UNIVERSITY *of York*

This is a repository copy of *Distinguishing between musical excepts learned by novices individually or in pairs*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/206242/

Version: Published Version

Article:

Schiavio, Andrea orcid.org/0000-0001-8109-9185, Popescu, Tudor, Kempf, Adrian et al. (1 more author) (2024) Distinguishing between musical excepts learned by novices individually or in pairs. Psychology of Music. 472–488. ISSN: 0305-7356

https://doi.org/10.1177/03057356231212408

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Article

Distinguishing between musical excerpts learned by novices individually or in pairs

Andrea Schiavio[1,2]* (iD), Tudor Popescu[3,4]* (iD), Adrian Kempf[5,2] and Renee Timmers[6]

Abstract

In a previous study, we found that musical learning of novices was equally accurate (in terms of playing the correct rhythms and pitches) when learning in a solo or duo setting. Intrigued by these findings, we conducted a follow-up experiment investigating whether the learning outcomes differed in subjective terms as perceived by listeners judging the performances. Here, expert musicians were asked to discern whether melodies learned under conditions of synchrony, turn-taking, or imitation were likely learned individually or with a partner. In addition, they evaluated the learners' proficiency in playing the melodies, assessing expressiveness, the clarity of their articulation and phrasing, as well as the overall coherence of the performances. Listeners showed differential responses to both the learning condition and the pairing (solo or duo) in which the melodies were learned. Although the outcome did not yield salient-enough results for significant clusters of responses to emerge, our findings could inspire future research to delve into the question of discernible signatures between individual and group musical learning by adopting a longitudinal approach.

Keywords

expressiveness, articulation, phrasing, coherence, music perception

Over the last few decades, several studies have investigated musicians' usage of a wide range of expressive strategies during performance, to make each interpretation unique. These strategies can consist of shaping parameters, such as phrasing, dynamics, and articulation, (see, e.g., Demos et al., 2016; Fabian et al., 2014; Palmer, 1996, 1997; Timmers, 2003, 2005). As

[1]School of Arts and Creative Technologies, University of York, York, UK
[2]Centre for Systematic Musicology, University of Graz, Graz, Austria
[3]Department of General Psychology, University of Padova, Padova, Italy
[4]Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria
[5]Department of Psychology, University of Graz, Graz, Austria
[6]Department of Music, University of Sheffield, Sheffield, UK

*Shared first authorship

Corresponding author:
Andrea Schiavio, School of Arts and Creative Technologies, University of York, Heslington, York, YO10 5DD, UK.
Email: andrea.schiavio@york.ac.uk

reported by Doğantan-Dack (2014), this research orientation has a long history, stretching back to early work on musical expressiveness in performance put forward by Seashore (1938) and Lussy (1874). This scholarly tradition considers expert musicians to be able to strike a remarkable balance between providing an accurate and recognizable performance of the music and delivering distinctive artistic outputs by manipulating different musical parameters (see Héroux, 2016). Indeed, beyond improved technical skills, expert musicians have been shown to also possess a wider range of interpretations than music students (Repp, 1990). That said, it should be noted that novice musicians and learners can also engage with, and cultivate, a variety of interpretative skills since their earliest learning experiences. Accordingly, a number of contributions have suggested that children display a natural tendency to develop a personal musical style (Imberty, 1983); that music students often conceive of expressiveness as an inherent quality of musicking (Schiavio et al., 2019); that young musicians can be taught to perform music expressively through a dialogic approach based on enquiry and discussion (Meissner & Timmers, 2020); and that, on a more general level, achieving independent expressive competences might be one of the main goals of musical learning, both in individual and in collaborative contexts (Borgo, 2007; Elliott, 2005; Griffiths, 2017).

Because of its focus on expressive creativity through interaction, learning music together with others has recently been explored as a particularly valuable pedagogical setting. This learning modality may help students enhance their creative and musical skills as well as benefit from an increased sense of social and musical inclusion (Burnard et al., 2008). Such insights have prompted researchers to reflect on the differences and continuities between musical pedagogies based on individual and group learning (Brandler & Peynircioglu, 2015; Schiavio, Biasutti, et al., 2020; see also Hanken, 2016; Nielsen et al., 2018). A recent study (Schiavio, Stupacher, et al., 2020) examined how well novice musicians, either solo or in pairs, were able to reproduce, on a digital piano, a set of short target melodies presented via dedicated instructional videos in three learning conditions: *synchrony*, *turn-taking*, and *imitation*. In the synchronization condition, participants reproduced the melody in synchrony with a video (solo group) or their assigned partner (duo group). The turn-taking condition involved the participant playing in alternation with the computer (solo group) or with the other participant (duo group). In the imitation condition, participants imitated a video performance shown on the computer (solo group) or imitated their partner who reproduced a video performance (duo group). Solo participants were instructed to learn the melodies in these conditions by engaging with the videos, whereas pairs were invited to perform the same tasks interacting with each other. Analyzing the temporal and melodic accuracy with which the target melodies were performed by participants, it was found that in both groups (*solo* and *duo*), learning in synchrony and by turn-taking gave rise to significantly better results when compared to imitation. Learning individually likewise did not significantly differ from learning with a partner when looking at the subjective ratings (e.g., enjoyment, satisfaction, confidence about learning, etc.) that participants reported after taking part in the study.

Since these results were obtained when analyzing the objective properties of the participants' musical outcomes, and their felt, subjective learning experiences (i.e., via questionnaires), it remains an open question whether differences between solo and duo musical learning might be differentially appreciated from a perceiver's standpoint. For instance, when listeners are invited to pay particular attention to expressiveness (Clarke, 1988), articulation and phrasing (Doğantan-Dack, 2012) or coherence (Zbikowski, 1999), new subtle nuances may be disclosed, potentially leading to a richer musical experience. Can this aid listeners to distinguish between musical excerpts learned individually and together with others? The present research addresses this question by reporting on a new exploratory study in which expert musicians

were invited to evaluate the short musical excerpts performed by the novice learners of the study by Schiavio, Stupacher, et al. (2020) and compare them to the target melodies these learners were invited to learn. We expected that relevant differences between solo and duo conditions might be observed, that is, that listeners would distinguish between musical outcomes learned individually or with a partner. We based this prediction on the hypothesis that interacting with others while learning shapes one's musicking. This perspective is prominent in contemporary music scholarship influenced by embodied and interactive approaches to cognition (see, e.g., Moran, 2014; Reybrouck, 2021; van der Schyff et al., 2022). These approaches emphasize how the reciprocal influence between individuals and their environment can foster creativity and positive experiences when making and learning music (e.g., Nijs, 2019; Sawyer, 2003). Specifically, a number of scholars have recently argued that prioritizing collaboration in contexts where learners can learn from each other as well as from a teacher can yield significant benefits in terms of facilitating skill acquisition, fostering trust, and promoting social understanding (see Borgo, 2007; Schiavio & Nijs, 2022). By contrast, in research looking at expert musicians, empirical evidence comparing expressivity in group versus solo conditions has pointed toward the potentially constraining aspects of performing in a group, leading musicians to reduce expressive variations (Bishop & Goebl, 2020).

Either way, we speculate that an interactive musical experience might foster distinctive stylistic characteristics that may not always be captured by the analysis of quantifiable parameters (such as pitch and temporal accuracy; though possibly of others which we have not evaluated), but which nevertheless may be picked up by attentive listeners.[1] Previous research has shown the challenges in predicting perceived similarity between performances based on measured characteristics (e.g., Timmers, 2005), indicating instead the relevance of communication of musical intentions and interpretations. Rather than perceiving exact deviations in timing, dynamics, rhythm or pitch, listeners have been shown to be sensitive to communicative aspects of performances, such as its degree of expressivity (Kendall & Carterette, 1990) and its manner of phrasing and articulation (Palmer, 1996). With respect to differences between duo versus solo performances, we expect these to emerge based on performers' engagement with both covert and overt forms of bodily activity (Brown & Martinez, 2007; Cross, 2010), and through motor resonance with the sounding performance (Overy & Molnar-Szakacs, 2009; Patel & Iversen, 2014). In addition, two studies conducted by De Poli et al. (2014) and Schubert & Kreutz (2014) explored participants' evaluations of computer-assisted music performances across various dimensions, such as accuracy, emotional content, and coherence of the performed style.[2] While the dimensions examined in these studies align with our own, the studies did not specifically address perceptual distinctions between stimuli generated individually versus those created collaboratively with another peer. For this reason, when designing the present study we were specifically interested in what qualitative differences may emerge between listening to musical excerpts performed independently or as a dyad. We see joint learning as a context where motivation and role-modeling play an important role, possibly leading to more coherence and stability (Bishop & Goebl, 2020) as well as creative emergence (Pennill & Breslin, 2022).

To examine these predictions, we asked our participants to rate all the melodies played by the learners of the previous study, in light of five questions. The first one explicitly asked whether, according to the listener, the melody had been learned alone or with a partner; the second referred to how well the melody had been learned when compared to the target stimulus. While these questions were posed to elicit answers that might be more easily examined in light of the data reported in the work of Schiavio, Stupacher, et al., 2020 (from now on labeled "the original study") and its experimental manipulations, the three following questions referred instead to the musical dimensions discussed above (i.e., expressiveness, phrasing and articulation, and

coherence). In asking these questions, we aimed at gaining richer insight into which set of subjective parameters may distinguish between musical outcomes learned individually versus with a partner.

Methods

Participants

Participants for the present experiment were recruited using the Prolific platform (www.prolific.co). Inclusion criteria were: (1) having more than 5 years of experience in playing a musical instrument and (2) holding an undergraduate, graduate, or doctorate degree in music. A total of 20 participants took part in the study (10 men, 10 women; age: $M = 24.9$ years, $SD = 4.23$). The number of years of musical training was recorded for each participant ($M = 14.1$, $SD = 4.56$). A sample size of 20 participants has been shown to be promising in reaching a high inter-rater reliability, even in complex and subjective assessment situations typical of creativity ratings (Ceh et al., 2022). All procedures were approved by the Ethical Committee at the University of Graz and were in accordance with the statements of the Declaration of Helsinki. Participants were monetarily compensated for their involvement in the study and provided written informed consent. None of the participants had participated in the original study.

Stimuli

Three short melodies were composed by author AS for the original study, all involving a limited set of pitches (F♯, G♯, and A♯) in different orders, styles, and rhythms. The melodies were performed by a professional pianist on a digital piano (*Yamaha Clavinova CLP370*) and were recorded using the software *Reaper64*. These recordings constituted the "target melodies" that participants were asked to reproduce under various conditions in the original study. Participants learned these melodies either on their own (*solo* group) or paired up (*duo* group). Furthermore, learning occurred under one of three learning conditions: *synchronization*, *turn-taking*, and *imitation.* Each condition had a unique short melody to learn and perform in that condition. Performances of 36 participants (18 groups) in the "duo" condition and 18 participants in the "solo" condition were recorded, for a total of 162 recordings collected (i.e., 54 per learning condition/melody). Participants of the rating study were asked to rate each of the recordings ("reproduced melodies") in a rating trial (162 trials in total). The average (mean $\pm$ *SD*) duration of a melody was 16.8 ± 4.29 s.

Procedure

The task was implemented using the jsPsych JavaScript framework (de Leeuw, 2015) and administered online using Pavlovia (www.pavlovia.org). Online participants were invited to listen to pairings of target melodies and reproduced melodies and rate each reproduced melody on a number of musical dimensions. Each trial consisted of a screen containing two playback buttons (for the current target melody and the current reproduced melody) and five questions. Between the different trials presented to any one participant, all combinations of available target and reproduced melodies were covered (presentation order: latter nested, in sequence, within the former). The five questions (and rating options) presented within each trial were:

(A) Do you think the melody was learned individually or with a partner? ("Individually" . . . "Together")

(B) How well do you think the performed melody was learned? ("Not well at all" . . . . "Extremely well")

(C) Please evaluate the expressiveness of the performed melody ("Very bad" . . . "Excellent")

(D) Please evaluate the musical phrasing and articulation of the performed melody ("Very bad" . . . "Excellent")

(E) Please evaluate the overall musical coherence of the performed melody ("Very bad" . . . . "Excellent")

Question (A) was rated on a quasi-continuous (21-step) confidence rating scale, presented to participants as horizontal line with a moving slider. Questions (B)–(E) were rated on 10-point Likert-type scales. We opted not to use a Likert-type scale for Question (A) as we expected this question to be hard for participants to decisively answer. With a quasi-continuous scale, we wanted to be able to measure more subtle variations in participant's responses. A linear transformation was applied prior to the analyses, to equalize scale ranges across all five questions. The order of questions was shuffled between trials. Each participant ran through 162 rating trials, one for each reproduced melody. Data analyses exclusively employed these five ratings as dependent variables. Trials' response times (RTs) were only used to define and exclude outliers (see the "Outlier removal" section). The entire rating task lasted around 1.5 hr on average (ca. 33 s per melody).

Data analyses

Outlier removal. We defined as "outliers" trials whose total RT was suspiciously low, measured from the onset of the trial screen to the onset of participants clicking on the "Next melody" button. Specifically, we defined the playback duration of the reproduced melody as the RT threshold. We did not include an upper bound on RTs (i.e., exclude participants who took too *long* with the task), since the task provided no opportunities for breaks, and we wished to allow for the possibility of participants taking arbitrary breaks within the duration of certain trials. Such outliers were excluded on a per trial basis, which amounted to 9.72% of the trials.

Linear mixed-effects models. To check for the effect of conditions on ratings, we fitted linear mixed-effects models, separately for each rating. We used the *fitlme* function in MATLAB (Mathworks Inc., MA), with the formula:

ratings ~ pairing + learning + repNumber + training + (1 | id_rater) + (1 | id_melody)

where the fixed-effects were the pairing condition (solo/duo), learning condition (imitation/synchronization/turn-taking), trial repetition number (1/2/3), and participant's number of years of musical training (continuous).

Dimensionality reduction. Given the high dimensionality of the data set (five different ratings per trial), we aimed to remove redundant information and arrive at a parsimonious explanation of our data. With dimensionality reduction, the aim is to "embed" high-dimensional data into lower-dimensional data (2D or 3D), which can be visualized and interpreted, for instance, by

observing emerging clusters that align with experimental manipulations. To this aim, we used tSNE (t-Distributed Stochastic Neighbor Embedding), a machine learning approach to dimensionality reduction which, unlike more traditional methods like principal component analysis (PCA), usefully retains local rather than global variance (Belkina et al., 2019; Hinton & Roweis, 2002). tSNE can be used together with several distance metrics, and by way of an exploratory analysis, we compared three different metrics (the Euclidean, Chebyshev, and Mahalanobis distances), aiming to find which yielded a better separation of our reproduced melodies, as classified (by the tSNE algorithm) in accordance with our two pairing conditions and our three learning conditions. We used for this the *tsne* function in MATLAB. The Euclidean, Chebyshev, and Mahalanobis distances are mathematical measures used to quantify the similarity or dissimilarity between objects or data points in various contexts. They respectively incorporate the straight-line distance, the maximum difference, and the covariance structure of the data.

Intra-class correlations. Intra-class correlations (ICCs) are descriptive statistics that can be used when measurements are made on units organized into groups (McGraw & Wong, 1996). They describe how strongly units in the same group resemble each other. We used ICCs to assess how consistent the participants' scores were to each other, for each of the five ratings. We treated all observations as coming from two different groups, on the basis of the pairing condition. To evaluate whether participants answered each question in a consistent manner, we computed the average ICC using the ICC(2, k) form (Koo & Li, 2016; Shrout & Fleiss, 1979), computed in MATLAB using the ICC function.[3] We considered its output to be a measure of inter-rater reliability, as it corresponds effectively to a two-way random effects model for consistency (McGraw & Wong, 1996). Since ICCs give a composite of intra- and inter-observer variability, their results are difficult to interpret when observers are not interchangeable.[4] Alternative measures, such as Cohen's kappa statistic are more suitable measures of agreement among non-exchangeable observers. However, in this case, we assume our participants are interchangeable.

Logistic regression. We created, separately for each of the three learning conditions, a logistic binomial model of the probability of a reproduced melody having been performed solo or duo, as a function of the five ratings, specified in the model individually (i.e., with no interaction terms). Using *fitglm* in MATLAB, the generalized linear regression model was specified as follows:

$$\text{logit(pairing)} \sim 1 + A + B + C + D + E$$

$$\text{Distribution} = \text{Binomial}$$

where A–E were ratings for the five questions.

Variance Inflation Factors. We computed variance inflation factors (VIFs) across the five questions, to check the degree to which these variables are multicollinear. This returns a single number per target. That is, for each of our five questions, its VIF estimates to what degree its data is accounted for by the data in the other four questions. VIFs range from 1 upward, and indicate what percentage of the variance (i.e., the standard error squared) is inflated for each coefficient. For example, a VIF of 1.9 means the variance of a particular coefficient is 90% greater than what would be expected if there was no multicollinearity at all, that is, no correlation with other predictors (VIF = 1). As a rule of thumb, VIFs between 1 and 5 indicate
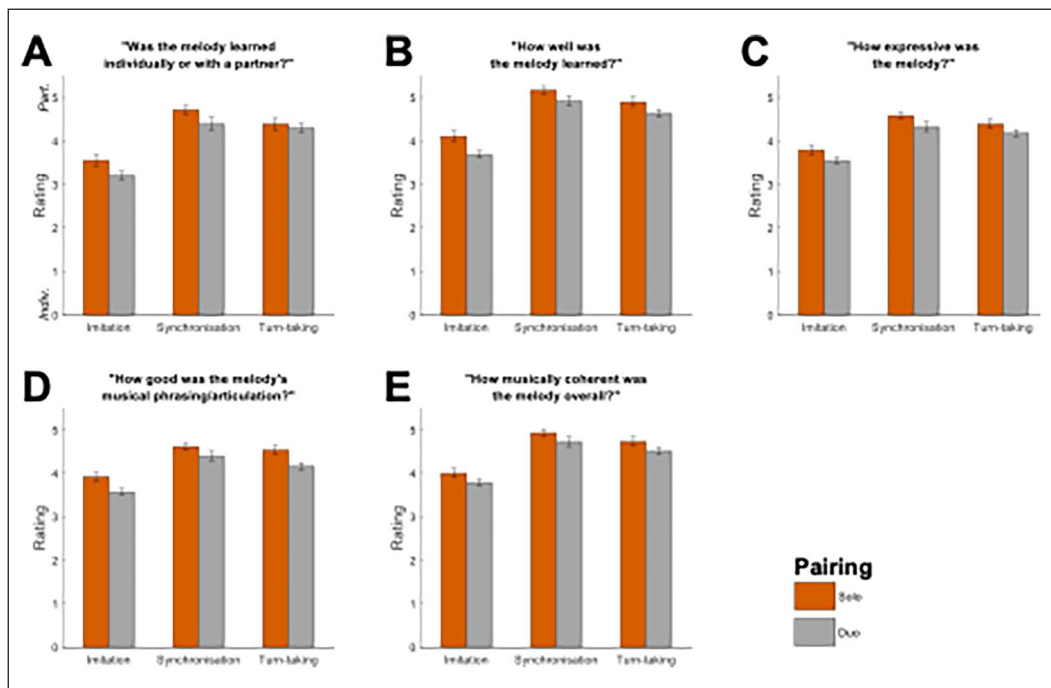
Figure 1. Ratings Across Conditions.
Note. Questions (A–E) solicited ratings relating to how the melodies were performed. Bar heights show means, error bars show standard errors. "Indiv." and "Part." in Panel A's *y*-axis refer to the ends of the scale that indicate a rating for pieces learned individually versus with a partner. Participants were explained that a higher rating indicates more of the respective trait (e.g., expressiveness) being present in the music.

moderately correlated variables, and VIFs greater than 5, highly correlated variables (Everitt & Skrondal, 2010).

Results

Rating distributions

Figure 1 shows the distributions of ratings for reproduced melodies under the various (learning and pairing) conditions, and in response to the five ratings concerning musical dimensions.

Linear mixed-effects models

Coefficients for the estimated linear mixed-effects models are given in Table 1. These results suggest that both the pairing and the learning conditions significantly influenced ratings given in response to each of the five questions, while musical training did not, likely due to a ceiling effect.

Dimensionality reduction

Figure 2 depicts the result of applying tSNE to our ratings, classifying by pairing condition and by learning condition, in two dimensions. Visually, it can be seen that no clusters form to

Table 1. Results of Linear Mixed-Effects Models.

Question/rating	Name	Estimate	SE	t	df	p	LB	UB
A	(Intercept)	6.13	0.76	8.05	2,920	<.001	4.64	7.63
	pairing	−1.12	0.21	−5.37	2,920	**<.001**	−1.53	−0.71
	learning	0.55	0.06	9.07	2,920	**<.001**	0.43	0.67
	repNumber	−0.57	0.12	−4.69	2,920	**<.001**	−0.81	−0.33
	training	0.02	0.03	0.56	2,920	.575	−0.05	0.08
B	(Intercept)	6.83	0.74	9.25	2,920	<.001	5.39	8.28
	pairing	−1.09	0.15	−7.13	2,920	**<.001**	−1.39	−0.79
	learning	0.43	0.04	9.77	2,920	**<.001**	0.35	0.52
	repNumber	−0.53	0.09	−5.95	2,920	**<.001**	−0.70	−0.35
	training	−0.02	0.04	−0.60	2,920	.547	−0.10	0.05
C	(Intercept)	6.75	0.79	8.58	2,920	<.001	5.21	8.29
	pairing	−0.90	0.13	−6.71	2,920	**<.001**	−1.16	−0.64
	learning	0.31	0.04	7.89	2,920	**<.001**	0.23	0.38
	repNumber	−0.45	0.08	−5.76	2,920	**<.001**	−0.60	−0.29
	training	−0.06	0.05	−1.34	2,920	.181	−0.15	0.03
D	(Intercept)	7.06	0.73	9.72	2,920	<.001	5.63	8.48
	pairing	−1.06	0.14	−7.73	2,920	**<.001**	−1.33	−0.79
	learning	0.30	0.04	7.48	2,920	**<.001**	0.22	0.38
	repNumber	−0.51	0.08	−6.42	2,920	**<.001**	−0.67	−0.36
	training	−0.05	0.04	−1.23	2,920	.220	−0.13	0.03
E	(Intercept)	7.25	0.77	9.43	2,920	<.001	5.74	8.75
	pairing	−1.08	0.14	−7.82	2,920	**<.001**	−1.35	−0.81
	learning	0.36	0.04	8.94	2,920	**<.001**	0.28	0.44
	repNumber	−0.59	0.08	−7.33	2,920	**<.001**	−0.74	−0.43
	training	−0.04	0.04	−0.95	2,920	.341	−0.13	0.04

Note. LB and UB, respectively, represent the lower and upper bounds of confidence intervals for the estimated coefficients. *p*-values for significant predictors are given in bold.

separate the melodies according to their ratings, a result we later also explore statistically. A three-dimensional embedding yielded similar results to the 2D embedding pictured.

Intra-class correlations

Table 2 reports confidence intervals for estimated ICCs, and corresponding hypothesis tests, performed for the null hypothesis that ICC = 0. The ICCs for Questions A and C are indicative of a good to excellent inter-rater reliability, and Questions B, D, and E reach an overall excellent inter-rater reliability (Koo & Li, 2016). Thus, our expert raters showed a very high consistency and coherence in their assessments.

Logistic regression

Only the model for the first learning condition, *imitation*, was significant ($p<0.05$), meaning it differed statistically from a trivial (constant) model, in its attempt to predict pairing condition (solo or duo) on the basis of the five regressors. Within this model, the two questions that constituted
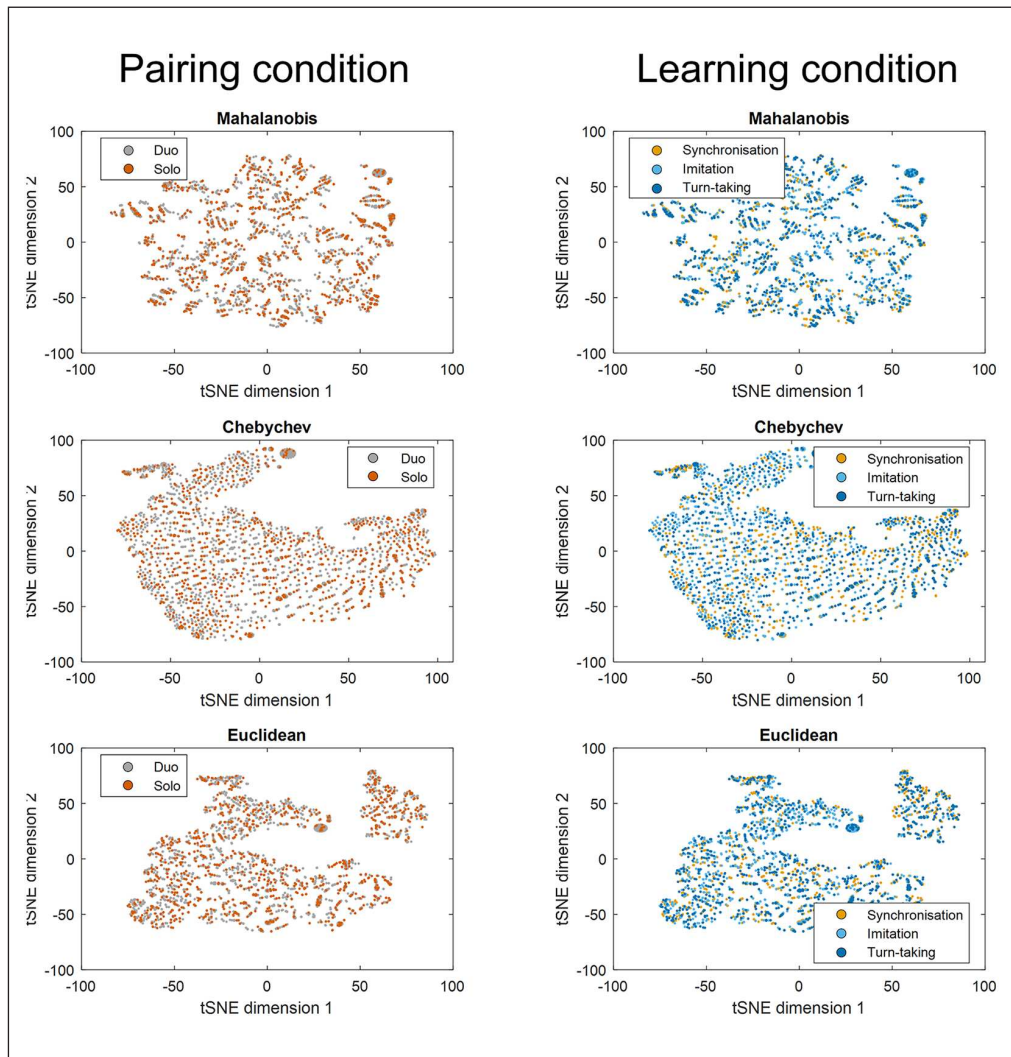
Figure 2. Dimensionality Reduction.

Note. In all cases, the two dimensions, *x* and *y* (which have no immediate interpretation or meaning) are reduced from the original five dimensions (the ratings) using t-Distributed Stochastic Neighbor Embedding (tSNE). Classified are either the pairing condition (left column) or the learning condition (right column). Note that *x*–*y* pairs (data points, representing trials) are distributed identically across the two columns, but colored differently, according to each trial's membership in the two respective sets of conditions. It can be seen that data points do not demarcate any obvious clusters, for any of the distance metrics used (Euclidean, Chebyshev, and Mahalanobis).

significant regressors were B and E, that is, the judgment regarding how well participants reproduced the melodies they were invited to learn and the overall coherence of their performances. Reproduction quality was negatively associated with pairing, whilst coherence showed a positive association. The model for the *turn-taking* condition approached significance ($p = .056$; see Table 3), Question D (which refers to phrasing and articulation) being its sole significant regressor showing a negative relationship with pairing.

Table 2. Intra-Class Correlations, of the Type ICC(2, *k*).

Question/ rating	*ICC (2, k)*	LB	UB	*F*	*df*1	*df*2	*p*
A	.87	0.84	0.90	7.77	161	3,059	<**.001**
B	.93	0.92	0.95	14.60	161	3,059	<**.001**
C	.91	0.89	0.93	10.98	161	3,059	<**.001**
D	.92	0.90	0.94	12.88	161	3,059	<**.001**
E	.92	0.90	0.94	12.76	161	3,059	<**.001**

Note. LB and UB, respectively, represent the lower and upper bounds of confidence intervals for the ICC estimated for each rating. The *F*-value, degrees of freedom, and the corresponding *p*-value for the hypothesis test ($\alpha = 0.05$) are also reported.

Variance Inflation Factors

For the five ratings, the obtained VIFs were, respectively: 1.7, 4.0, 3.5, 4.7, and 5.5. Since all factors are below 10 (and also considering the sample size), we did not worry about serious collinearity problems in the regression analyses (Miles, 2014). Moreover, this suggests that Rating A, concerning the pairing of the performed melody (learned individually versus with a partner), contained the most "signal," that is, the highest amount of unique subjective information.

Discussion and conclusion

The present study examined whether expert musicians, when focusing on subtle musical nuances that may not be captured by straightforward measurements of pitch and tempo accuracy (Schiavio, Stupache, et al., 2020), can distinguish between musical excerpts learned by novices individually or in pairs.

By asking five questions related to different musical dimensions, we found that our participants responded differentially to the learning condition (*synchrony*, *turn-taking*, and *imitation*) and pairing (*solo* and *duo*) with which the melodies were learned. However, this result was not salient enough to let us observe significant clusters of responses, as depicted in Figure 2. Distinguishing between musical outcomes performed in different learning conditions and pairings, in other words, was no easy task for our participants. This may be due to various factors. First, learners were asked to *reproduce* a melody, so that, it is possible that their creativity was not fully engaged by the task, restricting differences in group processes or individual performances to emerge. This still allows for the possibility that with a greater degree of creative freedom, the pairing effect would have been more salient in the ratings, potentially giving rise to distinguishable response clusters. Such a reasoning is supported by recent work in creativity research exploring how creative potential and flexibility are often limited when relying on habits, existing patterns, or when agents cannot move "past old ideas 'inside the box'" (Ibáñez de Aldecoa et al., 2021; see also Storm & Angello, 2010). Second, the interaction between musical partners in the original study was quite short (maximum 12 min per learning condition)—perhaps not enough to leave a significant "trace" in the way the melodies were played. A third reason for the lack of clear distinction between the two conditions could be that solo and duo contexts may both be conducive to development in their own ways. Individual contexts necessitate less to align performance with others and may therefore, at least initially, be less inhibiting.

Table 3. Results of Logistic Regression.

Learning condition	Model estimation				Question/rating	Estimated coefficients			
	Observations	Error *df*	χ^2	*p*		Estimate	*SE*	*t*	*p*
Imitation	961	955	14	**.0153**	(Intercept)	1.04	0.16	6.49	<.001
					A	−0.01	0.03	−0.41	.68
					B	−0.13	0.06	−2.19	**.03**
					C	0.02	0.07	0.38	.70
					D	−0.14	0.07	−1.87	.06
					E	0.17	0.08	2.14	**.03**
Synchronization	980	974	5.32	.378	(Intercept)	1.10	0.20	5.51	<.001
					A	−0.03	0.03	−0.94	.35
					B	−0.03	0.07	−0.50	.61
					C	−0.05	0.06	−0.84	.40
					D	0.01	0.08	0.10	.92
					E	0.03	0.08	0.36	.72
Turn-taking	984	978	10.8	.0565	(Intercept)	0.96	0.18	5.37	<.001
					A	0.03	0.03	0.88	.38
					B	−0.04	0.06	−0.62	.53
					C	0.06	0.06	0.91	.36
					D	−0.19	0.07	−2.62	**.01**
					E	0.08	0.08	1.02	.31

Note. Significant models and predictors (questions) are in bold. *df*: degrees of freedom. Chi-squared test (χ^2) is relative to a constant model.

These last two points align well with existing research that explores the efficacy of pedagogical settings emphasizing mutual trust, open discussion, and peer collaboration in fostering skill development over the course of a program (Robinson & Kakela, 2006; Schiavio & Nijs, 2022). With this in mind, it is crucial to recognize that the type of musical interaction that takes place in such contexts cannot be reduced to a passive process. Instead, it demands active engagement and cooperation with others, fostering the gradual development of trust and the establishment of connections over time. Instrumental pedagogy that involves peer-to-peer interaction can thus benefit from structured opportunities to sustain engaging interactions over extended periods. By doing so, the necessary conditions can be created to facilitate learning and foster expressive performance. This sustained interaction may give rise to musical stimuli with more distinct characteristics that can be associated with either individual or duo modes of learning. Indeed, if we consider the perception of music as an active process of creative re-creation, reliant on continuous bodily grounded and exploratory processes that aim to establish a novel and appropriate engagement with the environment (Schiavio et al., 2022; see also Baroni, 2006), then a more robust and prolonged interaction may promote modes of listening that consistently resonate with the intersubjective features inherent in the musical stimulus, allowing listeners to more consistently recognize said (social) relationships.

To gain richer insight into solo *versus* duo contexts for skill development, future studies may need to consider longitudinal approaches, where differences between individual and duo learning could be explored over the long term. This might involve looking at both quantifiable parameters (e.g., pitch and temporal accuracy, and expressive variation) and at the learners' subjective experiences, integrating quantitative and qualitative approaches in a rich variety of ways (see

West, 2014). In using a longitudinal approach, it will be of interest to examine the development over time in solo and duo learning across phases of rehearsal that may include exploration, transition, and consolidation (Pennill, 2019).

Probably, the most interesting finding of the present study emerges from the logistic regression: on one hand, we do not have hard evidence that our participants could discriminate directly whether a piece was learned individually or together. On the other hand, when it comes to musical outcomes learned via imitation, then certain rated characteristics do in fact bear an awareness of the pairing condition. In particular, with increasing scores in Rating B ("How well do you think the performed melody was learned?"), the probability decreases that the musical excerpt evaluated was learned in *duo*. Hence, the probability increases that the melody belonged to the *solo* group. Conversely, with increased ratings in E ("Please evaluate the overall musical coherence of the performed melody"), the probability increases that the melody belonged to the *duo* group. Note that this response pattern is associated with melodies learned via imitation only. In the original study, the melodies learned in this condition featured more errors when compared to those learned through the other learning conditions. Remarkably, this effect was almost always present for both solo and duo pairings: learning by imitation, in other words, generally yielded less accurate musical outcomes when novices were asked to reproduce target melodies both individually and collectively. It is therefore plausible that our raters were able to capitalize on such a difference between learning conditions, giving rise to results that resonate with those reported in the original study. And indeed, this would appear to be the case, as Questions B and E relate to how well the melodies were learned and perceived as coherent, respectively—parameters perhaps more associable with (yet not reducible to) general performance quality than others.

Looking more closely at the findings of the original study, some important details emerge that may further clarify the results obtained in the present contribution. There, temporal accuracy in the solo group was the only parameter not giving rise to a significant effect; consequently, no differences were found between learning conditions or pairings in that regard (see Schiavio, Stupacher, et al., 2020 for details). Said differently, when participants learned melodies individually, imitating musical and motor patterns provided by an instructional video was—for temporal accuracy—no less efficient than learning in turn-taking or in synchrony with the video. This might help explain why, in the present study, the increase of Rating B (*How well do you think the performed melody was learned?*), was associated with the probability that the melody was learned without a musical partner: raters may have been particularly focused on the temporal cues of the heard melodies (see Repp, 1992, 1999), linking fewer temporal deviations in the target melodies to more accuracy and less coherence (see also Quinn & Watt, 2006).

We now discuss how, for musical excerpts learned via imitation, an increase in rating E (*Please evaluate the overall musical coherence of the performed melody*) was simultaneously associated with the probability that the melody belonged to the *duo* group. Why would a category, such as musical coherence be linked to duo learning, in the imitation condition only? Let us first clarify that previous literature may have treated coherence slightly differently from how we have conceived of such a notion for the present study. Consider, for example, the study by Geake (1997), which demonstrated that perception of a sequence of pitch fluctuations involves some direct awareness of its structural *coherence*. In that study, coherence was defined in terms of the autocorrelation function of the melodic sequence—that is, its level of redundancy.[5] Such an understanding of coherence based on melodic variability, however, is not consistent with how the term was employed in the present study: evaluating the melodic patterns' "fixed" structural

properties was not the main goal of our raters, who were instead invited to focus on the uniqueness of a given musical excerpt—one that is achieved through the consistent use of expressive, dynamic, or phrasing performative strategies. In that specific sense, therefore, coherence is predicated at a lower level of granularity when compared to expressiveness as well as articulation and phrasing. This arguably places coherence on a continuum with the other "general" quality defined by Rating B. As mentioned earlier, indeed, these two dimensions (coherence and how well the melodies were performed) might speak better than others to more general performative dimensions.

With this in mind, we suggest that the temporal signature for the solo group in the imitation condition highlighted above (i.e., the absence of a significant difference between learning conditions) might have indirectly led our raters to note those temporal deviations that are more consistently present in the melodies reproduced by the duo group, taking such temporal cues as indicative of more coherence. Now, recall that in the original study the imitation condition generally gave rise to more mistakes when compared to others. It is possible that, since quantitative analyses offered in the original study only looked at the number of mistakes but not at their quality, qualitatively similar mistakes committed by the same learner during one performance could become particularly significant for our raters, fostering a perceived sense of coherence in the reproduced melody.[6] Although this interpretation remains quite speculative, it still appears consistent with the possibility that such temporal fluctuations represent genuine errors on part of the novices, rather than intentional performative choices. If that were the case, our evaluators could have classified them as expressive nuances or as specific articulation or phrasing solutions. Instead, these temporal deviations remain subtle enough to modulate the perceived coherence of the piece, without, however, becoming essential factors with respect to other parameters.

The present work has some limitations: because our participants were music experts, we tended to assume that our definitions of "expressiveness," "articulation and phrasing," and "coherence" would be unequivocally understood, albeit with expected individual differences essential to critical appraisal of such subjective dimensions. Future work that wishes to examine whether our findings could be replicated with novice participants as raters might therefore ensure that clear working definitions are provided in advance to all participants. Also, the inclusion of different types of differences between performances makes it hard to interpret the cause of observed effects. Moreover, there may be some adjustment in terms of expectation of what constitutes expressiveness or coherence in the context of performances by novice musicians. It should also be noted that in the original study the tasks through which melodies were learned involved instructional videos. As such, researchers interested in this topic could complement our findings by systematically exploring (in both solo and duos situations) more in-person learning contexts, examining whether the presence of the teacher could give rise to different outcomes. Such a possibility resonates well with recent work by Meissner and Timmers (2020), which suggests that instructional strategies based on teacher-pupil dialogue might be particularly useful in fostering expressiveness across weekly instrumental music tuition. The key role of teachers in supporting musical development and flourishing is also well documented in literature exploring how instructors' gestures and bodily movements facilitate learning in different contexts (Simones, 2019; Zhukov, 2012).

In conclusion, while we could not find a corroboration of our main hypothesis, the present study yields intriguing results that could inspire future work to investigate in further detail the complex dynamics of individual and collective musical skill acquisition.

Funding

ORCID iDs

Andrea Schiavio https://orcid.org/0000-0001-8109-9185
Tudor Popescu https://orcid.org/0000-0001-5890-1520

Notes

1. In other studies, expert raters were shown to be more sensitive than non-expert to the perceived valence and arousal emerging from changes in mode and tempo (Ramos et al., 2011). For this reason, we used trained musicians as raters in this study.
2. Based on these evaluations, participants were classified into two cognitive listening styles: music systemizers (those who assess a performance in technical and formal terms) and music empathizers (those who assess the performance primarily in terms of its emotional content).
3. https://www.mathworks.com/matlabcentral/fileexchange/22099.
4. In statistics, an interchangeable sequence is a sequence of random variables $X_1, X_2, X_3, \ldots$ whose joint probability distribution does not change when the positions in the sequence are permuted.
5. "A pitch sequence with high autocorrelation sounds like a drone, with little pitch variation. A pitch sequence with low autocorrelation sounds like random notes. In music, pitch sequences usually have an intermediate value of an autocorrelation, i.e., music has a balance between predictability and surprise" (Geake, 1997, pp. 201–202).
6. For example, imagine how some learners may have had particular problems playing a melody that involved the use and repetition of simultaneous notes. Some, perhaps, had difficulty in playing three notes at the same time, thus delaying the pressure of the keys at each repetition, while others could instead systematically anticipate them. This may have contributed to creating performances which, while may be imprecise (and not very expressive, or elegant from the point of view of articulation or phrasing), they nevertheless remain fundamentally coherent.

References

Baroni, M. (2006). Hearing musical style: Cognitive and creative problems. In I. Deliège & G. Wiggins (Eds.), *Musical creativity: Multidisciplinary research in theory and practice* (pp. 78–94). Psychology Press.

Belkina, A. C., Ciccolella, C. O., Anno, R., Halpert, R., Spidlen, J., & Snyder-Cappione, J. E. (2019). Automated optimized parameters for T-distributed stochastic neighbour embedding improve visualization and analysis of large datasets. *Nature Communications*, *10*(1), 1–12.

Bishop, L., & Goebl, W. (2020). Negotiating a shared interpretation during piano duo performance. *Music & Science*, *3*, 1–18. https://doi.org/10.1177/2059204319896152

Borgo, D. (2007). Free Jazz in the classroom: An ecological approach to music education. *Jazz Perspectives*, *1*(1), 61–88.

Brandler, B. J., & Peynircioglu, Z. F. (2015). A comparison of the efficacy of individual and collaborative music learning in ensemble rehearsals. *Journal of Research in Music Education*, *63*(3), 281–297.

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition*, *63*, 59–69. https://doi.org/10.1016/j.bandc.2006.08.006

Burnard, P., Dillon, S., Rusinek, G., & Sæther, E. (2008). Inclusive pedagogies in music education: International comparisons of music teachers' perspectives from four countries. *International Journal of Music Education*, *26*(2), 109–126.

Ceh, S. M., Edelmann, C., Hofer, G., & Benedek, M. (2022). Assessing raters: What factors predict discernment in novice creativity raters? *The Journal of Creative Behavior*, *56*(1), 41–54.

Clarke, E. F. (1988). Expression and communication in musical performance. In J. Sundberg, L. Nord & R. Carlson (Eds.), *Music, language, speech, and brain* (pp. 184–193). Oxford University Press.

Cross, I. (2010). Listening as covert performance. *Journal of the Royal Musical Association*, *135*, 67–77. https://doi.org/10.1080/02690400903414848

de Leeuw, J. R. (2015). JsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Demos, A. P., Lisboa, T., & Chaffin, R. (2016). Flexibility of expressive timing in repeated musical performances. *Frontiers in Psychology*, *7*, Article 1490. https://doi.org/10.3389/fpsyg.2016.01490

De Poli, G. D., Canazza, S., Roda, A., & Schubert, E. (2014). The role of individual difference in judging expressiveness of computer-assisted music performances by experts. *ACM Transactions on Applied Perception*, *11*(4), 1–20.

Doğantan-Dack, M. (2012). "Phrasing—the very life of music": Performing the music and nineteenth-century performance theory. *Nineteenth-Century Music Review*, *9*(1), 7–30.

Doğantan-Dack, M. (2014). Philosophical reflections on expressive music performance. In F. Fabian, R. Timmers & E. Schubert (Eds.), *Expressiveness in music performance. Empirical approaches across styles and cultures* (pp. 3–21). Oxford University Press.

Elliott, D. J. (2005). Musical understanding, musical works and emotional expression: Implications for education. *Educational Philosophy and Theory*, *37*(1), 93–103.

Everitt, B. S., & Skrondal, A. (2010). *The Cambridge dictionary of statistics*. Cambridge University Press.

Fabian, D., Timmers, R., & Schubert, E. (Eds.). (2014). *Expressiveness in music performance. Empirical approaches across styles and cultures*. Oxford University Press.

Geake, J. (1997). Individual Differences in the Perception of Musical Coherence. *Journal of New Music Research*, *26*(3), 201–226.

Griffiths, M. (2017). Fostering the development of expressive performance skills: A gestural approach within the reflective, one-to-one piano studio. *Australian Journal of Music Education*, *51*(1), 52–62.

Hanken, I. M. (2016). Peer learning in specialist higher music education. *Arts and Humanities in Higher Education*, *15*(3–4), 364–375.

Héroux, I. (2016). Understanding the creative process in the shaping of an interpretation by expert musicians: Two case studies. *Musicae Scientiae*, *20*(3), 304–324.

Hinton, G., & Roweis, S. T. (2002). Stochastic neighbour embedding. *Neural Information Processing Systems*, *15*, 833–840.

Ibáñez de Aldecoa, P., de Wit, S., & Tebbich, S. (2021). Can habits impede creativity by inducing fixation? *Frontiers in Psychology*, *12*, Article 683024. https://doi.org/10.3389/fpsyg.2021.683024

Imberty, M. (1983). *Les écritures du temps. Sèmantique Psychologique de la Musique*. Bordas.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception*, *8*(2), 129–163.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Lussy, M. (1874). *Traité de l'expression musicale: Accents, nuances, et mouvements dans la musique vocal et instrumentale*. Berger-Levrault & Heugel.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30–46.

Meissner, H., & Timmers, R. (2020). Young musicians' learning of expressive performance: The importance of dialogic teaching and modeling. *Frontiers in Education*, *5*, Article 11. https://doi.org/10.3389/feduc.2020.00011

Miles, J. (2014). Tolerance and Variance Inflation Factor. *Wiley Statsref: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.stat06593

Moran, N. (2014). Social implications arise in embodied music cognition research which can counter musicological "individualism." *Frontiers in Psychology*, 5, Article 676. https://doi.org/10.3389/fpsyg.2014.00676

Nielsen, S. G., Johansen, G. G., & Jørgensen, H. (2018). Peer learning in instrumental practicing. *Frontiers in Psychology*, 9, Article 339. https://doi.org/10.3389/fpsyg.2018.00339

Nijs, L. (2019). Moving together while playing music: Promoting involvement through student-centred collaborative practices. In S. Gies & J. H. Sætre (Eds.), *Becoming musicians: Student involvement and teacher collaboration in higher music education* (pp. 239–260). The Norwegian Academy of Music.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, 26(5), 489–504. https://doi.org/10.1525/mp.2009.26.5.489

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception*, 14, 23–56.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48, 115–138.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The action simulation for auditory prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8, Article 57. https://doi.org/10.3389/fnsys.2014.00057

Pennill, N. (2019). *Ensembles working towards performance: Emerging coordination and interactions in self-organised groups* [Doctoral dissertation]. University of Sheffield.

Pennill, N., & Breslin, D. (2022). Music ensembles as self-organized groups. In R. Timmers, F. Bailes & H. Daffern (Eds.), *Together in music: Coordination, expression, participation* (pp. 3–12). Oxford University Press.

Quinn, S., & Watt, R. (2006). The perception of tempo in music. *Perception*, 35, 267–280.

Ramos, D., Bueno, J. L. O., & Bigand, E. (2011). Manipulating Greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians. *Brazilian Journal of Medical and Biological Research*, 44(2), 165–172.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, 88, 622–641.

Repp, B. H. (1992). Probing the cognitive representations of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, 44, 241–281.

Repp, B. H. (1999). Relationships between performance timing, perception of timing perturbations, and perceptual motor synchronization in two Chopin preludes. *Australian Journal of Psychology*, 51, 188–203.

Reybrouck, M. (2021). *Musical sense-making: Enaction, experience, and computation*. Routledge.

Robinson, C. F., & Kakela, P. J. (2006). Creating a space to learn: A classroom of fun, interaction, and trust. *College Teaching*, 54, 202–207. https://doi.org/10.3200/CTCH.54.1.202-207

Sawyer, R. K. (2003). *Group creativity: Music, theater, collaboration*. Lawrence Erlbaum.

Schiavio, A., Biasutti, M., van der Schyff, D., & Parncutt, R. (2020). A matter of presence: A qualitative study on teaching individual and collective music classes. *Musicae Scientiae*, 24(3), 356–376.

Schiavio, A., Moran, N., Antovic, M., & van der Schyff, D. (2022). Grounding creativity in music perception? A multidisciplinary conceptual analysis. *Music & Science*, 5. https://doi.org/10.1177/20592043221122949

Schiavio, A., & Nijs, L. (2022). Implementation of a remote instrumental music course focused on creativity, interaction, and bodily movement. Preliminary insights and thematic analysis. *Frontiers in Psychology*, 13, Article 899381. https://doi.org/10.3389/fpsyg.2022.899381

Schiavio, A., Stupacher, J., Parncutt, R., & Timmers, R. (2020). Learning music from each other: Synchronization, turn-taking, or imitation? *Music Perception*, 37(5), 403–422.

Schiavio, A., van der Schyff, D., Biasutti, M., Moran, N., & Parncutt, R. (2019). Instrumental technique, expressivity, and communication. A qualitative study on learning music in individual and collective settings. *Frontiers in Psychology*, 10, Article 737. https://doi.org/10.3389/fpsyg.2019.00737

Schubert, E., & Kreutz, G. (2014). Open ended descriptions of computer assisted interpretations of musical performance: An investigation of individual differences. In Roda, A., & Canazza, S. (Eds.), *Proceedings of the 1st international workshop on computer and robotic Systems for Automatic Music Performance (SAMP14)*, Sound and Music Computing Group, Department of Information Engineering, University

of Padova, Italy, Presented at Proceedings of the 1st international workshop on computer and robotic Systems for Automatic Music Performance (SAMP14), Venice, Italy, 18 July 2014 - 19 July 2014, http://samp.dei.unipd.it/proceedings/12_WSa1_07.pdf

Seashore, C. (1938). *Psychology of music*. McGrow-Hill.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Simones, L. (2019). Understanding the meaningfulness of vocal and instrumental music teachers' hand gestures through the teacher behavior and gesture framework. *Frontiers in Education*, *4*, Article 141. https://doi.org/10.3389/feduc.2019.00141

Storm, B. C., & Angello, G. (2010). Overcoming fixation: Creative problem solving and retrieval-induced forgetting. *Psychological Science*, *21*, 1263–1265.

Timmers, R. (2003). On the contextual appropriateness of expression. *Music Perception*, *20*, 225–240.

Timmers, R. (2005). Predicting the similarity between expressive performances of music from measurements of tempo and dynamics. *Journal of the Acoustical Society of America*, *117*, 391–399.

van der Schyff, D., Schiavio, A., & Elliott, D. (2022). *Musical bodies, musical minds: Enactive cognitive science and the meaning of human musicality*. MIT Press.

West, C. (2014). First-generation mixed methods designs in music education research: Establishing an initial schematic. *Bulletin of the Council for Research in Music Education*, *199*, 53–67.

Zbikowski, L. M. (1999). Musical coherence, motive, and categorization. *Music Perception*, *17*(1), 5–42.

Zhukov, K. (2012). Interpersonal interactions in instrumental lessons: Teacher/student verbal and non-verbal behaviours. *Psychology of Music*, *41*, 466–483.