Journal of New Music Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nnmr20

AMMRI: a computational assessment tool for music novices' replication and improvisation tasks

Roger T. Dean, Anthony Chmiel, Madeleine Radnan, John R. Taylor & Jennifer MacRitchie

To cite this article: Roger T. Dean, Anthony Chmiel, Madeleine Radnan, John R. Taylor & Jennifer MacRitchie (26 Oct 2023): AMMRI: a computational assessment tool for music novices' replication and improvisation tasks, Journal of New Music Research, DOI: 10.1080/09298215.2023.2270973

To link to this article: https://doi.org/10.1080/09298215.2023.2270973

JOURNAL OF NEW MUSIC RESEARCH
https://doi.org/10.1080/09298215.2023.2270973

AMMRI: a computational assessment tool for music novices' replication and improvisation tasks

Roger T. Dean [a], Anthony Chmiel [a], Madeleine Radnan [a], John R. Taylor [a] and Jennifer MacRitchie [a,b]

[a]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia; [b]Department of Music, The University of Sheffield, Sheffield, UK

ABSTRACT

We present computational analyses of musical performances during 12-months study by novice participants aged 65–80. They learned two instruments (an electronic piano keyboard; the iPad app ThumbJam) each with two distinct approaches: replication by ear of melodies, and improvisation using specified methods. Here we present computational simulations and analyses of such processes and the corresponding R script. Using MIDI recordings from one participant group, we reveal diverse performance levels. Our tools are apt to analyse of our full dataset and potentially other assessments of early musical learning. The code can readily be developed for more advanced learners.

ARTICLE HISTORY

Received 7 September 2021
Accepted 10 October 2023

KEYWORDS

Computational analysis; MIDI; melody replication; improvisation; learning; AMMRI

Introduction

This paper aims to provide simple measurement tools to assess achievement in musical reproduction and improvisation during novice learning. We propose that these tools can be useful across music instrument education, but may be particularly so for novice learners across the lifespan.

To better understand the potential benefits of engagement with music later in life, we are presently engaged in an experimental longitudinal music education program called the *Active Minds Music Ensemble*, which focuses on older adult music novices aged between 65 and 80 (MacRitchie et al., 2022). As changing focal areas of music education develop and take hold, it is also necessary to produce novel methods for assessing these. Indeed there is a lack of quantitative computational measurement of novices' performance achievement on MIDI keyboards in the present literature (we can find no such computational assessment papers via Google Scholar searches). With this in mind, we developed the 'Automated Measures of Melodic Replication and Improvisation' (AMMRI).

AMMRI is a computational analysis tool developed to facilitate the large-scale analysis of the musical techniques presented to our participants. The primary aim of this paper is to explore the use of specific techniques of musical expression that are chosen by our participants using AMMRI. We have observed during our longitudinal study that many participants asked how they could determine their own progress. In this paper, we validate the use of AMMRI with data collected from one of our earliest classes in preparation for our forthcoming dataset that spans the entire ten classes (and which contains almost 2500 individual performances). In principle, both teachers and students could use AMMRI.

To demonstrate the wide range of possibilities for imprecision in melody replication, we simulated the main feasible replication limitations. Similarly, we simulated all 14 melodic improvisation methods that we taught (see below), as each is likely to be applied. We developed computational assessments of each simulated feature, both for replication and improvisation, and then compared the analyses of our simulations with an analogous set of results obtained from a subset of our participants, as detailed below. Our successful validation of the methods by means of the simulations allowed us to test some very simple hypotheses, in the present paper in reference to our first group of participants and to their diversity of achievement:

1. Older adult early learners display a range of performance aural skills applicable to the tasks of reproduction of melodies.
2. They display a range of attitudes and aptitudes towards systematic melodic improvisation.

CONTACT Roger T. Dean roger.dean@westernsydney.edu.au Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09298215.2023.2270973.

We assess these hypotheses using AMMRI and establish the wide applicability of this objective computational assessment. We want the assessments to be usable regardless of the mode of learning (e.g. aural vs. notation, of which we used only the former), and the context chosen for the improvisation or the expertise of the improvisers. The inclusion criteria assumed are that performances occur on MIDI instruments, are played with a single hand only, and that much of the material undergoing replication is already familiar to some degree to the performer.

Overview of the experimental approach

In the program, ten groups of participants receive music education from a professional teacher via aural training across a twelve-month period. Learning varies in terms of both the instrument that learning occurs on, and the task that the education focuses on.

Regarding instrument, participants spent half of the twelve-month period learning on a digital keyboard, and the other half of the twelve-month period learning on the iPad touch-screen app *ThumbJam* (Sonosaurus LLC, 2017). On each instrument participants undertake simple single-handed skills, utilising the right hand only. The two instruments are chosen to require substantially different kinds and levels of fine motor skills. Notably, the *ThumbJam* interface has all the keys in a single row (no elevated and recessed black keys as on the conventional music keyboard). Each key occupies essentially the whole of the iPad screen from base to top in landscape view. Control of note loudness is achieved on the conventional keyboard by key press velocity (the name of the corresponding MIDI, musical instrument digital interface encoding parameter). In contrast, on the *ThumbJam* interface, finger contact position on the key-length determines velocity (from low near the bottom of the screen to high at the top: the touch sensitivity of the iPad surface is not used). The choice of the two instruments was intended to enhance motor skill development, and brain-muscle coordination, as well as distinguish blocks for analytical purposes in work in the subsequent papers on the overall dataset we obtained.

The six-month education period for each instrument is split into two distinct tasks, run successively: *replicating* melodies, and *improvising* melodies. The replication task focuses on the accurate reproduction of pre-composed melodies, some of which are expected to be familiar to most if not all participants, and some which were specially composed for the course. The improvisation task focuses on generating new material using 3–5 note ‘prompts’ freely taken from the beginnings of taught or well-known melodies (or sometimes chosen anew by a participant). Overall, this equates to four distinct learning ‘blocks’ concerning instrument and task, with each block three months in duration. The order of the blocks is randomised and counterbalanced between groups.

At the end of each three-month ‘block’ of learning our participants were asked to complete replication and improvisation tasks taken in a systematic sequence, which we detail later. The sequence is split into the distinct sections of replication and improvisation, the latter of which contains 14 varied improvisation methods through which participants learn how to modify simple existing material. Jazz and commercial music improvisation, being arguably the main teaching foci for musical improvisers in Western culture, are routinely taught in the form of precise repetition of notated patterns, based closely on underlying defined tonal scales and harmonies (e.g. by well-known Jazz tutors such as Jamey Aebersold, since 1967, as recently discussed (Thibeault, 2021)). Contrasting this, we use a more ‘open’ approach based on the idea of transforming musical motives by simple but specific methods. This approach is partly based on Dean (1989) and aims to achieve self-chosen musical outcomes without particular reference to tonality or metricality.

Method

Participants

This paper was developed in preparation for the completion of our overall data acquisition during this pedagogic study (while each group studied for a year, the overall process took approximately 2.5 years, with practical staggering of the groups). A subset of our eventual dataset was taken from seven participants, who made up the first group from our overall set of 10 groups ($N = 68$) of older adult music novices. As detailed below, this particular group experienced three months of face-to-face group teaching, after which all teaching was online. (We have subsequently published a detailed analysis of some of the effects of and attitudes towards the transition to online learning (MacRitchie et al., 2022) but we find that nothing in the analysis suggests likely critical impacts on the measures and interpretations of the present paper). Participants in the examined class were aged between 65 and 78 ($M = 69.6$, $SD = 4.5$). Participant eligibility included a maximum of two years prior learning and/or playing experience on any instrument, apart from voice. All participants here reported having no prior formal training on a musical instrument; two participants reported having one year of prior experience playing an instrument, with the remaining participants reporting no prior playing experience, while two reported participating in a short introductory course on music theory.

Procedures

General procedures

Participants received fortnightly group lessons for one year, focusing on the two separate instruments (six months on a digital midi-keyboard, and six months on the *ThumbJam* application for the iPad) and the two separate learning tasks: see Overview above, and further details below. Each six-month section was split in half to focus on the two separate learning tasks of reproduction and improvisation, leading to four separate three-month blocks with a randomised, counterbalanced order. Classes were initially held face-to-face at our University campus. However, from March 2020, to comply with COVID-19 social distancing requirements, all lessons and data collection sessions were moved to an online-only format via *Zoom*. Thus, the participants in this subset experienced three months of face-to-face learning, followed by nine months of online-only learning, with one unchanging teacher.

The teacher's instructional language was designed to be as simple and practical as possible to support learning. For the six-month period of learning on a conventional keyboard, participants used an electronic *Yamaha* PSR-E363 USB MIDI keyboard. For the six-month period of learning on an iPad, participants used a 9.7-inch iPad and the *ThumbJam App* (version 2.6.7). Regardless of instrument, participants were asked to only use their right hands to perform and only the MIDI note range 60–79 (middle C to an octave and a fifth above). On *ThumbJam*, participants were given a custom preset that: (1) was configured with a 'Piano' sound; (2) showed a chromatic note layout with the note names hidden, and with coloured bars for C, F, and G; (3) had a visible note range of 60-79; and (4) had all advanced settings and parameters hidden to prevent participants from modifying the preset.

At the end of each three-month learning block (focusing on a unique combination of instrument and task, as just summarised), participants took part in an additional group 'celebration' performance session. At this performance session, participants were first asked to play either a replication or an improvisation of their choosing (depending on which of the two tasks the current block was focusing on), and then were also presented with a set of new replication tasks and improvisation tasks, as outlined in the systematic sequence below. All materials in the replication tasks were chosen or set to be in this pitch range. Additionally, all replication melodies had a requested maximum of 30 notes. Each task was performed in a *Zoom* breakout room with a supervising researcher or technical assistant. (For the initial performance session, which was done face-to-face, each participant performed their tasks simultaneously in the classroom, which was monitored by the teacher. Headphones were used in this face-to-face session so that participants could not hear the simultaneous performances of other participants.) Participant MIDI keyboard performances were captured using a prototype app called 'Session Recorder' developed in *Max/MSP* (Cycling '74, 2021). This was designed to (1) automatically de-identify and timestamp the data; (2) automatically connect to the participant's MIDI keyboard; (3) provide a simple diagnostic process that was easily readable and describable by participants, so that researchers could easily check the setup of each participant remotely (given that participants were responsible for setting up their own hardware remotely ahead of each session); and (4) to provide a simple way for participants to upload MIDI files of their performances directly to a secure server.

Participant MIDI *ThumbJam* recordings were captured using the app's 'Save set and mixdown' feature, which recorded both the MIDI and audio recordings of the performances by saving them as.mid and.wav files respectively. Participants then emailed these files as a compressed (.zip) file by exporting them from the *ThumbJam* app to a central university-hosted project email account.

Systematic sequence procedures

Participants were asked to complete replication and improvisation tasks in the 'celebration' at the end of each three-month block of learning. Each task included two minutes of preparatory playing time and concluded with a post-performance evaluation describing the participant's satisfaction with the performance on a 3-point Likert scale (this Likert data is not used in the present paper). For the first task, participants were asked to:

1. Play your favourite song, or improvise on your favourite fragment, from the last 3 months (i.e. the immediately preceding block of lessons and practice).

After this initial task, participants were asked to collectively pick a melody from the test bank (a separate bank of melodies not included in the lessons) for use in the subsequent tasks. Participants were asked to choose a melody they were confident in recognizing (i.e. they had heard it before) but had definitely not attempted to play before. Two *replication* tasks followed where participants were instructed to:

2. Replicate the melody as accurately as possible from memory (without having heard it at all during the class). Participants were encouraged to perform as

much of the melody as possible and asked to evaluate how familiar they were with the melody.

3. After auditioning a MIDI performance of the tune (performed by the teacher or played through sound sharing over *Zoom*), and following a two-minute practice interspersed with two further auditions, participants were then asked to replicate the tune and evaluate again.

These were followed by two *improvisation* tasks. If a class of participants had not undertaken a batch of lessons focused on improvisation at this juncture, the teacher would give a brief explanation of the improvisation techniques (listed in the following paragraph) at this point. Most techniques were fairly self-explanatory, as they were couched in terms of simple procedures.

4. Improvise on a short fragment (approximately three to five notes), chosen from any part of the selected test bank melody or if preferred, chosen freely. For this task, three of the specific improvisation methods were merely mentioned verbally as pointers or reminders of some of the available tools. Once our processes were online we allocated a different experimenter to each participant in a separate Zoom room, though the whole process was still led by the piano teacher. The experimenter could choose which of the three improvisation methods they mentioned on each occasion (with the full list in front of them, but at that point not normally shown to the participant). The participants did not have to specify the chosen method(s) they performed in this improvisation.
5. Improvise on the same short fragment as used in Task 4, but choose and specify one or a few methods from the complete list of 14, which was then displayed to the participant, and briefly explained as necessary by the researcher. Emphasis was placed on restricting the number of chosen methods.

The following list of simple improvisation methods was shown to the participant during the sessions:

1. Adding repeated notes
2. Adding passing notes (in between) or neighbour notes
3. Change the distance between the notes
4. Change the note lengths (Rhythm)
5. Make it louder or softer (Crescendo or Diminuendo)
6. Make it suddenly change volume (Accents)
7. Put silences in between notes (Rests)
8. Try playing at different speeds
9. Play smooth and joined (Legato)
10. Play bouncy and short (Staccato)
11. Different combinations of Legato and Staccato (Slurring)
12. Reverse the notes (Retrograde)
13. Repeat higher or lower (Sequence)
14. Change the set of notes / include some black notes (Modulation).

All musical terms were fully explained, although Modulation was only elaborated by the teacher when requested, and with a simple explanation based on dominant notes rather than any theories of tonality.

Analytical and simulation procedures

The first stage of the development of our analyses was to create simulations of all the specified replication and improvisation procedures. Our objective was that the simulations could represent a range of possible fulfilments of the procedures, from the very limited (as expected in our novices) to the fairly extensive (as might occur with expert musicians, perhaps when giving a demonstration). For the present purposes, moderate ranges amongst the extremes were chosen and turned out to be reasonably apt. These simulations are used in this paper alongside the set of performances from one group of our participants, to demonstrate the degree of diversity in their efforts.

The second stage was the creation of the analytical algorithms and code for assessing each of the specified tasks and we present assessments of the diversity of performance levels achieved among both the simulations and the novice performances. Given a battery of such assessments, in the discussion we consider briefly whether evaluations of different improvisation tasks can be compared, and how multiple assessments might be amalgamated into a single measure.

Analyses

The following description accompanies the provided AMMRI assessment code, which is located within the Supplementary Material. Code is provided as an extensively annotated *R* script that a reader can use if desired. Recorded MIDI files were imported into the statistical platform *R* (version 4.1, in the development environment RStudio 1.4) using the library tuneR, whose readMidi function constructs a data frame containing all its information. tuneR also provides getMidiNotes, which is intended to transform the resultant data frame so that it only contains the relevant information about the notes, and additional information is removed. While readMidi was effective with all files studied here, getMidiNotes failed with some, and so an alternative function was coded. This then allowed the resultant data frame to contain columns providing time in the performance, note, note duration, note attack velocity, inter-onset interval to the next note (hereafter abbreviated IOI), and MIDI-channel and track numbers. The latter two were used to check that there was only one performance in a file. This was necessary because *ThumbJam* is designed to permit overlaying of multiple tracks of performance. However, we spent considerable effort on instructing our participants to avoid overlaying tracks, and were successful in this respect, so 'cleaning' the file of extra tracks was rarely required. We also checked that chords were not retained (defined as two note attacks within 35msec of each other; see Dean et al., 2014; Pressing, 1988). Again, we trained participants to avoid this, and there were virtually no chords found. For keyboard recordings, any recorded pitches that were outside the 60–79 range were normally transposed down or up by an octave to bring them into range; for *ThumbJam* performances it was not possible to exceed the 60–79 range with our custom preset. In one exceptional analysis, concerning pitch contours, the transposition was instead to the nearest extreme value of the 60–79 range (see Supplementary *R* code).

Simulations and analyses were coded in *R*, and the simulations could be auditioned in a custom *Max/MSP* (Cycling '74, 2021) patch or saved as MIDI files (but they were never heard by participants), In general, we chose a moderate range of the possible features being studied. To represent this evenly, we chose the probabilistic parameters of the simulation from uniform distributions across the relevant range (rather than the typical Gaussian distribution).

Implications of context and purpose for our measurement methods

Replication

The methods by which different note velocities are produced differ dramatically between our two chosen instruments. As mentioned, *ThumbJam* is not touch velocity sensitive; rather, note velocity is determined by the vertical axis positioning of finger impact. In other words, the closer the impact of the finger is to the top of the iPad screen, the greater the note velocity and attendant loudness, and vice versa. Due to this, we could not reasonably expect measurements of velocity between the two instruments to be comparable. However, a comparison of within-instrument velocity variation could be made.

Replication fidelity was determined by measuring the proportion of the piece whose replication is attempted (the ratio of the number of performed notes to the number in the original tune), and then measuring the accuracy of the attempted portion (without penalising for its incompleteness). This was done with dynamic time warping (hence DTW; performed using the dtw package in *R*), which estimates the distance between the parameters of the performance and the original. We focused on pitch and IOI (timing) measures, and their combination in a single parameter. We did not emphasise velocity (i.e. loudness) replication precision, because it was not emphasised in the teaching of the pieces; rather velocity was considered more as an important in improvisation as a potentially expressive device. The DTW measure can integrate the assessment of alignment of one or more features of a replication with those of the target original. So for pitch and timing we could assess, accuracy both separately and jointly; and similarly for pitch and velocity. We made no consideration of metricality, since it was not taught; it would require a different, more complex, approach. However, such methods could readily be adapted from the measures we developed for the improvisations, and from prior work on beat detection such as MATCH (Dixon, 2001, 2005; Dixon & Widmer, 2005), which is available as a plugin for *Sonic Visualiser* (Cannam et al., 2010). Presented DTW values are all normalised (that is, expressed per note of the performance), so they do not penalise for note omissions, which are reported separately.

Improvisation

To determine the diversity of improvisations, and the degree to which they fulfilled any of the trained or pre-specified tasks, a set of measures relating to each method of improvisation was developed, as reported earlier in the *Systematic sequence procedures* section. Some of these are illustrated below in the *Results* section, and they are described in the Supplementary Annotated *R* code.

Rescaling IOI for use jointly with pitch

Because IOI is measured in milliseconds (msec) and may range up to 3000, whereas our permitted pitch values were only from 60-79, measures that simply combine both IOI and pitch would largely reflect IOI and hardly represent differences in pitch. To overcome this, both were rescaled to the range 1:20. In the case of IOI, the original range of IOI possibility was always assumed to be from 35 msec (the minimum to separate two notes as not belonging to a chord, which we did not permit in our data) to 3000 msec (the latter based on the observed data).

Results

Simulating replication and improvisation, and their assessment

Here we illustrate some of the individual simulations and the corresponding analyses. Initially, we chose to run 20 simulations of each type (replication and improvisation) and then assessed the corresponding variation in the parameters of replication or improvisation that were involved. Given success in this, we assembled a large battery of accumulated simulations of each of the two parent processes (replication and improvisation) so that these could all be analysed together with respect to multiple performance parameters, to further illustrate the diversity achieved. These batches of simulations were then compared with the complete batch of real performances from our examined class of participants (32 improvisation recordings and 67 replication recordings), to indicate to what degree they fulfilled the range of possibilities demonstrated by the simulations.

Replication

Figure 1 shows the results of assessing one of the simplest types of simulation of the replication task, with the 'Yesterday' (a song by Lennon and McCartney) main theme as the sole target melody. In this case, all melody events are represented but varied numbers of incorrect pitches replace their correct counterparts, and the analysis concerns pitch sequences only. The three panels compare the distribution of error frequencies (the proportion of pitches that were incorrect) with the DTW distances, to show how the different measures show different sensitivities.

Figure 2 shows the results from simulations of some more complex replication errors. In this case, repetitions and omissions were simulated, together with some pitch and timing errors. This was again using 'Yesterday' (29 events), and the transforms ranged from 14 to 48 notes in length. The displayed DTW value is normalised for transform length (as in Figure 1), and lengths in themselves are not penalised. Figure 2 reveals the much greater range of distances generated by the repetitions, omissions, and pitch and IOI transforms, and by taking account of both features: the maximum value shown is 216, whereas those in Figure 1 were < 5. This confirms the suitability of DTW for sensitive detection of the most likely kinds of replication errors expected in our participant learners.

Improvisation

To demonstrate the likely suitability of our assessments we again show results from simulating some simple improvising processes, as well as one more complex improvising process. Figure 3 confirms the separate simulation of simple but substantial variation in pitch intervals and note lengths, and their detection by the assessment measure (Panels A and B). Compared with the source materials, both increases and decreases are observed, as intended. Panel C illustrates a conceptually more complex improvising method, the introduction of retrogrades. This was included because it is easy to think of playing the fingers in reverse order without moving the hand (e.g. by playing fingers 1-3-5, 5-3-1), and so (fairly) precise retrogrades are also relatively simple to perform on a keyboard, and potentially also on a touchpad instrument such as *ThumbJam*.

The results so far illustrate some simple as well as some musically more complex errors in replication, and methods of improvisation. They confirm the diversity of obtainable results, and the effectiveness both of our simulation and assessment methods. As indicated earlier, this is our fundamental objective here, since we anticipate that such objective assessment values will be informative to learners, and potentially predictive of their own and others' estimation of the broader 'quality/accuracy' of their performances (and the supporting musical learning). We expect this because we assume that the degree to which a person can achieve a specified musical technique will be a strong influence on the degree to which they can harness it for musical expression; we do not claim that the measures are direct indicators of such expression.

In the space between technical precision and music expression lie matters of style and coherence. With this in mind, we added some more complex objective measures relating to these aspects, and that may be informative and helpful to learners and/or teachers. These involve considering the statistical distribution of certain musical features, with less regard for their specific sequencing, which has been central so far. Figure 4 illustrates this. The larger overall structure (Kullback-Liebler divergences; KL) and cohesion (entropy) values in Figure 4 confirm that the simulations, while not focused on these aspects, nevertheless change them detectably and diversely. In a limited sense, when there are substantial changes in entropy distributions, particularly of pitch, this may signal a stylistic change (i.e. incongruence in the case of attempted replications).

Supplementary analyses (not shown) confirmed the effectiveness of all simulations and assessments shown in the developed *R* Code (Supplementary Material).

Illustrating the diversity of performances by our novice keyboardists

Solely to judge the effectiveness of our assessment procedures in detecting differences in quality/accuracy when

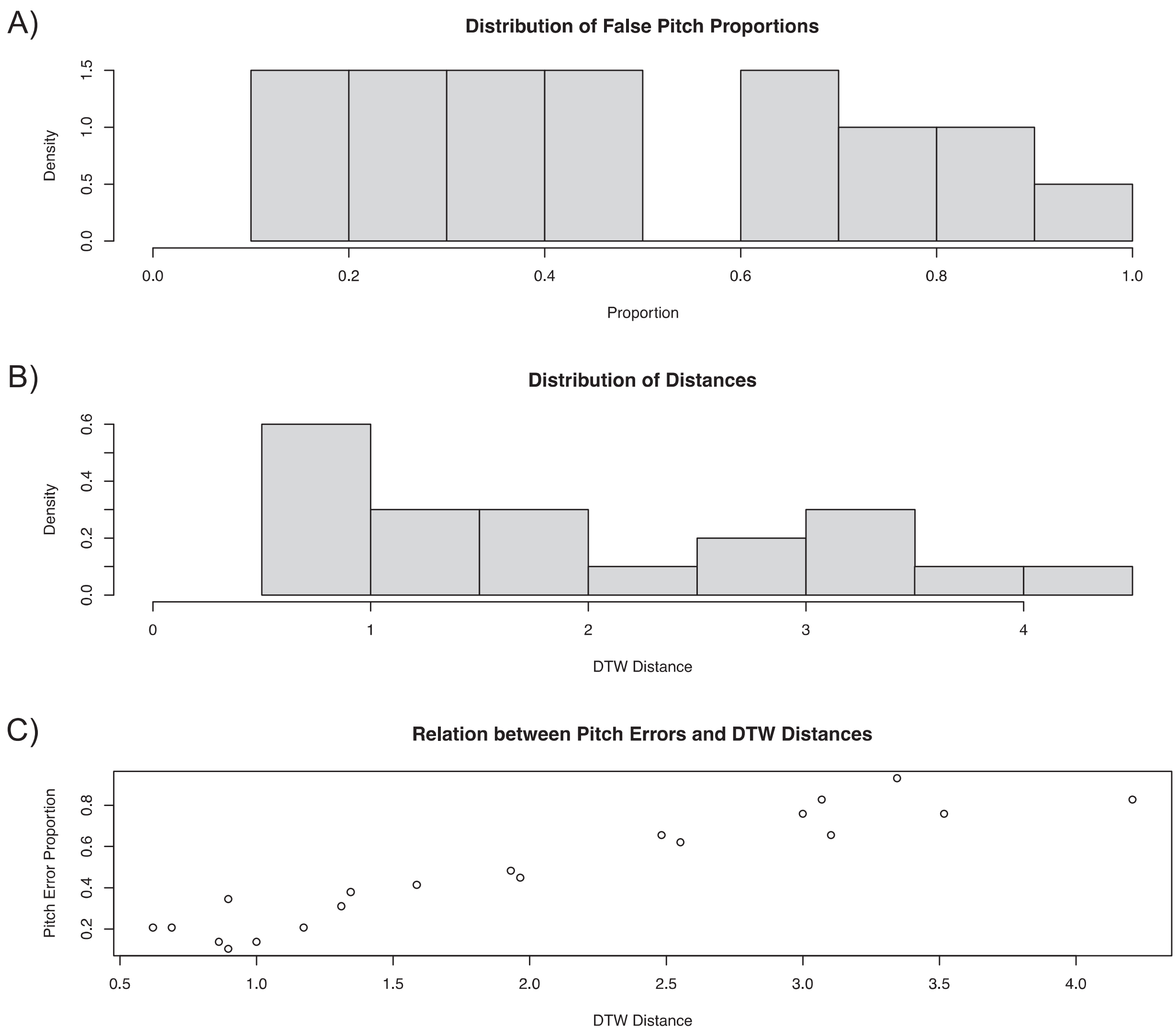

Figure 1. Comparing error frequency with normalized dynamic time warping (DTW) distance for 20 simulations of 'Yesterday' performances with wrong notes included. DTW is more sensitive, taking account of both the number and degree of note inaccuracies, as particularly revealed by Panel C. For example, there are three simulations with an error proportion of 0.2, and they are differentiated by their (normalised) DTW distances with respect to the reference. Note that in this simulation, performances and melody all contain the same number of notes, 29 (unlike the normal situation, where a performance is an abbreviated version of the melody).

faced with real novice performances, we took the complete set of performances available from an early group of participants (being the same subset class described in the *Method* section; at the time of initial submission of this paper the majority of our overall participants were still involved in classes). For the figures in this paper we did not distinguish participants among the performances and we used every individual performance. The performances were subjected to our developed assessment procedures. Figure 5 shows analyses of replications for their accuracy (DTW pitch distance, and joint pitch-IOI distance, using rescaled pitch and IOI as described in Methods) as well as their extent (the length in notes of the replication in comparison with that of the original).

Figure 5 demonstrates the diversity of performance levels observed among our participants, and the effectiveness of our assessments. Figure 5 also suggests that the pitch DTW measure is the more discriminating of the two measures shown (including 'perfect' performances and showing among positive values a maximum:minimum ratio that is much higher than that for the combined pitch-IOI measure). This coincides with the greater teaching/learning emphasis on reproducing melody pitch structure or contour than rhythmic contour. Turning to the improvisation performances, among the 14 previously listed Methods in the systematic sequence, the most popular were 'adding repeated notes', 'playing at different speeds' and 'change the note lengths'. Thus, to illustrate the diversity of improvisations in these

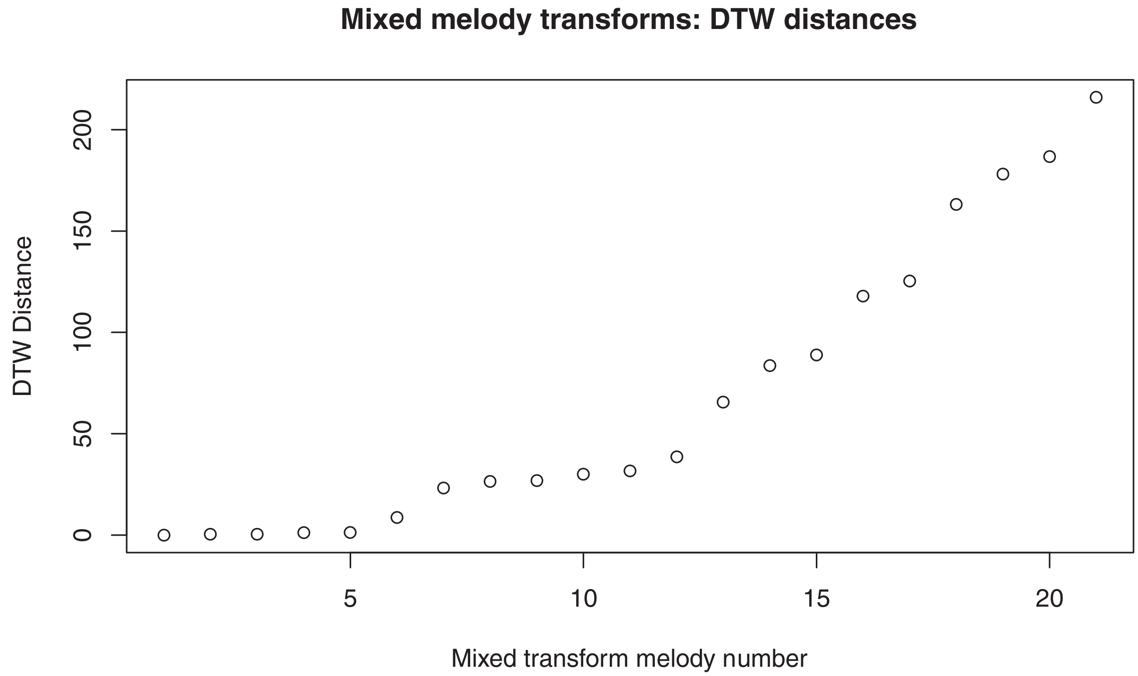

Figure 2. Normalised DTW alignment distances of mixed melody transforms (sorted by DTW value).

respects, all the improvisations from our examined class for tasks 4 and 5 were analysed for these three features and two contrasting others (Figure 6).

Assessing relations between set piece replication and improvisation performance distributional features

Figure 7 shows some distributional analyses, comparing the replications, improvisations and the 16 set pieces (referring to the initial teacher-made exemplar performances) attempted in the replications. These can give overall indices of congruence and disparity between the distributions, and potentially their 'styles'.

As might be expected, the entropies of both replications and improvisations mostly show a wider range than do those of the original melodies. The exception is the IOI entropies, where the original melodies include some with lower entropy than any performance. Ranges for replications and improvisations were fairly similar to each other. For our present purposes, the key observation is that a diversity of values is achieved in each case. Specifically, several minimum values for velocity entropies are shared by replications and improvisations, which overall are considerably lower in range as compared to the original performances' minima. This supports our anticipation that participants were commonly paying relatively little attention to velocity/loudness, which corresponds to the relative lack of emphasis on this parameter in the lessons. This also may be partly due to the non-intuitive dynamic control mechanics of the iPad app *ThumbJam*, where louder dynamics require the finger touch to be higher up the vertical axis of the screen. From a stylistic assessment perspective, these distributional results are also consistent with the others in indicating that participants achieve some significant commonality with the original melodies during their replications, and perhaps implying that their improvisations are also related in their intervallic and rhythmic senses. Possible distinctions between the distributions of replication and improvisation will be explored more fully in later work with our ultimately much larger complete dataset.

Discussion

The results presented show considerable diversity in all the assessment parameters (including those not shown in the presently reported results) both in our moderate range simulations and in the performances from our examined group of participants. In the case of the replications (reproductions of melodies learnt by ear), the implicit assumption of our computational analysis is that precision of replication is in itself an index of musical success, and potentially of expressivity *per se*, or at least of a participant's future capacity for expression. While a melodic reproduction could necessarily be more or less expressive – potentially due to a very wide range of factors – it remains reasonable that achieving good precision is an important positive step towards expressing its affective intent. Controlled variation therefore may allow for

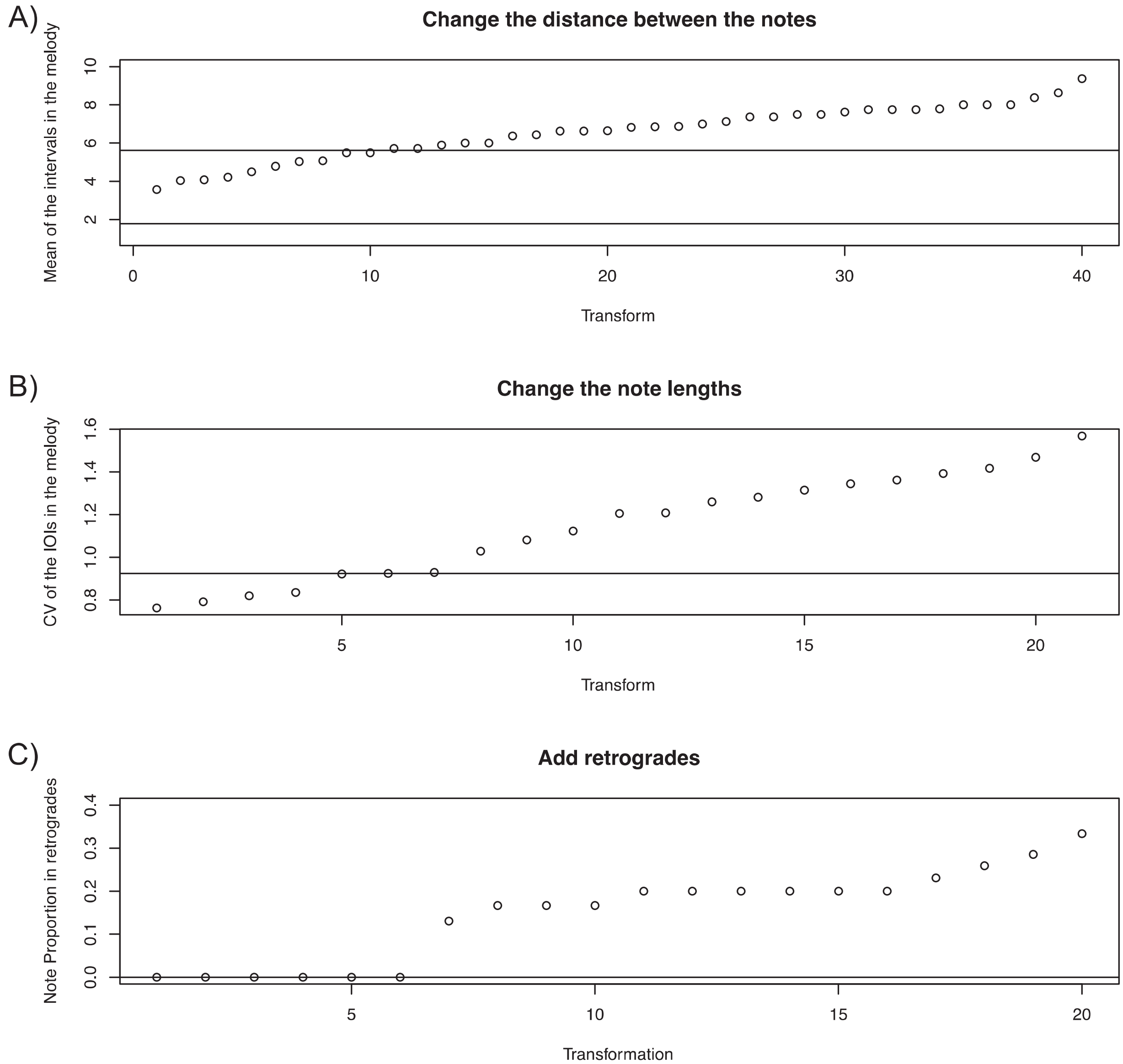

Figure 3. Simulating improvisation methods. Panel A: varying pitch intervals with 20 simulations on two melodies. The horizontal lines show the values for the two parent melodies. Panel B: varying note IOIs, 20 simulations (based on 'Yesterday'). The horizontal line shows the value measured for the source. CV: coefficient of variation (SD/M). Panel C: Adding retrogrades. The horizontal line shows that there were no retrogrades in the original material. The six other '0' values reflect the stringent exclusion of repetitions which we undertook (some retrograde components, especially note pairs, may repeat pre-existent patterns, and hence be excluded from the measure; furthermore if a sequence is retrograded, and then the original re-appears, the latter constitutes a retrograde of the added retrograde but should not be counted). N.B. the data are sorted by y value, and hence vertically aligned items are generally not the same transform.

optimising such impact. In the case of the improvisations, our implicit view is that controlled diversity of output is a key parameter that not only indicates musical success in the application of techniques but that also may well be necessary for achieving expressivity and creativity. The assessment of such relations of replication and improvisation techniques with musical expression, as judged particularly by the participants themselves, will be the subject of one of the analyses of our parent project, to be published later. In this context, an external audience of assessors (musically expert or otherwise) is among the least relevant.

Our Hypotheses 1 and 2 are supported by the data we obtained. Hypothesis 1 suggested that older adult early learners display a range of performance aural skills applicable to the tasks of reproduction of melodies. That our data show the ability to develop various degrees of replication precision in relation to the pitch sequences (and other features) of tunes that they hear (and without access to their notation) confirms this.

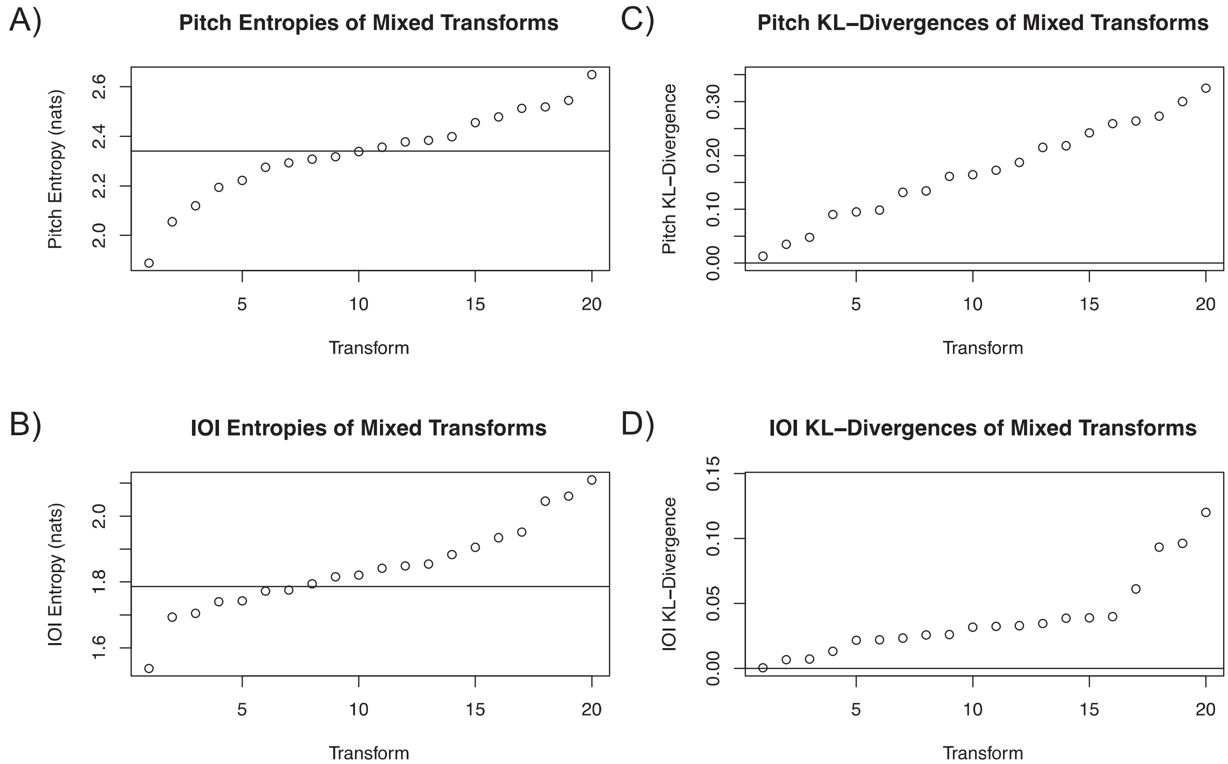

Figure 4. The mixed replication transforms studied in Figure 2 are analysed further, for pitch and IOI entropies (Panels A and B), and for Kullback-Liebler (KL) divergences from the starting melody (Panels C and D). Entropy is a measure of the 'disorder' of the overall distribution (disregarding the sequence of events), while KL-divergence assesses the difference between the reference distribution (the starting melody) and the transform. The horizontal lines in Panels A and B show the entropy of the reference melody (from which transformations differ both up and down). The horizontals in Panels C and D merely confirm that the divergence between a distribution and itself is '0'. N.B. KL divergences are asymmetric, so these are always assessed as query vs. reference. The values are sorted, so vertically aligned values do not usually represent the same transform.

Our second hypothesis was that they display a range of attitudes and aptitudes towards systematic melodic improvisation. The aptitude is demonstrated by their success in relation to learned improvisation methods (e.g. Figure 6). By 'attitude' we intended to imply that participants might show preferences for different approaches: their capacity for this is shown by Figure 6, but later analyses will be required to more directly test it.

For the improvisation tasks, we suggested that our participants begin with 3–5 prompt notes taken from familiar melodic material, which could be the subject of the improvisation. However, the researchers noted when supervising the performance sessions that participants may vacillate about (or perhaps forget) any chosen material as they consider the methods they will use (task 4) and as they practice the method(s) they will use (task 5). So our analyses here consider fulfilments that do not necessarily relate solely to the opening few notes. On the other hand, during the lessons in the blocks of teaching focused on improvising, significant time is spent on locating/identifying 2–3 note themes, which are then used. Thus, in the analysis of our forthcoming much larger dataset, we expect to carefully assess the fulfilment of the specified methods in relation to the 3–5 prompt notes, as well as in relation to all 3–5 note phrases (where applicable). The task demand is essentially intended to ensure that participants attempt thematic improvisation, rather than numerous other valid but much more diverse approaches that are possible. The present results suggest that this intent was achieved with the group studied here.

One alternative approach would have been to incorporate a geometric data compression algorithm such as COSIATEC for pattern discovery (see, e.g. Louboutin & Meredith, 2016; Meredith, 2013, 2014). However, in part due to the short durations of our recorded stimuli in comparison to the datasets that COSIATEC has been evaluated with, the decision was made that this approach was not optimal. Regardless, future examination with COSIATEC or a similar compression algorithm may prove fruitful. Another alternative approach for the computational assessment of melodic fragments would be to use the *R* program FANTASTIC (Müllensiefen, 2009).

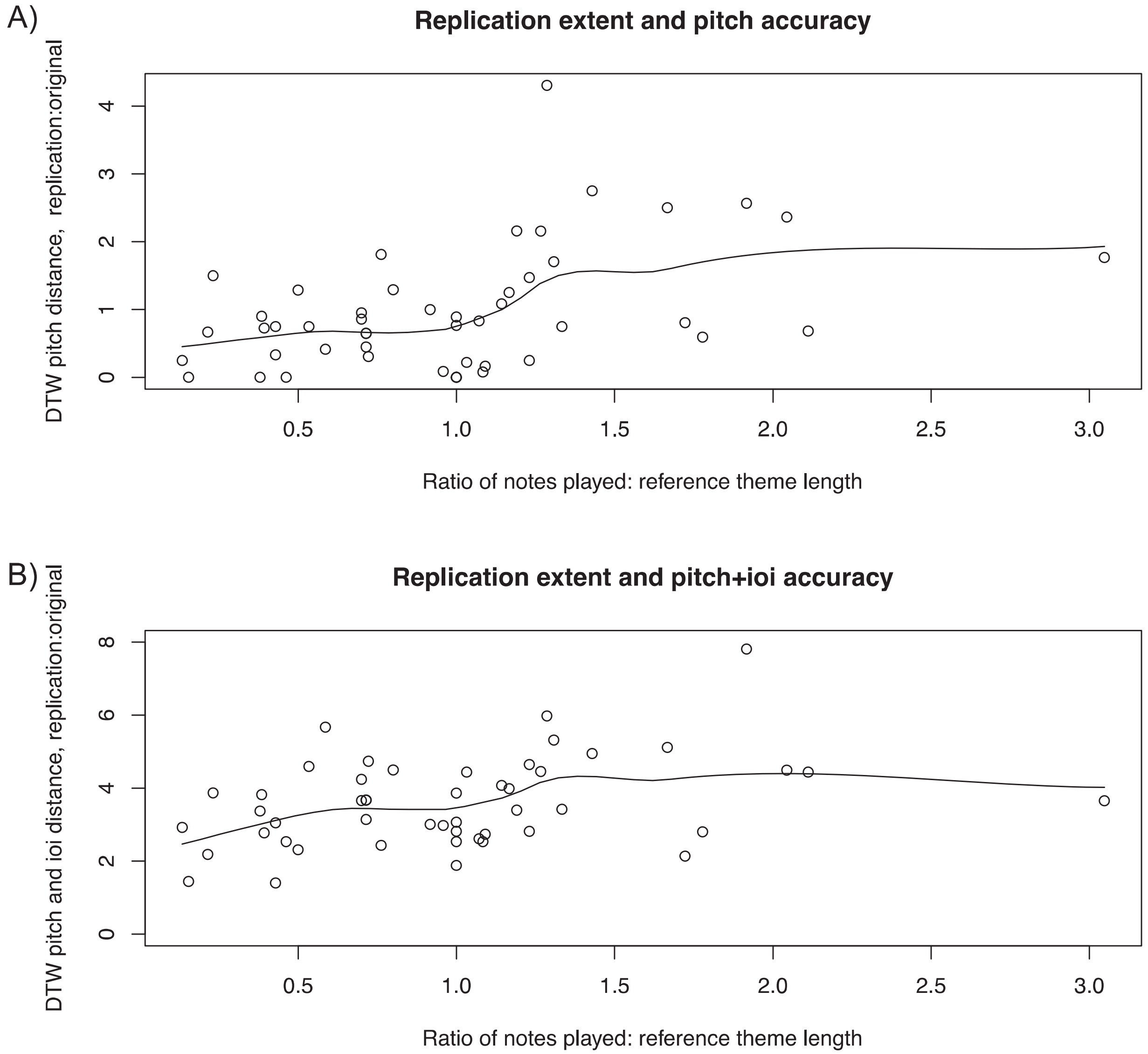

Figure 5. Sixty-seven replications by the seven participants in our examined class, representing 15 melodies. The DTW pitch distance (Panel A) is essentially an accuracy measure on the pitch sequence of whatever part is performed, and the '0' (perfect) scores are mostly short chunks. The superimposed loess-smooth plot suggests that accuracy worsens slightly as notes played increase. In Panel B the joint pitch-IOI DTW distance confirms this. There are no perfect scores in the lower panel. Note that to approximately balance the contributions of pitch and IOI for the lower panel, both sets of pitch and IOI values were rescaled to 1-20, performance by performance, prior to the DTW determination (see Method section).

FANTASTIC analyzes pitch and duration aspects, with an extensive list of assessments including entropy, range, descriptive statistics, and the like. However, FANTASTIC was not apt for some of our intended assessments, particularly DTW and to some degree for automated analysis of retrogrades. While FANTASTIC was not ideal for our particular set of tasks, one potential application by researchers in related areas may be the inclusion of both AMMRI and FANTASTIC in tandem.

Each of our measures can be used individually, whenever it is relevant to participant learning. More broadly, we can ask how an overall current aural replication and improvisation skill level assessment might be made. In other words, how can one most fairly combine the several different assessments of each? In the case of replication, our DTW assessment of joint pitch/IOI precision is one such approach, but it may not necessarily be the most appropriate. In the case of improvisation, and when the method(s) used are known, then their specific assessment parameters are appropriate. But in cases where the method(s) are not specified beforehand, the question remains as to how we best proceed.

We have prepared the present simulation and assessment techniques so that we will be in a good position to assess these issues fully with our large complete dataset. For that purpose, it is desirable if not necessary to make

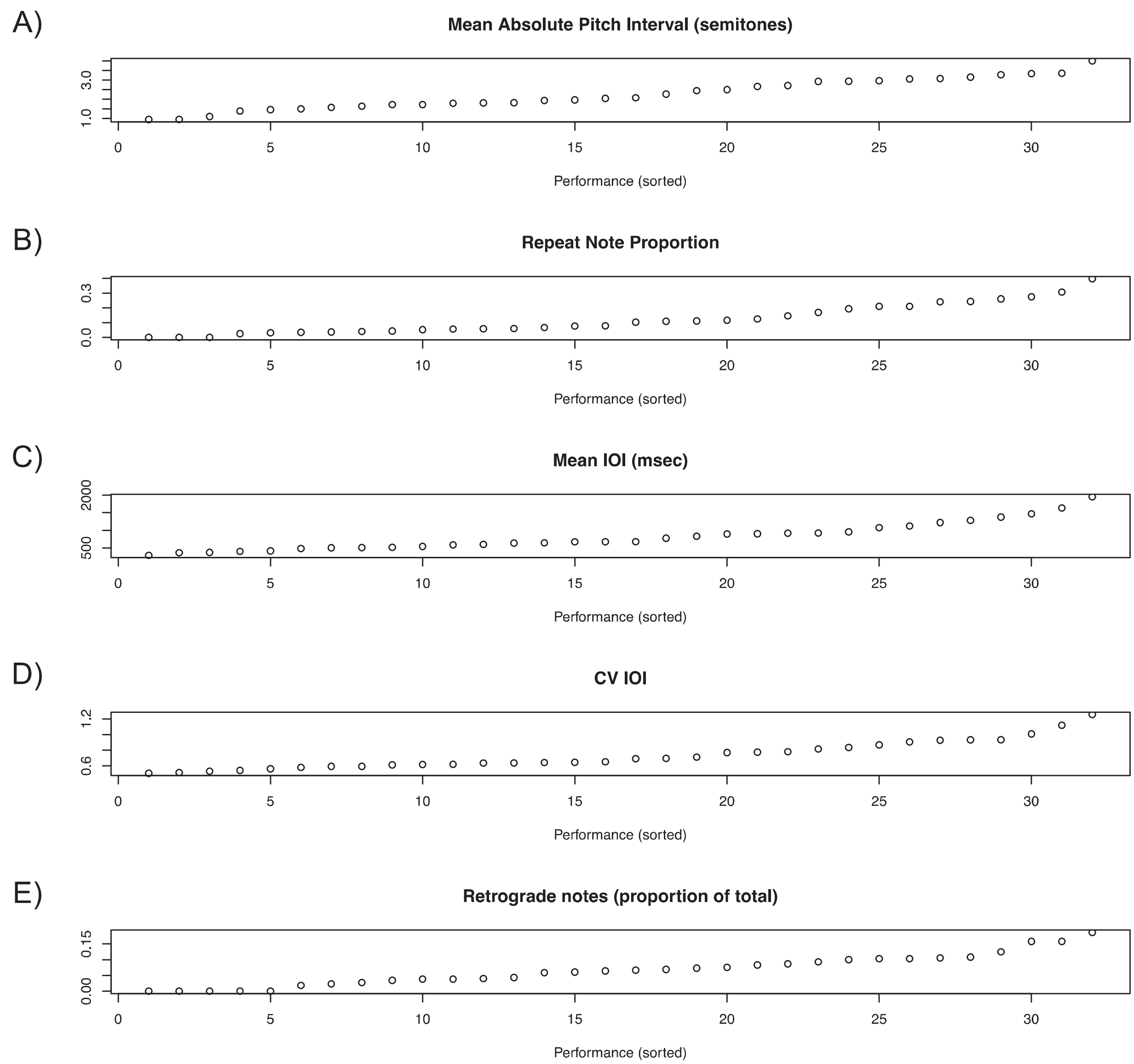

Figure 6. Measures of improvisations; 32 improvisations by members of our examined group are assessed for five parameters, and in each case, the results are sorted (thus vertically aligned points are not usually the same performance). Mean absolute pitch interval reflects the 'change the pitch intervals' measure; repeat note proportion the 'repeat notes' measure; mean IOI the 'vary the speed' measure; CV IOI (coefficient of variation of IOI values) the variability of IOIs, which should change in response to the 'vary the note lengths' method; and retrograde notes (the proportion of notes which belong to a retrograde). As expected, there are some zero values among the retrogrades. The non-zero values for the parameters all vary by a factor > 2, supporting their suitability for our performance evaluation purposes.

the measures monotonic, in the sense that as the fulfilment extent increases all measures should also increase. This means for example that distance measures should be converted into similarity measures (where similarity $\sim$ 1 – distance). In addition, measures that can readily change upwards or downwards (such as some timing measures consequent on changing tempo) might be converted into absolute change measures (in which the sign of the change is ignored). Given such considerations, one common approach that we consider is to weigh the assessment components in inverse proportion to the measure variances: when measure precision is low this may likely represent error as well as purposive behaviour. Such principles are commonly used in model selection and weighting procedures during regression modelling and machine learning. In any case, individual or merged-weighted assessments that are included in a model – for example, participant performance self-evaluation – can have their coefficients optimised during the modelling procedure. Note also that we do not assume there will

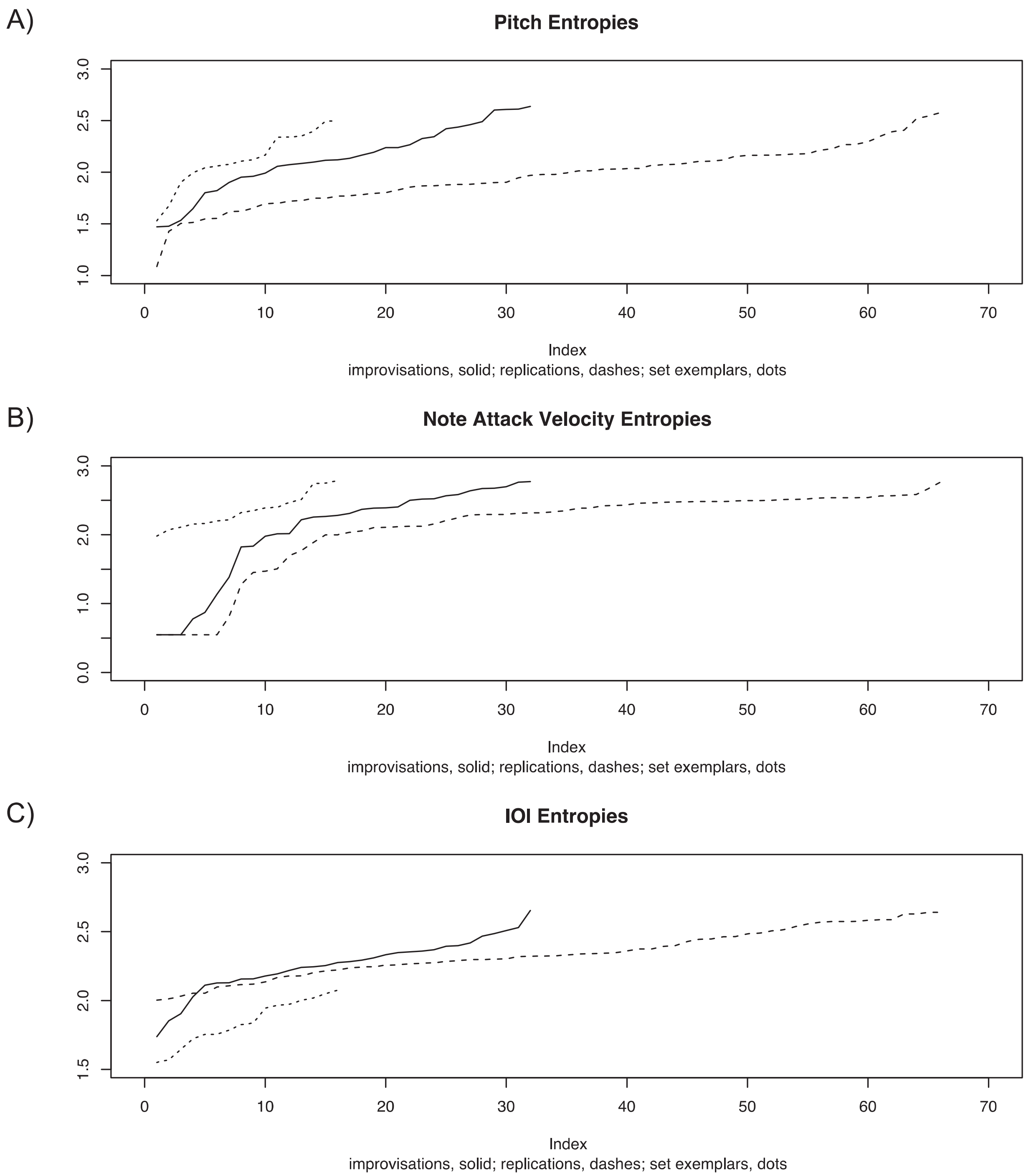

Figure 7. Pitch, velocity, and IOI entropies of all the set exemplar melodies, their replication performances, and the improvisations by the participants in our examined class, in their performance sessions. The results are sorted by entropy in each case.

be a linear response of perceived performance success in relation to extent of achievement of some particular improvisation method; it seems more likely that there might be an increase, as basic competence in the method is achieved, followed by a decrease if it is overused. This can potentially be captured with quadratic terms.

As mentioned in the Introduction, it is very difficult to find objective automated or even computational assessments of replication and improvisation ability in the literature. A brief consideration of two extensive edited books focusing on relevant music teaching will illustrate the point as neither contains any discussion of computational assessment, nor extensive consideration of the coherence (or otherwise) of expert evaluations. The volume edited by Heble and Laver concerns improvisation in and beyond the classroom, and does discuss the early stages of learning to improvise (Heble & Laver, 2016): there is no consideration of comparative assessment. Similarly, even in a volume focused on 'evaluation' (albeit primarily at the University level), there is no consideration of objective or computational assessment of improvisation or replication techniques, in spite of a wide and progressive set of ideas and authors (Encarnacao & Blom, 2020). Here, most discussion of evaluation, appropriately at this level, focuses on ensemble interaction and creativity.

It seems that assessment methods and simulations such as those presented in the present work will potentially be useful in many learning or practising contexts, formal and informal. Even though the literature is very sparse, there has been prior work towards manual or semi-automated quantitative assessment of both expert and novice performance. For example, in the study of the utility of isomorphic keyboard pitch layouts (quite distinct from those of the piano, our conventional electronic keyboard, or of *Thumbjam*), MIDI data was used to assess the accuracy of the performance of scales and simple melodies by students with at least 5 years of musical training (MacRitchie & Milne, 2017). A simplified version of combined pitch-time accuracy was developed and successfully exploited. In the case of expert professional musicians, a 'Note-time playing path' analysis has been developed to allow graphical comparison of score-implied pitch and timing with that executed, for the purpose of providing objective help during practice (de Graaff & Schubert, 2011). This approach requires the notation of the original score in an application (such as Sibelius or NoteAbility) that can then convert it into MIDI; and the complementary conversion of audio files of the ongoing performance (in this case using Melodyne) into MIDI. The two MIDI streams can then be compared graphically (or otherwise). Both these approaches provide data that could also be analysed with the suite of methods we have developed. It seems that assessment methods and simulations such as those presented in the present work will therefore be potentially useful in many learning or practising contexts, formal and informal.

In the future analyses of our own complete dataset, the most important musical issues will likely be: (1) to what degree did participants progress in their capacities; (2) how did participants improve their assessment scores with time as teaching proceeded; and (3) what aspects of the teaching and learning procedures were most influential on this. This latter question is particularly valid, as there is a dearth of literature examining musicians' specific uses of and attitudes towards digital technology in learning (Waddell & Williamon, 2019). Analyses on these questions will include the establishment of optimised weighting procedures, and can also be conducted with unweighted assessment inputs, since quite possibly the impact of some assessment measures on the overall progress will be insignificant, and best discarded from the models (which is not readily congruent with weighting them according to variance or entropy). It would also be interesting to make some simple comparisons of participant replication and improvisation with that of highly trained musicians; trained participants may not necessarily be higher in their abilities with any or all of these tasks. Rather, it may be that some minimum level of the present performance abilities can suffice for any high-level professional application. In other words, given a solid base, quite different performance learning (such as expressivity technique *per se*, or psychology of performance under pressure) is necessary to make the transition to being an effective professional performer.

The implications of the present study include the possible utility of the methods for learners' self-assessment as they seek to develop. This objective assessment might be moulded to the individual learner's expressive intents, by means of their own self-evaluations (made blind to the computational assessment). This would allow them to determine what features are most important for them to achieve the personal musical expression they wish; or even for them to adjust the emphasis of their ongoing learning to judgments made by their intended audience (be it family, friends, performing group, or beyond). Similarly, computational assessments such as these may be useful for professional musicians who wish to develop specified skills on which they have previously never focused.

Research on andragogy, referring to methods and practices of teaching older adults, has received a much smaller focus than teaching aimed at younger people. This disparity occurs in many creative research areas, including music (Lehmberg & Fung, 2010). However, older adults are able to learn new skills to a high degree of proficiency, as suggested by brain plasticity across the lifespan (Freitas et al., 2013). Based on emerging evidence suggesting that interaction with music in later life can lead to a range of potential benefits – such as in older adults' cognition (Bugos et al., 2007; Schneider et al., 2019), well-being (Creech et al., 2013; Krause et al., 2018), and motor function (Biasutti & Mangiacotti, 2018) – additional research on music andragogy has begun to take a foothold in recent years (e.g. Bugos & Cooper, 2019; MacRitchie et al., 2020; MacRitchie et al., 2022;

Seinfeld et al., 2013). Our intent is that our methods will contribute to this.

Single-handed performance (as examined here) is undoubtedly far more limited than two-handed performance. Nevertheless, important literature exists concerning musical compositions for one-handed piano performance, and also concerning notable improvisers with less than normal physical two-handedness, such as the eminent jazz pianist Horace Parlan (Lubet, 2010). But we can ask whether the approaches developed here can be extended to two-handed performance: they can, and multistrand and strand-interactive features can be analysed by more highly developed code, with due attention to strand segregation (Bregman et al., 2001; Meredith, 2013). Among other issues, the definition of motivic units, melodic elements that may be shared between strands (such as the TECs proposed by Meredith, 2013), and the corresponding harmonic considerations, together with those of tonality and post-tonality, metricality and post-metricality (George & Bregman, 1989) are complex, but not completely opaque.

Conclusion

In this paper, we presented a novel approach and accompanying code (via an annotated *R* script) for evaluating single-handed performances in the *replication* and *improvisation* of melodic fragments. We used fragments recorded from our earliest chronological participant group to determine the feasibility of the analyses, in anticipation of the larger dataset now our music education program for older adult novices concluded. Based on our findings we suggest that AMMRI appears to be an effective method, both for teachers and students wishing to perform self-assessment. In addition to the eventual examination of our overall dataset, subsequent steps in the development process can examine the suitability of the enclosed tools for broader application in music evaluation, such as the use of chords, a second hand, longer fragments, and the use of fragments produced by expert performers.

AMMRI can also function as an informative tool for both students and teachers, such that they can undertake assessment or self-assessment of performances. In this way, it can be seen as a tool to enhance musical learning both in and outside the classroom. AMMRI produces broad and objective estimations of the 'quality' or 'accuracy' of performances, with quality or accuracy in this context referring to the level that a performance is able to reproduce an existing melody (regarding precision of replication) or is able to successfully achieve a specified technique of musical improvisation. We assume, in common with most music pedagogy, that the achievement of effectiveness in such tasks (and hence the availability of objective assessment tools for them) is a prerequisite to musical expression and ultimately creativity on a keyboard instrument, and that this, in turn, contributes to overall musical appreciation and understanding. The relative lack of study in musical andragogy gives additional importance to our current study.

At the conclusion of our education program, the development of these analysis tools will also allow us to broadly quantify the level of musical development that has occurred for our participants, such as by identifying which techniques were most successfully performed.

We plan to use these methods in the later full study to determine whether learning is successful and to assess how the different learning contexts (the two instruments, and the two musical activities of reproduction and improvisation) compare in their effects.

After the completion of the whole training period for all participants and subsequent analysis the following hypotheses will be assessed:

1. These skills can be developed within a one-year period of group study with a professional teacher using two differing keyboard instruments, and two different approaches to making music.
2. Fine motor and some cognitive skills can be enhanced as a result of the year's study.

Acknowledgements

The authors would like to acknowledge the important contributions of Catherine J. Stevens and Patrick O'Donnell to this ongoing project.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by an Australian Research Council Discovery Project Grant (*Maintaining active minds and bodies through older adult music education*, DP190102012).

Ethics statement

Prospective participants all agreed to participate and completed a written consent form. The study received Human Research Ethics Approval (Western Sydney University, Approval Number H13206).

ORCID

Roger T. Dean http://orcid.org/0000-0002-8859-8902
Anthony Chmiel http://orcid.org/0000-0003-3294-0534
Madeleine Radnan http://orcid.org/0000-0003-3836-1162

John R. Taylor http://orcid.org/0000-0002-4435-0657
Jennifer MacRitchie http://orcid.org/0000-0003-4183-6552

References

Biasutti, M., & Mangiacotti, A. (2018). Assessing a cognitive music training for older participants: A randomised controlled trial. *International Journal of Geriatric Psychiatry*, *33*(2), 271–278. https://doi.org/10.1002/gps.4721

Bregman, A. S., Ahad, P. A., & Van Loon, C. (2001). Stream segregation of narrow-band noise bursts. *Perception & Psychophysics*, *63*(5), 790–797. https://doi.org/10.3758/BF03194438

Bugos, J. A., & Cooper, P. (2019). The effects of mallet training on self-efficacy and processing speed in beginning adult musicians. *Research Perspectives in Music Education, 20*(1), 21–32.

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging & Mental Health*, *11*(4), 464–471. https://doi.org/10.1080/13607860601086504

Cannam, C., Landone, C., & Sandler, M. (2010). *Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files*. Paper presented at the proceedings of the 18th ACM international conference on multimedia, New York.

Creech, A., Hallam, S., McQueen, H., & Varvarigou, M. (2013). The power of music in the lives of older adults. *Research Studies in Music Education*, *35*(1), 87–102. https://doi.org/10.1177/1321103X13478862

Cycling '74. (2021). Max/MSP. Walnut, CA.

Dean, R. T. (1989). *Creative improvisation: Jazz, contemporary music, and beyond: how to develop techniques of improvisation for any musical context*. Open University Press.

Dean, R. T., Bailes, F., & Drummond, J. (2014). Generative structures in improvisation: Computational segmentation of keyboard performances. *Journal of New Music Research*, *43*(2), 224–236. https://doi.org/10.1080/09298215.2013.859710

de Graaff, D. L., & Schubert, E. (2011). Analysing playing using the note-time playing path. *Behavior Research Methods*, *43*(1), 278–291. https://doi.org/10.3758/s13428-010-0041-0

Dixon, S. (2001). An *interactive beat tracking and visualisation system*. Paper Presented at the ICMC.

Dixon, S. (2005). *Live tracking of musical performances using online time warping*. Paper Presented at the Proceedings of the 8th International Conference on Digital Audio Effects.

Dixon, S., & Widmer, G. (2005). *Match: A music alignment tool chest*. Paper Presented at the ISMIR.

Encarnacao, J., & Blom, D. (Eds.). (2020). *Teaching and evaluating music performance at university. Beyond the conservatory model*. Routledge.

Freitas, C., Farzan, F., & Pascual-Leone, A. (2013). Assessing brain plasticity across the lifespan with transcranial magnetic stimulation: Why, how, and what is the ultimate goal? *Frontiers in Neuroscience*, *7*, 42. https://doi.org/10.3389/fnins.2013.00042

George, M. F.-S., & Bregman, A. S. (1989). Role of predictability of sequence in auditory stream segregation. *Perception & Psychophysics*, *46*(4), 384–386. https://doi.org/10.3758/BF03204992

Heble, A., & Laver, M. (Eds.). (2016). *Improvisation and music education. Beyond the classroom.* Routledge.

Krause, A. E., Davidson, J. W., & North, A. C. (2018). Musical activity and well-being. *Music Perception, 35*(4), 454-474. https://doi.org/10.1525/mp.2018.35.4.454

Lehmberg, L. J., & Fung, C. V. (2010). Benefits of music participation for senior citizens: A review of the literature. *Music Education Research International*, *4*(1), 19–30.

Louboutin, C., & Meredith, D. (2016). Using general-purpose compression algorithms for music analysis. *Journal of New Music Research*, *45*(1), 1–16. https://doi.org/10.1080/09298215.2015.1133656

Lubet, A. (2010). *(Paralyzed on one) Sideman: Disability studies meets jazz, through the hands of Horace Parlan.* Paper presented at the Guelph jazz festival colloquium, Guelph, CA.

MacRitchie, J., Breaden, M., Milne, A. J., & McIntyre, S. (2020). Cognitive, motor and social factors of music instrument training programs for older adults' improved wellbeing. *Frontiers in Psychology*, *10*, 2868. https://doi.org/10.3389/fpsyg.2019.02868

MacRitchie, J., Chmiel, A., Radnan, M., Taylor, J. R., & Dean, R. T. (2022). Going online: Successes and challenges in delivering group music instrument and aural learning for older adult novices during the COVID-19 pandemic. *Musicae Scientiae*, 10298649221097953.

MacRitchie, J., & Milne, A. J. (2017). Exploring the effects of pitch layout on learning a new musical instrument. *Applied Sciences*, *7*(12), 1218. https://doi.org/10.3390/app7121218

Meredith, D. (2013). *COSIATEC and SIATECCompress: Pattern discovery by geometric compression.* Paper presented at the Music Information Retrieval Evaluation eXchange (MIREX 2013).

Meredith, D. (2014). *Compression-based geometric pattern discovery in music.* Paper Presented at the 4th International Workshop on Cognitive Information Processing (CIP).

Müllensiefen, D. (2009). *Fantastic: Feature Analysis Technology Accessing Statistics (in a Corpus): Technical Report v1.* http://www.doc.gold.ac.uk/isms/m4s/FANTASTIC_docs.pdf

Pressing, J. (1988). Improvisation: Methods and models. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 129–178). OUP.

Schneider, C. E., Hunter, E. G., & Bardach, S. H. (2019). Potential cognitive benefits from playing music among cognitively intact older adults: A scoping review. *Journal of Applied Gerontology*, *38*(12), 1763–1783. https://doi.org/10.1177/0733464817751198

Seinfeld, S., Figueroa, H., Ortiz-Gil, J., & Sanchez-Vives, M. V. (2013). Effects of music learning and piano practice on cognitive function, mood and quality of life in older adults. *Frontiers in Psychology*, *4*, 810. https://doi.org/10.3389/fpsyg.2013.00810

Thibeault, M. D. (2021). Aebersold's mediated play-A-long pedagogy and the invention of the beginning jazz improvisation student. *Journal of Research in Music Education*, *70*(1), 66–91. 00224294211031894.

ThumbJam. (2017). Sonosaurus LLC [Mobile Application Software].

Waddell, G., & Williamon, A. (2019). Technology use and attitudes in music learning. *Frontiers in ICT*, *6*(11). https://doi.org/10.3389/fict.2019.00011