journal of interdisciplinary music studies
season 2020, volume 10, art. #20101205, pp. 73-98

Score-dependency: Over-reliance on performing music from notation reduces aural pitch replication skills

Christopher Corcoran[1] & Neta Spiro[2]

[1] Faculty of Music, University of Cambridge
[2] Royal College of Music & Faculty of Music, University of Cambridge

Background in music performance. Most music is performed or improvised by ear, but Western classical musicians primarily perform music from notated scores.

Background in music perception. Classical musicians have greater difficulties playing melodies by ear than musicians with other backgrounds (Woody & Lehman, 2010). This ties in with Harris and de Jong's (2015) finding that non-improvising musicians exhibit less activation in the right auditory cortex than improvising musicians. The right auditory cortex is known to play a central role in pitch processing (Peretz & Zatorre, 2005).

Aims. (1) To investigate score-dependency (SD) in a behavioural study as a tendency for classical musicians to rely on notation over aural engagement in a music-learning scenario, and quantify SD levels for research. (2) To establish whether SD affects pitch perception. (3) To establish whether SD is a result of long term engagement in a score-focussed performance culture that precludes or minimizes participation in ear-playing scenarios.

Main contribution. Through a behavioural experiment we explore how score-dependency (SD) affects aural reproduction mechanisms, especially with regards to pitch. We introduce the SDR (score-dependency rating) measure for establishing musicians' individual levels of SD in relation to that of their peers. 20 notationally literate classical musicians were played a number of melodies and were asked to reproduce them on their instruments while simultaneously referring to provided music notation. By manipulating the degree of pitch and rhythm information shown in the music notation, we controlled the amount of pitch/rhythm information participants had to reproduce by ear alone. Counting how many times they needed to hear each melody again before task completion let us quantify their individual levels of dependence on notation. More score-dependent musicians showed a significant effect of struggling to reproduce pitch content—but not rhythm content—by ear. This effect was not found for non-dependent musicians. As pitch and rhythm are likely processed separately, SD may selectively affect aural pitch perception mechanisms, explaining reportedly limited activations in the right auditory cortex among SD musicians. Participants' age and years of music experience correlated with their SD levels, suggesting that long-term reliance on notation may increase this effect. Our results therefore indicate that SD may be an extreme form of overlearning, stemming from a long-term involvement in score-focussed performance practice without engaging in mitigating ear-playing scenarios. This posits SD as an effect of extreme task specialisation that creates a dependency on a specific technology (notation). We argue that this may limit a wider embodied engagement with music in favour of developing skill-specific cognitive mechanisms.

Implications. Our data suggests that a score-specific musical focus as often found in classical music education can endanger aural reproduction skills for pitch and associated perception/action mechanisms. Consequently, results have implications for music education and performance, as well as for cognitive and neuroscientific research into perception of music.

Keywords: Pitch perception, score-dependency, playing by ear, classical musicians

•*Correspondence:* Christopher Corcoran, Faculty of Music, University of Cambridge, CB3 9DP Cambridge, UK; E-mail: cc790@cam.ac.uk

•*Received:* 30 January 2020*; Revised:*9 June 2020*; Accepted:* 16 September 2020

•*Available online:* 29 January 2021

•doi: 10.25364/24.10:2020.1.5

One notable difference between Western classical music and other musical styles is the manner in which it is rehearsed and performed. While the vast majority of music has historically been played "by ear", i.e. learned and performed primarily on the basis of aural engagement with music (Liliestam, 1996), classical music is primarily read from musical scores and therefore learned and performed to a large degree on the basis of visual engagement with music. Although many other music cultures have their own forms of music notation, these are usually noticeably less prescriptive than scores of Western classical music and instead tend to fulfil a more mnemonic function (Bennett, 1983), providing visual aid in aural performance contexts (e.g. jazz charts, several forms of Asian tablature). It is unclear whether the majority of classical musicians receive early training that discourages existing ear-playing skills or whether many of them never learn such skills in the first place. However, aural replication is centrally important to early musical behaviour and expert performance in most music cultures, and it is a skill strongly reminiscent of other aural and physical imitation mechanisms rooted deeply in human behaviour, for example early language learning (Cross, 2013). Therefore, it is striking that aural replication plays little role in classical music, where it is largely replaced by notational literacy—especially since aural replication and transformation skills such as improvisation were previously part of classical performance culture, but disappeared over time (Moore, 1992). Consequently, classical music culture's shift away from learning music by aural imitation and instead learning by reading visual representations of music could be speculatively interpreted as moving from an embodied imitative understanding of music to a more abstract symbolic understanding. As Lilliestam (1996) writes:

> Unfortunately there are few discussions and analyses of how music and musical practice change when notation is introduced. Does the form of music and the way music is made change? Do note-reading musicians think about and conceptualise music differently than those who [don't] read and write music? Changes in these respects undoubtedly *do* appear, but the question is which changes and how do they come about? (198).

In the current behavioural study, we sought to identify whether strong reliance on music notation does indeed change music perception or action mechanisms associated with aural replication skills. We used the paradigm of "score-dependency"—defined as the tendency to rely on notation instead of aural replication mechanisms—and used it to test how score-dependent musicians may "think about and conceptualise music differently", as Lilliestam (1996) puts it.

Notational literacy and aural discrimination skills

Due to their score-focussed performance practice, classical musicians rely heavily on notational literacy and sight-reading abilities. Notational literacy is considered a psychomotor decoding skill, which allows for separately extracting pitch and rhythm information from written instructions and realising them with appropriate motor skills (Gudmundsdottir, 2010). Sight-reading is a particularly specialised form of notational literacy, as it requires executing notated instructions during a performance in real time, placing additional demands on performers' speed of information processing and psychomotor speed (Kopiez & Lee, 2006). In many areas of professional classical performance practice (e.g. orchestral rehearsals, film music recordings, piano accompaniment), sight-reading is an essential skill.

Likely as a result of these divergent modes of performance, studies have found several perceptual differences between classical performers and those of other styles. Musicians from an aural performance background more easily identify melodic contour changes (Tervaniemi et al., 2001; Seppänen, Brattico & Tervaniemi, 2007) as well as harmonically deviant notes and pitch slides, while classical musicians are stronger at identifying timbre (Vuust et al., 2012). Aural and classical musicians seem to conceptualise harmonic structures differently (Goldman, Jackson & Sajda, 2018), responding with different latency times to surprising harmonies (Bianco, et al., 2018).

Classical musicians and aural musicians also seem to engage with music differently in an aural reproduction context, which makes sense given that classical musicians' performance culture is less rooted in aural learning. Exploring the effects of genre experience on playing by ear, Woody and Lehman (2010) played classical and "vernacular" musicians (notationally literature music majors at university with background in jazz/popular music) several melodies and asked them to reproduce these melodies on their instruments. They found that classical musicians needed to hear the melodies significantly (nearly four times) more often than the vernacular musicians before they could reproduce them correctly. Notably, while half of the vernacular musicians described the melodies as predictable or typical, five of twelve (>40%) of the classical musicians experienced them as unpredictable or difficult to memorise. Three quarters of the classical musicians later reported having had to actively consider fingerings during the experiment, while a single vernacular musician reported this. This led the authors to speculate that the classical musicians may have experienced a "cognitive bottleneck" when trying to play by ear, inhibiting them from transforming perceived pitch information into motor action at the stage of motor goal imaging. While this is possible, research with violinists found that melodic goal imaging, though significant, is outweighed by hours of instrumental performance as a predicting factor for successful aural replication (Hakim & Bullerjahn, 2018). This implies that other domain-related mechanisms may be at work.

Score-dependency may affect aural perception skills

Instead of representing a problem with goal imaging, comparatively weaker aural reproduction skills among score-playing musicians could also indicate an inhibition in their aural perception of certain musical structures. Harris and de Jong (2015) proposed that many classical musicians are affected by score-dependency (from hereon: SD), defined as the inability to learn music without notation. Dividing their sample of classical keyboard players absolutely into improvisation-capable and "score-dependent", Harris and de Jong ran a functional resonance imaging (fMRI) scan while asking musicians both to imagine playing along and to critically assess familiar and unfamiliar music recordings. The improvisers exhibited significantly stronger activations in the bilateral auditory cortex during these tasks, with score-dependent musicians exhibiting no significantly stronger activations in the right auditory cortex than musically unskilled control subjects. Specifically, the right auditory cortex plays an important role in pitch perception and encoding, while rhythm is likely processed bilaterally and across more widely distributed regions (see Peretz & Zatorre, 2005 for an overview). As a result, it is possible that SD specifically affects mechanisms for pitch perception. This insight motivated the behavioural study presented here on score-dependency's effect on aural pitch and rhythm reproduction.

Harris and de Jong (2015) further found that both types of musicians shared several activations typically associated with pitch-space rotations among keyboard performers (see also Stewart et al., 2003). However, only improvising musicians displayed additional right-hemisphere activations (incl. posterior-superior parietal areas and dorsal premotor cortex), which the authors suggest may indicate that only improvisers perform additional audio-spatial transformations when hearing music. This would mean that score-dependent musicians (from hereon: SDMs—to be contrasted with score-independent musicians, SIMs) struggle to apply top-down benefits—in this case aural experiences gleaned from previous musical activity—when asked to learn or appraise music by ear alone. That would make SD an example of extreme specialisation of literacy and overlearning, supporting only select cognitive mechanisms to the detriment of other, more widely distributed mechanisms.

That is not to say that learning to read music notation is in itself detrimental to aural skills. In fact, facility in sight-reading—as a heightened form of fluent notational literacy—is often dependent on building up inner auditory representations of music and visually matching them with musical information displayed in notation (Waters, Townsend & Underwood, 1998), and has often been correlated with improvisation (Kopiez & Lee, 2006; Lehmann & Ericsson, 1993 & 1996; Lehmann & Kopiez, 2016; McPherson, Bailey & Sinclair, 1997; Mishra, 2014). However, statistically, improvising skills alone cannot predict sight-reading skills sufficiently (Lehmann & Ericsson, 1996) and so this link is perhaps better explained by ear-playing skills as a hidden third factor, as these have been shown to improve both improvisation and sight-reading skills (see Musco, 2010, for an overview).

In fact, research shows that sight-reading may be simply the result of long-term domain-related activities, reflecting experience in notational literacy rather than innate musical talent or even a special talent for sight-reading (Gudmundsdottir 2010; Lehmann & Ericsson, 1996; Mishra, 2014; Waters, Townsend & Underwood, 1998; Wristen, 2005). While it may interact with other musical skills, in and of itself it is likely a learnable psychomotor decoding skill that in the absence of other factors neither contributes to nor damages aural perception skills. This would explain why sight-reading is a skill found in both many improvising musicians (e.g. many professional jazz players) as well as SDMs.

If sight-reading is merely a challenging psychomotor decoding skill, this would suggest that the limited activations observed in SDMs may be the result of a near-exclusive notation-focussed background, which is common in classical music education contexts (Feichas, 2010). In extreme cases, it has been suggested, a notation-focussed music education that does not sufficiently take aural training into account may blind musicians to musical details that are not represented in the score and therefore strengthen an eye-hand connection that allows performers to play notes from notation without internalising them first (Mills & McPherson, 2006). Mills and McPherson suggest that such an eye-hand connection can cause musicians to execute pitch instructions without internalising them, but that rhythm instructions cannot be executed without first forming an inner representation of them. This suggests that classical musicians may find pitch more difficult to reproduce aurally than rhythm, since they possibly internalise pitch instructions less. Unlike improvising musicians, classical musicians are discouraged from deviating from notated pitch sequences and instead express themselves by microrhythmic and articulation-based deviations from the score (Goldman, 2016). Consequently, a score-dependent performance practice in which aural replication is not trained could act as a behavioural feedback loop, in which reliance on notation leads to ever increasing fluency in notational literacy, in turn further encouraging reliance on notation.

Score-dependency as a tendency rather than absolute

However, extreme overspecialisation on reading from notation should not be assumed as universal among classical musicians. Many musicians participate in a variety of musical experiences and play music in multiple styles, and since musicians tend to rely both on aural and non-aural cues in performance, it can be assumed that the vast majority of classical musicians do not categorically shut out aural information during musical performances. In fact, the inverse is just as true: Many improvising musicians use visual representations of music (e.g. jazz lead sheets, drummers' charts, or guitarists' chord charts) to aid them in their learning or performance process.

This means that most musicians are likely open to both aural cues and visual representations when performing music, with some relying more on the former and other more on the latter. Therefore, it would be a conceptual fallacy to conceive of SD as one half of two binary *absolutes* (dependent or non-dependent) as suggested by Harris and de Jong's division of their sample into improvisers and non-improvisers (2015), since likely very few musicians are either fully score-dependent or independent. Instead, defining SD as a *tendency* to rely more strongly on musical notation than on aural replication skills is more likely to reflect the practical reality. As a result, it may be useful to develop a measure for SD that takes this tendentious nature into account.

Here we describe an experiment that tests the degree to which classical musicians rely on the aid of notated materials over aural information during a music reproduction task, quantifying the proposed tendentious nature of SD with the newly established measure SDR (score-dependency rating). The basic experimental setup was based on Woody and Lehmann's (2010) study. Just as in that experiment, musicians here were played a melody twice, were asked to reproduce it on their instrument, and were assessed in how many more times they needed to hear a melody before reproducing it correctly. However, unlike in Woody and Lehman's earlier experiment, we included five conditions of visual support in order to establish a measure for score-dependency, using music notation to display the melodies' information content to varying degrees of completeness. The musicians relied on these representations as well as on aural discrimination while attempting to reproduce the melodies. Quantifying the musicians' task-performance across the five conditions formed an empirical basis for determining how much they relied on visual over aural information in their learning process and indicated their level of dependency on notation. By weighting task performance in the different conditions based on their difficulty, it was possible to create a nuanced estimation of every musician's individual level of SD expressed by the SDR measure as a rating on a scale indicating low to high dependency, which in turn can be operationalized in other experiments.

Given that SD may affect classical musicians' activations in the right auditory cortex, which is associated with pitch perception, we also wished to investigate what behavioural consequences this might have in the domain of aural reproduction. Consequently, here we also investigate behaviourally to what degree relatively SDMs and SIMs differed in their ability to reproduce pitch and rhythm by ear. We used the visual aids to control for how much rhythm/pitch information was displayed on the page and consequently how much missing pitch/rhythm information participants had to reproduce by ear alone.

This procedure was used to investigate three hypotheses:

1. Based on their enculturation in a notation-focussed performance culture, the majority of classical musicians are relatively score-dependent, meaning that classical musicians overall are more score-dependent than score-*in*dependent.
2. SDMs find pitch more difficult to reproduce by ear than rhythm, given their limited activations in the right auditory cortex, but this effect depends on the amount of information to determine (i.e. *how much* pitch and rhythm information had to be reproduced aurally would be a greater factor for task difficulty than *which* element had to be reproduced).
3. SD is a result of long-term engagement in a notation-focussed performance practice, but it can be mitigated by participating in musical activities that engender playing by ear.

Methods

Participants

We recruited professional classical instrumentalists (N=20) living in Germany, Denmark, the UK, and Ireland (mean age: 39 years, std. dev.: 12.06, range: 23-75; 9 female) using a mixture of advertisements on social media and snowball sampling among professional classical musicians. "Professional" is here defined as working or studying full time as a performer, i.e. primarily earning a living from performing or teaching as instrumentalist or studying as an actively performing musician at a conservatory. Among participants, there were ten keyboard (8 x piano, 1 x organ, 1 x accordion), eight string (4 x cello, 2 x violas, 2 x violins), and two wind instruments (1 x recorder, 1 x saxophone in E flat). Singers were deliberately not recruited, since different cognitive mechanisms may be involved in externalising musical intentions to an instrument compared to expressing them through a physical feature of one's body (Gudmundsdottir 2010; Fine, Berry, and Rosner 2006). All participants were screened for strong notational literacy by providing them with an excerpt of a complex score and asked if they could sight-read it. No participants had to be excluded based on this screening.

Two additional participants were recruited for a pilot study in order to test and adjust methods, after which their data was discarded. In a preliminary effort to explore whether musical expertise in the absence of a professional context could be assessed similarly, three experienced amateur instrumentalists (i.e. strong instrumentalists who did not work or study as musicians full time) were also recruited. However, their results differed noticeably from those of professional instrumentalists, and so their data was discarded. Due to the limited sample (n=20) any conclusions have to be treated as indicative rather than as absolute, even at high significance levels.

Materials

Before the experimental session, participants were provided with an online questionnaire using the Google Forms (Google LLC, 2008) questionnaire platform, asking them to provide free-text answers to biographical questions and self-ratings for musical skills and experience levels.

Eight monophonic melodies (seven for the experiment, one for task practice; c. 4 bars in length, diatonic, in 4/4, 100-120 BPM) were composed by the first author for the experiment and fully notated in the notation software Finale 25 (MakeMusic, 2016). Melodies were kept monophonic in anticipation of recruiting participants whose main instrument might not be capable of polyphony. Although performers of polyphonic instruments may develop additional skills to deal with polyphony, this should not invalidate results by performers of monophonic instruments (Sloboda, 1977). A MIDI performance for each melody was exported in the WAV file format using the software's Steinway piano sound and was then converted to the MP3 format. Each melody was deliberately designed to include specific difficulties for aural reproduction, e.g. an unusual interval skip or a rhythmic irregularity. All melodies were checked by a senior ear-training professor working at a national conservatoire, who made suggestions for equalising difficulty levels between the melodies, all of which were implemented. Five notational representations were created for each melody, one each for the five experimental conditions numbered C1 through C5 (see Figure 1 for an example).

Figure 1. Example of Visual Representations of a Melody for Experimental Conditions C1-C5

Note: C1 = blank bars (no pitch or rhythm content); C2 = rhythmic outline only; C3 = melodic outline only; C4 = full rhythm only; C5 = full pitches only; Original = melody as originally scored.

In order to assess our hypothesis regarding SD affecting pitch reproduction, we used these visual aids to control for the degree of rhythm/pitch information that participants had to reproduce by ear alone in each condition. Since all our participants were proficient score-readers, only the information absent from the notation had to be reproduced aurally, meaning that we could use task performance in each condition to infer how difficult it was for participants to reproduce any non-displayed pitch/rhythm information by ear.

Conditions were labelled C1-C5 in anticipation of task difficulty levels, assuming that participants would struggle most with C1 and struggle least with C5. This ordered labelling was based on the information content provided in each condition: C5 and C4 provided full transcriptions of pitch and rhythm respectively, while C3 and C2 provided merely partial outlines of each, with C1 providing no pitch/rhythm information. The ordering therefore reflected two assumptions: (1) that pitch would be more difficult for participants to reproduce aurally than rhythm; but also (2) that the pitch-rhythm hierarchy would be outweighed in task difficulty by the amount of missing information content to reproduce, i.e. that it would still be easier to aurally reproduce all pitch information (C4) than some pitch and all rhythm information together (C3).

The visual aids were presented in the G-clef at natural pitch for all non-transposing instruments traditionally playing from this clef. The materials were transposed as relevant for transposing instruments, ensuring that the aurally provided pitch information would match the information presented in the notation at relative pitch. Octave-transposed versions showing the melody in the bass clef were created for bass instruments.

Procedure

Each participant was randomly allocated five of the seven melodies[1] created for the experiment and each melody was randomly allocated one of the five notational conditions, such that all conditions were represented for each participant.[2] In advance of the experiment, every participant was emailed a link to the questionnaire and asked to complete it before the experimental session.

Each experimental session was conducted with mobile equipment in a participant's practice environment, with the aim of helping performers feel more comfortable than in a laboratory setting and ensuring a naturalistic environment. Times and settings were agreed in advance on the premise of ensuring an uninterrupted quiet environment suitable for concentration and recording. Participants were told that their identity would remain anonymous and signed the consent forms required for experiments and interviews by the Faculty of Music, University of Cambridge. The experimental process was explained and a summary of the procedure was also provided in writing in a Participant Information Sheet. However, in order to prevent bias, participants did not receive a full explanation of the study's aims until after completion of the experimental session. A practice run of the task was conducted with the practice melody to ensure that participants had understood and were comfortable with the procedures; this was also used to establish suitable volume levels for playback and recording.

Each condition followed the same procedure: A participant was handed the randomly allocated notation for a melody and was then played the MIDI-piano recording of the melody twice. Then she or he was told the melody's first pitch (transposed for transposing instruments) and asked to reproduce the melody in pitch and rhythm in a suitable range on their instrument, using the visual representation as needed.

---

[1] After testing ten of twenty participants, the data was reviewed and the two melodies that led to best/worst task performance were determined as potentially too easy or difficult. For the remaining experimental sessions, these melodies were replaced at random with the two remaining melodies in the pool.

[2] The list function on the website random.org [accessed between 01 March and 30 June 2019] or more often its mobile app for the Android platform, "Certified True Randomisers" version 1.2.12, were used for all randomisation processes. Random.org claims to make use of atmospheric noise in order to produce orderings that are more random than those based on computer-generated algorithms.

In addition to hearing two playbacks of each melody as part of the experimental setup, participants could request to hear the playback again if necessary, but only up to a maximum of five requested playbacks (from hereon: RPs) in order to avoid task fatigue. The number of RPs required for successful task completion was noted. The number of playing attempts was not noted. If correct reproduction in pitch and melody was not achieved after five RPs (i.e. seven playbacks, including those two provided as part of the setup), the final number of RPs was noted as 6 to indicate that the participant would require at least six RPs (and possibly many more). If a participant was certain she or he could not achieve task completion within five RPs, they could skip the task after a few attempts in order to avoid unnecessary discomfort and associated task fatigue; in this case the number of RPs for this condition was also noted as 6 to indicate that the participant would require to hear the melody six or more times. All experimental sessions were audio-recorded with permission of each participant in order to verify the noted number of RPs per condition.

The audio was recorded and played back on an Apple MacBook Pro 2011 laptop, AKG K451 headphones, a Sennheiser E835 microphone mounted on a portable stand, and a Behringer U-Phoria UMC404HD audio interface. The sessions were recorded using the software Logic Pro 9 (Apple Inc., 2004-2012). Each participant was free to choose whether they heard the audio stimuli via headphones or laptop speakers, but good listening conditions were ensured for each. The same researcher was present for all experimental sessions. Ethical approval was granted by the University of Cambridge Music Department's Ethics Board on 17 May 2018.

Results

In this section, results will first be elaborated for the entire sample, then for each subsample (relatively score-dependent or -*in*dependent) in order to demonstrate specific effects of SD. All statistics were computed in Microsoft Excel for Mac 2011 or StataIC 15.1 for Mac.

General differences between pitch and rhythm reproduction

The differences between the five visual aid conditions overall were found to be highly significant (Friedman = 51.9200, Kendall = 0.5465, $p<0.001$).[3] We expected that our labels for conditions (C1-C5, expressing conditions' anticipated difficulty levels), would match actual difficulty levels as expressed by ordering conditions by mean RPs. The data mainly matched expectations but with one important difference: As Table 1 shows, an ordered ranking from fewest to most RPs (lowest to highest difficulty) was not C5-C1, but rather C5, *C3, C4*, C2, C1.

[3] Each condition's data was analysed for normality using the Shapiro-Wilkes test, and since assumptions of normality were not confirmed for all datasets, non-parametric tests were employed going forward.

Table 1. Conditions Sorted by Mean Number of Requested Playbacks (RPs)

Condition	Mean RPs	Standard Deviation	Minimum	Maximum
C5	1.55	1.64	0	6
C3	2.6	1.88	0	6
C4	3.15	1.98	1	6
C2	3.8	2.09	1	6
C1	4.2	1.91	1	6

Consequently, aurally reproducing pitch *always* required more RPs than rhythm, regardless of how much rhythm or pitch content had to be reproduced, i.e. even when only partial pitch but all rhythm content had to be reproduced aurally. Wilcoxon matched-pairs signed-rank tests were used to compare all conditions pairwise (see Table 2).

Table 2. Comparisons of Results per Condition using Wilcoxon Matched-Pairs Signed-Rank Tests

Conditions	C1	C2	C3	C4	C5
C1	– –				
C2	z = 1.02 p = 0.3080	– –			
C3	z = 2.92 p = 0.0035	z = 2.69 p = 0.0071	– –		
C4	z = 2.27 p = 0.0234	z = 2.31 P = 0.0209	z = -1.85 p = 0.0644	– –	
C5	z = 3.48 p = 0.0005	z = 3.15 p = 0.0016	z = 2.21 p = 0.0268	z = 3.09 p = 0.002	– –

As Table 2 shows, two pairs of conditions were not significantly different from one another: C1 and C2 (p=0.31) and C3 and C4 (p=0.0644). This information modifies the order of difficulty as suggested by Table 1. Table 3 summarises all relevant information and presents the modified ranking.

Table 3. Ranked Conditions for All Participants

<table>
<tr><td>Difficulty ranking (low to high)</td><td>1</td><td colspan="3">2 & 3
(shared)</td><td colspan="2">4 & 5
(shared)</td></tr>
<tr><td>Mean RPs (rounded) low to high</td><td>1.55</td><td colspan="3">2.6 / 3.15</td><td colspan="2">3.8 / 4.2</td></tr>
<tr><td>Condition</td><td>C5</td><td colspan="3">C3 / C4</td><td colspan="2">C2 / C1</td></tr>
<tr><td>Displayed information</td><td>Full pitch</td><td colspan="3">Melodic outline /
Full rhythm</td><td colspan="2">Rhythmic outline /
Blank bars</td></tr>
<tr><td>Information to be identified aurally</td><td>Full rhythm</td><td colspan="3">Partial pitch, full rhythm /
Full pitch</td><td colspan="2">Partial rhythm, full pitch / Full rhythm, full pitch</td></tr>
<tr><td>Wilcoxon
p-value (rounded)</td><td colspan="2">C5 & C3 p=0.0268
C5 & C4 p=0.0020</td><td>C3 & C4 p=0.0644</td><td colspan="2">C3 & C2 p=0.0071
C3 & C1 p=0.0035
C4 & C2 p=0.0209
C4 & C1 p=0.0234</td><td>C2 & C1 p=0.3080</td></tr>
</table>

Note: Conditions are ranked by difficulty from low to high, as indicated by mean requested playbacks (RPs) and grouped according to significance levels resulting from Wilcoxon matched-pairs signed-rank tests. The table further shows what pitch or rhythm information was displayed in notation for each condition, but also which information was purposefully not displayed and instead had to be identified aurally by participants.

As summarised in Table 3, even after taking the Wilcoxon tests into account in order to account for differences between conditions, conditions where pitch information had to be reproduced aurally were generally ranked higher. This suggests that participants found the aural reproduction of pitch more difficult than the aural reproduction of rhythm.

Assessing levels of score-dependency

In order to explore how this effect may differ between relatively SDMs and SIMs, Table 3's results were used to determine each participant's level of SD by producing a newly developed score-dependency rating (from hereon: SDR). The SDR represents each participant's tendency to rely on visual over aural information in relation to the entire sample's performance as a group. Since we assumed SD to exert a tendentious rather than an absolute effect, the SDR formula is designed to create a nuanced evaluation of each musician set against the performance of her or his peers. It does this by using the difficulty ranks established in Table 3 as weightings, thereby taking the significance of differences between conditions into account.

For the formula, each participant's RP number per condition was multiplied by that condition's difficulty rank, then results were added together and divided by 15 to produce a result on a scale of 0-6, then elevated to a scale of 1-7 by adding 1. Since the Wilcoxon tests had shown that conditions C2 and C1 are not significantly different from each other and consequently shared difficulty ranks 4 and 5, they were each assigned the factor 4.5 for weighting. Since conditions C3 and C4 also were not significantly different from each other and therefore shared ranks 2 and 3, they were assigned the weighting factor 2.5. The resulting formula for each participant was:

$$SDR = (RP_{C5}*1 + RP_{C3}*2.5 + RP_{C4}*2.5 + RP_{C2}*4.5 + RP_{C1}*4.5)/15 + 1$$

As a result of this calculation, the extremes of the resulting scale represent the following:

- 1 = no RPs needed in any condition, so labelled as "very score-independent"
- 7 = at least 6 RPs needed in every condition, so labelled as "very score-dependent".

The use of difficulty ranks as weightings in the SDR formula meant that RPs had a different effect on a participant's score in different conditions: An RP in condition C5 elevated a participant's final SDR by 0.067 on the 1–7 scale, while an RP in C1 or C2 elevated the SDR by 0.3, counting almost four times as much. Therefore, participants would gain a higher final SDR if they requested more playbacks in more difficult conditions. This was a deliberate part of the formula's design, since it accounted for participants' differing abilities: Those who required few playbacks in conditions with little notational support showed they could reproduce any non-displayed information by ear quite quickly, logically making them less score-dependent and therefore deserving of a lower final SDR; in contrast, those who requested many playbacks in conditions with little notational support showed that they struggled more to reproduce the music by ear, logically making them more score-dependent and so deserving of a higher SDR. By assessing participants based on difficulty levels established by their own performance as a group, we could assess their individual tendency to rely on notation over aural faculties in contrast to the performance of their peers. Table 4 shows SDRs for all participants and Figure 2 shows how the SDRs are distributed.

Table 4. Participants Ranked by Score-Dependency Rating (SDR)

Participant rank	**C1**	**C2**	**C3**	**C4**	**C5**	**SDR**
1	1	2	0	1	0	2.07
2	2	1	1	1	3	2.43
3	3	1	1	1	2	2.67
4	1	2	3	2	1	2.77
5	3	1	1	2	1	2.8
6	4	2	1	1	1	3.2
7	1	4	4	1	0	3.33
8	6	2	1	1	0	3.67
9	3	1	2	6	2	3.73
10	5	3	2	2	0	4.07
11	3	5	2	4	4	4.67
12	5	5	2	3	0	4.83
13	5	6	1	4	0	5.13
14	6	5	2	4	2	5.43
15	6	6	3	3	1	5.67
16	6	6	3	3	2	5.73
17	6	6	5	6	1	6.5
18	6	6	6	6	1	6.67
19	6	6	6	6	4	6.87
20	6	6	6	6	6	7
Mean	**4.20**	**3.80**	**2.60**	**3.15**	**1.55**	**4.46**
						Range: 2.07–7.0

Note: Variables C1-C5 show participants' number of requested playbacks (RPs) for task completion in each condition (min. 0; max. 6, indicating 6 or more RPs required for task completion). SDR shows score-dependency ratings on a scale of 1–7. The data is sorted by SDR from low to high and participants are ranked accordingly.

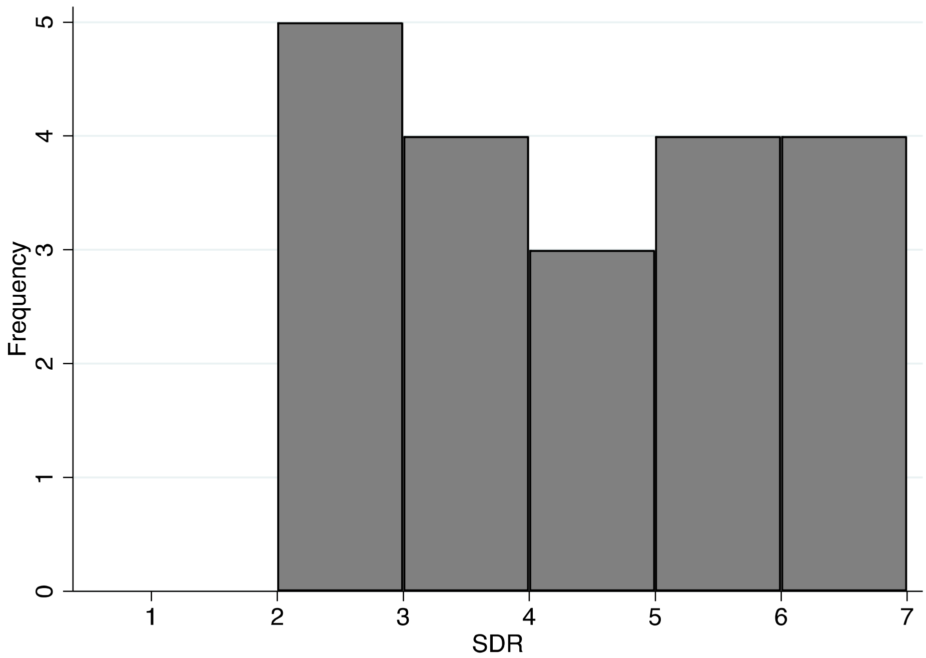

Figure 2. Distribution of Participants by Score-Dependency Ratings (SDRs)

Note: SDR scale of 1–7, with bins set to 1. SDRs are shown on the x-axis and numbers of participants on the y-axis. No participants scored in the range 1–2, so that bin is empty.

Hypothesis 1 stated that classical musicians overall are more score-dependent than score-independent. As seen in Table 4, which provides the distribution of participants along the SDR scale, the sample's mean SDR was 4.46 (range: 2.07–7, std. dev.: 1.6, SE: 0.36) with a 95% confidence interval estimated at [3.71–5.21]. Adjusting the confidence interval levels showed that a population mean equal to or greater than the scale median 4 could only be predicted with 79% confidence, which is too low for scientific confidence. Consequently, a population of classical musicians cannot with sufficient confidence be estimated to be more score-dependent than -independent according to the parameters of the SDR scale. Therefore, Hypothesis 1 could not be confirmed conclusively. However, as seen in Figure 2, no participant scored between 1 and 2 on the SDR scale. This means that no one approached the lower extreme "very score-independent" (which represents correct aural reproduction of pitch and rhythm information without any RPs, so after only hearing a melody twice as part of the experimental setup).

Score-dependency's effect on rhythm and pitch reproduction

To test the second hypothesis—that SDMs find pitch more difficult to reproduce by ear than rhythm—we used the SDR scale. The SDR variable expresses each participant's tendency to rely on notation over aural information and so we divided the participants into two subsample groups:

- Those who scored in the *lower* half of the scale (SDR <= 3.99; n = 9) and so were deemed relatively score-*independent* (and therefore are referred to as SIMs).
- Those who scored in the *upper* half of the SDR scale (SDR >= 4; n = 11), and so were deemed relatively score-*dependent* (and therefore are referred to as SDMs).

As for the entire sample earlier, we ran Friedman and Wilcoxon matched-pairs signed-rank tests for each subsample group separately. For SIMs there was no significant difference between conditions overall (Friedman = 4.6800; Kendall = 0.1170; p=0.7912). This indicates that their superior ear-playing skills allowed them to perform equally well across conditions overall, regardless of any notation shown to them. Only the extreme conditions C1 & C5 were significantly different from each other (p=0.0269). Table 5 shows the resulting ordered difficulty ranking for SIMs.

Table 5. Conditions Ranked in Difficulty for SIMs (SDR = 1.0–3.99)

<table>
<tr><td>Difficulty ranking (low to high)</td><td colspan="2">1-4</td><td colspan="2">1-5</td><td colspan="2">2-5</td></tr>
<tr><td>Mean RPs (rounded) low to high</td><td colspan="2">1.11</td><td colspan="2">1.56 / 1.78 / 1.78</td><td colspan="2">2.67</td></tr>
<tr><td>Condition</td><td colspan="2">C5</td><td colspan="2">C3 / C4 / C2</td><td colspan="2">C1</td></tr>
<tr><td>Displayed information</td><td colspan="2">Full pitch</td><td colspan="2">Melodic outline
/ Full rhythm
/ Rhythmic outline</td><td colspan="2">Blank bars</td></tr>
<tr><td>Information to be identified aurally</td><td colspan="2">Full rhythm</td><td colspan="2">Partial pitch, full rhythm / Full pitch / Partial rhythm, full pitch</td><td colspan="2">Full pitch & full rhythm</td></tr>
<tr><td>Wilcoxon p-value</td><td>C1 & C5, p=0.0296</td><td colspan="2">C1-C4
= no significant differences for any pairs</td><td colspan="2">C2-C5
= no significant differences for any pairs</td><td>C1 & C5, p=0.0296</td></tr>
</table>

Note: Only C1 and C5 were significantly different from each other, resulting in rankings spread over multiple conditions. Conditions are ranked by difficulty from low to high, as indicated by mean requested playbacks (RPs) and grouped according to significance levels resulting from Wilcoxon matched-pairs signed-rank tests. The table also shows what pitch or rhythm information was purposefully not displayed in each condition and instead had to be identified aurally by participants.

Since only C1 and C5 could be ranked for SIMs, we can only say with confidence that SIMs found it easier to aurally reproduce *rhythm alone* than *pitch and rhythm together*. Since this might be a result of amount of information to reproduce (i.e. having to reproduce full information for both factors rather than just for one of them), and as differences between conditions overall are not significant, there are no clear indications for SIMs finding pitch more difficult to reproduce aurally than rhythm.

The results for SDMs stand in contrast to those for SIMs. Differences between conditions overall were highly significant (Friedman = 26.5909; Kendall = 0.5318; p=0.003) and the Wilcoxon tests showed that all conditions were significantly different from each other, regardless of ranking, except for conditions C2 and C1 (see Table 6).

Table 6. Conditions Ranked in Difficulty for SDMs (SDR = 4.0-7.0).

Difficulty ranking (low to high)	1	2	3	4 & 5 (shared)
Mean RPs (rounded) low to high	1.91	3.45	4.27	5.45 / 5.45
Condition	C5	C3	C4	C2 / C1
Displayed information	Full pitch	Melodic outline	Full rhythm	Rhythmic outline / Blank bars
Information to be identified aurally	Full rhythm	Partial pitch, full rhythm	Full pitch	Partial rhythm, full pitch / Full pitch, full rhythm
Wilcoxon p-value	C5 & C3, p=0.0269	C3 & C4, p=0.027	C4 & C2, p=0.01; C4 & C1, p= 0.0385; C2 & C1 = p=1.0 (same mean RPs)	

Note: Conditions are ranked by difficulty from low to high, as indicated by mean requested playbacks (RPs) and grouped according to significance levels resulting from Wilcoxon matched-pairs signed-rank tests. The table also shows what pitch or rhythm information was purposefully not displayed in each condition and instead had to be identified aurally by participants.

As Table 6 shows, it is noticeable that task difficulty for SDMs increased with the pitch content to be aurally reproduced: *No pitch – partial pitch (with full rhythm) – full pitch – full pitch with partial/full rhythm*. In Hypothesis 2, we predicted that SDMs would find pitch more difficult to reproduce by ear than rhythm. However, we added the caveat that where elements of both pitch and rhythm had to be reproduced aurally, the factor with the greatest information load to reproduce would control difficulty levels. The ranking in Table 6 shows that SDMs found reproducing pitch more difficult than reproducing rhythm *regardless* of information load: For them reproducing a melody was more difficult when having to aurally identify full pitch without any rhythm information (C4) than when having to identify partial pitch and full rhythm content (C3). Rhythm content appeared to add to difficulty only when full pitch content already had to be identified (C2/C1 compared to C4). As a result, the results confirm our Hypothesis 2—so much so that our assumed caveat is shown as erroneous. In summary, SDMs found it significantly more challenging to reproduce pitch than rhythm content. This strongly supports our hypothesis that SD affects perception or action mechanisms involved in aural pitch replication.

Correlations for SD and biographical factors

To test our third hypothesis—that SD results from long-term engagement in a notation-focussed performance practice, but that it can be mitigated by participating in musical activities that engender playing by ear—we used the questionnaire answers about musical and biographical background. Text answers were converted to numerical values and correlated with participants' SDRs. Since several resulting variables were not normally distributed, non-parametric Spearman correlations (or, in the case of categorical variables, Kruskal-Wallace tests) were computed. Significant correlations are shown in Table 7.

Table 7. Biographical Factors Significantly Correlated with SDR

Variable	Spearman correlation	Significance
Age	r_s= 0.63	p = 0.003
Years of playing music	r_s= 0.58	p = 0.0071
Experience performing jazz	r_s= -0.46	p = 0.0396
Absolute pitch	r_s= -0.45	p = 0.0491

Note: The table shows the results of Spearman's correlation between SDR and participants' self-ratings on musical and biographical factors. For all correlations N=20.

Both the variables "age" (mean: 38.85, std. dev.: 12.106, range: 23–75) and "years of playing music" (mean 32.55, std. dev.: 11.11886, range: 15-65) produced moderate-to-strong positive correlations with SDR. This may possibly indicate that long-term involvement in notation-focussed practice is associated with SD. "Experience performing jazz" (mean 2.3, std. dev.: 1.174286, range: 1–5 on scale of 1–7) produced a moderate negative correlation, which is likely due to jazz being a genre that is traditionally played by ear. Similarly, "absolute pitch" (ordinal variable showing 13 without, 7 with) produced a moderate negative correlation, likely as the skill of absolute pitch facilitates pitch identification. "Experience improvising" merely approached significance (p=0.0632). "Primary instrument category" (coded as Bowed Strings, Winds, or Keyboards), "secondary instrument category" (coded as Bowed Strings, Plucked Strings, Winds, or Keyboards), "sex", "nationality", "highest attained level of music education", "experience listening to groove-based music", "experience performing groove-based music", and "experience improvising" were all not significantly correlated. Taken together, this data supports our hypothesis that SD develops through long-term engagement with a score-focussed performance practice but that experience in the ear-playing domain (e.g. jazz) can counteract this development.

Discussion

In this experiment, we tested aural music reproduction abilities of classically trained performers in order to establish the SDR measure for score-dependency. We specifically tested aural reproduction skills for pitch and rhythm information in order to establish whether score-dependency may be linked to limited abilities in perceiving or reproducing pitch content. In summary, our findings strongly suggest that SD negatively affects aural replication skills for pitch, which indicates a negative effect of SD on task-relevant pitch perception or action mechanisms. Less definitively, these results also suggest that SD is associated with long-term participation in score-focussed performance practice without engaging in scenarios that involve playing by ear.

Classical musicians may tend towards score-dependency

Although the results lend suggestive support to our first hypothesis that classical musicians are relatively score-dependent due to their notation-focussed performance culture, the findings were ultimately inconclusive on this issue. However, other research supports our data in suggesting a tendency for classical musicians to be score-dependent. In their study on ear-playing, Woody and Lehmann (2010) skills found that classical musicians required an average of 10.58 RPs (in addition to two playbacks provided by the experimental setup) before accurately reproducing heard melodies. This was done without any visual support and so—in terms of pitch/rhythm content—corresponds to our condition C1 (blank bars). As an exploratory analysis, we ran this result through the SDR formula, inputting 10.58 for C1 and 0 for conditions C2–C5, which resulted in SDR = 4.174. This result falls into the 95% confidence interval for of our results and lies within one standard error from our sample mean. It provides support for our results being indicative of classical musicians' ear-playing skills. However, as this remains a theoretical extrapolation, it should not be taken as scientific replication of Woody and Lehman's data for classical performers.

The SDR range for the data collected by us was 2.07–7.0. It is noticeable that none of our classical musicians approached the lower extreme of 1. This extreme signifies that no playbacks (besides the two given by the experimental setup) are required for task completion across all conditions. While our findings could point towards the lower scale extreme representing an unreasonably high skill level in aural replication, Woody and Lehmann's results suggest otherwise: for vernacular musicians, they recorded a mean of 3.83 RPs. In another exploratory analysis, we input this as 3.83 RPs for C1 and 0 for conditions C2–C5, generating an SDR of 2.15. Since this is a mean figure, it suggests that vernacular musicians are likely to have scored both above and below, i.e. closer to 1. This indicates that our lower scale extreme does not represent an unrealistic extreme for this skill. As a result, we maintain that there is a realistic possibility that classical musicians are more score-dependent than not.

However, it should be stressed again that such exploratory extrapolations of other studies' data are theoretical. While both experiments tested for ear-playing abilities, our study differs from Woody and Lehmann's in several important points. We tested specifically for the tendency to rely on visual over aural information in order to establish levels of SD, and in doing so further differentiate by pitch and rhythm information. As such, our study has a distinct focus and a different experimental design, using several innovative parameters compared to the earlier study. Therefore, Woody and Lehmann's data provides a useful context, but was not replicated here, and so the extrapolations can only offer suggestive contextualisation to our findings.

Score-dependency negatively affects aural reproduction of pitch

Our second hypothesis was that pitch would be more difficult to reproduce aurally than rhythm for SDMs, but that an excessive information load for one factor would outweigh any pitch/rhythm divisions. While the sample of classical musicians overall found pitch more difficult to reproduce than rhythm, when dividing it into subsamples for score-dependent musicians (SDMs) and score-independent musicians (SIMs) this effect held true only for SDMs. The observation that for SIMs only the two extreme conditions were significantly different from each other (reproducing rhythm alone vs. rhythm with pitch) indicates that their superior ear-playing skills mitigated much of the effect of the (lack of) visual aids. These musicians played so well by ear that they barely required any notation to help them in an aural learning scenario. Very different results were obtained for SDMs, for whom task performance indicated a more defined difficulty ordering of conditions. For them pitch was clearly the primary factor in difficulty, even when less pitch than rhythm information had to be reproduced. Rhythm only played a role in increasing difficulty levels when full pitch information already had to be reproduced aurally (conditions C2 and C1), presumably adding to the amount of information to determine as a secondary difficulty factor.

These results confirm our hypothesis that pitch would be more difficult to reproduce aurally than rhythm for SDMs—so much so that our assumed caveat of information load mitigating this effect were shown as erroneous. As a result, the evidence strongly supports the assumption that SD—or a hidden factor related to SD—modulates perception and/or action processes required for aurally reproducing pitch. Since Harris and de Jong (2015) found limited right-hemisphere activations in the auditory cortex of SDMs compared to SIMs, this may be a perception rather than action issue, given the right auditory cortex's documented role in pitch perception (e.g. Peretz & Kolinsky, 1993) and memorization (Peretz & Zatorre, 2005).

One possible explanation for such an issue in pitch perception could be that playing from notation removes the need for in-depth pitch processing during performance. Mills and McPherson's (2006, 181-82) suggest that playing from notation does not require the same levels of audiation (i.e. creating an inner musical representation) for pitch as it does for rhythm. They propose that on many instruments it is possible to develop a strong eye-hand connection that makes it possible to rapidly execute notated pitch instructions without necessarily having to audiate them first. Notated rhythm, they argue, always requires audiation before executing it, since one must first consider how the rhythm sounds. We would add that this could also be explained by many instruments' visuo-spatial layout: On some instruments, pitch can be produced in one motion (e.g. piano); rhythm, on the other hand, by definition requires multiple events and therefore multiple motions, demanding greater motor planning. An additional factor could be that the rubato approach towards rhythm in classical music means that classical music culture expects less precision in executing notated rhythms than it does in executing notated pitch.

Both interpretations (perception or action issue) are supported by the results of a positron emission tomography (PET) conducted by Thaut, Trimarchi and Parson (2014), who found that rhythmic activities evoked more widely spread sensory-related activation patterns than pitch activities. They propose that temporal features may be more multi-sensory in nature than melody or harmony, with pitch being more closely related to vocalisation while rhythm is more closely related to body motions. Therefore, it is possible that SDMs have developed such strong psychomotor decoding skills from long-term score-focussed practice that a strong eye-hand connection removes their need to audiate pitch in performance. Over time, this may contribute to rhythmic audiation being continuously exercised when playing from notation, whilst pitch audiation is not. Long-term, this lack of audiation could contribute to reduced aural pitch discrimination skills and corresponding activations in the auditory cortex. The idea of a long-term effect is further supported by the results presented in the next section.

In this context, it is striking to consider notation as a technology: Notation is an invaluable communication technology in classical music culture, allowing composers to script and disseminate complex music and allowing performers to coordinate en masse in large ensembles (e.g. in orchestras)—both of which would not be possible in purely aural settings. However, over-reliance on any technology can also lead to dependency and hyper-specialisation in task performance over time, limiting the ability to engage more flexibly with situations. Notably, Western classical music seems to be the only music culture that follows notationally scripted performances to this degree. Several music psychologists and educators have compared the classical music practice of learning to read notation fluently but not to play by ear with learning to read a language but not to speak it (see McPherson & Gabrielsson, 2002 for an overview)—though perhaps it is more aptly compared to theatre actors who cannot improvise their lines (see Sawyer, 1999, on improv theatre).

This poses questions about how increasingly complex use of Western staff notation may represent concepts of task specialisation and scripting modes of action that reflect wider Western philosophies of efficiency and standardisation—however, these questions are beyond the scope of our article here. More to the point, from a music action perspective, one might argue that notation-dependent eye-hand coordination represents an effect of overlearning that negates more holistic experiences of music, relying on symbolic representations instead of social interactions, as can also be the case in other arenas of orality vs. literacy (see Ong, 2002, for an in-depth discussion). This is not to negate the many social aspects found in classical music performance, but rather to suggest that SD may represent a subtle cognitive shift in conceptualising music—away from an improvisational, imitative and therefore more embodied form of music-making (as demonstrated by, for example, autodidact garage bands) to a more instruction-driven, cerebral and therefore formalised understanding of music.

Score-dependency may result from long-term reliance on playing from notation

The data supports our third hypothesis that SD results from long-term engagement in a notation-focussed performance practice, but can be mitigated by participating in musical activities that demand playing by ear—though questions on the causality of SD remain. Results of correlating SDR with participants' self-ratings on biographical factors and music skills showed that both "age" and "years playing music" were correlated positively with SDR. This supports our hypothesis and the proposed model of SD as an eye-hand connection that increases with time. The results also show that "experience performing jazz" correlated negatively with SDR. Jazz learning and performance traditions—though sometimes involving simple forms of notation as mnemonic devices—demand strong improvisational and ear-playing abilities from performers. Together with the near-significant negative correlation for SDR and "experience improvising" ($p=0.0632$), it appears that engagement in ear-playing activities can mitigate the effects of SD, as expected. Indeed, research shows that playing by ear can even significantly improve music literacy by developing audiation skills (see Musco, 2010, for an overview).

Without such activities, the results suggest, long-term enculturation and participation in a score-focussed performance culture may engender SD and possibly develop into a self-sustaining process, in which notational literacy becomes increasing fluent while ear-playing skills decrease. Since extended use of notation is likely to lead to increased notational literacy, which in turn makes future use of notation more likely, this points towards a behavioural feedback loop. Over time, this may engender changes in neural processing and associated cognition, affecting (pitch) perception mechanisms. However, it is likely that this feedback loop can be disrupted by engagement in performance traditions that require greater ear-playing skills, which stimulates relevant neural areas as it broadens levels of musical expertise.

However, it should be noted that these are suggested mechanisms only: correlations can show only relationships between variables, but not causality. An alternative explanation for the positive correlations between both "age" and "years playing music" with SDR could be that educational measures have changed over time and that performers from earlier generations were educated to participate more in score-based performance practice than younger musicians. Results may also differ by instrument type, given the different visuospatial and sensorimotor feedback demands of different instruments. While we found no indication at all for an effect by instrument, our sample may have been too small to produce results when subdividing participants in this way. A study with a larger sample size—ideally combining neuroimaging techniques with behavioural observations—is called for in order to provide more evidence that could be used to confirm causality.

Conclusion

This study demonstrates that score-dependency affects musicians, but does so tendentiously rather than absolutely, and is likely linked to their performance habits. The data presented here shows that score-dependent musicians struggle in replicating pitch but not rhythm by ear, while there is no clear difference between the two factors for score-*in*dependent musicians. This may be the result of long-term engagement in a score-focussed performance practice, which—without the counterbalance of activities that demand playing by ear—perhaps engenders a behavioural feedback loop in which notational literacy increases as ear-playing skills decrease. If so, performers of different instrument types may be affected differently, given the varying visuospatial and sensorimotor feedback demands of different instruments.

Score-dependency may be posited as a form of overlearning that limits a wider embodied awareness of music by creating a dependency on a particular technology, in this case music notation, due to extreme task specialisation. It is noticeable that SD seems limited to Western staff notation in a classical performance context, as most other performance cultures use their notation systems more as visual mnemonic aids in aural performance contexts. While our limited sample size (n=20) means that results can only be treated as indicative, our behavioural observations match findings generated by existing behavioural and neuroimaging research.

This study suggests that the reportedly notation-focussed curriculum in classical music education (Feichas, 2010) may affect musicians throughout their lives, enabling them to become specialist performers in a score-focussed culture, but also possibly limiting their aural engagement in group performances and social musical settings, especially in non-classical settings. The possibility of worsening pitch perception with age may also be relevant to researchers studying changes in musical activity over a lifetime and music-making in old age. Therefore, the results presented here have implications not only for cognitive and neuroscientific research into perception of music, but also for research on musical practice and education.

Ethics statement & acknowledgements

The authors declare that there was no conflict of interest. Participants took part freely and did not receive remuneration for their participation. The article presents work from the first author's PhD thesis. The authors wish to thank all the participants, for generously providing their time and talents, as well as Prof Erik Højsgaard for his aid in preparing the notational materials, and the anonymous reviewers for their helpful feedback.

References

Bennett, H. S. (1983). Notation and identity in contemporary popular music. *Popular Music*, 3, 215. doi:10.1017/s026114300000163x

Bianco, R., Novembre, G., Keller, P.E., Villringer, A., Sammler, D. (2018). Musical genre-dependent behavioural and EEG signatures of action planning. A comparison between classical and jazz pianists. *NeuroImag*e, *169*, 383-394. doi: 10.1016/j.neuroimage.2017.12.058

Cross, I. (2003). Music and evolution: Consequences and causes. *Contemporary Music Review*, 22.3: 79–89. doi: 10.1080/0749446032000150906

Feichas, H. (2010). Bridging the gap: Informal learning practices as a pedagogy of integration. *British Journal of Music Education,* 27(1), 47-58. doi: 10.1017/S0265051709990192

Fine, P., Berry, A., & Rosner, B. (2006). The effect of pattern recognition and tonal predictability on sight-singing ability. *Psychology of Music 34*(4), 431 47. doi: 10.1177/0305735606067152

Goldman, A. (2016). Improvisation as a way of knowing. *Music Theory Online*, 22(4). Retrieved from: http://mtosmt.org/issues/mto.16.22.4/mto.16.22.4.goldman.html

Goldman, A., Jackson, T., & Sajda, P. (2018). Improvisation experience predicts how musicians categorize musical structures. *Psychology of Music*, 48(1), 18-34. doi: 10.1177/0305735618779444

Gudmundsdottir, H. R. (2010). Advances in music-reading research. *Music Education Research*, *12*(4), 331-338. doi: 10.1080/14613808.2010.504809

Hakim, A.-M., & Bullerjahn, C. (2018). Spiel nach Gehör auf der Violine – Wie beeinflusst musikalische Vorerfahrung die Imitation kulturell vertrauter und fremder Melodiemuster?. *Jahrbuch Musikpsychologie* 28, paper e39. doi: https://doi.org/10.5964/jbdgm.2018v28.39

Harris, R., & de Jong, B. M. (2015). Differential parietal and temporal contributions to music perception in improvising and score-dependent musicians, an fMRI study. *Brain Research*, *1624*, 253-264. doi: 10.1016/j.brainres.2015.06.050

Kopiez, R., & Lee, J. I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, *8*(1), 97–120. doi: 10.1080/14613800600570785

Lehmann, A.C., & Ericsson, K.A. (1993). Sight-reading ability of expert pianists in the context of piano accompanying, *Psychomusicology*, 12.2: 182–195. doi: 10.1037/h0094108

Lehmann, A.C., & Ericsson, K.A. (1996). Performance without preparation: structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, *15*(1-2), 1–29. doi: 10.1037/h0094082
Lilliestam, L. (1996). On playing by ear. *Popular Music*, *15*, 195–216. doi: 10.1017/S0261143000008114
McPherson, G., Bailey, M., & Sinclair, K. (1997). Path analysis of a model to describe the relationship among five types of music performance. *Journal of Research in Music Education*, *45*, 103-129. doi: 10.2307/3345469
McPherson, G. E. & Gabrielsson, A. (2002). From sound to sign. In R. Parncutt & G. McPherson (Eds.), *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 99-115). Oxford University Press.
Mills, J., & McPherson, G. E. (2006). Musical literacy. In G.E. McPherson (Ed.), *The child as musician: A handbook of musical development*, pp. 155-72. New York, NY: Oxford University Press.
Mishra, J. (2014). Factors related to sight-reading accuracy: a meta-analysis. *Journal of Research in Music Education*, *61*, 452–465. doi: 10.1177/0022429413508585
Moore, R. (1992). The decline of improvisation in Western art music: An interpretation of change. International *Review of the Aesthetics and Sociology of Music*, 23(1), 61–84. doi: 10.2307/836956
Musco, A. (2010). Playing by Ear: Is Expert Opinion Supported by Research? *Bulletin of the Council for Research in Music Education,* (184), 49-64. Retrieved from http://www.jstor.org.ezp.lib.cam.ac.uk/stable/27861482
Ong, W. J. (2002). *Orality and literacy: the technologizing of the word.* London: Routledge. First published in 1982 by Methuen & Co. Ltd.
Peretz, I. & Kolinsky, R. (1993) Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *The Quarterly Journal of Experimental Psychology, 46*, 301–325. doi: 10.1080/14640749308401048
Peretz, I.; Zatorre, R.J. (2005.) Brain organisation for music processing. *Annu. Rev. Psychol*., *56*, 89–114.
Sawyer, K.R. (1999). Improvised Conversations: Music, Collaboration, and Development. *Psycholog of Music, 27*, 192–216. doi: 10.1177/0305735699272009
Seppänen, M., Brattico, E., & Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiology of Learning Memory*, *87*(2), 236–247. doi: 10.1016/j.nlm.2006.08.011
Sloboda, J.A. (1977). Phrase units as determinants of visual processing in music reading. *British Journal of Psychology*, 68(1), 117-124. doi:10.1111/j.2044-8295.1977.tb01566.x
Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *NeuroImage, 20*(1), 71–83. doi: 10.1016/S1053-8119(03)00248-9
Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory*, *8*, 295–300. doi: 10.1101/lm.39501

Thaut, M.H., Trimarchi, P.D., & Parson, L.M. 2014. Human Brain Basis of Musical Rhythm Perception: Common and Distinct Neural Substrates for Meter, Tempo, and Pattern. *Brain Sciences*, 4, 428-452. doi:10.3390/brainsci4020428

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia, 50*(7), 1432–1443. doi: 10.1016/j.neuropsychologia.2012.02.028

Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in musical sight reading: A study of pianists. *British Journal of Psychology, 89*(1), 123-149. doi:10.1111/j.2044-8295.1998.tb02676.x

Wristen, B. (2005). Cognition and Motor Execution in Piano Sight-Reading: A Review of Literature. *Update: Applications of Research in Music Education, 24*(1), 44–56. doi: 10.1177/87551233050240010106

Woody, R.H., and Lehmann, A.C. (2010). Student musicians' ear-playing ability as a function of vernacular music experiences. *Journal of Research in Music Education, 58*(2), 101–115. doi: 10.1177/0022429410370785

Biographies

Christopher Corcoran

PhD Student in Composition & Music Psychology, Faculty of Music, University of Cambridge

MA, MMus, BA

I am a composer and researcher. My research experimentally investigates questions of perception and behaviour based on aural vs. notation-based learning. In my PhD thesis, I explore interactions of different music notation styles with aural priming on the performance of score-reliant performers, with particular application to the performance of groove and swing rhythms. Therefore, my research touches on a wide range of topics, including: Notation-based learning vs. aural learning; cognitive factors in sight-reading and score-dependency; swing and groove cognition, techniques, and performer interactions.

Dr Neta Spiro

BA/MUS, Music OXON

MSc, Cognitive Science and Natural Language, University of Edinburgh
PhD, Music Psychology, Co-supervised by Prof Rens Bod, University of Amsterdam & Dr Ian Cross, University of Cambridge

My recent research has focussed on two areas: music and wellbeing, and communication and collaboration in music making.