downs). For example, a short melody with three tones, where the second pitch is higher than the first, and the third is higher than the second, creates an ascending pattern. If the second and third pitches are lower than the first one, then the pattern will be a descending one. A melody usually contains more complex patterns of contour, which can be both ascending and descending, and change as the melody progresses.

Studies have shown that contour is an essential feature that contributes to the encoding and remembering of melodies (e.g., Dowling & Hollombe, 1977; Peretz & Babaï, 1992; Rosner & Meyer, 1986). For example, in short-term memory recognition paradigms, participants can more easily identify a comparison melody as being different from the standard when the two melodies differ in contour than when the contour remains the same (Dyson & Watkins, 1984). In long-term memory paradigms, participants are generally able to identify previously learnt or familiar melodies when the contour is preserved, even when the exact intervals between the tones are distorted relative to the original melody (Dowling & Fujitani, 1971; Idson & Massaro, 1978; Schubert & Stevens, 2006). Pitch contour appears to have a specific neural substrate, with studies showing dissociations between interval and contour perception following partial removal of the cerebral cortex (i.e., cortectomy, Liégeois-Chauvel et al., 1998) or brain lesions (Schuppert et al., 2000). Specific neural sensitivity to pitch contour has also been reported in animal models (e.g., cats, Weinberger & McKenna, 1988).

Although contour is recognized as an essential feature of music perception and memory, the precise mechanisms underlying contour perception remain poorly understood. One key question is whether contour perception is a music-specific or a domain-general mechanism. If it is domain-general, this mechanism would be responsible for encoding sequences of various materials based on the changes between elements along a dimension, regardless of the type of material. For example, a recent study has proposed that one mechanism might be responsible for encoding the relative magnitude

Music Perception, pp. 1–20, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2026.2431582

information in sequences of items, which should not be specific to any modality or stimulus (Bonn & Cantlon, 2017). Previous research has shown that contour can be perceived in sequences of items containing other-than-pitch variations, even for nonmusical or visual stimuli (e.g., Marks, 1987; Prince et al., 2009; Talamini et al., 2022), suggesting that contour might indeed be perceived by simply encoding a series of changes along a given dimension, and that this dimension may be different from pitch.

In tone sequences, contour can also be elicited by varying dimensions such as loudness or brightness of timbre while keeping the pitch constant (i.e., isotonic) (Graves et al., 2019; Marks, 1987; McDermott et al., 2008). Similarly, contour can emerge by varying pitch and loudness also in non-musical sounds, namely in fractional white noises, flicker noises, and brown noises (Schmuckler & Gilden, 1993). Changes in loudness and timbre can also interfere with pitch, and with each other (e.g., a physical change along one dimension can produce a perceptual change in the other dimension) (Allen & Oxenham, 2014; Grau & Kemler Nelson, 1988; Melara & Marks, 1990a, 1990b). Thus, it is possible that auditory contour perception involving non-pitch variations is influenced by the interactions between different auditory dimensions.

For visual material, contour perception can be elicited with line drawings appearing on a screen at different heights (e.g., Prince et al., 2009), and with luminance variations (Aizenman et al., 2018; Hubbard, 1996; Marks, 1989; Marks et al., 2003; Talamini et al., 2022). Creating contour with luminance variations is particularly interesting because it is based on non-musical visual stimuli that are unrelated to spatial positions, which could remind of musical notation. Contour based on luminance thus appears to be quite different from the classical melodic contour based on pitch variations.

Some studies have investigated whether there could be a crossmodal experience of contour, and whether individuals can perceive a correspondence between contour sequences created in different modalities. Crossmodal correspondences refer to the perception of associations between specific stimulus dimensions across different modalities (Spence, 2011). For example, pitch height has been found to be associated with spatial positions (Ben-Artzi & Marks, 1995; Evans & Treisman, 2010), where higher pitches tend to be associated with higher spatial positions, possibly reminiscent of the position of notes in musical notation. The spatial representation of pitch has been labeled as "SMARC effect" (Lidji et al., 2007; Rusconi et al., 2006).

Another crossmodal equivalent of pitch height is color luminance (i.e., the higher the pitch, the lighter the color, the lower the pitch, the darker the color). This association appears to be universal, and does not occur only in synesthetic individuals, who automatically perceive the crossmodal equivalent when exposed to one stimulus (Hubbard, 1996; Marks, 1987, 1989; Martino & Marks, 1999; Melara, 1989). Interestingly, pitch height has also been found to be associated with object size (Evans & Treisman, 2010; Gallace & Spence, 2006), with higher pitches being associated with smaller objects and lower pitches associated with larger objects. Additionally, there also seems to be a correspondence between pitch height and direction of motion (Connell et al., 2013), with higher pitches associated with upward motion and lower pitches associated with downward motion.

Investigating the crossmodal perception of contour, Prince et al. (2009) showed that participants could perceive similarities between pitch contour (i.e., melodies) and line drawings that followed (or not) the same contour of the melody. Participants were able to identify when the contours matched (i.e., an ascending line paired with an ascending pitch sequence and vice versa). Balch and Muscatelli (1986) tested contour perception for both auditory (i.e., pitch) and visual stimuli (i.e., different marker positions on vertical lines), both in unimodal trials (comparing two same-modality contours presented sequentially, with the starting tone or visual position changing between the two contours) and in crossmodal trials (comparing visual and auditory contours presented sequentially). Performance in crossmodal trials was superior to that in auditory-only trials (though not superior to that in visual-only trials), suggesting that we can indeed perceive and recognize contours crossmodally, and use this correspondence to reinforce contour memorization under certain circumstances.

The memory advantage for crossmodal stimuli over unimodal ones when the stimuli of two different modalities are presented simultaneously may be due to dual encoding (multimodal encoding), which has been shown to enhance memorization of stimuli (Heikkilä et al., 2015; Wammes et al., 2019). Multimodal encoding can also be referred to as multisensory integration (or multisensory binding, Green, 2021), which is the process by which the brain combines information from multiple sensory modalities to create a unified and coherent perception of the environment. Music perception can be a multisensory experience (e.g., in live performances), requiring constant integration of information coming from multiple senses, and this integration can fundamentally change the experience of music (Schutz, 2008). Sometimes the integration is so strong that it alters perception (e.g., not detecting asynchronies in audiovisual stimuli, Chuen & Schutz, 2016; hearing

a different sound based on what would be coherent with a simultaneous visual stimulus, Saldaña & Rosenblum, 1993). Multimodal encoding seems also to strengthen priming effects. For example, a study by Johanis and Schmuckler (2021) investigated the role of audiovisual information in harmonic priming. They employed a classic semantic priming paradigm where participants were presented with either auditory, visual, or audiovisual prime chords followed by target chords that had to be judged as being in tune or out of tune. Findings revealed that when the prime was presented audiovisually, a significant priming effect was observed (i.e., faster reaction times and lower error rates), whereas the prime effect was not observed with visual primes and auditory primes alone.

Although not directly investigating contour, a crossmodal study by Aizenman et al. (2018) confirms that the simultaneous presentation of melodies (auditorily presented) with matched visual stimuli (in this case, luminance variations) facilitates the encoding of a sequence, regardless of the modality. Specifically, participants were asked to judge whether the last four items of a sequence (audiovisual, auditory-only, and visual-only) repeated the four first items of the same sequence. Performance with audiovisual sequences was superior to unimodal ones, but only when the pitch and luminance stimuli were crossmodally matched. Interestingly, participants with prior music training performed better than individuals without music training with the unimodal sequences (i.e., both with pitch and luminance sequences, respectively), and with audiovisual sequences only when they were crossmodally matched.

Our study also revealed an association between music training and memory for luminance-based contour (Talamini et al., 2022). Musician and nonmusician participants completed a series of short-term memory recognition tasks in two modalities (visual and auditory), for three different types of stimuli, namely (1) verbal, (2) nonverbal with contour (i.e., luminance or loudness variations), and (3) nonverbal without contour. Musicians outperformed nonmusicians in the contour tasks not only with auditory stimuli, but also with visual stimuli. For auditory material, this musicians' advantage extends previous findings on contour perception based on pitch (Wayland et al., 2010). To further investigate the observed musicians' advantage in the non-pitch contour tasks, we explored participants' reported use of strategies during the loudness and luminance tasks (Talamini et al., 2022). The best-performing participants reported imagining (some even mentally "hearing") a melody while looking at the sequences with luminance variations (luminance contour). This association was more frequently reported by musician participants, suggesting that music training might facilitate the use of crossmodal correspondences with pitch to perceive and memorize contour with other-than-pitch stimuli. This observation, based on the participants' strategy reports, suggests that associating a melody with a luminance sequence is a successful memory strategy. This leads to the hypothesis that individuals who do not spontaneously use this strategy might be taught to do so to improve the memorization of non-pitch contour sequences.

In the present study, we investigated whether memory for non-pitch contour sequences—specifically sequences based on luminance variations—could be enhanced by preceding the visual sequence with a contour-matched pitch sequence (i.e., a melody). This investigation was motivated by previous findings indicating that (1) there is a crossmodal correspondence between luminance variations and pitch variations (e.g., Marks, 1987, 1989), and (2) nonmusicians tend to experience difficulties in memorizing luminance sequences, often relying on verbal mnemonic strategies or none at all, whereas pitch-related strategies are associated with superior performance (Talamini et al., 2022). Building upon these studies, we used a delayed matching-to-sample task to test whether perceptual contour correspondences between pitch and luminance could extend to short-term memory and aid memory for non-pitch sequences. We hypothesized that emphasizing the association between luminance sequences and melodies, by presenting preceding contour-matched pitch sequences, would promote the use of melodic mnemonic strategies, thereby facilitating recognition performance.

Moreover, as only few past studies have investigated music training in association with contour processing, it is not clear whether the ability to perceive/recognize no-pitch contours is a product of the music training or whether it can be present in nonmusicians too, we assessed participants' music perception skills, pitch perception, and pitch memory abilities. We expected that individuals with better music perception abilities (especially pitch-related) would perform better in the luminance contour task and could benefit more from the presence of a preceding contour-matched pitch sequence, similar to musicians who performed better in this type of task and reported to use pitch-based strategies even for non-pitch contours (Talamini et al., 2022).

Method

PARTICIPANTS

Forty-two young adults participated in the study. Two participants did not meet the inclusion criteria (i.e., limited

music training, see description below), and they were excluded from further description and analyses. The final dataset included 40 nonmusician participants (22 females). They were all young adults, with an average age of 23.32 years ($SD = 3.70$). All participants were university students recruited via a university mailing list, with a mean of 15.15 years of education ($SD = 2.25$). All of them reported to have normal hearing and normal or corrected-to-normal vision, and no history of neurological disorders. Participants had only minimal formal (extra-school) music training and were not regularly playing/singing at the time of testing. This was assessed during the recruitment phase. Follow-up questionnaires revealed that 36 participants reported to have never received any music training beyond the compulsory music classes at school, one participant reported to have played guitar for two months, one to have played the piano for three years, and two participants reported to sing as a hobby, but had never received any formal training.

MATERIALS

Visual Contour Tasks

Luminance Stimuli. The sequences of luminance variations (of varying length depending on the task, see below) were created from a pool of six squares with different luminance levels, based on white, ranging from 2.14 cd/m^2 to 26.3 cd/m^2, in steps of $\sim$ 5 cd/m^2 (as in Talamini et al., 2022). Luminance was measured with a Chroma Meter CS-100A luminometer. The square size was 490 x 490 pixels, and stimuli were presented at the center of a 23" DELL U2312Hm monitor with a resolution of 1920 x 1080 (approximately 60 cm away from the participants' eyes), on a black background. Squares were presented with a SOA of 600 ms without ISI.

Pitch Stimuli. Pitch sequences (of varying length depending on the task, see below) were created with six tones from the F major scale, namely F3, G3, A3, Bb3, C4, D4 (i.e., F0 ranging from 174.61 Hz to 293.66 Hz). Tones were generated using the Cubase software (as in Albouy et al., 2013; Schulze & Tillmann, 2013) with a piano timbre. The SOA was 600 ms (each tone was 500 ms long, with an ISI of 100 ms).

Crossmodal Contour Familiarization Task. Participants had to listen to a pitch sequence, repeated twice, followed by a luminance variation sequence. The contour of the pitch sequence either matched or mismatched the contour defined by the different luminance levels (see Figure 1). In the present study, the six possible luminance levels were thus paired with the six possible pitches in an ascending order (i.e., the darkest luminance level was paired with the lowest pitch, the lightest luminance level was paired with the highest pitch). The pairing did not account for having equally perceived steps between pitch levels and between luminance levels, but they were chosen to be easily discriminable, as assessed after pilot testing. Participants were also instructed beforehand about the direction of this association. This correspondence was based on previous studies (Hubbard, 1996; Marks et al., 2003), where high-pitched tones were associated with lighter colors, whereas low-pitched tones were associated with darker colors. In a mismatch trial, the luminance sequence differed in contour from the pitch sequence (e.g., the pitch was ascending and the luminance was descending). The mismatch was created by replacing one item of the luminance sequence with a new one, and this replacement always changed the contour of the sequence. The block started with sequences of two items (both for pitch and luminance), and the sequence length was increased by one after 4 trials, up to 5 items (i.e., 20 trials in total). Half of the trials contained matching sequences. Between the pitch and luminance sequences, there was a silent delay of 600 ms (the same as between the two presentations of the pitch sequence), during which a fixation cross appeared in the center of the screen to warn participants of the upcoming visual sequence. A fixation cross also appeared before the first pitch sequence (i.e., between trials) and lasted for 1,000 ms. Participants had to press one of two keys on the keyboard to indicate whether the luminance sequence matched the pitch sequence or not (see Figure 1). Participants could take as much time as they wanted to respond, and after pressing a key the next sequence was automatically presented but always preceded by the fixation cross (1,000 ms). Participants received written feedback indicating whether the answer was correct or incorrect. The familiarization task consisted of one block and lasted approximately 5 minutes.

Contour Recognition Task. The main task consisted of the presentation of a luminance variation sequence (S1), followed by a second luminance variation sequence (S2), which could be either the same or different. All sequences contained five luminance levels each (i.e., items). S1 contained at least one contour reversal (e.g., a rising change followed by a falling change), and S1 and S2 did not include repeated items (i.e., a unique item was present only once within each sequence). When S2 differed from S1, the difference was created by replacing one of the items with a new item, and this replacement always changed the contour. There was an equal number of changes for each possible item position (i.e., items 1, 2, 3, 4, 5 were replaced equally often across different trials). To generate S1 and S2, a pool of 120 unique numerical sequences was generated using the Matlab

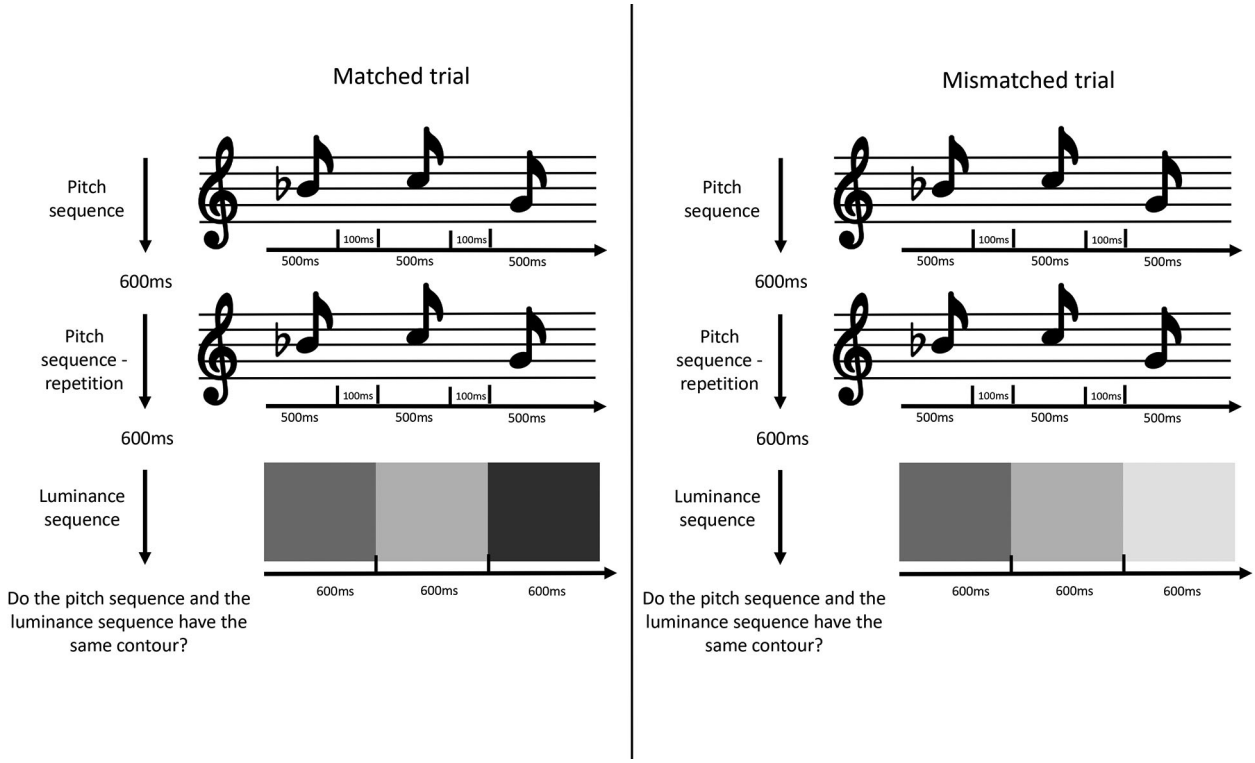

FIGURE 1. Example of a matched and mismatched trial for three-item sequences (sequences could be 2, 3, 4, or 5 items long) in the familiarization task. The levels of gray of the luminance sequence here depicted were chosen to maximize illustration clarity and do not have the same cd/m^2 as in the original task, although they appear similar.

software according to the above-listed constraints. Sequences were distributed into six lists, which were assigned to the six blocks in a counterbalanced order across participants (following a Latin square), so that all participants were presented with the same sequences, but in different blocks. In three of the six blocks, the to-be-compared luminance sequences were preceded by a pitch sequence (referred to as *melody* condition, see Figure 2), and in the other three blocks, the two luminance sequences were presented directly (referred to as *no-melody* condition). In blocks of the *melody* condition, the pitch sequence was presented twice before S1 (with a silent delay of 600 ms) and always had the same contour as S1. *Melody* and *no-melody* blocks were always alternated; moreover, half of the participants started with a *melody* block and the other half started with a *no-melody* block. In each block, half of the trials were "same" trials (S1 = S2) and half were "different" trials (S1 ≠ S2). Participants were asked to compare S1 and S2 and decide whether the two luminance sequences were the same or different. The answer was provided by pressing one of two keys on the computer keyboard. Participants could take as much time as they wanted to respond, and after pressing a key, the next trial was automatically presented, always preceded by a fixation cross (1,000 ms). Participants were instructed to pay attention to the melody when it was present, as it could help them to memorize the following luminance sequence. They were informed that the contour of the melody always matched the contour of S1, and before each block they received written information on the screen as to whether the block contained melodies or not. The first pitch sequence and its repetition were presented with a simultaneous written text on the screen (for 2,900 ms), "melody" and "melody repetition," respectively, to guide participants through the trials. The delay between the two repetitions of the melody and between the melody and S1 was 600 ms (silent delay with a black screen). In the *no-melody* condition, each luminance sequence was preceded by a fixation cross that was presented at the center of the screen for 1,000 ms. In the *melody* condition, the fixation cross appeared before the first pitch sequence (1,000 ms) and between the repeated pitch sequence and the luminance sequence for 600 ms. In each condition, between S1 and S2 sequences, there was a silent delay of 120 ms with

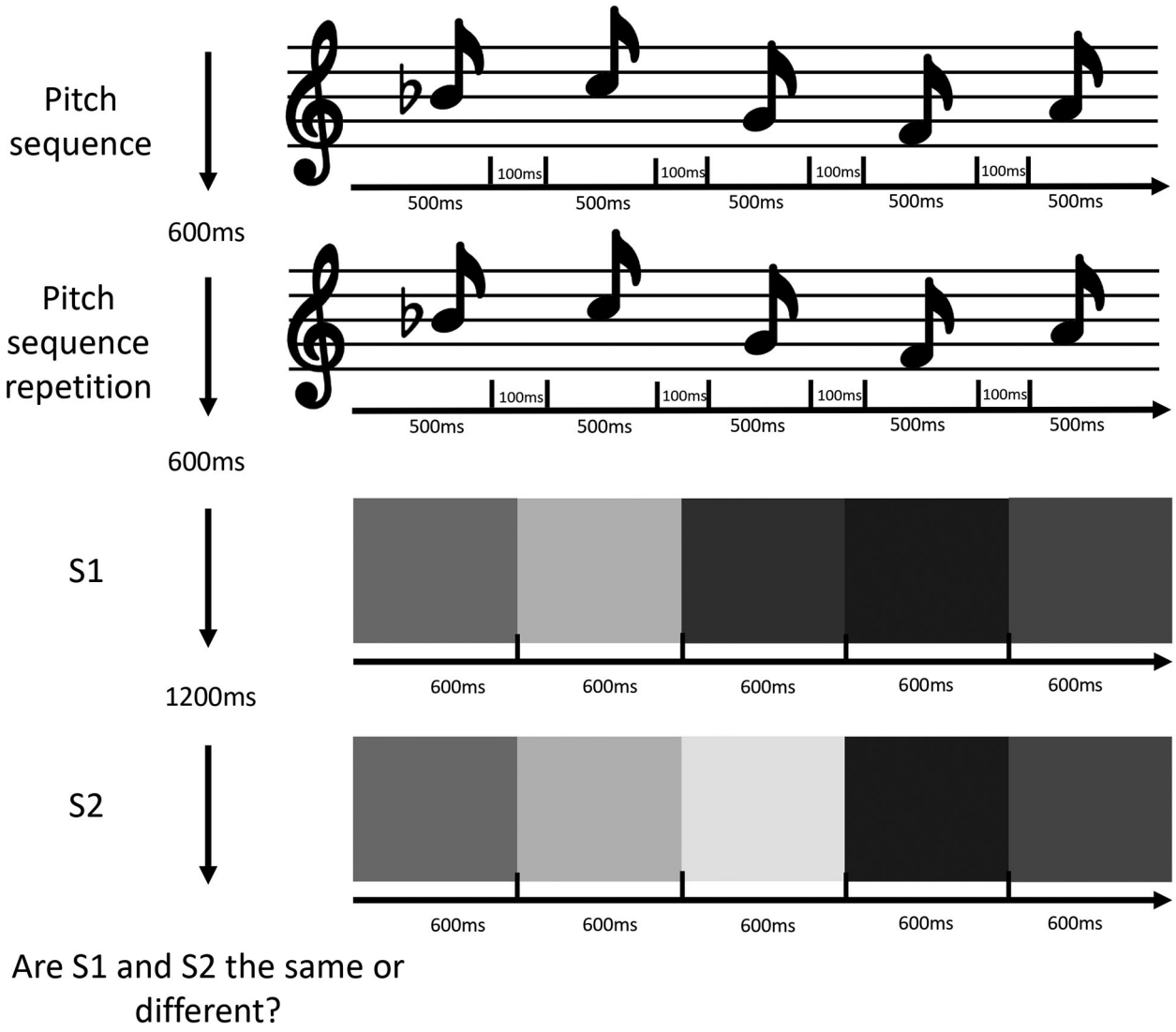

FIGURE 2. Example of a trial in the *melody* condition with a change between S1 and S2 (luminance sequences to be compared, expected answer: different). In the *no-melody* condition, only S1 and S2 were presented during the trials. The levels of gray of the luminance sequence here depicted were chosen to maximize illustration clarity and do not have the same cd/m^2 as in the original task, although they appear similar.

a black screen. Participants had an example block before the experimental task, with 4 *melody* trials and 4 *no-melody* trials, always in this order. During the example block, written feedback was displayed on the screen indicating whether the answer was correct or not. This task took approximately 35 minutes to complete.

Pitch Perception and Memory Skills (Pitch Battery)

To test pitch perception and pitch memory skills, three subtests were selected from a battery assessing non-verbal listening skills (Pralus et al., 2021). The subtests were Pitch Change Detection (PCD), Direction of Change Identification (DCI), and pitch-sequence Short-Term Memory (STM), administered in this order (see details of each task below). The tone stimuli used in these three tasks were the same, and those used in the PCD and DCI tasks had a roving of frequency across trials. All stimuli were synthetic harmonic tones consisting of twelve harmonics. Each tone lasted 500 ms, and the inter-tone interval was 100 ms. All subtests were preceded by a short example to familiarize participants with the task. The tasks were delivered on an iPad touch tablet, and answers were recorded by touching the screen. Performance was assessed by computing the percentage of correct answers for each of the subtests, and the scores for the three subtests were also averaged

into a composite score of pitch perception and memory skills. The three subtests took approximately 15 minutes to complete.

Pitch Change Detection (PCD, Adapted From Hyde & Peretz, 2004). Participants listened to five tones, four of which were always identical (i.e., standards). Only the fourth tone could differ in frequency from the others. There could be four different frequencies for the standards (i.e., 165, 196, 262, 392 Hz). The fourth tone, when different, could be either higher or lower than the standard, and could differ from the standard by different degrees (i.e., from 1/16 to 2 tones). Participants had to judge whether the fourth tone was the same as or different from the others. There were 64 trials in total, out of which 16 were identical trials and all the others were different, evenly distributed across the different standard tones and size/direction of change.

Pitch Direction of Change Identification (DCI). Two tones of different frequency were presented to participants. The frequency of the tones ranged from 123 to 523 Hz, and the difference between two tones ranged from 0.5 to 3.5 tones (7 different sizes of change). The task required judging whether the second tone was higher than the first (ascending pattern) or whether it was lower (descending pattern). There were 56 tone pairs in total, half of which had an ascending pattern, and half of which had a descending pattern, evenly distributed across the 7 different sizes of pitch change.

Pitch Short-Term Memory (STM). Participants listened to two four-tone sequences (i.e., melodies). Tones' F0 ranged between 262 and 440 Hz (corresponding to the six notes between C4 and A4 of the C major scale); each tone was presented for 500 ms, with an ISI of 100 ms ISI, and the two sequences were separated by a delay of 1,000 ms. The second sequence (S2) was either the same as or different from the first sequence (S1). Participants had to judge whether S2 was the same or different from S1. The test had a total of 32 trials, half of which were same trials (S1 = S2) and half of which were different trials (S1 $\neq$ S2). When S2 was different, the difference was created by changing the second or third tone (by shifting the frequency by 1.5, 2, 2.5, 3.5, or 4.5 tones): this always changed the contour of S2 relative to S1.

Music Aptitude

Mini-PROMS (Zentner & Strauss, 2017). The Mini-PROMS is a shorter version of the Profile of Music Perception Skills (PROMS, Law & Zentner, 2012), an online test that assesses music perception abilities. In nonmusicians, it has been used as an index of music aptitude. The Mini-PROMS includes four different subtests that assess discrimination skills in the areas of melody, tuning, accent (also called "beat"), and speed (i.e., tempo). Before the perceptual test started, participants answered questions about their music-related experiences, such as having received formal and/or theoretical music training, years of experience, instruments played and practice hours, and listening habits. After this first questionnaire part, the perceptual test began. The structure of each subtest was the same: a standard stimulus is presented twice, followed by a comparison stimulus. Participants had to judge whether the standard and the comparison were the same or not, using a 5-point rating scale, ranging from "definitely different" to "definitely the same," with different levels of confidence (i.e., including the options "probably the same/different" and "I don't know"). The Melody subtest (10 trials) requires participants to decide whether a standard melody and a comparison melody are the same or not. The Tuning subtest (8 trials) consists of comparing chords made up of 4 different tones. In the case of different trials, one of the tones composing the chord is shifted (from 10 to 50 cents) so that it sounds more or less mistuned. The Accent subtest (8 trials) consists of comparing rhythmic patterns, created by changing the accent (by manipulating the loudness) of different "clicks." The Speed subtest (10 trials) consists of judging whether or not a standard and a comparison rhythmic pattern have the same tempo (i.e., beats per minute, BPM). The Mini-PROMS takes approximately 20 minutes to complete. Correct answers with maximum confidence (i.e., "definitively the same," "definitely different") score 2 points, correct answers with partial confidence (i.e., "probably the same," "probably different") score 1 point, all other options (i.e., "I don't know," and incorrect answers with partial or maximum confidence) score 0 points. The score for each single subtest corresponds to the sum of the raw scores for each trial divided by 2 (so the maximum score per test is the number of trials). The total Mini-PROMS score corresponds to the sum of the four subtest scores.

PROCEDURE

The experiment took place in a quiet laboratory room. Participants first signed the informed consent and completed a paper questionnaire about their demographic information, namely gender, age, years of education, and subject studied at university. The contour tasks (*crossmodal contour familiarization task* and *contour recognition task*) were then administered on a computer using the Psychopy software. Following the contour tasks, participants completed the three pitch perception and memory tasks (Pralus et al., 2021), delivered on an

iPad touch tablet. Finally, to assess music perception skills and self-reported musical habits and education, participants completed the Mini-PROMS test (Zentner & Strauss, 2017), which was delivered on the same computer as the contour tasks via LimeSurvey. For all tasks, participants wore a pair of AKG K371 headphones and the volume was kept constant at a comfortable level for all participants. At the end of the experiment, participants were asked whether they used any strategy for the contour task, separately for trials with and without the pitch priming sequence. Their free responses were first written down by the experimenter on the paper questionnaire completed at the beginning of the session and then transcribed by the experimenter on a computer. The experimental session lasted approximately 1 hour and 30 minutes. Participants were compensated for their participation with a choice of either 15 euros or 1.5 hours of credit (only for undergraduate psychology students). The study protocol was approved by the ethics committee of the University of Innsbruck, Austria, and was conducted in accordance with the Declaration of Helsinki.

STATISTICAL ANALYSES

Performance in Contour Tasks

Sensitivity, Response Bias, and Chance-Level Performance. We assessed participants' performance in the *crossmodal familiarization task* and in the *contour recognition task* by computing *d'* and *c* (see Signal Detection Theory; Green & Swets, 1974). To compute the *d'*, raw scores were first categorized into "hits" (here the proportion of correctly discriminated different sequences), and "false alarms" (the proportion of *same* sequences incorrectly judged to be different). *d'* values were then computed using the function "dprime" in the R package "psycho" (Makowski, 2018), which automatically applies the Hautus adjustment (Hautus, 1995) for extreme values. Performance (*d'*) in the *crossmodal familiarization task* and in the *contour recognition task* (separately for *melody* and *no-melody* conditions), was tested against chance level by comparing the mean *d'* against 0, via a one-sample, two-tailed *t*-test. The response criterion *c* was computed to assess potential response bias and was tested against 0 via a one-sample, two-tailed *t*-test. In the *contour recognition task* the criterion *c* was computed and tested against 0 separately for *melody* and *no-melody* conditions. Positive *c* values reflect the tendency to answer "same," whereas negative *c* values reflect the tendency to answer "different."

Melody vs No-melody Comparison. To test our hypothesis of whether luminance sequences preceded by a contour-matched melody would be better memorized, performance (*d'*) between *melody* vs *no-melody* conditions was compared with a two-tailed paired *t*-test. Criterion *c* was also analyzed with a two-sided paired *t*-test to examine potential differences in response bias between *melody* and *no-melody* conditions.

Correlations Between Contour Tasks. To explore whether the performance in the *contour recognition task* when the melody was present was related to the ability to understand the crossmodal correspondence between pitch and luminance, Pearson correlations were run between the performance (*d'*) in the *crossmodal contour familiarization task*, and the performance (d') in the *contour recognition task* in the *melody* condition, *no-melody* condition and the *d'* difference between the two conditions (a positive *d'* difference indicated better performance in the *melody* condition, and a negative *d'* difference indicated the opposite pattern). Furthermore, in the Supplementary Materials (accompanying this paper at online.ucpress.edu/mp) we provide an analysis of performance (*d'*) separately for starting order (i.e., those who started with *melody* blocks, and those who started with *no-melody* blocks), see Figure S1.

Additional Exploratory Analyses on Contour Recognition Performance

Sequence Complexity. Sequence complexity (defined here as the number of contour reversals present in S1) was analyzed to explore whether the presence of the melody might have different effects depending on the complexity of the sequence to be memorized. A repeated-measures ANOVA was performed on *d'* with condition (*melody* vs *no-melody*) and complexity (1, 2, or 3 contour reversals) as within-subjects factors.

Strategies. The strategies that participants reported using in the *contour recognition task* were first grouped into different categories, initially done independently by three of the authors (F.T., J.V., D.W.). Where there was disagreement, the authors discussed which category was most appropriate. These categories were then merged into three macro-categories (see Table 5 in the results section), to allow for statistical analysis. The choice of the macro-categories was discussed and agreed on by all authors. The strategies used and their macro-categories, separately for *melody* and *no-melody* conditions, and their description are reported in the results section in Table 5. We then explored whether the accuracy (*d'*) was associated with specific strategies used, separately for *melody* and *no-melody* conditions (because strategies could change in the same subject depending on the condition) via two one-way ANOVAs, with strategies as between-subjects factor. We also assessed whether the

frequency of their use was different in the two conditions, via a marginal homogeneity test (for all strategies), and McNemar's test (separately for each strategy).

Relationship Between Contour Recognition Performance and Pitch/Musical Abilities

To assess whether performance could be related to pitch perception and memory skills and music perception skills, Pearson correlations were run between *d'* (separately for *melody* condition, *no-melody* condition, and the difference between the two) of the *contour recognition task*, the Pitch Battery mean score (i.e., the mean of the three subtests), the three Pitch Battery subtests, the total PROMS score, and the scores of the four PROMS subtests. Furthermore, several linear mixed models were constructed to further understand which variables predicted performance (*d'*) in the *contour recognition task*. Potential predictors included Condition (*melody*, *no-melody*), the Pitch Battery mean score, and PROMS total score in all possible combinations, with and without interactions. Each model included a random intercept for participants' IDs. Models were fitted using restricted maximum likelihood (REML). To select the best-fitting model, models were compared via the "anova" function in R, and the Akaike Information Criterion (AIC) was evaluated. The model with the lowest AIC value was preferred.

In all analyses, *p* values of post hoc tests and correlations were corrected for multiple comparisons via the False Discovery Rate (FDR). Effect sizes were reported as *g* of Hedges for *t*-tests, and as eta-squared for ANOVAs. Analyses were performed with the software R (R Core Team, 2018) and JASP (JASP Team, 2024).

Statistical Power. An a priori power analysis for a two-tailed paired *t*-test, for a medium effect size and with a power of 1-$\beta = .80$, indicated that the minimum number of participants to be achieved was 34.

DATA AVAILABILITY

The dataset is available in aggregated form (mean score per condition per participant) on OSF (https://osf.io/cm486).

Results

PERFORMANCE IN CONTOUR TASKS

Response Bias, and Chance-Level Performance. Performance (*d'*) in the *crossmodal familiarization task* was above chance level, $t(39) = 6.94$, $p < .001$, $g = 1.08$ ($M = 1.46$, $SD = 1.33$). Criterion *c* analysis revealed the presence of some response bias, $t(39) = 2.57$, $p = .014$, $g = 0.40$, namely, the positive mean c criterion ($M = 0.16$, $SD = 0.40$) indicated that participants tended to answer "same" more often than "different". Performance (*d'*) in the *contour recognition task* (see Table 1 and Figure 3), was above chance level for both *melody* and *no-melody* conditions, respectively: *melody*, $t(39) = 11.20$, $p < .001$, $g = 1.74$, *no-melody*, $t(39) = 12.57$, $p < .001$, $g = 1.95$. Analysis of response bias (see Table 1) revealed a positive response bias: *melody*, $t(39) = 6.46$, $p < .001$. $g = 1.00$, *no-melody*: $t(39) = 6.41$, $p < .001$, $g = 0.99$. This indicates that participants tended to answer "same" more often than "different," and thus failed to identify different sequences to a greater extent (see Table 1), which is often observed in difficult tasks. The difference in response bias between *melody* and *no-melody* conditions was not significant, $t(39) = 1.48$, $p = .147$, $g = 0.23$.

TABLE 1. *d' and Criterion c per Condition in the Contour Recognition Task*

	Melody	No-Melody
d'	1.97 (1.11)	1.73 (0.87)
c	0.35 (0.34)	0.28 (0.27)

Note. *M* (*SD*)

Melody vs. No-melody Comparison. Comparison of the *melody* condition with the *no-melody* condition revealed a marginally significant difference in performance, $t(39) = 1.93$, $p = .060$, $g = 0.30$, with the sequences preceded by a melody being better recognized ($M = 1.97$, $SD = 1.11$) than the luminance sequences alone ($M = 1.73$, $SD = 0.87$) (see Figure 3).

Correlations Between Contour Tasks. The performance (*d'*) in the *melody* and *no-melody* conditions was correlated with *d'* in the *crossmodal contour familiarization task*, respectively, significantly for the *melody* condition: $r(38) = .40$, $p = .009$, and marginally significant for the *no-melody* condition: $r(38) = .30$, $p = .057$. Note that in the *no-melody* condition the marginal significance disappears after correcting for multiple comparisons. The difference between *melody* and *no-melody* performance was positively but not significantly correlated with the performance in the *crossmodal familiarization task*, $r(38) = .24$, $p = .133$. See Table 2 for the full correlation matrix with corrected *p* values, and Supplementary Materials Figure S2 for scatterplots.

ADDITIONAL EXPLORATORY ANALYSES ON CONTOUR RECOGNITION PERFORMANCE

Sequence Complexity. The additional analysis on the *contour recognition task* integrating sequence complexity revealed no significant effect of condition,

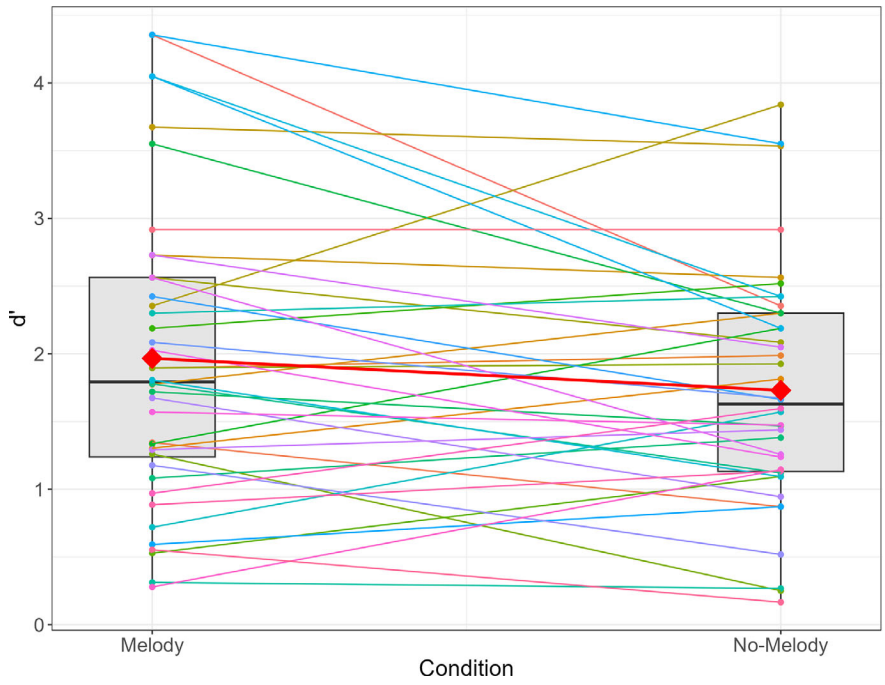

FIGURE 3. Boxplots representing the performance (*d'*) as a function of Condition (i.e., *melody*, *no-melody*). Black horizontal lines within each box represent the median, red diamonds represent the mean. Colored lines connecting the two boxes represent the performance of individual participants and the red line represents the overall group mean ($N = 40$).

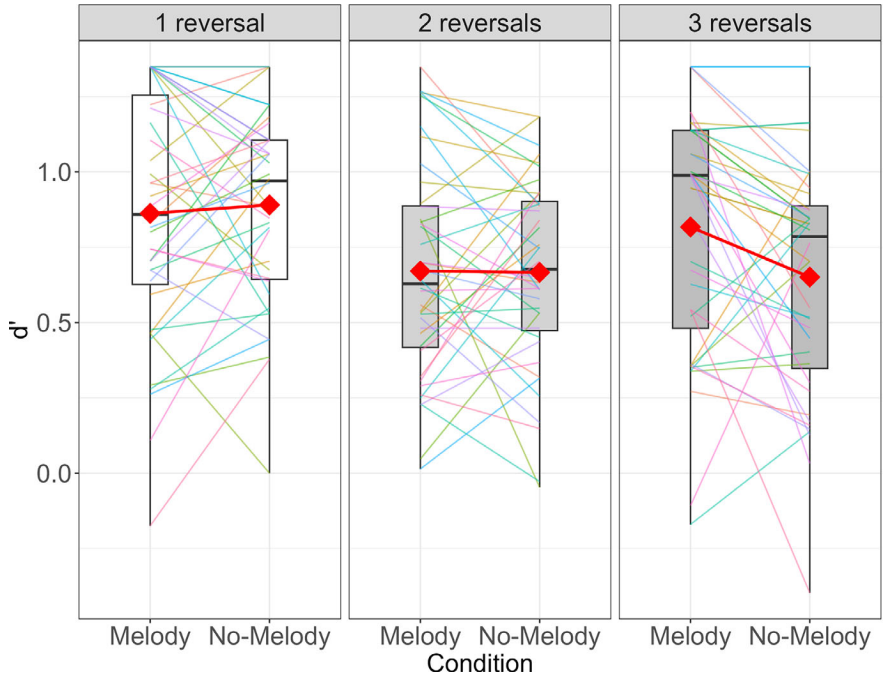

FIGURE 4. Boxplots representing participants' performance (*d'*) as a function of *melody* and *no-melody* conditions and the level of complexity (i.e., 1, 2 or 3 contour reversals) in S1. For each participant, there were 29 trials with one contour reversal, 61 with two contour reversals, and 30 with three contour reversals. Black horizontal lines within each box represent the median, red diamonds represent the mean. Colored lines connecting the two boxes represent the performance of individual participants and the red line represents the overall group mean ($N = 40$).

TABLE 2. *Pearson Correlations Between Performance in the Contour Recognition Task, Familiarization Task, Pitch Battery, and Mini-PROMS Test*

	d' Melody	d' No-Melody	d' Difference	d' Familiarization
d' Melody	–	–	–	–
d' No-Melody	**.72****	–	–	–
d' Difference	**.64****	−.08	–	–
d' Familiarization	**.41***	.30	.24	–
Pitch Battery PCD	**.33**	.20	.24	.26
Pitch Battery DCI	**.35**	**.33**	.13	.27
Pitch Battery STM	**.37**	**.32**	.18	.28
Pitch Battery mean	**.43***	**.36**	.22	**.34**
PROMS Melody	**.43***	**.41***	.16	.23
PROMS Tuning	**.32**	**.32**	.10	.30
PROMS Accent	**.36**	**.35**	.13	**.31**
PROMS Speed	−.03	−.04	.01	.12
PROMS total	**.40***	**.38**	.15	**.34**

Note. **$p \leq .01$, *$p \leq .05$, $N = 40$. Asterisks denote correlations that were significant after correction for multiple comparisons (FDR). Bold-font correlations without asterisks were significant before correction for multiple comparisons. Accuracy (d') is analyzed separately for *melody* condition, *no-melody* condition, and the difference in accuracy between *melody* and *no-melody* conditions. PCD: Pitch Change Detection, DCI: Direction of Change Identification, STM: pitch Short-Term Memory, Pitch Battery mean: average of the three Pitch Battery subtests.

$F(1, 39) = 1.75$, $p = .193$, $n^2 = .010$, but a significant effect of sequence complexity $F(2, 78) = 16.55$, $p < .001$, $n^2 = .127$, and a significant interaction between condition and complexity, $F(2, 78) = 3.740$, $p = .028$, $n^2 = .031$ (see Figure 4). To further analyze the interaction, specifically whether the difference between *melody* and *no-melody* was influenced by the level of complexity, we performed two-tailed paired *t*-tests between *melody*/*no-melody* for each level (FDR corrected). For the most complex S1 sequences (with 3 contour reversals), performance in the *melody* condition was significantly better than in the *no-melody* condition, $t(39) = 2.61$, $p = .038$, $g = 0.40$. For 1- and 2- contour reversals the difference between *melody* and *no-melody* was not significant, $p = .895$, $g = -0.08$ and $p = .933$, $g = 0.01$ respectively.

Strategies. As the strategies were numerous and some were used by only a few individuals, we created macro-categories to statistically explore whether the performance was influenced by strategy type, and whether their use differed by condition. We thus grouped all strategies into three categories: 1) strategies based solely on contour (regardless of whether they were visual or auditory), referred to as "contour-based strategies"; 2) other strategies that were not based on contour, referred to as "noncontour-based strategies"; 3) strategies that were a mix of contour- and noncontour-based strategies, referred to as "mixed contour-noncontour" (see Table 3).

Two one-way ANOVAs on the *contour recognition task* performance (*d'*) with the between-subjects factor Strategy (contour-based, noncontour-based, and mixed) were run separately for *melody* and *no-melody* conditions. For the *melody* condition, the analysis revealed no statistically significant effect of type of strategies on *d'*, $F(2, 37) = 1.05$, $p = .360$, $n^2 = .054$. In the *no-melody* condition, the effect of strategies on *d'* was also not statistically significant, $F(2, 37) = 0.175$, $p = .840$, $n^2 = .009$. However, the frequency of strategy use in *melody* and *no-melody* conditions (see Table 3) was statistically different, marginal homogeneity test: $p < .001$. Specifically, McNemar tests revealed that the frequency of use of contour-based strategies was significantly different in *melody* and *no-melody* conditions, $p = .008$ (i.e., more frequently used in the *melody* condition, see Table 3), however, the frequency of use of noncontour-based and mixed contour-noncontour strategies did not differ significantly between conditions: $p = .07$ and $p = .221$, respectively.

It is worth mentioning that beyond the use of strategies, 9 participants reported that the presence of the pitch sequence was distracting and was not helpful. For these individuals, however, performance in the *no-melody* condition was on average no higher than in the *melody* condition (melody: $M = 1.40$, $SD = 0.80$; no-melody: $M = 1.17$, SD = 0.77). However, their performance in both melody and no-melody condition appears to be overall worse than that of participants ($N = 31$) who did not report being disturbed by the melody (*melody*: $M = 2.13$, $SD = 1.14$; *no-melody:* $M = 1.89$, $SD = 0.84$).

RELATIONSHIP BETWEEN CONTOUR RECOGNITION PERFORMANCE AND PITCH/MUSICAL ABILITIES

Pearson correlations were run to assess a potential relationship between performance in the contour task,

TABLE 3. *Strategies Used and Accuracy, Separately for the Melody and No-melody Conditions*

		Melody		No-Melody	
	Strategy	***N***	**Mean *d'* (*SD*)**	***N***	**Mean *d'* (*SD*)**
Contour-based	Pitch contour	13	1.86 (1.14)	6	1.77 (1.20)
	Visual contour	5	1.22 (0.66)	4	1.06 (0.60)
	Mixed contour	4	2.36 (1.52)	3	2.10 (0.63)
	Total	*22*	*1.80 (1.14)*	*13*	*1.63 (0.97)*
Noncontour- based	Verbal	9	2.01 (0.97)	12	1.77 (0.84)
	Visual memory	/	/	2	2.06 (0.34)
	Other	1	1.18 (NA)	1	0.52 (NA)
	Total	*10*	*1.93 (0.95)*	*15*	*1.73 (0.83)*
Mixed contour - noncontour		8	2.46 (1.20)	12	1.84 (0.88)

Note. Frequencies (*N*), and accuracy (*d'*) associated with the various strategies reported, separately for *melody* and *no-melody* conditions. The strategies were categorized as follows: (1) contour-based strategies: pitch contour = imagining the melody; visual contour = imagining a line following the contour; mixed contour = use of both pitch and visual contour; (2) noncontour-based strategies: verbal = assigning a verbal label to each level of luminance (e.g., "dark" "light"); visual memory = retaining the luminance levels only visually; other = use of other strategies not based on contour; (3) Mixed contour - noncontour strategies = use of both contour and noncontour strategies.

TABLE 4. *The Three Best Linear Mixed Models Based on AIC*

Model Formula	AIC
d' ~ Condition + Pitch Battery + PROMS + (1|ID)	**196.55**
d' ~ Condition * Pitch Battery + Condition * PROMS + (1|ID)	197.99
d' ~ Condition + Pitch Battery + PROMS + Condition * Pitch battery + Condition * PROMS + (1|ID)	198.39

Notes. The dependent variable is the performance in the *contour recognition task* (*d'*). Pitch Battery = mean of all subtests. PROMS = sum of all subtests. (1|ID) = random intercept by participants' IDs. The best model is indicated by bold characters.

TABLE 5. *Fixed Effects Estimates for the Best Linear Mixed Model*

Predictor	**Estimate**	**Std. Error**	**df**	***t***	***p***
Intercept	−2.56	1.24	37.18	−2.06	0.046*
Condition	0.24	0.12	39.00	1.93	0.060
Pitch Battery mean	0.03	0.01	37.00	2.28	0.029*
PROMS total	0.08	0.04	37.00	2.19	0.035*

Note. *$p < .05$. Best model's formula: *d'* ~ Condition + Pitch Battery mean + PROMS total + (1|ID).

separately for *melody* and *no-melody* conditions, the difference between the two, pitch perception and memory skills, and music aptitude. Results are reported in Table 2. In addition, we report as Supplementary Materials the scatterplots of the correlations between the difference between *melody* and *no-melody* conditions and the performance in the contour recognition task (separately for *melody* and *no-melody*) (Figure S3), as well as the correlation matrix between the mini-PROMS and the Pitch Battery (Table S1).

When investigating the possible predictors of the performance (*d'*), the model with the best fit was the model that included the fixed effects of Condition, Pitch Battery mean, and PROMS total score without any interaction, along with the random intercept for participants' IDs (AIC = 196.55 see Table 4 for the three best models and Table S2 in the Supplementary Materials for all models' fits). The marginal R^2 value, which represents the proportion of variance explained by the fixed effects alone, was $R^2m = 0.239$, indicating that approximately 23.9% of the variance of the performance in the *contour recognition task* (*d'*) can be explained by the predictors included in the model. The conditional R^2 value, which accounts for the variance explained by both the fixed and random effects (i.e., our predictors plus the random intercept for participants' IDs) was $R^2c = 0.712$, suggesting that the model explains approximately 71.2% of the variance in the performance. Random effects analysis revealed a variance of 0.49 ($SD = 0.70$) for the intercept by ID and a residual variance of 0.30 ($SD = 0.55$). Fixed effects results are reported in Table 5.

As the benefit of the melody condition was more evident for the most complex sequences (i.e., three contour reversals), the same linear mixed models were also computed only on the performance for most complex sequences. The best model included as fixed effects two interaction terms: one between Condition and Pitch Battery mean, and one between Condition and PROMS total score, along with the random intercept for participants' IDs (AIC = 64.21), see Table 6 for the three best models. The marginal R^2 value was $R^2m = 0.258$, indicating that approximately 25.8% of the variance of the performance for most complex sequences can be

TABLE 6. *The Three Best Linear Mixed Models for Most Complex Sequences Based on AIC*

Model Formula	AIC
d' ~ Condition * Pitch Battery + Condition * PROMS + (1| ID)	**64.21**
d' ~ Condition + Pitch Battery + PROMS + Condition * Pitch battery + Condition * PROMS + (1|ID)	65.40
d' ~ Condition + Pitch Battery + PROMS + (1|ID)	66.53

Notes. The dependent variable is the performance in the *contour recognition task* (*d'*) calculated only on most complex sequences (i.e., 3 contour reversals, $N = 30$). Pitch Battery = mean of all subtests. PROMS = sum of all subtests. (1|ID) = random intercept by participants' IDs. The best model is indicated by bold characters.

TABLE 7. *Fixed Effects Estimates for the Best Linear Mixed Model for Most Complex Sequences*

Predictor	Estimate	Std. Error	df	*t*	*p*
Intercept	−1.04	0.47	37.00	−2.21	0.033*
Condition *Melody**Pitch Battery	0.02	0.01	48.55	3.48	0.001*
Condition *No-Melody**Pitch Battery	0.01	0.01	48.55	1.63	0.109
Condition *Melody**PROMS	0.01	0.016	60.71	0.36	0.719
Condition *No-Melody**PROMS	0.04	0.02	60.71	2.49	0.015*

Note. *$p < .05$. Best model's formula: *d'* ~ Condition * Pitch Battery + Condition * PROMS + (1| ID).

explained by the fixed effects. The conditional R^2 value was $R^2c = 0.570$, suggesting that the model, including random effects for IDs, explains approximately 57% of the variance in the performance. Random effects analysis revealed a variance of 0.05 ($SD = 0.23$) for the intercept by ID and a residual variance of 0.07 ($SD = 0.27$). Fixed effects results are reported in Table 7.

Discussion

The present study aimed to investigate whether short-term memory for luminance contour could be boosted by the prior presentation of a contour-matched pitch sequence, given the well-established crossmodal correspondences between these two dimensions (e.g., Spence, 2011). Forty nonmusicians completed a short-term memory recognition task in which they had to compare two sequences of luminance variations and to decide whether they were the same or different. Half of the trials were preceded by a melody whose contour matched the contour of the luminance sequence. It was hypothesized that the melodic contour would reinforce the encoding and memorization of the luminance sequence and thus improve performance.

Results revealed that when the luminance sequences were preceded by a melody, average performance tended to be higher than without melody, but this was significant only for the most complex sequences (i.e., with three contour reversals). Moreover, regardless of the accuracy, participants reported to use contour-based strategies more often when the melody was present than when it was not, indicating that having a melody preceding the luminance sequence may prompt the use of contour-based strategies. Finally, participants with higher music and pitch perception abilities performed better in the luminance *contour recognition task*, particularly when the melody was present. These abilities seem to contribute more strongly to the performance in the *contour recognition task* than the presence of the melody alone, as music and pitch perception abilities were stronger predictors in our mixed model. First, we will discuss the melody effect on short-term memory for luminance sequences, in relation to other studies on crossmodal contour perception, then we will discuss the possible influence of inter-individual differences (i.e., music and pitch perception/memory abilities) and of strategy use.

CROSSMODAL EFFECTS IN CONTOUR PERCEPTION AND MEMORY

The luminance short-term memory performance was improved by the presence of a matching pitch contour when the sequences to be remembered were more complex. The inclusion of the melody was hypothesized to help the encoding and recognition of a luminance sequence, by suggesting the use of mental strategies based on contour (e.g., imagining a melody). In our previous study (Talamini et al., 2022) we observed that the most successful performances were those of individuals (particularly musicians), who reported encoding the non-pitch contour sequences (i.e., luminance and loudness variations) as being a melody, thus implicitly using crossmodal correspondences to mentally reinforce the abstraction of contour from these sequences. However, in our present implementation, the use of strategies did not seem to affect performance in the contour recognition task. Strategies were assessed at the end of the experiment, separately for *melody* and *no-melody* conditions. Overall, there was a more frequent use of contour-based strategies when the melody was present than when there was no melody, but performance did not significantly differ between those who used a contour-based strategy to remember the

luminance sequence, and those who used a different strategy (e.g., verbal labels).

When present, the benefit of the melody preceding the to-be-judged S1-S2 pair, could be referred to as a priming effect. Being exposed to a melody that has the same contour of the luminance sequence just before the latter appears, would facilitate its encoding as it would "prepare" the participant to encode the luminance sequence by paying attention to its contour. Although classical priming studies use slightly different paradigms (e.g., using congruent and incongruent primes, or comparing congruent and incongruent targets), they also include the use of paradigms similar to the present study, in which the compared conditions were "prime present" vs. "prime absent" (e.g., Reinitz et al., 1989; Schacter & Buckner, 1998). Schacter and Buckner (1998) describe types of priming tasks, such as the word-stem completion, which show how the information that is presented/learnt before the task can influence the subsequent processing, even in the absence of congruent/incongruent conditions.

Another possible reason that explains why there was an advantage of the *melody* vs. the *no-melody* condition is that the contour percept was reinforced because of multisensory or dual-modality encoding effects, which are known to enhance memorization (Heikkilä et al., 2015; Wammes et al., 2019). For example, Heikkilä and colleagues (2015) assessed whether memorizing information could be facilitated by presenting congruent stimuli in two different modalities. They presented pairs of congruent and incongruent audiovisual stimuli during encoding, which could be either verbal or nonverbal. During a recognition phase, unimodal old and new stimuli were presented. Results suggest that the use of two modalities in a congruent manner (i.e., conveying the same information) reinforces memorization, albeit with slightly different effects depending on the types of pairs (e.g., visual object and sound, written and spoken word, etc.). It is thus possible that participants in our present study generally memorized the sequences preceded by a melody better because they were congruent, crossmodal (audiovisual) stimuli. Even if our sequences were presented sequentially, gathering the information about contour from two different (and congruent) sensory modalities might have reinforced its abstraction and memorization. Similarly, the present results could be also explained in light of the Intersensory Redundancy Hypothesis (IRH), which posits that information presented redundantly across multiple sensory modalities (e.g., sight and sound) facilitates perceptual processing and encoding (e.g., Bahrick & Lickliter, 2000; Bahrick et al., 2015; Johanis & Schmuckler, 2021). This enhanced processing occurs because the overlapping sensory inputs draw attention to the shared features (in our case, the contour), making them more salient compared to when presented in a single modality. It has to be noted, though, that studies on IRH usually present stimuli simultaneously, and not sequentially.

The fact that only the most complex sequences (i.e., sequences with three contour reversals) seemed to have benefitted from the presence of the melody, might be explained by the relative easiness of the task for sequences with only one or two contour reversals, which were possibly not challenging enough to profit from the preceding pitch sequence. This suggests that under higher cognitive load, the presence of a crossmodal cue could help encoding the sequence, whereas under low cognitive load it is possible that the sequence can be processed unimodally and without any additional benefit of using crossmodal information. Similar results, although on a target detection paradigm, were found in the study by Marucci et al., (2021), where participants had better target detection when the target was multisensory (presented simultaneously) than unisensory, but only under a high cognitive load condition.

The principle of inverse effectiveness in multisensory integration could also explain why the effect emerged especially for complex sequences. It posits that the performance benefits resulting from the combination of stimuli across different senses are greatest when the individual sensory stimuli are not easily perceivable (i.e., weaker percept) (Gillmeister & Eimer, 2007; Senkowski et al., 2011). Conversely, when the unimodal stimuli are already strong and easily perceivable, the additional benefit from multisensory integration tends to be smaller. In our case, the more complex sequences might have been processed with more difficulty, thus increasing the benefit of having a second modality (i.e., the pitch sequence). It must be noted, though, that these effects (i.e., facilitation due to multisensory integration/multimodal encoding) usually emerge when there is a simultaneous presentation of multimodal stimuli (e.g., Frassinetti et al., 2002). When having an asynchronous presentation of multimodal stimuli, the facilitation appears to be less consistent (Gillmeister & Eimer, 2007). Whether the observed effects are indeed resulting from a facilitation because of multimodal encoding / multisensory integration / IRH is thus difficult to ascertain, as most studies seem to investigate and/or observe these effects with simultaneous presentation of stimuli from different sensory modalities. In other domains, the manipulation of complexity of the target stimulus in priming paradigms seems to produce mixed results. In a masked semantic priming

experiment, the benefit of the prime increased together with the task difficulty (Hines, 1993). In an affective priming task, priming was found to be more efficient when the target had lower complexity (Sassi et al., 2014). Future research could further investigate the effects of task difficulty in the context of crossmodal contour encoding by systematically manipulating the complexity, and possibly the length of the sequences to have participants memorize increasingly difficult sequences.

Finally, although the present study was not designed to directly test whether participants could match contours crossmodally, the above-chance performance in the crossmodal familiarization task suggests that participants were indeed able to perceive the crossmodal correspondence between luminance and pitch contour, which is consistent with previous studies (Aizenman et al., 2018; Marks, 1989). Moreover, our findings also confirm that contour does not define only melodies and can be perceived with other-than-pitch variations (Balch & Muscatelli, 1986; Graves et al., 2019; Küssner et al., 2014; McDermott et al., 2008; Morrongiello & Roes, 1990; Prince et al., 2009). The comparison of performance between *melody* and *no-melody* conditions suggests that participants were able to perceive, match, and transfer the contour information from pitch to luminance. The ability to transfer and match the contour information from one dimension to another was also observed in McDermott et al. (2008), where participants had to listen to a pitch sequence, followed by another auditory sequence of constant pitch, but with loudness or timbre brightness variation, and had to judge whether the contour of the pitch and no-pitch sequence was the same or not. These previous findings showed that participants were successful in matching the contours emerging from different dimensions, but only with a certain mapping across dimensions (i.e., higher pitch = louder tone; higher pitch = brighter tone). Similar results were found in Prince et al. (2009), where participants were asked to compare the contour of melodies with that of line drawings (i.e., stair plots), presented sequentially in varying order. Results indicated that participants were better at rating contour similarity when the two sequences were crossmodally matched (i.e., higher pitches with higher spatial positions of the line drawings and vice versa) rather than when not matched. In their case, comparing pitch and visual contour was easier when the visual contour preceded the pitch contour, unless participants were musically trained. Balch and Muscatelli (1986) also observed that contour abstraction was easier for visual contour (a bar changing vertically on a horizontal line) than for melodic contour. It is possible that, in particular for nonmusicians, varying spatial positions on a vertical axis (e.g., lines, but also hand movements) makes contour more salient, and/or easier to abstract, than listening to pitch sequences. In the present study, the choice to use pitch sequences to facilitate the encoding of luminance sequences was a direct consequence of our previous study reporting that pitch-based strategies were associated with the best performance in the non-pitch contour tasks (Talamini et al., 2022). Future studies could aim to promote the use of different types of strategies (not only pitch-based) when having to abstract and compare contour of two luminance sequences, to understand whether specific types of materials are more efficient in facilitating contour encoding, separately for musicians and nonmusicians.

INTER-INDIVIDUAL VARIABILITY IN CONTOUR PERCEPTION

When comparing the accuracy between melody and no-melody conditions in the contour recognition task, some inter-individual variability was observed (see Figure 2), which we will discuss in relation to previous findings. In our previous study (Talamini et al., 2022), musically trained participants performed better than nonmusicians in a similar luminance contour task (which did not include any melody) and reported using contour-based strategies, such as imagining a melody, more often. Based on this observation, we explored whether music and pitch perception/memory skills and strategy were associated with the ability to memorize luminance variation sequences and benefit from the presence of the melody. Although we have no direct comparisons with previous studies assessing music and pitch perception skills in relation to contour perception, previous links between music training/experience and contour perception have already been observed (e.g., in cochlear implant users, Galvin et al., 2009; in pitch contour identification, Wayland et al., 2010). The relationship between music and pitch perception/memory skills with luminance contour perception was explored here by running a set of correlations. It should be noted that the correlations that were between $r = .30$ and $r = .39$ (thus considered *small*, but close to *moderate* according to the conventional interpretation, Schober et al., 2018), did not maintain statistical significance after correction for multiple comparisons (see Table 2). Below, we will interpret our findings, albeit cautiously, preferring the magnitude of the correlation to its significance.

Performance in the contour task, particularly in the *melody* condition, was weakly to moderately correlated with the PROMS total score and three of its subtests, and with the Pitch Battery and its subtests. The PROMS

Melody subtest and the short-term memory for pitch subtest both include sequences of tones that create a contour. The correlations with these subtests are thus not surprising, and the PROMS Melody subtest showed the strongest correlation between the two and was also the strongest correlation with the PROMS subtest (i.e., the only statistically significant after correction). This subtest might be more sensitive to melody discrimination abilities than the short-term memory subtest of the Pitch Battery, given the greater variety of melody patterns and types of changes present in the former. The PROMS Accent subtest, which was weakly correlated with the performance in the contour recognition task, could also be seen as having a contour, as the rhythmic patterns are created by changing the loudness of certain "clicks," and loudness variations have already been observed to induce a contour percept (McDermott et al., 2008; Talamini et al., 2022). It should be noted, however, that here too the correlation was no longer significant after correction for multiple comparisons. The most puzzling correlation (although weak and no longer significant after correction) is that between the PROMS Tuning subtest and the *contour recognition task* regardless of the condition. This subtest does not clearly elicit a contour percept as the Melody and Accent subtests, as the stimulus played is a four-note chord. It is possible that the observed correlation is simply due to the short-term memory component of these subtests or may reflect that contour perception is intrinsically linked to pitch perception and memory. The only subtest that did not correlate with the contour task was the PROMS Speed; tempo perception might rely on different cognitive mechanisms than contour perception; for example, distinct neural substrates between tempo and melody perception have been observed, with specific areas for tempo that are usually associated with somatosensory/body representations (Lee et al., 2011; Thaut et al., 2014). Moreover, the lack of a correlation here could be interpreted by assuming that the correlations observed between the PROMS and the *contour recognition task* may be explained by their common use of a recognition paradigm, which involves short-term memory. In fact, musical tempo can be perceived and retained by hearing just a few seconds of a musical excerpt (and then mentally "tapping" on the tempo); this might reduce the load on short-term memory. However, it must be noted that the Tuning subtest (which was correlated with the *contour recognition task*) presents the shortest-duration stimuli among all subtests, thus, short-term memory skills may be less important for success in this task than accurate sensory encoding of the stimulus. Interestingly, in the Pitch Battery, the Pitch-Change Detection subtest was weakly correlated (but no longer statistically significant after correction) with the contour task in the *melody* condition only. This suggests that participants who performed well in the *melody* condition also had higher pitch discrimination abilities, which may have helped them to better encode the pitch sequence and thus reinforce contour perception. The Pitch Battery mean score was also correlated with the performance in the *contour recognition task*, however, only the correlation in the *melody* condition survived correction for multiple comparison.

Overall, our findings suggest that pitch discrimination and memory skills, as well as music perception abilities, appear to correlate to some extent with the ability to perceive contour also in non-musical stimuli, especially when a congruent melody is present. This is also confirmed by the results of the linear mixed model comparison, which revealed that the best model for predicting the variance in the *contour recognition task* performance included both the Pitch Battery mean score and the PROMS total score (note that these two variables were only weakly correlated, so there was no concern for multicollinearity) and the *melody/no-melody* condition as predictors. Moreover, when examining the individual contribution of each single predictor, it appears that pitch and music perception/memory skills contributed more to the performance in the *contour recognition task* than the presence of the melody, although the latter was also a marginally significant predictor. This suggests that the contour perception and memory for non-musical stimuli (at least for luminance) may depend more on the individual's set of music- and pitch-related skills than on the inclusion of a crossmodal pitch cue to facilitate task performance. In any case, it should be noted that the largest part of the variance was explained by the random effects at the subject's level, suggesting that there is some variability between participants that could not be identified by our predictors alone. Interestingly, the best-fitting model for the most complex sequences (i.e., those involving three contour reversals)—where the effect of melody inclusion clearly emerged—did not include any main effects. Instead, it included two interaction terms: one between condition and the Pitch Battery mean score, and another between condition and the PROMS total score. These findings suggest that, even for highly complex sequences, individual differences in pitch perception/memory and general music perceptual abilities seem to influence task performance, however differently depending on the condition. Notably, enhanced pitch perception and memory were associated with better performance especially when a melody was present, whereas higher general music

perceptual abilities were linked to improved performance especially in the absence of a melody (see Table 7). It should be noted, however, that this analysis was based on a relatively limited number of trials ($N = 30$). Future research should therefore aim to include a greater number of complex sequence trials to better elucidate these interaction effects.

APPLICATIONS AND FUTURE STEPS

A question that might emerge is whether contour is an amodal property emerging from the relationship between objects (e.g., higher - lower spatial positions, pitches, loudness levels, etc.), or whether it is specific to certain stimuli, such as music and/or speech prosody. Studying contour crossmodally can help shed light on this matter. For example, comparing discrimination performance between pitch and non-pitch contour might give indications on whether the mechanisms involved are similar or not: in the case of serial position effects and accuracy rates being similar (e.g., primacy/recency/mid-list effects), this might suggest that contour is an amodal property that follows the same rules across different types of stimuli. Comparing participants with and without music training, and/or exploring the correlations between musical abilities and contour perception could also suggest possible underlying mechanisms. Musicians were previously found to outperform nonmusicians in no-pitch contour recognition tasks (Talamini et al., 2022). In the present study, we observed several correlations between the performance in the contour tasks (also without the melody) and music/pitch perception skills. This pattern might be seen as a support for the hypothesis that contour is a pitch-specific process (and pitch perception abilities are needed to encode it) that might extend to other stimuli because of uni-modal and crossmodal correspondences with pitch.

Another hypothesis that might explain our data pattern is that contour processing is not specific to pitch (i.e., it is amodal), but it is intrinsically linked with it, as the brain is likely to encode contour for the first time by being exposed to pitch changes in speech prosody and melodies. In this scenario, contour would not necessarily need pitch changes to be perceived but could be facilitated by pitch because of prior exposure to it.

The crossmodal study of contour could be a starting point to develop rehabilitation techniques that can be adopted in different clinical populations to improve music perception (e.g., in congenital amusia, in cochlear implanted patients). Future studies could also explore whether the present results could be replicated when using different stimuli that can induce a contour, such as loudness variations, spatial positions etc.

CONCLUSIONS

The present study investigated whether the (contour) memory of luminance variation sequences can be enhanced when preceded by a contour-matched pitch sequence. This was tested in a population of nonmusicians, who have previously been found to have difficulty in perceiving contour with other-than-pitch dimensions (Talamini et al., 2022). Our results revealed that the presence of a pitch sequence before the luminance sequence improved performance for sequences that were more complex and thus more difficult to memorize. We also observed variability at the individual level that could be partially explained by pitch and music perception/memory skills, which were associated with overall better performance in the *contour recognition task*. The strategies used by the participants, on the contrary, did not seem to influence their performance in the task. Indeed, the strategies remained quite heterogeneous, although when the melody was present there was a larger use of contour-based strategies, confirming that the inclusion of a contour-matched melody could indeed suggest the use of such strategies. The present study contributes to the understanding of crossmodal contour perception and offers perspectives for new possibilities of training programs aiming at improving auditory processing and/or sequential memory (e.g., by reinforcing perception via crossmodal/dual-modal encoding).

Published online: August 14, 2026

Author Note

Author contributions are as follows—Conceptualization: FT, AC, and BT; Methodology: FT, JV, AC, and BT; Software (psychopy): FT; Formal analysis: FT; Investigation: FT, JV, and DW; Resources: MZ, AC, and BT; Data curation: FT; Writing – original draft: all authors; Writing – review & editing: FT, JV, MZ, AC, and BT; supervision: AC, BT, and MZ; project administration: FT.

Anne Caclin and Barbara Tillmann share the last authorship.

Funding: Barbara Tillmann was supported by the ANER grant of the Region Bourgogne-Franche Comté (MusiC).

Correspondence concerning this article should be addressed to Dr. Francesca Talamini (Francesca.Talamini@uibk.ac.at), Department of Psychology, Universitätsstraβe 15, University of Innsbruck, Innsbruck 6020, Austria.

References

Aizenman, A. M., Gold, J. M., & Sekuler, R. (2018). Multisensory integration in short-term memory: Musicians do rock. *Neuroscience*, *389*, 141–151. https://doi.org/10.1016/j.neuroscience.2017.04.031

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*, *136*(5), 1639–1661. https://doi.org/10.1093/brain/awt082

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*(3), 1371–1379. https://doi.org/10.1121/1.4863269

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., Lickliter, R., Castellanos, I., & Todd, J. T. (2015). Intrasensory redundancy facilitates infant detection of tempo: Extending predictions of the intersensory redundancy hypothesis. *Infancy*, *4*, 377–404. https://doi.org/10.1111/infa.12081

Balch, W. R., & Muscatelli, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception and Psychophysics*, *40*(5), 351–358. https://doi.org/10.3758/BF03203027

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception and Psychophysics*, *57*(8), 1151–1162. https://doi.org/10.3758/BF03208371

Bonn, C. D., & Cantlon, J. F. (2017). Spontaneous, modality-general abstraction of a ratio scale. *Cognition*, *169*, 36–45. https://doi.org/10.1016/j.cognition.2017.07.012

Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception, and Psychophysics*, *78*(5), 1512–1528. https://doi.org/10.3758/s13414-016-1088-5

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, *81*(1), 124–130. https://doi.org/10.1016/j.bandc.2012.09.005

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*(2, Pt. 2), 524–531. https://doi.org/10.1121/1.1912382

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics*, *21*(1), 60–64. https://doi.org/10.3758/BF03199469

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception and Psychophysics*, *35*(5), 477–488. https://doi.org/10.3758/BF03203924

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6. https://doi.org/10.1167/10.1.6

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. https://doi.org/10.1007/s00221-002-1262-y

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics*, *68*(7), 1191–1203. https://doi.org/10.3758/BF03193720

Galvin III, J. J., Fu, Q.-J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, *1169*(1), 518–533. https://doi.org/10.1111/j.1749-6632.2009.04551.x

Gillmeister, H., & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Research*, *1160*, 58–68. https://doi.org/10.1016/j.brainres.2007.03.041

Grau, J. W., & Kemler Nelson, D. G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology. General*, *117*(4), 347–370. https://doi.org/10.1037//0096-3445.117.4.347

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Caclin, A., & Tillmann, B. (2019). Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and Cognition*, *136*, 103614. https://doi.org/10.1016/j.bandc.2019.103614

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Robert E. Krieger.

Green, E. J. (2021). Binding and differentiation in multisensory object perception. *Synthese*, *198*(5), 4457–4491. https://doi.org/10.1007/s11229-019-02351-1

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values ofd′. *Behavior Research Methods, Instruments, and Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental Psychology*, *62*(2), 123–130. https://doi.org/10.1027/1618-3169/a000279

Hines, D. (1993). Effect of masked picture primes on semantic priming for easy- and difficult-to-name words. *The Journal of General Psychology*, *120*(2), 149–164. https://doi.org/10.1080/00221309.1993.9921190

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, *109*(2), 219–238. https://doi.org/10.2307/1423274

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

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, *24*(6), 551–565. https://doi.org/10.3758/BF03198783

Jasp Team (2024). *JASP* (Version 0.18.3) [Computer software].

Johanis, T. C., & Schmuckler, M. A. (2021). Investigation of multisensory harmonic priming: Audiovisual integration in chord perception. *Auditory Perception and Cognition*, *4*(1–2), 74–96. https://doi.org/10.1080/25742442.2021.1982584

Küssner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00789

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLOS One*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Lee, Y.-S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, *57*(1), 293–300. https://doi.org/10.1016/j.neuroimage.2011.02.006

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology. Human Perception and Performance*, *33*(5), 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, *121*(10), 1853–1867. https://doi.org/10.1093/brain/121.10.1853

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

Marks, L. E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(3), 586–602. https://doi.org/10.1037//0096-1523.15.3.586

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, *50*(1–2), 125–145. https://doi.org/10.1016/S0167-8760(03)00129-6

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*(7), 903–923. https://doi.org/10.1068/p2866

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., Babiloni, F., Betti, V., & Aricò, P. (2021). The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports*, *11*(1), 4831. https://doi.org/10.1038/s41598-021-84196-8

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263–1271. https://doi.org/10.1111/j.1467-9280.2008.02235.x

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 69–79. https://doi.org/10.1037/0096-1523.15.1.69

Melara, R. D., & Marks, L. E. (1990a). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics*, *48*(2), 169–178. https://doi.org/10.3758/bf03207084

Melara, R. D., & Marks, L. E. (1990b). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(2), 398–414. https://doi.org/10.1037//0096-1523.16.2.398

Morrongiello, B. A., & Roes, C. L. (1990). Developmental changes in children's perception of musical sequences: Effects of musical training. *Developmental Psychology*, *26*(5), 814–820. https://doi.org/10.1037/0012-1649.26.5.814

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, *30*(3), 277–292. https://doi.org/10.1016/0028-3932(92)90005-7

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. *Journal of Clinical Medicine*, *10*(10), Article 10. https://doi.org/10.3390/jcm10102093

Prince, J., Schmuckler, M., & Thompson, W. (2009). Cross-modal melodic contour similarity. *Canadian Acoustics*, *37*, 35–49.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reinitz, M. T., Wright, E., & Loftus, G. R. (1989). Effects of semantic priming on visual encoding of pictures. *Journal of Experimental Psychology: General*, *118*(3), 280–297. https://doi.org/10.1037/0096-3445.118.3.280

Rosner, B. S., & Meyer, L. B. (1986). The perceptual roles of melodic process, contour, and form. *Music Perception*, *4*(1), 1–39. https://doi.org/10.2307/40285350

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, 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

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception and Psychophysics*, *54*(3), 406–416. https://doi.org/10.3758/BF03205276

Sassi, F., Campoy, G., Castillo, A., Inuggi, A., & Fuentes, L. J. (2014). Task difficulty and response complexity modulate affective priming by emotional facial expressions. *Quarterly Journal of Experimental Psychology*, *67*(5), 861–871. https://doi.org/10.1080/17470218.2013.836233

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*(2), 185–195. https://doi.org/10.1016/S0896-6273(00)80448-1

Schmuckler, M. A., & Gilden, D. L. (1993). Auditory perception of fractal contours. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(3), 641–660. https://doi.org/10.1037/0096-1523.19.3.641

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5), 1763. https://doi.org/10.1213/ANE.0000000000002864

Schubert, E., & Stevens, C. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research*, *35*(2), 161–174. https://doi.org/10.1080/09298210600835000

Schulze, K., & Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory (Hove, England)*, *21*(3), 377–395. https://doi.org/10.1080/09658211.2012.731070

Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: Evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, *123*(3), 546–559. https://doi.org/10.1093/brain/123.3.546

Schutz. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review*, *3*(3), 83–108. https://kb.osu.edu/items/78348d42-101b-571c-aaac-5f4e00e92e10

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2022). Auditory and visual short-term memory: Influence of material type, contour, and musical expertise. *Psychological Research*, *86*(2), 421–442. https://doi.org/10.1007/s00426-021-01519-0

Thaut, M. H., Trimarchi, P. D., & Parsons, L. M. (2014). Human brain basis of musical rhythm perception: Common and distinct neural substrates for meter, tempo, and pattern. *Brain Sciences*, *4*(2), 428–452. https://doi.org/10.3390/brainsci4020428

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition*, *191*, 103955. https://doi.org/10.1016/j.cognition.2019.04.024

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

Weinberger, N. M., & McKenna, T. M. (1988). Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception. *Music Perception*, *5*(4), 355–389. https://doi.org/10.2307/40285407

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410