
Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise

Bruno Gingras[¶]§, Tamara Lagrandeur-Ponce, Bruno L Giordano, Stephen McAdams
CIRMMT, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montréal, Quebec
H3A 1E3, Canada; § also Department of Cognitive Biology, University of Vienna, Vienna, Austria
Received 16 December 2010, in revised form 11 September 2011; published online 14 November 2011

Abstract. Can listeners distinguish unfamiliar performers playing the same piece on the same instrument? Professional performers recorded two expressive and two inexpressive interpretations of a short organ piece. Nonmusicians and musicians listened to these recordings and grouped together excerpts they thought had been played by the same performer. Both musicians and nonmusicians performed significantly above chance. Expressive interpretations were sorted more accurately than inexpressive ones, indicating that musical individuality is communicated more efficiently through expressive performances. Furthermore, individual performers' consistency and distinctiveness with respect to expressive patterns were shown to be excellent predictors of categorisation accuracy. Categorisation accuracy was superior for prize-winning performers compared to non-winners, suggesting a link between performer competence and the communication of musical individuality. Finally, results indicate that temporal information is sufficient to enable performer recognition, a finding that has broader implications for research on the detection of identity cues.

1 Introduction

Humans possess a remarkable and well-documented ability to identify individuals on the basis of their physical characteristics, such as facial features (Carey 1992), movements (Blake and Shiffrar 2007), or voice (Belin et al 2004). This ability, which is essential for survival in species that value kin recognition (Tang-Martinez 2001) and social interaction (Thompson and Hardee 2008), is not specific to humans: monkeys can recognise faces or vocalisations (Rendall et al 1996; Sugita 2008), while seals (Charrier et al 2001) and chickadees (Charrier and Sturdy 2005) can reliably identify conspecifics based on their vocal production.

Voice is not the only culturally universal auditory stimulus capable of carrying identity information; music appears to be another likely candidate. Indeed, certain famous musicians, such as John Coltrane or Sonny Rollins, need only to play a few notes to be unequivocally recognised (Benadon 2003). Yet, whereas the processes by which speaker identification is achieved have been extensively analysed, no published study has focused explicitly on the communication of identity information through music performance in a controlled experimental setting. This article aims to fill that void by examining the influence of listeners' musical training, performers' expertise, and performance expressiveness on the listeners' ability to associate different performances by the same performer in a sorting task.

A great deal of research on voice recognition has been devoted to the perception of characteristic attributes of individual speakers. The classical model, supported by substantial evidence, posits a perceptual dissociation between the linguistic features of an utterance, related to prosody and higher-level phonetic structures, and its indexical or 'paralinguistic' features, described as speaker-specific acoustical properties related to vocal timbre (Abercrombie 1967). However, more recent accounts have shown that speakers can be identified from sinewave replicas of their voices lacking natural intonation and

[¶] Please address all correspondence to Bruno Gingras, Althanstrasse 14, A 1090 Vienna, Austria; e-mail: bruno.gingras@univie.ac.at

voice quality, implying that other prosodic properties convey sufficient speaker-specific information to allow identification (Remez et al 1997, 2007). Further converging evidence indicating that acoustically based identification is possible in the absence of timbral cues is provided by studies showing that temporal information is sufficient for recognising one's own clapping (Flach et al 2004), and that pianists recognise their own performances better than those of other pianists from modified recordings in which only timing information is available (Repp and Knoblich 2004; Repp and Keller 2010). However, self-recognition may be explained by greater experience with one's own productions.

Artificial intelligence experts have also attempted to build computational models that could identify music performers (Saunders et al 2008; Stamatatos and Widmer 2005). For instance, Stamatatos and Widmer presented a learning ensemble that achieved a 70% recognition rate, using a database of piano performances of 22 pianists playing two pieces by Chopin. The authors noted that their model displayed a level of accuracy "unlikely to be matched by human listeners".

The principal objective of the present study was, therefore, to assess whether human listeners could accurately group together excerpts played by the same (non-self) performer in the absence of timbral or dynamic differentiation by relying exclusively on temporal cues. Organ music, which offers only limited timbral and dynamic differentiation, was chosen to address this question. Furthermore, all recordings were realised on the same instrument and using the same registration (configuration of organ stops controlling the timbre), thus ensuring that timbre and intensity were identical for all performances.

A secondary objective was to identify the acoustical parameters used by listeners to discriminate between performers. Given the lack of timbral or dynamic differentiation between the excerpts, we sought to estimate the relative contribution of available temporal cues, such as tempo and articulation (the degree of overlap between two successive notes), to the listeners' solutions in the sorting task.

Domain-specific expertise has been shown to affect one's ability to identify specialised stimuli in both the visual and auditory domains. This is particularly true in the case of musical stimuli, as several studies have demonstrated a timbre-processing and pitch-processing advantage for musicians versus nonmusicians (for a review, see Chartrand et al 2008). On the other hand, a number of responses to musical stimuli do not seem to be affected to a great extent by musical training (Bigand and Poulin-Charronnat 2006). More specifically, both musicians and nonmusicians are able to distinguish among different levels of expressiveness in performances of the same piece (Kendall and Carterette 1990), to recognise the emotions that performers intend to communicate (Juslin 2000), and to discriminate between familiar and novel performances of the same piece (Palmer et al 2001). Especially relevant to the topic at hand is the finding by Timmers (2005) that predictive models of perceptual similarity between performances were highly similar for nonmusicians and musicians: for both groups, models based on absolute values of tempo and loudness were better predictors of perceptual similarity than models based on normalised variations. These studies suggest that the effect of musical expertise could be task-dependent. In the present study, the role of musical expertise was assessed by comparing musicians' and nonmusicians' ability to discriminate between performers.

Because earlier research has shown that recordings from performers with a higher level of expertise tend to be perceived by listeners as exhibiting more individuality than those of graduate students (Repp 1997), we also sought to explore a potential link between the performers' level of expertise and listeners' ability to correctly identify their performances. Therefore, performers were divided into two groups: organists having previously won one or more prizes at national organ competitions and non-prize-winners. We surmised that prize-winners would be easier to identify, either because their superior technical proficiency would result in more consistent performances or because their

success in competitions could be related to their ability to successfully convey their musical individuality.

A further objective of this study was to assess the effect of musical expressiveness on the ability of listeners to discriminate between performers. Previous research on movement-based recognition from point-light displays has shown that subjects recognise actors more accurately when they perform expressive actions, such as boxing or dancing, instead of less expressive actions, such as walking (Loula et al 2005); similarly, exaggerated temporal differences between arm movements enhance recognition (Hill and Pollick 2000). On the basis of these findings, we predicted that a performer's musical individuality would be conveyed more clearly when performing a piece in an expressive manner, rather than in an inexpressive rendition. In order to evaluate whether listeners' ability to accurately group together recordings from the same performer was higher for expressive performances than for inexpressive ones, performers were asked to record expressive and inexpressive interpretations of the piece.

2 Methods

2.1 *First phase: Recording organ performances*

2.1.1 *Participants.* Eight professional organists participated in exchange for financial compensation. Their mean age was 26 years (range: 19 to 30 years). All performers identified themselves as right-handers. They had received organ instruction for a mean duration of 9 years (range: 3 to 13 years) and had 4 to 21 years of experience playing the organ. All of them held or had held a position as church organist for an average of 8 years (range: 1 to 21 years). Three of them had previously won one or more prizes in national organ competitions.

2.1.2 *Materials.* A chorale setting of *Wachet auf, ruft uns die Stimme* by Samuel Scheidt (1587–1654) was selected for this experiment. This piece, which is representative of a typical genre of the Baroque organ repertoire, is not technically demanding and does not require the use of the pedals. None of the performers was familiar with the piece prior to the experiment.

2.1.3 *Procedure.* The score of the piece was given to the performers 20 min before the recording session began, in order to give them time to practice. Performers were asked to practice two interpretations of the piece: an expressive interpretation and an inexpressive interpretation, for which they were instructed to play as mechanically as possible without adding any expressiveness beyond what was notated in the score (Palmer 1989). Two versions of the expressive interpretation were recorded, followed by two versions of the inexpressive interpretation. Performances were recorded on the Casavant organ of the Church of St-Andrew & St-Paul in Montreal (Canada), which is equipped with a MIDI console (Solid State Organ Systems) that allows precise recording of the moments at which organ keys are pressed and released. The scanning rate of the MIDI system was estimated at 750 Hz (1.33 ms), the on and off points being determined by key-bottom contact.⁽¹⁾ For the experiment, the stops used were the Spitz Principal 8', the Spitz Principal 4', and the Fifteenth 2' on the *Great* manual. All performers used the same registration (configuration of organ stops controlling the timbre), thus ensuring that timbre and intensity were identical for all performances. Because the organ used was an electro-pneumatic organ, performers had no direct control over the attack and decay phases (the valve is either open or closed). The audio signal was recorded through two Boehringer ECM 8000 omnidirectional microphones. The microphones were located 1.20 m behind the organ bench, at a height of 1.70 m, and were placed 60 cm apart. Audio and MIDI signals were sent to a PC computer through a MOTU audio interface.

⁽¹⁾ Information provided by Mark Gilliam, sales manager of Solid State Organ Systems.

2.2 Second phase: Sorting task

2.2.1 Participants. Twenty students with 2 or fewer years of musical training and limited exposure to organ music (mean age = 22.1 years, SD = 2.8 years; mean number of years of training = 0.6 years, SD = 0.8 years, range: 0 to 2 years), henceforth referred to as nonmusicians, and twenty students having completed at least a year in an undergraduate degree in music performance or musicology (mean age = 21.6 years, SD = 1.9 years; mean number of years of training = 13.8 years, SD = 4.1 years, range: 7 to 20 years), henceforth referred to as musicians, participated for course credit or financial compensation.⁽²⁾

2.2.2 Materials. The first musical phrase of the piece, representing a syntactically coherent musical unit, was used in the sorting task (figure 1). In order to reduce the number of excerpts to a manageable number, the recordings of the performers with the fastest and slowest mean tempi were discarded. All four recordings from the remaining six performers (three prize-winners and three non-prize-winners) were used. In addition, an identical duplicate of the first expressive performance of each organist was added as an experimental control. Thus, there was a total of 30 excerpts ranging in duration from 10 to 14 s. The task was practiced with a reduced training stimulus set using three recordings (two expressive, one inexpressive) by two different performers of a phrase of similar character and duration from the same piece, for a total of six excerpts.



Figure 1. Excerpt from Samuel Scheidt's chorale setting of *Wachet auf, ruft uns die Stimme* used for the main phase of the listening experiment (boxed section).

2.2.3 Procedure. The experimental interface consisted of a computer monitor on which the excerpts were represented by 30 randomly numbered icons which would play the corresponding audio recordings when clicked. Participants were instructed to play each excerpt in any order they liked and to sort them by moving the icons into one of six boxes, corresponding to each of the six performers. The sorting task was constrained: listeners were informed that the excerpts had been recorded by six different performers and that each performer had recorded the piece five times. However, they were not made aware that performers had recorded inexpressive and expressive versions and that some excerpts were duplicates. Listeners were free to rearrange their categorisation and listen to the excerpts as many times as necessary. Prior to the sorting task, listeners practiced the task in a familiarisation phase using six alternate excerpts, consisting of a musical phrase similar in character and duration to the one used in the sorting task, recorded by two performers. No feedback was provided on the listeners' performance in the familiarisation phase. The experiment took place in a sound-attenuated booth. Listeners wore Sennheiser HD 280 Pro headphones and were screened for peripheral hearing problems by a standard audiometric procedure. Following the experiment, participants filled out a questionnaire concerning their musical background and listening habits.

⁽²⁾Note that musicians admitted to university-level training at McGill University are required to demonstrate a high level of proficiency in music performance.

3 Results

3.1 Characterisation of the musical features of the excerpts

In order to compare the excerpts on the basis of their musical features, an analysis was conducted on the following parameters: mean tempo (expressed as mean quarter-note duration), tempo variability (expressed as the coefficient of variation of the quarter-note duration, which corresponds to the standard deviation of the quarter-note duration normalised by the mean duration), articulation (expressed as the degree of overlap between successive notes), and onset asynchrony (referring to the difference in onset times between notes that are attacked simultaneously in the score, such as notes belonging to the same chord). These parameters essentially comprise the range of expressive factors that are controlled by the performer in Baroque organ music (excluding registration effects, which were controlled for in this experiment). Table 1 lists the mean values for these parameters averaged over the four recordings of each performer (identified by the letters A to F). Since the purpose of this analysis was to compare the excerpts both on the basis of their respective performers and of their expressive intent, mixed-model analyses of variance were conducted for each of the aforementioned parameters, with performer as a random factor and expressive intent as a fixed factor, on the 24 excerpts that were used in the sorting task (table 2). A-posteriori tests (Tukey HSD) were conducted to identify the parameters on the basis of which individual performers could be significantly differentiated.

Table 1. Mean values for the expressive parameters, averaged for each performer.





Expressive parameters	Performer					
	A*	B	C	D*	E	F*
Mean quarter-note duration/ms	637 (5)	716 (41)	723 (14)	839 (11)	832 (63)	656 (23)
Mean coefficient of variation of quarter-note duration/%	8.54 (0.76)	5.91 (0.68)	7.15 (0.76)	8.02 (3.18)	7.03 (1.65)	6.27 (2.66)
Mean overlap/ms	-63 (61)	-106 (40)	-44 (10)	-160 (37)	-109 (34)	-161 (9)
Mean onset asynchrony/ms	9.0 (0.9)	10.7 (1.2)	7.8 (0.4)	8.5 (2.6)	8.5 (0.8)	7.2 (0.4)

Note: Prize-winners are indicated with an asterisk. Standard deviations of the mean values obtained for each of the four recordings per performer are given in parentheses.

3.1.1 Mean tempo. The inter-onset interval (IOI) was computed for each quarter note for all excerpts, and the mean value obtained for each excerpt was used as a measure of mean tempo. Significant differences in mean quarter-note duration were observed between performers, but no effect of expressive intent was observed. A-posteriori tests confirmed that performers could be divided into three distinct groups on the basis of the mean tempo, comprising D and E (slowest tempi), B and C (medium tempi), and A and F (fastest tempi).

3.1.2 Tempo variability. The coefficient of variation of the IOI, obtained by dividing the standard deviation of the IOI by the mean quarter-note duration and expressing the result as a percentage of the mean quarter-note duration, was used as a measure of the degree of tempo variability. An effect of expressive intent was observed, indicating that tempo variations were significantly smaller for inexpressive excerpts than for expressive ones. These results are congruent with previous studies reporting that expressive performances typically exhibit more pronounced tempo variability than mechanical or 'deadpan' performances (Palmer 1989). In contrast to mean tempo, performers did not vary significantly with respect to the amount of tempo variation they used.

Table 2. Analyses of variance for the expressive parameters of the excerpts used in the sorting task.

Expressive parameters	Factors			A-posteriori tests
	performer $F_{5,12}$	expressive intent $F_{1,5}$	performer × expressive intent $F_{5,12}$	comparison by performer
Mean quarter-note duration	59.88***	2.22	3.91*	
Mean coefficient of variation of quarter-note duration	2.73	7.19*	2.62	
Mean overlap	44.57***	0.07	20.18***	
Mean onset asynchrony	5.54**	0.56	3.11*	

Note. For a-posteriori tests, performers whose means did not differ significantly are circled together; means are ranked from the highest to the lowest. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

3.1.3 Articulation. Mean overlap was defined as the time interval between two consecutive notes, and calculated as the offset of note event n minus the onset of note event $n + 1$. A positive overlap indicates a legato articulation, while a negative value represents a detached or staccato articulation. Significant differences in the amount of overlap were found between performers, but no effect of expressive intent was observed. A-posteriori tests confirmed that performers could be divided into three distinct groups on the basis of the mean overlap, with D and F using a very detached articulation, while A and C played quasi-legato. The highly significant interaction between performer and expressive intent reflects the fact that some organists performed the mechanical excerpts in a more staccato fashion than the expressive ones, while others did the exact opposite. In contrast, Palmer (1989) reported that pianists played unexpressive excerpts in a more detached way than expressive ones.

3.1.4 Onset asynchrony. Onset asynchrony was measured as the standard deviation of the difference in onset times between notes of a chord (Palmer 1989; Rasch 1979). As with other expressive parameters analysed here, the degree of synchronisation differed significantly among performers. However, unlike results reported for piano performance (Palmer 1989), onset asynchronies were not significantly larger in expressive performances than in mechanical ones. Given the lack of dynamic differentiation on the organ, these results are perhaps not unexpected, in light of more recent studies suggesting that onset asynchrony is related to dynamic differentiation between voices (Gingras 2006; Goebel 2001; Repp 1996). It should also be noted that asynchronies across all excerpts averaged 9 ms (SD = 2 ms), which is noticeably less than the asynchronies of 15 to 20 ms which are typically observed in piano performance (Palmer 1989). It is therefore unlikely that excerpts from different performers could have been segregated on the basis of differences in onset asynchrony, since the reported threshold for detecting onset asynchronies is around 20 ms (Hirsh 1959).

From these analyses, we may conclude that performers could be statistically differentiated on the basis of mean tempo, mean overlap, and amount of onset asynchrony, although the latter may not have been a perceptually relevant parameter given the small size of the asynchronies observed here.⁽³⁾ Importantly, these analyses indicate that some performers, such as D and E, could not be discriminated solely on the basis of differences in mean tempo. Furthermore, we observe that the only statistically significant difference between expressive and inexpressive excerpts lies in the magnitude of the tempo variability.

3.2 *Analysis of the sorting data*

Sorting data were analysed in terms of listeners' categorisation accuracy and underlying perceptual space. From the standpoint of the performer, we analysed the effects of performer expressiveness and expertise on categorisation, as well as differences in categorisation accuracy for individual performers.

3.3 *General assessment of the listeners' categorisation accuracy*

A measure of the listeners' categorisation accuracy can be obtained by comparing their partitioning of the excerpts with the correct solution, which corresponds in this case to a partition in which all the excerpts recorded by the same performer are grouped together. Categorisation accuracy was assessed on the basis of the adjusted Rand index (Hubert and Arabie 1985), a chance-corrected measure of the agreement between the correct categorisation solution and that which was arrived at by the participants. Positive adjusted Rand index values, indicating above-chance categorisation accuracy, were obtained for 20 musicians (100% of the participants) and 18 nonmusicians (90%). Furthermore, the adjusted Rand index was significantly above chance for 15 musicians (75%) and 13 nonmusicians (65%) ($p < 0.05$) estimated with bootstrap methods (Efron and Tibshirani 1993). Although musicians fared slightly better than nonmusicians, no significant difference was observed between the adjusted Rand index values of the two groups ($t_{38} = 1.24$, $p = 0.22$).

3.4 *Representing the listeners' perceptual space*

A multidimensional scaling (MDS) analysis was conducted on the listeners' partitioning of the excerpts in order to uncover the main dimensions of their perceptual space. We assumed that excerpts that were often grouped together were perceived to be more similar than excerpts that were not (Arabie and Boorman 1973). In a group analysis, the INDSCAL procedure (Carroll and Chang 1970) was used to estimate the relative weight of each spatial dimension for each individual listener. A two-dimensional group space (shown in figure 2) provided the best fit as determined by fit-by-dimensionality analyses taking into account both the stress measure (Kruskal stress-I = 0.15; range of participant-specific values: 0.06–0.22) and the proportion of variance explained (RSQ = 0.87; range of participant-specific values: 0.71–0.98).

3.4.1 *Interpretation of the dimensions.* As a means of interpreting the listeners' perceptual space recovered from the MDS analysis shown in figure 2, regression models were used to predict the MDS dimensions (Kruskal and Wish 1978), based on the four parameters used to characterise the musical features of the excerpts, namely mean tempo, tempo variability, overlap, and onset asynchrony. Importantly, this analytical approach assumes that the timing information extracted from the MIDI data corresponds closely to the perceived overlap or onset asynchrony. To that end, the organ registration used here only made use of pipes 8' or shorter, avoiding longer bass organ pipes, which may need more time to reach full amplitude (Braasch and Ahrens 2000). In addition, the music spanned a limited pitch range comprising slightly more than two octaves in

⁽³⁾ Although the analyses presented here refer only to the first phrase of Scheidt's chorale setting which was used in the sorting task, similar results were obtained for performances of the entire piece.

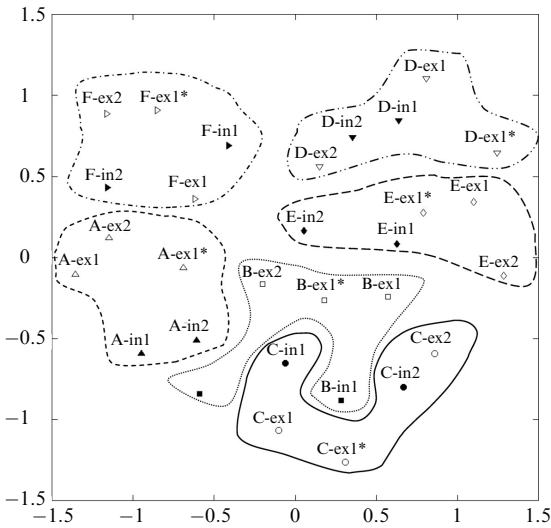


Figure 2. Multidimensional scaling (INDSCAL) of the perceptual distances between excerpts, based on the co-occurrence matrices from the sorting task. Letters A to F identify individual performers; ex (open symbols) refers to expressive excerpts and in (filled symbols) to inexpressive ones. Numbers refer to the order of recording. Asterisks indicate duplicate excerpts. Excerpts from the same performer are enclosed in the same region.

the excerpts selected for both the training phase and the main phase of the listening experiment, thus minimising potential differences in temporal envelope between lower- and higher-pitched organ tones. Finally, the analyses presented here only involve comparisons on mean values computed over the entire excerpt, thus balancing out any envelope differences that would affect these measures.

Mean tempo was the sole parameter that correlated significantly with coordinates on the first MDS dimension (abscissa) ($r_{28} = 0.89$, $p < 0.001$, $F_{1,28} = 102.94$, $p < 0.001$), whereas only mean overlap correlated significantly with coordinates on the second dimension (ordinate) ($r_{28} = -0.78$, $p < 0.001$, $F_{1,28} = 31.12$, $p < 0.001$). These observations suggest that tempo and articulation were the most important parameters used by listeners to discriminate between performers.

3.4.2 Individual weights. Musicians (mean weights for the first dimension: 0.70, $SD = 0.03$; for the second dimension: 0.69, $SD = 0.03$), and nonmusicians (mean weights for the first dimension: 0.70, $SD = 0.02$; for the second dimension: 0.68, $SD = 0.01$) ascribed nearly identical importance to both dimensions, indicating that tempo and articulation were of equal perceptual relevance in the sorting task for both these groups.

3.5 Effect of expressiveness on categorisation accuracy

To analyse the effect of the expressiveness of the performances on the listeners' ability to sort them correctly, the participants' partitions must be decomposed by comparing the performer's identity and the expressive intent of excerpts that were grouped together. Such analyses typically involve comparisons of pairs of excerpts (Daws 1996; Miller 1969). The proportion of pairs of excerpts grouped together (observed pairs) out of the total number of possible pairs was then computed for each of the following types of pairs (figure 3):

- (a) One inexpressive and one expressive excerpt from different performers ($I \neq E$ in figure 3).
- (b) Two expressive excerpts from different performers ($E \neq E$).
- (c) Two inexpressive excerpts from different performers ($I \neq I$).
- (d) One inexpressive and one expressive excerpt from the same performer ($I = E$).
- (e) Two inexpressive excerpts from the same performer ($I = I$).
- (f) Two expressive excerpts from the same performer ($E = E$).
- (g) Two identical expressive excerpts from the same performer (duplicates).

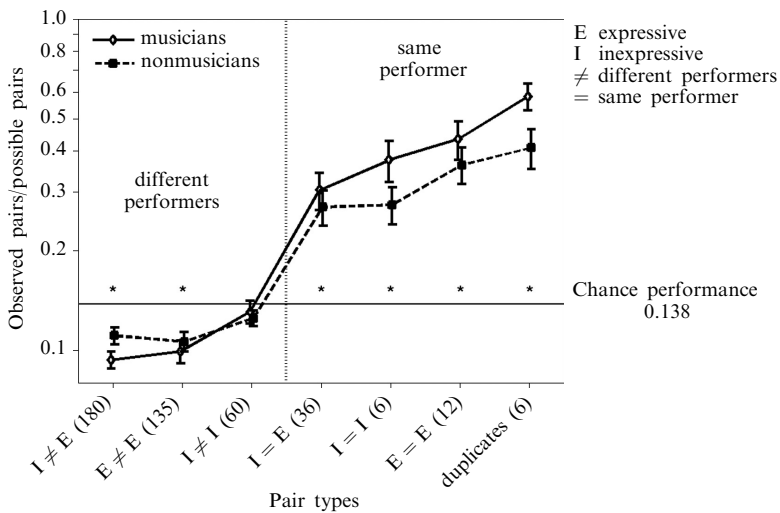


Figure 3. Proportion of observed pairs compared to the total number of possible pairs (indicated in parentheses) for all pair types. Error bars indicate standard errors of the mean. Asterisks indicate values that are significantly different from chance performance (for both musicians and nonmusicians) as determined by two-tailed single-sample *t* tests (Bonferroni-corrected, $p < 0.02$ in all cases). Points are connected to help distinguish the two participant groups visually.

Combinatorial probabilities are used to determine the chance performance level. In the present case, a partition of 30 excerpts into six groups of five excerpts contain 60 pairs of excerpts $\{6 \times [5!/(3! \times 2!)]\}$. The total number of possible pairs is given by $30!/(28! \times 2!)$, yielding 435. The chance performance level corresponds to the probability of randomly assigning a pair to a given partition, which is equivalent to $60/435$, or $p = 0.138$.

The proportion of observed pairs was significantly above chance for all types of correct pairs (representing excerpts from the same performer), with expressive–expressive pairings occurring more frequently than pairs comprising at least one inexpressive excerpt. For incorrect pairs (corresponding to excerpts played by different performers), the proportion of observed pairs involving at least one expressive excerpt was significantly below chance, while the frequency of inexpressive–inexpressive pairings did not differ significantly from chance.

A repeated-measures logistic regression analysis was conducted on the proportion of pairs of excerpts grouped together with the following factors: listener’s musical training (musician or nonmusician), and, for each pair of excerpts, performer identity (same for both excerpts/different for each excerpt) and expressiveness (inexpressive/inexpressive, inexpressive/expressive, and expressive/expressive). Duplicate pairs were included with the same-performer expressive–expressive pairings for this analysis.

Listeners grouped pairs of excerpts played by the same performer more often than those from different performers ($\chi^2_1 = 25.91$, $p < 0.001$). Expressive excerpts were categorised more accurately than were inexpressive excerpts ($\chi^2_2 = 19.57$, $p < 0.001$). Whereas pairs of expressive excerpts from the same performer were more likely to be grouped together than pairs of inexpressive excerpts, the reverse was observed with pairs from different performers, as indicated by the interaction between expressiveness and performer identity ($\chi^2_2 = 15.17$, $p < 0.001$). Again, no significant effect of musical training was observed. A separate model was built for each level of the expressiveness factor, showing a significant effect of musical training only for the expressive pairs played by the same performer ($\chi^2_1 = 3.92$, $p < 0.05$). This effect seems largely explainable by the lower accuracy of the nonmusicians on the duplicate pairs (see figure 3).

3.6 Predicting the categorisation accuracy for individual performers

We also sought to predict the categorisation accuracy for individual performers based on the musical parameters extracted from the MIDI data. Because mean tempo and mean overlap were found to be reliable predictors of the perceptual similarity between performers (as shown by the MDS analysis of the listeners' partitioning of the excerpts), we surmised that the degree of similarity in note-by-note expressive patterns between performances would likely be a relevant parameter for predicting the categorisation accuracy for different performers. Specifically, we hypothesised that performers exhibiting more consistency in their use of expressive patterns, as well as greater distinctiveness from the patterns of other performers, would be sorted more accurately. The degree of similarity between the expressive patterns of different excerpts was evaluated by computing tempo correlations on a quarter-note basis and overlap correlations on a note-by-note basis. These correlations were then averaged across all excerpts from the same performer, providing a measure of consistency. Excerpts from each performer were also compared with excerpts from other performers, and the averaged correlation coefficients yielded a measure of distinctiveness. The correlation ratio (η), a measure of statistical dispersion that is computed by dividing the sum of the between-group variance by the total variance, was used to quantify a consistency-to-distinctiveness ratio between the correlation coefficients obtained for excerpts played by the same performer and those obtained with excerpts from other performers. As shown in table 3, the value of η for tempo was generally higher for prize-winners (performers A, D, and F), who were also the performers who were sorted most accurately; however, no such trend was observed for overlap. Performer B, by far the most poorly identified, scored very low on the consistency-to-distinctiveness scale for both tempo and overlap. A repeated-measures logistic regression analysis was conducted on the proportion of pairs of excerpts correctly grouped together for each performer with the correlation ratios for tempo and overlap as continuous covariates. The correlation

Table 3. Correlation coefficients for tempo and overlap and proportions of pairs correctly grouped for each performer.

	Performer					
	A*	B	C	D*	E	F*
<i>Tempo</i>						
Mean within-performer correlation	0.89 (0.05)	0.43 (0.29)	0.80 (0.08)	0.76 (0.09)	0.64 (0.18)	0.59 (0.12)
Mean between-performer correlation	0.46 (0.23)	0.35 (0.29)	0.41 (0.23)	0.24 (0.33)	0.35 (0.29)	0.19 (0.28)
η	0.20	0.01	0.17	0.15	0.07	0.13
<i>Overlap</i>						
Mean within-performer correlation	0.46 (0.26)	0.44 (0.18)	0.57 (0.25)	0.48 (0.22)	0.71 (0.10)	0.87 (0.05)
Mean between-performer correlation	0.40 (0.20)	0.35 (0.20)	0.24 (0.18)	0.37 (0.20)	0.43 (0.13)	0.38 (0.21)
η	0.01	0.01	0.18	0.02	0.24	0.27
Proportion of correct pairs	35.3%	21.8%	33.0%	37.0%	31.0%	42.5%

Note. *Mean within-performer correlation*: mean correlation coefficients between excerpts played by the same performer. *Mean between-performer correlation*: mean correlation coefficients between excerpts played by a given performer and excerpts from the other performers. *Proportions of correct pairs*: proportion of pairs of excerpts played by the same performer that were correctly grouped together by listeners. Prize-winning performers are marked with an asterisk. Standard deviations are given in parentheses.

ratios for both tempo ($\chi^2_1 = 23.80$, $p < 0.001$) and overlap ($\chi^2_1 = 8.97$, $p < 0.05$) were significantly predictive of the number of pairs correctly classified, thus supporting the hypothesis that consistency and distinctiveness constitute the main determinants of musical individuality.

3.7 Predicting the categorisation accuracy for individual listeners

Although no significant difference was observed in the categorisation accuracy of musicians and nonmusicians as a group, large differences were observed between individual listeners. In order to identify the factors responsible for these differences, each participant's log file was examined. Listening activity, defined by the total number of times a participant listened to each excerpt, was found to be significantly correlated with categorisation accuracy (musicians: $r_{18} = 0.46$, $p < 0.05$; nonmusicians: $r_{18} = 0.47$, $p < 0.05$). Musicians listened to more excerpts than nonmusicians on average (mean number of excerpts listened to for musicians: 207.6; for nonmusicians: 178.6). Although this difference did not reach significance, it may account for the musicians' slightly higher accuracy. As can be seen in figure 4, the slope of the linear regression for the proportion of correct pairs versus listening activity was indeed very similar for both groups.

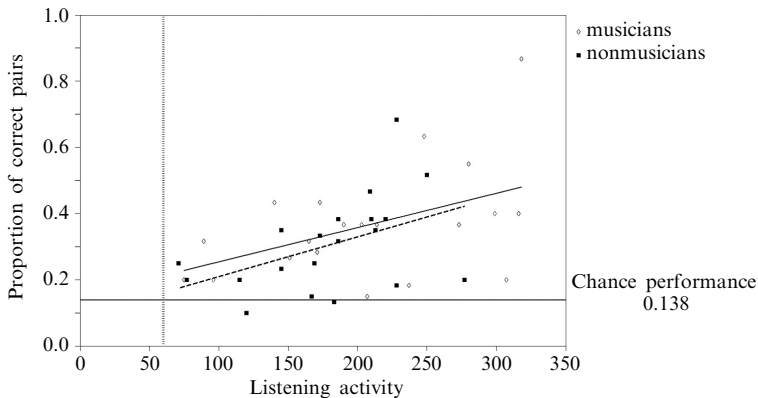


Figure 4. Proportion of correct pairs versus listening activity for musicians and nonmusicians. Regression lines are indicated for musicians (solid line) and nonmusicians (dashed line). Participants minimally had to listen to each of the 30 excerpts twice in order to complete the sorting task (vertical dotted line).

3.8 Effect of performer expertise on categorisation accuracy

In order to model the effect of the performers' expertise on the participants' performance in the sorting task, a repeated-measures logistic regression analysis on the proportion of correct pairs was conducted, with participants' musical training as a between-subjects factor, performer expertise (prize-winners versus non-prize-winners) and expressiveness (inexpressive/inexpressive, inexpressive/expressive, and expressive/expressive) as within-subject factors, and listening activity as a continuous covariate. In line with prior analyses, significant effects were observed for listening activity ($\chi^2_1 = 27.36$, $p < 0.001$), and expressiveness ($\chi^2_2 = 11.32$, $p < 0.001$). Performer expertise was also found to be a significant predictor ($\chi^2_1 = 9.71$, $p < 0.01$), with prize-winners being sorted more accurately than non-prize-winners. As with previous analyses, no main effect of musical training was observed. No two-way interactions between the categorical factors (musical training, performers' expertise, and expressiveness) reached significance, with p values above 0.5 in all cases. However, a significant three-way interaction emerged ($\chi^2_2 = 4.76$, $p < 0.01$), which can be explained by the better performance of musically trained participants on expressive/expressive pairs recorded by prize-winning performers, relative to nonmusicians (see also the analysis in section 3.5).

4 General discussion

We evaluated the combined effects of listener expertise, performer expertise, and performance expressiveness on listeners' ability to recognise different performances of the same piece as being played by the same performer in the absence of timbral and dynamic differentiation. Most listeners, whether musicians or nonmusicians, performed significantly above chance in this task, suggesting that sufficient information to identify a performer's individual style is contained in a short (10–14 s) recording. In contrast to experiments that require the explicit detection of localised pitch or temporal modifications, in which musicians generally outperform nonmusicians (Rammsayer and Altenmüller 2006; Tervaniemi et al 2005), implicit knowledge of, or passive exposure to, a specific musical culture may be more relevant to the present task than formal musical training (Bigand and Poulin-Charronnat 2006; Honing and Ladinig 2009). The agreement between musicians and nonmusicians may also be explained on the basis that the judgments can be made without much prior knowledge, suggesting a bottom-up process based on global and local characteristics. Indeed, the perception of individuality in general, and in the particular task employed in this study, are in all likelihood based on a general process of similarity estimation; there is a tight empirical link between similarity, categorisation and identification, as evidenced by previous research (Ashby and Lee 1991).

Both musicians and nonmusicians grouped excerpts mainly on the basis of tempo and articulation, a finding which is in agreement with earlier results (Repp and Knoblich 2004). Whereas mean tempo and mean overlap correlated reliably with the dimensions of the listeners' perceptual space as obtained from a MDS analysis, the consistency-to-distinctiveness ratio in note-by-note expressive patterns (for both tempo and articulation) was a good predictor of the listeners' categorisation accuracy for individual performers. Additionally, categorisation accuracy was superior for prize-winning performers compared to non-prize-winners, suggesting a link between performer competence and the communication of musical individuality. These findings imply that both superior consistency and the use of distinctive expressive features may be closely associated with the projection of a well-defined musical personality (Sloboda 2000).

Expressiveness affected categorisation accuracy: expressive interpretations from the same performer were more likely to be grouped together than inexpressive ones, and expressive performances from different performers were less likely to be grouped together than inexpressive ones. These observations provide evidence that performer individuality was conveyed more efficiently through expressive recordings, thus corroborating earlier findings on movement-based recognition (Hill and Pollick 2000; Loula et al 2005). Because expressiveness is linked to the magnitude of tempo variations (Palmer 1989), it may be surmised that musical individuality is conveyed, at least in part, through expressive variations in tempo patterns. Indeed, the fact that listeners were less accurate in grouping inexpressive excerpts than expressive ones points to an important role for tempo variation, because the only statistically significant difference between inexpressive and expressive excerpts was that the former exhibited a smaller tempo variability (table 2).

Although this study has shed some light on the transmission of musical individuality in music performance, it also leaves several questions unanswered. For instance, even though we have provided evidence that a performer's individual style could be recognised across varying degrees of expressiveness on several recordings of the same piece, it remains to be seen whether listeners could recognise an unfamiliar performer's style across different pieces or even different genres. The fact that artificial intelligence approaches have achieved high recognition rates in such a task (Saunders et al 2008; Stamatatos and Widmer 2005) suggests that some characteristics associated with a performer's specific style remain more or less invariant across various pieces and

genres, potentially enabling listeners to recognise it. Possible associations between specific musical features or acoustical parameters of the performances and perceived personality traits should also be investigated. Because emotions (Juslin 2000) and even person-related semantic dimensions such as male–female (Watt and Quinn 2007) have been shown to be reliably transmitted through music performance, it is not unreasonable to suppose that some aspects of the performers' personalities may be conveyed as well.

The ability of listeners to accurately discriminate between unfamiliar performers using only timing cues points towards the existence of generalised perceptual mechanisms for the detection of identity cues conveyed by biological movement, possibly based on a tight coupling between action and perception (eg Aziz-Zadeh et al 2006; Knoblich and Prinz 2001). Time appears to play a fundamental role in such a system as an information-bearing dimension common to all types of biological motion (Wohlschläger et al 2003), whether it concerns the visual identification of individuals from their movements (Hill and Pollick 2000), the aural identification of individuals from their voice (Remez et al 1997) or hand clapping (Repp 1987), the offline recognition of one's own actions such as handwriting (Knoblich and Prinz 2001), or the identification of music performers. The existence of such a general identity-detection mechanism would predict that these tasks share common neural processes, which remain to be characterised.

Acknowledgments. This research was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams. We thank Caroline Palmer for her assistance in designing this study and the Church of St-Andrew & St-Paul (Montreal) for permission to use their Casavant organ. We also wish to thank Lauren Stewart and three anonymous reviewers for their comments on an earlier version of this manuscript.

References

- Abercrombie D, 1967 *Elements of General Phonetics* (Chicago: Aldine)
- Arabic P, Boorman S A, 1973 "Multidimensional scaling of measures of distance between partitions" *Journal of Mathematical Psychology* **10** 148–203
- Ashby F G, Lee W W, 1991 "Predicting similarity and categorization from identification" *Journal of Experimental Psychology: General* **120** 150–172
- Aziz-Zadeh L, Wilson S M, Rizzolatti G, Iacoboni M, 2006 "Congruent embodied representations for visually presented actions and linguistic phrases describing actions" *Current Biology* **16** 1818–1823
- Belin P, Fecteau S, Bédard C, 2004 "Thinking the voice: Neural correlates of voice perception" *Trends in Cognitive Sciences* **8** 129–135
- Benadon F, 2003 "Spectrographic and calligraphic cues in the identification of jazz saxophonists", in *Proceedings of the 5th Triennial ESCOM Conference, Hanover, 8–13 September* Eds R Kopiecz, A C Lehmann, I Wolther, C Wolf (Osnabrück: epOs-Music) pp 245–249
- Bigand E, Poulin-Charronnat B, 2006 "Are we experienced listeners? A review of the musical capacities that do not depend on formal musical training" *Cognition* **100** 100–130
- Blake R, Shiffrar M, 2007 "Perception of human motion" *Annual Review of Psychology* **58** 47–73
- Braasch J, Ahrens C, 2000 "Attack transients of free reed pipes in comparison to striking reed pipes and diapason pipes" *Acta Acustica united with Acustica* **86** 662–670
- Carey S, 1992 "Becoming a face expert" *Philosophical Transactions of the Royal Society of London, Series B* **335** 95–103
- Carroll J D, Chang J J, 1970 "Analysis of individual differences in multidimensional scaling via an *N*-way generalization of Eckart–Young decomposition" *Psychometrika* **35** 283–319
- Charrier I, Sturdy C B, 2005 "Call-based species recognition in black-capped chickadees" *Behavioural Processes* **70** 271–281
- Charrier I, Mathevon N, Jouventin P, 2001 "Mother's voice recognition by seal pups" *Nature* **412** 873
- Chartrand J P, Peretz I, Belin P, 2008 "Auditory recognition expertise and domain specificity" *Brain Research* **1220** 191–198

- Daws J T, 1996 "The analysis of free-sorting data: Beyond pairwise cooccurrences" *Journal of Classification* **13** 57–80
- Efron B, Tibshirani R, 1993 *An Introduction to the Bootstrap* (New York: Chapman & Hall)
- Flach R, Knoblich G, Prinz W, 2004 "Recognizing one's own clappings: The role of temporal cues" *Psychological Research* **69** 147–156
- Gingras B, 2006 "Emphasizing voices in polyphonic organ performance: issues of expressive performance on an instrument with fixed tone intensity", in *Proceedings of the 9th International Conference on Music Perception and Cognition, Bologna, 22–26 August* Eds M Baroni, A R Addessi, R Caterina, M Costa (Bologna: Bologna University Press) pp 1712–1720
- Goebel W, 2001 "Melody lead in piano performance: Expressive device or artifact?" *Journal of the Acoustical Society of America* **110** 563–572
- Hill H, Pollick F E, 2000 "Exaggerating temporal differences enhances recognition of individuals from point-light displays" *Psychological Science* **11** 223–228
- Hirsh I J, 1959 "Auditory perception of temporal order" *Journal of the Acoustical Society of America* **31** 759–767
- Honing H, Ladinig O, 2009 "Exposure influences expressive timing judgments in music" *Journal of Experimental Psychology: Human Perception and Performance* **35** 281–288
- Hubert L, Arabie P, 1985 "Comparing partitions" *Journal of Classification* **2** 193–218
- Julsin P N, 2000 "Cue utilization in communication of emotion in music performance: Relating performance to perception" *Journal of Experimental Psychology: Human Perception and Performance* **26** 1797–1812
- Kendall R A, Carterette E C, 1990 "The communication of musical expression" *Music Perception* **8** 129–164
- Knoblich G, Prinz W, 2001 "Recognition of self-generated actions from kinematic displays of drawing" *Journal of Experimental Psychology: Human Perception and Performance* **27** 456–465
- Kruskal J B, Wish M, 1978 *Multidimensional Scaling* (Beverly Hills, CA: Sage Publications)
- Loula F, Prasad S, Harber K, Shiffrar M, 2005 "Recognizing people from their movement" *Journal of Experimental Psychology: Human Perception and Performance* **31** 210–220
- Miller G A, 1969 "A psychological method to investigate verbal concepts" *Journal of Mathematical Psychology* **6** 169–191
- Palmer C, 1989 "Mapping musical thought to musical performance" *Journal of Experimental Psychology: Human Perception and Performance* **15** 331–346
- Palmer C, Jungers M K, Jusczyk P W, 2001 "Episodic memory for musical prosody" *Journal of Memory and Language* **45** 526–545
- Rasch R A, 1979 "Synchronization in performed ensemble music" *Acustica* **43** 121–131
- Rammesayer T, Altenmüller E, 2006 "Temporal information processing in musicians and nonmusicians" *Music Perception* **24** 37–48
- Remez R E, Fellowes J M, Nagel D S, 2007 "On the perception of similarity among talkers" *Journal of the Acoustical Society of America* **122** 3688–3696
- Remez R E, Fellowes J M, Rubin P E, 1997 "Talker identification based on phonetic information" *Journal of Experimental Psychology: Human Perception and Performance* **27** 456–465
- Rendall D, Rodman P, Emond R E, 1996 "Vocal recognition of individuals and kin in freeranging rhesus monkeys" *Animal Behaviour* **51** 1007–1015
- Repp B, 1987 "The sound of two hands clapping: An exploratory study" *Journal of the Acoustical Society of America* **81** 1100–1109
- Repp B H, 1996 "Patterns of note onset asynchronies in expressive piano performance" *Journal of the Acoustical Society of America* **100** 3917–3931
- Repp B H, 1997 "The aesthetic quality of a quantitatively average music performance: two preliminary experiments" *Music Perception* **14** 419–444
- Repp B H, Keller P E, 2010 "Self versus other in piano performance: detectability of timing perturbations depends on personal playing style" *Experimental Brain Research* **202** 101–110
- Repp B H, Knoblich G, 2004 "Perceiving action identity: How pianists recognize their own performance" *Psychological Science* **15** 604–609
- Saunders C, Hardoon D R, Shawe-Taylor J, Widmer G, 2008 "Using string kernels to identify famous performers from their playing style" *Intelligent Data Analysis* **12** 425–440
- Sloboda J A, 2000 "Individual differences in music performance" *Trends in Cognitive Sciences* **4** 397–403
- Stamatatos E, Widmer G, 2005 "Automatic identification of music performers with learning ensembles" *Artificial Intelligence* **165** 37–56
- Sugita Y, 2008 "Face perception in monkeys reared with no exposure to faces" *Proceedings of the National Academy of Sciences of the USA* **105** 394–398

-
- Tang-Martinez Z, 2001 “The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: a critical re-evaluation” *Behavioural Processes* **53** 21–40
- Tervaniemi M, Just V, Koelsch S, Widmann A, Schröger E, 2005 “Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study” *Experimental Brain Research* **161** 1–10
- Thompson J C, Hardee J E, 2008 “The first time I ever saw your face” *Trends in Cognitive Sciences* **12** 283–284
- Timmers R, 2005 “Predicting the similarity between expressive performance of music from measurements of tempo and dynamics” *Journal of the Acoustical Society of America* **117** 391–399
- Watt R, Quinn S, 2007 “Some robust higher-level percepts for music” *Perception* **36** 1834–1848
- Wohlschläger A, Haggard P, Gesierich B, Prinz W, 2003 “The perceived onset time of self- and other-generated actions” *Psychological Science* **14** 586–591

ISSN 0301-0066 (print)

ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 40 2011

www.perceptionweb.com

Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.