Proceedings of the 9th Conference on Interdisciplinary Musicology – CIM14. Berlin, Germany 2014

PERCEPTION AND IDENTIFICATION OF PIANO TIMBRE NUANCES IN DIGITAL PERFORMANCE SIMULATIONS VS. AUDIO RECORDINGS

Michel Bernays[1], Philippe Depalle[1], Marcelo M. Wanderley[1]
[1] *IDMIL, SPCL, CIRMMT, Schulich School of Music, McGill University, Montreal, Canada*
Correspondence should be addressed to: michel.bernays@mail.mcgill.ca

Abstract: **Digital models of the piano sound have become increasingly realistic, either with deep-layered sample-based solutions or physical models of the instrument. Now, it becomes crucial for a digital piano model to provide an expressive response to performance control parameters. Those are widely carried by the MIDI communication protocol, which essentially limits performance information to timing and velocity. While it may suffice for adequately conveying broad traits of expressive piano performance, finer features like timbral nuances, achieved through subtle and largely intuitive playing variations, may be lost. Consequently, digitally reproducing performer-controlled timbral nuances may be impaired by the limitations of both the MIDI protocol and the responsiveness of piano-modeling algorithms. This study examines this potential issue, by comparing the perception and identification of piano timbre nuances by expert listeners, between audio recordings and digital simulations of piano performances. For this aim, three pianists were asked to perform two short pieces with different timbral intentions (bright, dark, dry, round, and velvety). Audio and MIDI recordings were made. Digital simulations were created from the MIDI recordings, with the physical piano-modeling software Pianoteq, and also with the basic GarageBand sample-based model. Participants were then asked for a comparative identification of the five timbral nuances, for each type of audio stimuli (direct recordings and both digital simulations). The results indicated that expert listeners could successfully identify the timbral nuances in both the direct audio recordings and the high-quality physical-model simulations, but not so much in the coarser, sample-based simulations. Identification rates also varied between timbral nuances, with the patterns between the three types of audio stimuli correspondingly varying in range and spread. These results suggest that the limitations of the MIDI protocol may not impede the expression of piano timbre nuances, as long as digital performance simulations involve sufficiently realistic-sounding and responsive piano models.**

1. INTRODUCTION

In classical piano music, the role of the performer is crucial, for the expressive interpretation can enliven the composed work and convey emotions to the audience. Accordingly, the art and technique of piano performance is illustrated by an extensive, empirical body of knowledge, such as [1, 2]. Meanwhile, the characteristics of expressive piano performance have also been studied scientifically, especially with regard to the performance parameters of timing and dynamics. Indeed, those can be controlled directly by pianists, and were shown to bear a significant influence on the perception of emotional expression [3]. Furthermore, these expressive parameters can be indicated only coarsely on scores (with traditional notation), which leaves further leeway for a performer's own musical expression [4]. Accordingly, computational models of expressive piano performance [4] have applied deviations to the score in terms of timing and dynamics, or in terms of higher-level concepts (such as *ritardando*, *legato*, or voicing) that are reduced to timing or dynamics. Those deviations are applied according to hierarchical models of musical structure [5, 6], to general low-level rules [7], or with machine learning algorithms [8]. Noticeably, most expressive models rely on MIDI files, as data to be modified by the model and as training sets for machine learning models. Conversely, the MIDI protocol has also been largely used for acquiring piano performance data [9], which in turn could be used for building expressive performance models (in a synthesis-by-measurement paradigm) and for 're-performances' [10]. The ubiquitous MIDI communication protocol [11] can indeed convey musical information about pitch, timing (*note-on* and *note-off*) and dynamics (note velocity), as well as pedalling.

And MIDI is used for input in nigh all digital models of the piano as instrument. Many commercially available digital piano software rely on samples, with increasingly deep layers (different samples for different velocities for each note) that can provide a more realistic response (at the cost of ever larger sample libraries). However, these sample-based solutions are usually oriented toward popular music production, and the adequacy of their response in rendering the expressive nuances of classical music performance is debatable [12]. On the other hand, there has been considerable improvement in the physical modeling of the acoustic piano, following the increase in computational power of personal computers. Different strategies (finite-difference [13], waveguides [14], modal synthesis [15]) have been employed to model the different parts of the piano (vibrating strings and soundboard, hammers, etc.) and their interactions with increasing accuracy. These strategies propose or simulate approximate solutions to the differential equations that govern the physical dynamic behaviour of the piano parts with increasingly higher order of non-linearity.

However realistically the acoustic piano and its sound may be digitally simulated, it is yet crucial that digital piano models can convey the expressiveness of a performance, whether the expressiveness is digitally modelled or stems from the MIDI recording of an actual human performance. To this end, digital piano models must offer a fast and accurate dynamic response to expressive performance control parameters. In relation to our study is raised the research question of whether a high-grade digital piano model can faithfully reproduce the finer expressive subtleties that can be conveyed through the MIDI protocol.

Nevertheless, with regard to expressiveness in piano performance, the reliance on MIDI for models and simulations of expressive piano performance can only remain adequate in so far as one assumes that all the relevant subtleties of piano performance can be reduced to timing and dynamics (and pedalling), within the limits temporal accuracy (on the order of a few milliseconds) and velocity resolution (7-bit) offered by MIDI [16]. In particular, this assumption may undermine the role of another musical attribute that pianists deem essential to expressive performance [17]: timbre. Further than the inherent characteristic of a sound source or instrument, pianists envision timbre as the subtle tone quality that can emerge from the expressive nuances of their performances [18]. The different timbral nuances that are familiar to pianists, as empirically-defined abstract concepts, are described with an extensive vocabulary of adjectival descriptors [19, 20], and are used in piano performance to convey specific expressive intentions [21].

On the other hand, timbre control is closely associated with the controversial concept of piano touch. In fact, with regard to the actual sound production, timbre control for a single, isolated note was shown long ago to only rely on keystroke velocity at the instant of hammer let-off, thus inseparable from intensity with no independent influence of touch [22]. However, in a polyphonic, musical context, piano touch was shown to provide a kinaesthetic feedback helpful to the expressive control of timbral nuances [23].

In our previous research, a quantitative study of the semantic similarity between verbal descriptors of piano timbre [24] showed that the semantic space of timbre descriptors that pianists commonly employ could be best represented by a minimal set of five adjectives: bright, dark, dry, round, and velvety. The production of these five different timbral nuances was then shown to involve fine-grained performance subtleties in terms of touch and articulation (in addition to more MIDI-suitable subtleties) [25]. Moreover, a preliminary study revealed that the five timbral nuances could be reliably identified by pianists, upon listening to performances guided by these timbral intentions. More precisely, the timbral nuances appeared to be easily distinguished into four quadrants: bright, dry, round, and dark/velvety.

Yet it remains undetermined whether the nuances of performance and touch bear a direct effect on sound production or are only indirectly involved in helping pianists reach their expressive intentions (which would then be instantiated solely in terms of polyphonic timing, dynamics and pedalling). This issue brings forth the following research question: can expressive nuances of piano timbre be effectively conveyed by the MIDI protocol, with performance data reduced to note timing, velocity and pedalling, or is there more (piano touch) to it?

Consequently, digital simulations of expressive piano performances could be confronted with two potential information bottlenecks in the rendering of expressive timbral nuances: (1) the reduction of the complete features of expressive performance to only the parameters which can be conveyed with the MIDI protocol, and (2) the limits in the response of digital piano models to such control parameters, in terms of speed, accuracy and finesse. In the face of these potential issues, this study thus aims at determining whether the expressive timbral intentions of pianists in their performances (following instructions given by the five adjectival descriptors bright, dark, dry, round, and velvety) can be identified once the performance data is reduced to MIDI parameters then used as input to audio simulations by digital piano modeling. For this aim, the identification of piano timbre nuances by expert pianists listeners is compared between direct audio recordings of the performances and their corresponding MIDI-driven digital simulations.

2. METHOD

The design of the perception and identification test of piano timbre nuances in digital performance simulations vs. audio recordings took the following steps, within the framework of the larger research project on piano timbre:

- selection of the timbral nuances to examine, by way of their verbal descriptors;
- conception of musical pieces to be expressively performed according to these different timbral nuances;
- audio and performance-data recordings of timbre-coloured performances by several pianists;
- creation of digital performance simulations based on the performance-data recordings reduced to MIDI;
- implementation of a graphic user interface for the perception test;
- and the testing of 19 participant pianists.

2.1. *Timbral nuances*

Five verbal descriptors were selected: **dry**, **bright**, **round**, **velvety** and **dark**, previously identified as the most salient and representative in the semantic space of piano timbre description [24]. They were used first as timbral instructions to performers, then as the answer choices for participants to the perception test.

2.2. *Musical pieces*

All five timbral nuances had to be expressed by performers over the same piece, with equally good fit of each timbral nuance to the musical material. Consequently, original, short solo piano pieces were written by local composers, under this timbre-fit constraints. Two pieces were selected for this study (cf. Fig. 1), as they were considered the best suited to the expression of each of the five timbral nuances. They were composed by Ana Dall'Ara-Majek and Frédéric Chiasson, respectively.

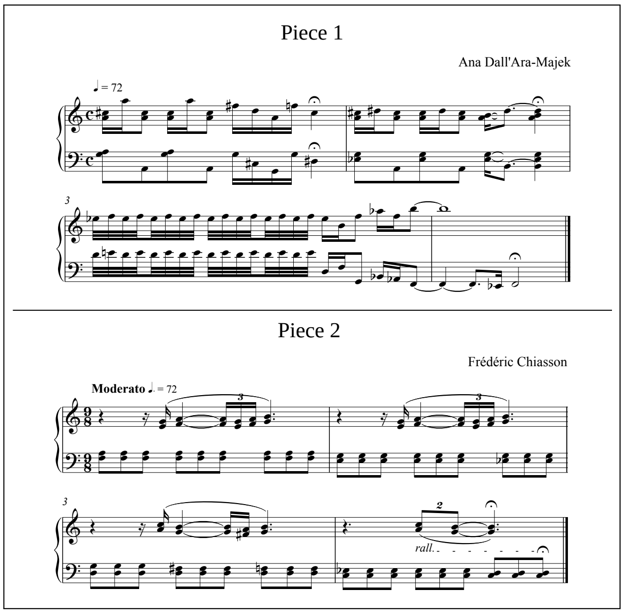



Figure 1: Scores of the two pieces composed and selected for the study.

2.3. *Performance recordings*

Four pianists, all with extensive professional experience and advanced-level piano performance diploma, were recruited to perform the pieces. They were first given time to practice in the days prior to the recording session. The timbral instructions were provided with only the five adjectival descriptors (dry, bright, round, velvety, and dark), to which they confirmed their familiarity as a piano timbre nuance. They were asked to perform each piece with each of the five timbral nuances, separately. The sequence was repeated three times. The recording session took place at BRAMS (International Laboratory for Brain, Music and Sound Research, Montreal) in a professional recording studio. The performances were played on a Bösendorfer Model 290 grand piano.

For the audio recording of the performances, two DPA 4011-TL professional cardioid microphones were used. They were set in XY type, in a 'sweet spot' about 1 m off the open-side ream at mid-frame, at a height of about 1.20 m (diving toward the far side of the soundboard). The microphone signals were routed to a Millenia HV-3D preamp, and were recorded at 44.1 kHz and 24-bit resolution. Unfortunately, during the recording of one pianist's performances, a failure of one of the two microphones occurred, leaving only monophonic recordings. Those could not be used in the perception test, but were instead used in the familiarization and training phase preceding the actual perception test. Ninety recorded performances thus remained (3 pianists × 2 pieces × 5 timbres × 3 same-condition repetitions). Out of the three performances that featured the same pianist, piece and timbral nuance (*i.e.* the 3 same-condition repetitions), one recording was selected, according to criteria of performance quality and fidelity to the score, quality of recording (especially the absence of background noise, such as a creaking chair or a singing pianist), and consistency and reliability of timbre expression. In total, 30 audio recordings were thus selected for the perception test, featuring every combination of the five timbral nuances, the two pieces and the three performers. Ten additional recordings of the fourth pianist were likewise selected for the training phase to the perception test.

In parallel, all performances were recorded with the Bösendorfer CEUS system [26]: high-sample-rate (500 Hz) tracking of key and pedal positions and recording of peak hammer velocities, with 8-bit resolution. These high-accuracy piano performance recordings were useful for a previous study of the individual fine-grained performance features involved in expressive timbre production [27]. However, for the purpose of this study, they were used as input to digital piano models. Thirty high-accuracy CEUS recordings were thus selected, corresponding to the 30 performances whose audio recordings were to be used for the perception test.

2.4. Performance simulations

In order to create digital simulations of the 30 performances from their CEUS recordings, the data first had to be scaled as a suitable input for digital piano models, *i.e.*, MIDI. Each CEUS recording was thus converted and reduced to MIDI-format data. First, the PianoTouch Matlab toolbox [28] was used to identify the notes within the streamline of CEUS data, and to extract their relevant characteristics (key number, onset, offset, hammer velocity). Pedal use was also retrieved for both the soft and sustain pedals. The note characteristics were then scaled and converted into MIDI note information, while the pedals data were scaled and converted into MIDI control changes. Lastly, the MIDI performance information was formatted according to MIDI specifications, and output as a MIDI file.

After careful consideration of different digital piano modeling solutions, the commercial Pianoteq® 4.5 software [29] was chosen. Using a detailed physical model of the piano and its different parts (hammers, strings and soundboard), Pianoteq indeed indeed allows for the fine-tuning of numerous physical parameters (including, but far from limited to, string length, sympathetic resonance, soundboard impedance, hammer hardness and strike point, piano tuning, sustain pedal noise, as well as velocity curves and virtual microphone position).

The ability to control so many physical parameters allowed for the fine-tuning of the Pianoteq digital piano model, in order to match as precisely as possible the sound of the Bösendorfer grand piano that was used for the performances and their audio recordings. The Pianoteq 'Grand K1' digital piano model was thus carefully fine-tuned, slightly differently for each pianist in order to compensate for small variations in the piano tuning and in microphone position between recording sessions (the pianists were recorded on three different days spanned over three weeks). The same fine-tuned model was otherwise used for the performances of all pieces and all timbres by one pianist. Although the fine-tuning of Pianoteq was essentially conducted empirically, the goodness-of-fit evaluation was systematically assessed with spectral analyses of several short sound samples corresponding to specific notes and chords. Furthermore, great attention was paid in this process of evaluation to balancing the goodness of fit of the sound response between timbral nuances. Thirty digital performance simulations, corresponding to the 30 selected audio recordings, were thus created. The MIDI files converted from high-accuracy CEUS performance recordings were used as input to the Pianoteq physical model of a grand piano. The model was fine-tuned in its physical parameters to fit the general sound of the piano captured in the audio recordings, independently of the expressive inflections infused by the performers. The digital simulations were recorded as 44.1 kHz, 24-bit wav files.

Lastly, another type of digital performance simulations was created, using the grand piano samples from GarageBand®. With 23 piano keys sampled (and pitch-shifted to fill the whole piano register) and three velocity layers (three samples per key corresponding to different dynamics, along with different samples for the same keys played with pedals), this sample-based model was expected to be of relatively low quality. The 30 MIDI files of the performances to render were used to create 30 other digital performance simulations, this time with a low-resolution, sample-based piano model. Those were included in the experiment in the aim of testing whether the quality of a digital piano model would matter for the perception of expressive timbral nuances.

In total, 90 audio stimuli were created: 30 direct audio performance recordings, 30 performance simulations created from MIDI-reduced performance recordings with the Pianoteq physical piano model, and 30 performance simulations of expected lower quality created from the same MIDI-reduced performance recordings with the GarageBand low-resolution sample-based model. Each of these sets of 30 stimuli corresponded to the performances of two pieces by three pianists with five different timbral nuances expressed.

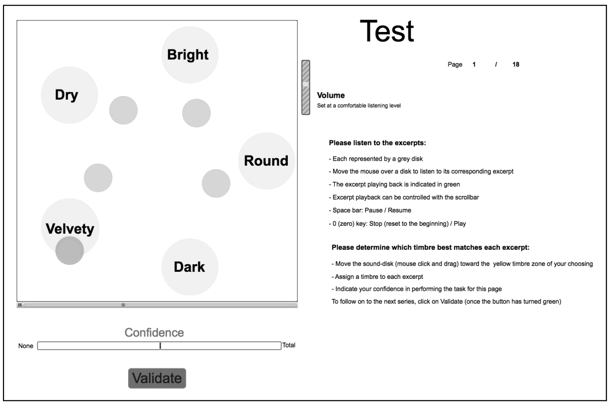
2.5. Perception test

Perceptions tests were conducted at CIRMMT (Centre for Interdisciplinary Research in Music Media and Technology, Montreal), in the acoustically treated Critical Listening Lab, on a Mac Pro computer, with stereo sound output through a RME MADI HDSPe audio interface on two Bowers & Wilkins 802D speakers.

The Max/MSP graphical user interface used for the test was developed by Sébastien Bel in collaboration with the first author. Through the interface, each participant was presented with 18 successive pages, each corresponding to the performances of one piece by one pianist for one type of recording (direct audio, Pianoteq simulation, or GarageBand simulation). Each page thus contained five audio stimuli corresponding to the five different timbral intentions. The five audio stimuli were represented by grey disks, unidentified and placed at random. Participants could choose which stimulus to listen to by moving the mouse over the corresponding grey disk. As all five sounds played continually in a loop (with four of them muted in the background), it was possible to switch seamlessly between stimuli. Participants could also control the audio playing loop, with pause, stop, start, resume and scrolling commands. They could thus listen to the audio stimuli in any manner and order, for as long as deemed necessary.

The main frame of the page also featured the five timbre descriptors (dry, bright, round, dark, velvety), each written within a static yellow disk. The task for participants was then to match each audio stimulus with the most fitting timbre descriptor, by dragging and dropping the grey disks (audio) onto the yellow disks (timbre descriptors).

Additionally, participants were asked to rate with a slider their confidence in assessing correspondence between audio stimuli and timbre descriptors for the page. An example page of the GUI is presented in Fig. 2.



Please listen to the excerpts:

- Each represented by a grey disk
- Move the mouse over a disk to listen to its corresponding excerpt
- The excerpt playing back is indicated in green
- Excerpt playback can be controlled with the scrollbar
- Space bar: Pause / Resume
- 0 (zero) key: Stop (reset to the beginning) / Play

Please determine which timbre best matches each excerpt:

- Move the sound-disk (mouse click and drag) toward the yellow timbre zone of your choosing
- Assign a timbre to each excerpt
- Indicate your confidence in performing the task for this page

To follow on to the next series, click on Validate (once the button has turned green)

Figure 2: Example page of the Max/MSP graphical user interface used in the piano timbre identification test. The small, darker disks represent the audio excerpts. Each is to be dragged towards the timbre descriptor judged the most fitting.

2.6. Participants

Nineteen participants have taken this perception test (age range: 19 to 37; 13 female, 6 male; six Canadian, three American, two French, two Iranian, two Russian, one Colombian, one Dutch, one German, one Vietnamese). They were all pianists and either were professional, had obtained an advanced diploma no more than five years prior, or were currently pursuing advanced studies in piano performance. They had followed between 10 and 25 years of piano performance training.

3. RESULTS

The average timbre identification rates over all participants, performers, pieces and timbres, were, for each type of recording: **0.591** for audio recordings (standard deviation between participants: 0.164); **0.616** for Pianoteq physical-modeling digital simulations (S.D. 0.191); and **0.454** for GarageBand low-resolution sample-based digital simulations (S.D. 0.147). All three timbre identification rates are highly significantly ($p < 10^{-3}$) above chance level (0.2).

3.1. Factor effects

The Generalized Estimating Equations [30] statistical procedure (a variant of generalized linear modeling) revealed significant effects on timbre identification (binary dependent variable) for the within-subjects factors of Recording type (Wald $\chi^2(2) = 25.92$, $p < 10^{-3}$) and Timbre ($\chi^2(4) = 57.89$, $p < 10^{-3}$).

Post-hoc pairwise comparisons indicated significant difference between timbres [dry and bright] vs. [round, dark, and velvety] ($p < 10^{-3}$), as well as between round and dark ($p = 0.026$). Timbre identification was significantly lower ($p < 10^{-3}$) for the GarageBand simulations than for either the audio recordings or the Pianoteq simulations. However, no significant difference in timbre identification was found between Pianoteq simulations and audio recordings ($p = 0.447$).

The effect of the recording-type×timbre interaction was also found significant for identification rates ($\chi^2(8) = 31.81$, $p < 10^{-3}$). In particular, dry and bright timbral nuances were correctly identified significantly more frequently ($p < 10^{-3}$) in audio recordings and Pianoteq simulations than in GarageBand simulations. Timbre identification rates were higher for Pianoteq simulations than audio recordings for dry, bright, round, and velvety timbres, as well as overall, but not for the dark timbre. In all cases though, the differences are not significant. These interactions, and the corresponding timbre identification rates, are presented in Fig. 3.

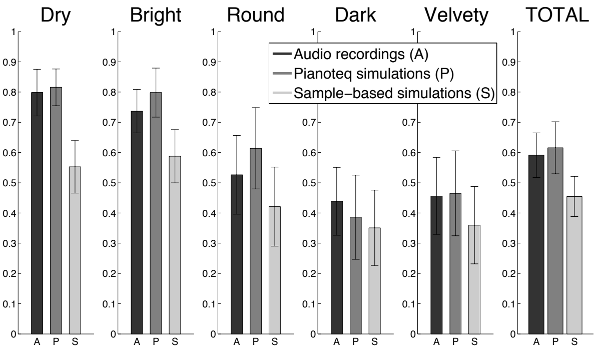


Figure 3: Timbre identification rate, by timbre and type of recording/simulation. Error bars: ±2 S.E. between participants.

3.2. Confidence ratings

The ratings of confidence in timbre identification (evaluated for each of the 18 combinations of recording type × performer × piece) were normalized per participant (z-scores). A 3-way repeated measures ANOVA revealed significant effects of recording type ($F(2,36) = 92.88, p << 10^{-4}$), as well as of performer ($F(2,36) = 15.46, p < 10^{-4}$) and piece ($F(1,18) = 16.27, p < 10^{-3}$), on the normalized confidence ratings. In particular, the confidence ratings for GarageBand digital simulations were significantly lower, pairwise, than for either audio recordings or Pianoteq simulations. Confidence ratings for the latter two did not significantly differ (cf. Fig. 4).

However, the correlation between all normalized confidence ratings and the corresponding timbre identification rates, although significant, is fairly weak (Spearman's $\rho = 0.24$). In other words, the participants have not been very good at assessing their ability to correctly identify timbres.

3.3. Confusion matrices

The three timbre-wise confusion matrices, corresponding to audio recordings, Pianoteq simulations and GarageBand simulations, are presented in Tab. 1 , Tab. 2 and Tab. 3 (resp.). In addition to identification rates per timbre (diagonal elements), they also indicate the patterns of errors made by the participants. Rows correspond to the timbre nuances expressed in the audio excerpt, while columns correspond to the participants' answers. Thus, the same-row elements outside the diagonal indicate the patterns of false negatives, while the same-column elements outside the diagonal indicate the patterns of false positives.

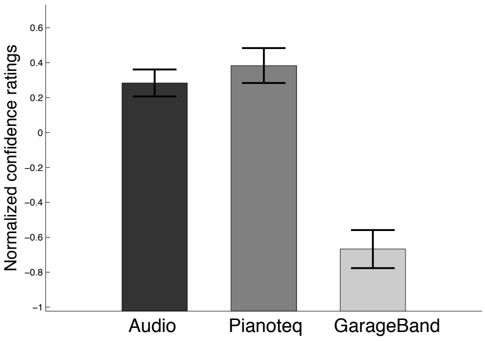


Figure 4: Normalized ratings of confidence in timbre identification, for each type of recording/simulation (error bars: ±2 S.E. between participants).

Table 1: Matrix of confusion between timbral nuances for audio recordings. Rows: timbres expressed in the performances. Columns: timbre descriptors chosen by participants. Diagonal elements: correct identification rates.

	Dry	**Bright**	**Round**	**Dark**	**Velvety**
Dry	**0.798**	0.123	0.044	0.018	0.018
Bright	0.132	**0.737**	0.079	0.026	0.026
Round	0.026	0.105	**0.526**	0.149	0.193
Dark	0.000	0.018	0.237	**0.439**	0.307
Velvety	0.044	0.018	0.114	0.368	**0.456**

Table 2: Matrix of confusion between timbral nuances for Pianoteq performance simulations.

	Dry	**Bright**	**Round**	**Dark**	**Velvety**
Dry	**0.816**	0.140	0.035	0.000	0.009
Bright	0.114	**0.798**	0.061	0.026	0.000
Round	0.018	0.061	**0.614**	0.149	0.158
Dark	0.018	0.000	0.228	**0.386**	0.368
Velvety	0.035	0.000	0.061	0.439	**0.465**

Table 3: Matrix of confusion between timbral nuances for GarageBand low-resolution sample-based performance simulations.

	Dry	**Bright**	**Round**	**Dark**	**Velvety**
Dry	**0.553**	0.263	0.105	0.035	0.044
Bright	0.246	**0.588**	0.088	0.035	0.044
Round	0.096	0.114	**0.421**	0.167	0.202
Dark	0.035	0.018	0.246	**0.351**	0.351
Velvety	0.070	0.018	0.140	0.412	**0.360**

Timbre identification was more accurate, for all timbral nuances but dark, with Pianoteq simulations than with audio recordings. Timbre identification rate was always the lowest for GarageBand sample-based simulations. Timbre-wise, dry and bright timbres were the most accurately identified in all three cases. The accuracy in the identification of dry and bright timbres is especially salient, relative to the other three timbres, in audio recordings and Pianoteq simulations, whereas the difference is much smaller with sample-based simulations.

The patterns of errors are essentially similar for all three types. Dry and bright timbres, in case of errors, were mostly confused or misused with each other. Meanwhile, dark and velvety timbres were highly confused with each other, and to a lesser extent with round. Round as a descriptor was especially misused for dark-timbre audio excerpts. Now for some slight differences in error patterns, the confusion between dark and velvety was a little less pronounced with the audio recordings. The confusion and misuse of round (both toward dark/velvety and dry/bright) was also more frequent with sample-based simulations.

4. DISCUSSION

First, the results suggest that advanced-level pianists are able to identify mostly correctly the timbral nuances highlighted in performances of short piano pieces. In particular, the timbre identification rate for audio recordings is on par with the results of a previous preliminary study (0.627 identification rate) which involved more pieces but the same five timbre descriptors [25].

With regard to the different types of audio stimuli, the consistently lower timbre identification rate for the GarageBand low-resolution sample-based digital performance simulations was expected. This result essentially confirms that conveying timbre expression by digital piano synthesis is non-trivial, and requires a certain amount of audio fidelity and responsiveness to expressive control parameters for timbral nuances to remain identifiable by expert listeners. Indeed, the more sophisticated Pianoteq physical piano model, finely tuned over its many physical instrument parameters (in a non timbre-specific way) so as to match as closely as possible the sound of the acoustic piano featured in the audio recordings, was used to create higher-fidelity digital performance simulations. The timbre identification rate for those simulations was comparable to that of the audio recordings, indicating that digital performance simulations with a sufficiently realistic and responsive digital piano model are able to convey expressive timbral nuances, even after reducing the original performance control parameters to MIDI data. Timbre identification rates were surprisingly even better on average in Pianoteq performance simulations than in the audio recordings, although not significantly so. As a possible explanation, the Pianoteq simulations might present a 'cleaner' sound, unaffected by potential sources of perturbation in sound recording such as extraneous noise, complex reverberation, or microphone response. Timbral nuances would then possibly become easier to identify. However, the difference in identification rate between audio recordings and Pianoteq simulations is slight, and may disappear under a larger sample size.

Identification rates were also dependent on the timbral nuances. Dry and bright timbres were highly correctly identified, especially in audio recordings and Pianoteq simulations. For all three types of audio stimuli, identification rates of dry and bright were significantly higher than for round, dark and velvety. In case of errors, dry and bright were essentially confused with each other (or, to a much lesser extent, with round, mostly in sample-based simulations). On the other hand, dark and velvety timbres were sometimes confused with round, but essentially showed high confusion between each other. In fact, we can infer that participants could highly identify four quadrants of timbral nuances, corresponding to dry, bright, round, and the association of dark and velvety. With the results for dark and velvety assimilated (including their confusion with each other), the identification rates are indeed noticeably improved, at 0.712, 0.764, and 0.575 (resp.) for audio recordings, Pianoteq simulations and sample-based simulations. The identification patterns between types of audio stimuli remain unchanged.

Other interesting conclusions may be gained from the effects of the interactions between types of audio stimuli and timbral nuances. The observation that Pianoteq simulations showed lower identifications rates than audio recordings for, and only for, the dark timbre, might suggest (with all reservations due to non-significant differences) that Pianoteq simulations were less effective in reproducing certain acoustical characteristics of the dark timbre (presumably a predominance of the lower register). Furthermore, the low-resolution sample-based performance simulations were most unsuccessful (relative to the other two stimulus types) in reproducing the dry and bright timbres, possibly due to a lesser dynamic range or a lack of energy in high frequencies.

Lastly, the confidence ratings present the same average pattern, with regard to each type of stimulus, as the average timbre identification rates.

5. CONCLUSION

The results of this study suggest that the timbral intentions of a pianist in expressive performances can be reproduced in digital performance simulations, using a sufficiently realistic digital piano model controlled with MIDI performance information.

The question of what constitutes a 'sufficiently realistic' digital piano model, between the extremes (fine-tuned Pianoteq model vs. coarse sample-based model) envisioned in this study, should deserve further inquiry.

Second and foremost, the results of this study indicate that the control information required for the expression of piano timbre nuances is essentially, if not only, contained within the MIDI information: note onset and duration (with an accuracy on the order of a millisecond), velocity (with 7-bit resolution), and pedalling. As it was shown that, for the expression of different timbral nuances, pianists use distinctive subtleties of key depression (attack, release, touch) [25] that cannot be contained in standard MIDI messages, the results of this study would suggest that such additional, beyond-MIDI control information may stem from the ancillary gestures that can help pianists (by providing kinaesthetic feedback or by working around physiological constraints) obtain the actual effective playing nuances, which can be conveyed as MIDI information.

Finally, in light of the conclusions of this study, a computational model may be designed, relying only on the processing and modification of MIDI data, to reproduce the performance control parameters required for the expression of different piano timbre nuances.

6. ACKNOWLEDGMENTS

We wish to thank all the participants to the test; the composers of the pieces and the performers; Sébastien Bel for his collaboration and the testing interface; Bernard Bouchard (BRAMS) for his invaluable technical and pianistic help; Julien Boissinot and Yves Méthot (CIRMMT) for their technical support in setting up the perception test. This work was made possible by funding from FRQSC (*Fonds de recherche du Québec - Société et culture*).

REFERENCES

[1] J. Hofmann: *Piano Playing with Piano Questions answered*. Theodore Presser Co., Philadelphia, PA, 1920.

[2] H. Neuhaus: *The art of piano playing*. Barrie and Jenkins, London, UK, 1973. Translated from Russian by K.A. Leibovitch.

[3] A. Bhatara, A. Tirovolas, L. Duan, B. Levy, and D. Levitin: *Perception of emotional expression in musical performance*. In *Journal of Experimental Psychology: Human Perception and Performance*, volume 37(3):921–934, 2011.

[4] A. Kirke and E. Miranda: *Guide to Computing for Expressive Music Performance*. Springer, London, UK, 2013.

[5] E. Narmour: *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press, Chicago, IL, 1992.

[6] N. Todd: *The kinematics of musical expression*. In *Journal of the Acoustical Society of America*, volume 97(3):1940–1949, 1995.

[7] A. Friberg, V. Colombo, L. Frydén, and J. Sundberg: *Generating musical performances with Director Musices*. In *Computer Music Journal*, volume 24(3):23–29, 2000.

[8] G. Widmer, S. Flossmann, and M. Grachten: *YQX plays Chopin*. In *AI Magazine*, volume 30(3):35–48, 2009.

[9] W. Goebl, S. Dixon, G. DePoli, A. Friberg, R. Bresin, and G. Widmer: *'Sense' in expressive music performance: Data acquisition, computational studies, and models*. In P. Polotti

and D. Rocchesso (eds.), *Sound to Sense, Sense to Sound: A State of the Art in Sound and Music Computing*, pages 195–242. Logos, Berlin, Germany, 2008.

[10] N. Seaver: *A Brief History of Re-performance*. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 2010.

[11] G. Loy: *Musicians make a standard: the MIDI phenomenon*. In *Computer Music Journal*, pages 8–26, 1985.

[12] J. Masko, J. F. Friberg, and A. Friberg: *Software tools for automatic music performance*. In *Proceedings of the 1st International workshop on computer and robotic Systems for Automatic Music Performance (SAMP14), Venice, Italy*. 2014.

[13] A. Chaigne and A. Askenfelt: *Numerical simulations of piano strings. I. A physical model for a struck string using finite difference methods*. In *Journal of the Acoustical Society of America*, volume 95(2):1112–1118, 1994.

[14] J. O. Smith and S. A. Van Duyne: *Commuted piano synthesis*. In *Proceedings of the 1995 International Computer Music Conference, Banff, AB*, pages 319–326. 1995.

[15] B. Bank, S. Zambon, and F. Fontana: *A modal-based real-time piano synthesizer*. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 18(4):809–821, 2010.

[16] F. R. Moore: *The dysfunctions of MIDI*. In *Computer music journal*, volume 12(1):19–28, 1988.

[17] P. Holmes: *An exploration of musical communication through expressive use of timbre: The performer's perspective*. In *Psychology of Music*, volume 40(3):301–323, 2012.

[18] M. Bernays: *Expression et production du timbre au piano selon les traités: Conception du timbre instrumental exprimée par les pianistes et professeurs dans les ouvrages à vocation pédagogique*. In *Recherche en éducation musicale*, volume 29:7–27, 2012.

[19] M. Bellemare and C. Traube: *Verbal description of piano timbre: Exploring performer-dependent dimensions*. In *Digital proceedings of the second Conference on Interdisciplinary Musicology (CIM05)*. Observatoire interdisciplinaire de création et de recherche en musique (OICRM), Montreal, QC, 2005.

[20] P. Cheminée: *"Vous avez dit 'clair' ?" Le lexique des pianistes, entre sens commun et terminologie*. In *Cahiers du LCPE: Dénomination, désignation et catégories*, volume 7:39–54, 2006.

[21] G. Sandor: *On piano playing*. Schirmer Books, New York, 1995. First edition: 1981.

[22] H. Hart, M. Fuller, and W. Lusby: *A precision study of piano touch and tone*. In *Journal of the Acoustical Society of America*, volume VI:80–94, 1934.

[23] J. MacRitchie and M. Zicari: *The Intentions of Piano Touch*. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, and K. Pastia (eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition*, pages 636–643. 2012.

[24] M. Bernays and C. Traube: *Verbal expression of piano timbre: Multidimensional semantic space of adjectival descriptors*. In A. Williamon, D. Edwards, and L. Bartel (eds.), *Proceedings of the International Symposium on Performance Science, Toronto*, pages 299–304. European Association of Conservatoires (AEC), Utrecht, Netherlands, 2011.

[25] M. Bernays: *The expression and production of piano timbre: gestural control and technique, perception and verbalisation in the context of piano performance and practice*. Ph.D. thesis, Université de Montréal, 2013.

[26] Bösendorfer: *CEUS reproducing system webpage, http://www.boesendorfer.com/en/ceus-reproducing-system.html*, 2006. Retrieved 2014-10-30.

[27] M. Bernays and C. Traube: *Investigating pianists' individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics and pedaling*. In *Frontiers in Psychology*, volume 5(157):157, 2014.

[28] M. Bernays and C. Traube: *Piano touch analysis: A MATLAB toolbox for extracting performance descriptors from high-resolution keyboard and pedalling data*. In T. Dutoit, T. Todoroff, and N. d'Alessandro (eds.), *Proceedings of Journées d'Informatique Musicale (JIM2012), Gestes, Virtuosité et Nouveaux Medias*, pages 55–64. UMONS/numediart, Mons, Belgium, 2012.

[29] Modartt: *Pianoteq webpage, http://www.pianoteq.com/*, 2006. Retrieved 2014-10-12.

[30] J. A. Hanley, A. Negassa, and J. E. Forrester: *Statistical analysis of correlated data using generalized estimating equations: an orientation*. In *American journal of epidemiology*, volume 157(4):364–375, 2003.