

## Communicating voice emphasis in harpsichord performance

Bruno Gingras,<sup>\*#1</sup> Pierre-Yves Asselin,<sup>#2</sup> Meghan Goodchild,<sup>#3</sup> and Stephen McAdams<sup>#4</sup>

<sup>\*</sup>*Department of Computing, Goldsmiths College, University of London, United Kingdom*

<sup>#</sup>*Schulich School of Music and CIRMMT, McGill University, Canada*

<sup>1</sup>brunogingras@gmail.com, <sup>2</sup>pierre-yves.asselin@sympatico.ca,  
<sup>3</sup>meghan.goodchild@mail.mcgill.ca, <sup>4</sup>smc@music.mcgill.ca

### ABSTRACT

We investigated the means used by harpsichordists to communicate voice-specific melodic emphasis. Twelve harpsichordists were asked to perform a short Baroque polyphonic piece by Frescobaldi (1583-1643) on a harpsichord equipped with a MIDI console. Three conditions were tested, each requiring performers to emphasize a particular voice in the piece (soprano, alto, tenor). Four parameters were analyzed: velocity, note onset asynchrony, timing deviations, and articulation. Mean onset asynchronies were much larger than those observed in organ or piano performance, averaging more than 100 ms between outer voices. However, asynchronies did not vary significantly between conditions, suggesting that they do not play a major role in communicating voice emphasis. On average, notes belonging to the upper three voices were struck with a higher velocity and played with a more detached articulation when they were emphasized. Although timing deviation patterns were globally similar across all conditions, we observed local differences that corresponded to passages in which a specific melodic gesture was accentuated by means of *ritenuto* (tempo deceleration).

### I. INTRODUCTION

The communication of melodic emphasis provides a useful experimental model for the empirical study of the communication of expressive intentions in musical performance for several reasons. First, melodic emphasis is a musically meaningful expressive strategy: while it may be somewhat artificial to request a performer to emphasize a specific voice for the duration of a piece, performers often seek to emphasize a particular theme, motive, or melodic passage. Therefore, studying the expressive strategies that they use for this purpose is a musically valid research question. Second, the performer's expressive intentions are explicitly defined in this context, thus facilitating analytical comparisons between different experimental conditions, such as performances of the same piece played with different expressive intentions. Last, the relative efficiency of the expressive strategies used by the performer can be assessed by means of perceptual experiments.

A substantial body of research has been conducted on piano performance in order to identify and characterize the expressive strategies used to communicate melodic emphasis. Musical expressivity in piano performance is essentially conveyed by manipulating three parameters: the inter-onset interval between successive notes (local variations of tempo such as *rubato* and *accelerando*), the intensity of the notes (dynamics), and the offset-to-onset intervals (articulation effects, such as *legato* and *staccato*). Regarding the expressive strategies used by pianists to emphasize a given voice or melody in a polyphonic texture, several studies have shown that the notes of the principal melody are played somewhat louder, and also 20 to 30 ms earlier, than nominally

simultaneous notes in other voices (Goebel, 2001; Palmer, 1989, 1996; Repp, 1996b). This onset asynchrony between the melody notes and notes in the remaining voices has been termed "melody lead." Whereas Palmer (1996) claims that pianists intentionally play the melody notes somewhat earlier, other researchers such as Repp (1996b) and Goebel (2001) have suggested that melody lead may be an artifact due to the fact that when a note is played louder, its key is pressed faster and strikes the hammer earlier than another key that is struck at the same time but softly.

Although the piano can justifiably be seen as a model instrument for such experiments, due to its relative ease of use in a laboratory setting, its widespread practice among the general population, and the large amount of music written for this instrument, few studies have sought to determine whether these findings are applicable to other keyboard instruments. In particular, it is interesting to consider the case of the organ or the harpsichord, for which it is virtually impossible to differentiate individual notes on the basis of intensity (ignoring registration effects or the use of the swell and crescendo pedals on the organ). For instance, although the organ keyboard action is superficially similar to that of the piano, a pipe valve is either open or closed, meaning that dynamic differentiation is impossible on the organ. In this context, organists must, presumably, use expressive strategies which do not entail dynamic differentiation as a means to separate voices (Goebel, 2001), and articulation may be expected to become more important for distinguishing parts in a polyphonic setting for organ than it is on the piano. Indeed, we have recently shown that contrasting articulation patterns between voices was the main expressive strategy used to emphasize specific voices in performances of a Baroque organ piece (Gingras, 2006). We also observed that note onset asynchronies were much smaller than those reported in the literature on piano performance research. These results tend to corroborate the velocity artifact hypothesis, which proposes a direct link between these asynchronies and the use of dynamic differentiation between voices.

However, to our knowledge, no published empirical study has investigated the means used by performers to emphasize individual voices on the harpsichord. We sought to address this issue by conducting an experiment in which harpsichordists were invited to record different interpretations of a harpsichord piece in which they emphasized different voices. Since dynamic differentiation is also very limited on the harpsichord (Penttinen, 2006), we hypothesized that harpsichordists would rely on alternative expressive strategies, such as onset asynchrony and articulation differentiation between voices to a greater extent than pianists.

The investigation of expressive strategies in harpsichord performance is also relevant to a comparison of expressive

strategies in piano, organ and harpsichord. Such a comparison is of interest because, although these instruments share obvious similarities in terms of playing technique, their timbre and acoustical properties are markedly different. While the organ is essentially a wind instrument controlled by a keyboard, the piano is a struck string instrument and the harpsichord a plucked string instrument. Thus, there are important differences in terms of the attack and decay parameters of the sound. In contrast to piano and harpsichord tones, which are characterized by a short rise time followed by a decay (Fletcher, 1977; Palmer & Brown, 1991), organ tones typically reach peak amplitude 50 to 100 ms after note onset and maintain a quasi-constant intensity while the key is pressed (Braasch & Ahrens, 2000). These characteristics may affect the relative efficiency of expressive strategies such as asynchrony or articulation.

## II. METHODS

### A. Piece

The third variation from the *Partita No. 12 sopra l'aria di Ruggiero* by Girolamo Frescobaldi (1583-1643) was chosen for this experiment as being representative of the Baroque harpsichord repertoire (Figure 1). As is typical of the Baroque contrapuntal writing style, the piece contains four distinct melodic lines (parts or voices): these are, from the highest to the lowest, the soprano, alto, tenor, and bass parts. In contrast to the Classical and Romantic piano repertoire, this piece has no obvious principal melodic line and thus lends itself well to multiple interpretations. Another motivation behind the choice of this particular piece is the fact that the four voices are active throughout the piece, and the melodic and rhythmic content of the three upper voices is relatively similar (the bass voice is, however, markedly different).



Figure 1. Girolamo Frescobaldi, *Partita No. 12 sopra l'aria di Ruggiero*, Variation III.

### B. Participants

Twelve skilled harpsichordists, five female and seven male, participated in the experiment. They were professional harpsichordists from the Montreal (Canada) area, or harpsichord students at McGill University in Montreal. Their average age was 39 years (the youngest was 21 years old, the oldest 61). They had played the harpsichord for a mean duration of 22 years (minimum 6, maximum 40). Seven of them had previously won prizes in regional, national, or international harpsichord competitions.

### C. Procedure

Harpsichordists performed the Variation III using the score shown in Figure 1. They were instructed to emphasize a

different voice in each version (respectively, the soprano, alto, and tenor parts). Each version was recorded twice, for a total of six recordings per harpsichordist. Performers were allowed to do three recordings per version and choose the two most satisfactory. The order of the instructions was randomized according to a Latin square diagram.

Performances took place in an acoustically treated studio, on a Italian-style Bigaud harpsichord (Paris, Heugel) with two 8-foot stops. Only the back stop was used for the experiment. This harpsichord was equipped with a MIDI console, allowing precise measurement of performance parameters. MIDI velocity values for each note event were coded in a range between 16 (slowest) and 100 (fastest).

The audio signal was recorded through two omnidirectional microphones MKH 8020. The microphones were located 1 m above the resonance board and were placed 25 cm apart. The audio and MIDI signals were sent to a PC computer through a RME Fireface audio interface. Audio and MIDI data were then recorded using Cakewalk's SONAR software and stored on a hard disk.

### D. Data analysis

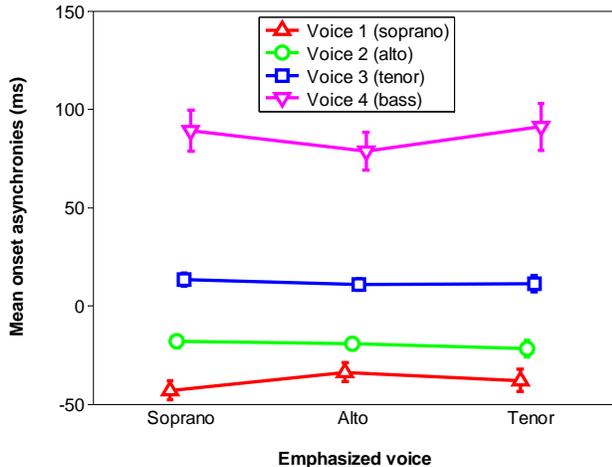
Note onsets and offsets were extracted from the MIDI data of the performances and matched to the score using an algorithm developed by Gingras & McAdams (2007). Four parameters were analyzed from the MIDI data: note onset asynchrony, velocity, articulation, and local tempo variations. For all performances, the rate of errors, defined as the proportion of wrong notes or missing notes relative to the total number of score notes, was very low: 0.75% (of  $n_{\text{total}} = 11,016$ ), comparable to the error rates observed by Repp (1996a), Palmer (1996), and Goebel (2001).

## III. RESULTS

### A. Onset asynchrony

Note onset asynchrony, or chord asynchrony, is defined as the difference in onset time between note onsets that are notated in the musical score as synchronous (Palmer, 1989). Several measures of onset asynchrony have been constructed. Rasch (1979) proposed to use the root mean square, or standard deviation of the onset times of nominally simultaneous notes. Palmer (1989, 1996) used the difference in onset times between the notated melody and the mean onset of the remaining voices, whereas Repp (1996b) presented a measure of asynchrony in which the lag time for each individual note in a chord was obtained by subtracting from its onset time that of the highest note in the chord. The choice of the highest note as a reference note for the computation of asynchronies seemed inappropriate for the experiment discussed here because the main melody was not necessarily located in the uppermost voice. Asynchronies were thus computed for each note as the difference between its onset time and the mean onset of the remaining notes in the chord, with a positive asynchrony referring to a lead, as described in Palmer (1989). One potential disadvantage of using this definition is that the sum of those differences, when computed for all the notes, necessarily equals zero. Consequently, the asynchronies computed for all voices are not independent variables and a multivariate approach is required.

As Figure 2 shows, harpsichordists tend to arpeggiate chords by playing the bass note first, followed by the tenor, the alto, and the soprano. Asynchronies were much larger than those observed on the piano (Goebel, 2001; Palmer, 1996; Repp, 1996b) or on the organ (Gingras, 2006). Indeed, chord asynchronies, measured using Rasch's definition, averaged 31 ms. In comparison, Palmer (1989) reported chord asynchronies of 18 ms for musical performances at the piano, and Gingras (2008) found chord asynchronies of 9 ms on the organ. Furthermore, a large proportion of these asynchronies, comprising 50.6% of all nominally simultaneous note pairs, exceeded 20 ms, which is considered the minimum threshold for listeners to be able to discriminate between onsets (Hirsh, 1959), and were thus likely to be perceptible. These asynchronies can be seen as an expressive strategy which aims at "staggering" synchronous note onsets.

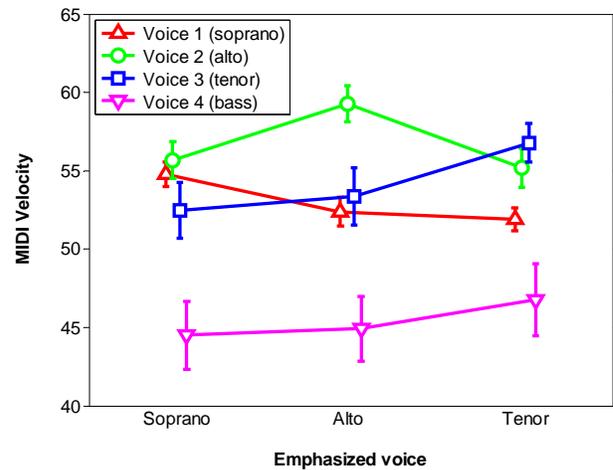


**Figure 2. Mean onset asynchrony for all voice/emphasis combinations, averaged across performers. Error bars represent standard errors of the mean. Positive onset asynchronies indicate that a note was played before the mean onset of other notes belonging to the same chord.**

A repeated-measures MANOVA was conducted on the mean asynchronies for each voice, with emphasis as within-subject factor. No main effect of emphasis was found,  $F(8,40) = 1.87, p > .05$ , suggesting that interpretation did not significantly affect the onset asynchrony pattern. The melody lead, measured using Palmer's (1989) definition and treating the emphasized voice as melody for each interpretation, was actually negative, averaging  $-17 \pm 3$  ms across all performances. In comparison, average melody leads of 20-30 ms were reported on the piano (Goebel 2001; Palmer, 1989, 1996), and of 2 ms on the organ (Gingras, 2006).

### B. Velocity

MIDI velocities are estimated by a mechanical double contact located underneath the keys and from which the travel time of the keys are measured; a high velocity corresponds to a shorter travel time (faster attack). As Figure 3 shows, MIDI velocities were, on average, higher for notes belonging to a voice that was emphasized than for those belonging to a voice that was not emphasized. Moreover, for the same voice, velocity was lower when it was emphasized than when it was not.



**Figure 3. Mean MIDI velocity for all voice/emphasis combinations, averaged across performers. Error bars represent standard errors of the mean.**

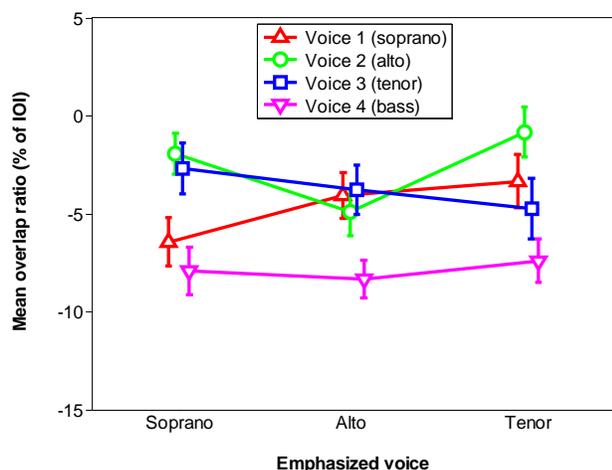
A mixed-model repeated-measures ANOVA was conducted on the mean velocity for each voice, with emphasis as within-subject factor. There was a significant effect of voice,  $F(3,33) = 77.74, p < 0.0001$ , and a significant interaction between emphasis and voice,  $F(6,66) = 4.81, p < 0.001$ . The bass voice may be played softer because it was not emphasized in any of the conditions. However, a mixed-model repeated-measures ANOVA on the mean velocity for voices 1 to 3 also resulted in a significant effect of voice,  $F(2,22) = 18.14, p < .0001$ , and in a significant interaction between emphasis and voice,  $F(4,44) = 10.32, p < .0001$ .

Although these results indicate that harpsichordists tend to attack a note with a greater velocity when they attempt to emphasize it, it should be noted that the extent of the variation in MIDI velocity was relatively modest. Moreover, it is not known whether such a variation was perceptible by listeners.

### C. Articulation

Articulation refers to the amount of overlap between two consecutive note events belonging to the same voice. A legato articulation corresponds to a positive overlap (when the offset of note  $n$  occurs after the onset of note  $n+1$ ), while a staccato articulation corresponds to a negative overlap. The offset of a note is defined as the time at which a key is released (as measured by the MIDI system) and the onset is the time at which a key is pressed. Because the amount of overlap varies with tempo (Repp, 1995) we prefer to use the overlap ratio, defined as the ratio between the overlap between two adjacent tones and the interonset interval between these tones as a measure of articulation (Bresin & Battel, 2000).

Figure 4 shows the mean overlap ratio for all voice/emphasis combinations. We observe a tendency to play notes belonging to a given voice with a more staccato articulation (lower overlap ratio) when this voice is emphasized than when it is not.



**Figure 4.** Mean overlap ratio for all voice/emphasis combinations, averaged across performers. Error bars represent standard errors of the mean.

A mixed-model repeated-measures ANOVA on the mean overlap ratio for each voice, with emphasis as within-subject factor, indicated a significant effect of voice,  $F(3,33) = 13.73$ ,  $p < 0.0001$ , and a marginal interaction between emphasis and voice,  $F(6,66) = 2.02$ ,  $p < 0.1$ . Because the bass voice was not emphasized in any of the conditions, a mixed-model repeated-measures ANOVA was also conducted on the mean overlap ratio for voices 1 to 3, resulting in a significant effect of voice,  $F(2,22) = 3.79$ ,  $p < .05$ , and a significant interaction between emphasis and voice,  $F(4,44) = 4.08$ ,  $p < .01$ . These results indicate that harpsichordists tend to modify their articulation patterns so as to play a voice more staccato when it is emphasized than when it is not. However, we should again note that the effect is modest: the overlap ratio varies by less than 5% on average.

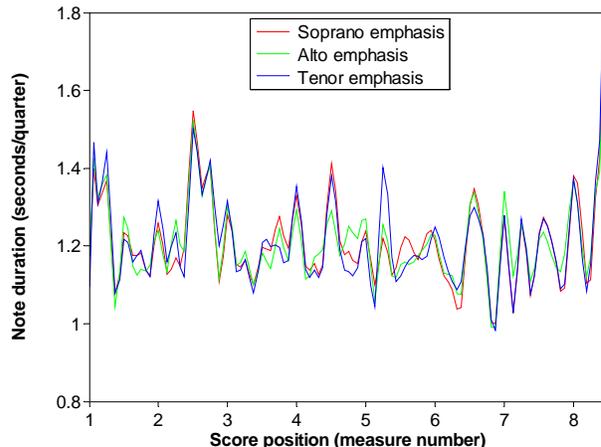
#### D. Tempo patterns

Although note onsets are discrete events, tempo is more accurately described as a continuous parameter. As such, it cannot be treated adequately using “traditional” multivariate analysis. The statistical technique of functional data analysis (Ramsay & Silverman, 1997) was recently developed to facilitate the treatment of continuous variables, and this technique has been suggested for the analysis of tempo and tempo changes (Vines, Nuzzo, & Levitin, 2005).

In the present study, we used functional data analysis to investigate whether tempo patterns differ according to which voice was emphasized. For each performance, onset times for each sixteenth-note were estimated by linear interpolation using the onset times for each score event obtained from the matching algorithm. The discrete data for each performance was then modeled by a basis expansion of 60 sixth-order B-splines (Deboor, 1978), using custom software (Ramsay, 2003), thus generating a derivable function representing the elapsed time at score position  $s$  for each performance.

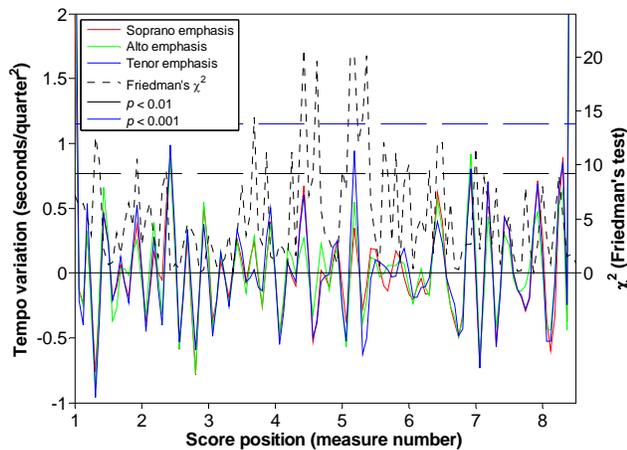
The first derivative of this function corresponds to the sixteenth-note duration, which can be used to compute the average quarter-note duration for each condition (i.e., emphasized voice). A comparison of the quarter-note duration profiles for each condition reveals subtle differences (Figure 5).

For instance, we can see that, in the first few beats of measure 5, the average tempo is much slower (as shown by the longer note durations) for the tenor emphasis condition than for the other conditions.



**Figure 5.** Mean note duration for each voice-emphasis condition, averaged across performers.

However, we are more concerned with differences in tempo variation than with differences in absolute tempo or note duration: two performances could have different tempi but similar acceleration patterns (for instance, if the tempo slows down at the same points in the score). Figure 6 shows the second derivative of the elapsed-time function, which corresponds to tempo variation (we prefer to use the term tempo variation rather than tempo velocity as proposed by Vines & al. [2005] because tempo may already be considered a measure of velocity). We observe that the differences between the three conditions become more obvious when looking at tempo variation patterns instead of absolute note durations. These differences can be assessed quantitatively by using functional analogues of traditional inferential statistical tests. Because the data do not satisfy the assumption of normality, a functional equivalent of the non-parametric repeated-measures ANOVA (Friedman’s test) was used in the present case. The  $\chi^2$  was thus computed for each data point. This approach obviously raises issues related to family-wise adjustments for significance, an issue that has not yet been adequately covered in the functional literature; for this reason, we will restrict ourselves to a qualitative analysis of the statistical profile obtained. This analysis confirms that there are significant differences in local tempo patterns between the three conditions, and that these differences correspond to specific melodic gestures. For instance, the sixteenth-note passage in the tenor voice at the beginning of m.5 is emphasized by means of a significantly larger tempo deceleration (ritenuto) when the tenor is emphasized than when other voices are emphasized. A similar analysis holds for m.4, where the soprano melody is much more active than the other voices, an aspect that performers may attempt to emphasize by using a markedly different acceleration profile at this juncture when emphasizing the soprano voice (the score of the piece is provided in Figure 1).



**Figure 6. Mean tempo variation for each voice-emphasis condition, averaged across performers.**

#### IV. DISCUSSION

This study sought to identify the expressive means used by harpsichordists to emphasize a specific voice in a polyphonic piece. Four parameters were analyzed: note onset asynchrony, velocity, articulation (overlap), and tempo pattern. Mean onset asynchronies were much larger than those observed in organ or piano performance, averaging more than 100 ms between outer voices. However, asynchronies did not vary significantly across conditions, suggesting that they do not play a major role in communicating voice emphasis. On average, notes belonging to the upper three voices were struck with a higher velocity and played with a more detached articulation when they were emphasized. Furthermore, functional data analysis revealed local differences in tempo patterns which corresponded to passages in which a specific melodic gesture was accentuated by means of *ritenuto*. To our knowledge, this is the first reported evidence of the use of local modification of tempo patterns as an expressive strategy to emphasize a voice or melody.

The results presented here indicate that harpsichordists use several strategies in order to emphasize a specific voice. Because dynamic differentiation is limited on the harpsichord, this finding was not unexpected. It remains to be seen whether these strategies are successful, that is, whether listeners can actually perceive the performers' intentions.

#### ACKNOWLEDGMENTS

This research was supported by grants from NSERC and a Canada Research Chair (Stephen McAdams). We wish to thank Bennett Smith, Julien Boissinot and Harold Kilianski (CIRMMT) for their technical assistance, McGill Sound Recording for permission to use their equipment, and Martha de Francisco for her assistance with sound recording.

#### REFERENCES

- Braasch, J., & Ahrens, C. (2000). Attack transients of free reed pipes in comparison to striking reed pipes and diapason pipes. *Acustica*, 86, 662-670.
- Bresin, R., & Battel, G. U. (2000). Articulation strategies in expressive piano performances. *Journal of New Music Research*, 29(3), 211-224.
- Deboor, C. (1978). *A practical guide to splines*. Berlin: Springer-Verlag.
- Fletcher, N. (1977). Analysis of the design and performance of harpsichords. *Acustica*, 37(2), 139-147.
- Gingras, B. (2006). Emphasizing voices in polyphonic organ performance: issues of expressive performance on an instrument with fixed tone intensity. In: Baroni, M., Addessi, A.R., Caterina, R., and Costa, M., editors, *Proceedings of the 9<sup>th</sup> International Conference on Music Perception and Cognition* (pp. 1712-20). Bologna, Italy.
- Gingras, B. (2008). Expressive strategies and performer-listener communication in organ performance. Unpublished doctoral dissertation, McGill University, Montreal, Canada.
- Gingras, B., & McAdams, S. (2007). Improved score-performance matching using both structural and temporal information from MIDI recordings. *Society for Music Perception and Cognition*. Montreal, Canada (July 30-August 3).
- Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, 110(1), 563-572.
- Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31(6), 759-767.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance*, 15(12), 331-346.
- Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception*, 14(1), 23-56.
- Palmer, C., & Brown, J. C. (1991). Investigations in the amplitude of sounded piano tones. *Journal of the Acoustical Society of America*, 90(1), 60-66.
- Penttinen, H. (2006). On the dynamics of harpsichord and its synthesis. In Verfaillie, V., editor, *Proceedings of the 9<sup>th</sup> International Conference on Digital Audio Effects* (pp. 115-120). Montreal, Canada.
- Ramsay, J. O. (2003). FDA Analysis Toolkit. Unpublished software application for Matlab, available for download from ftp site [ego.psych.mcgill.ca/pub/ramsay/](http://ego.psych.mcgill.ca/pub/ramsay/).
- Ramsay, J.O. & Silverman, B.W. (1997). *Functional data analysis*. New York: Springer-Verlag.
- Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, 43(2), 121-131.
- Repp, B. H. (1995). Acoustics, perception, and production of *legato* articulation on a digital piano. *Journal of the Acoustical Society of America*, 97(6), 3862-2874.
- Repp, B. H. (1996b). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, 100(6), 3917-3931.
- Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception*, 23(2), 137-152.