

“Life, Color, Warmth and Truth”: Expressive timing in German late Romantic organ music

Yulia Draginda

<https://orcid.org/0000-0002-0905-7631>

McGill University/CIRMMT

yulia.draginda@mail.mcgill.ca

Ichiro Fujinaga

<https://orcid.org/0000-0003-2524-8582>

McGill University/CIRMMT

ichiro.fujinaga@mcgill.ca

SCIENTIFIC ARTICLE

Submitted date: 31 Oct 2020

Final approval date: 22 mar 2021

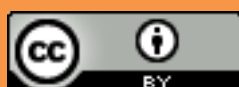
Abstract: This paper proposes a mathematical model for expressive timing in German late Romantic organ music based on Riemannian phrasing principles. The model is built following the symmetric hierarchical motivic scheme and then applied to the organ work of Max Reger. It was shown that music expressive content is carried by the model parameter defined as temporal elasticity. Along with the standard analytical and empirical evaluation, the model was also evaluated in terms of its naturalness. We found that the correlation between model and human performance is comparable with the correlation between distinct professional human interpretations.

Keywords: Hugo Riemann; Music information retrieval; Expressive timing; Organ performance.

“Vida, Cor, Calor e Verdade”:Tempo Expressivo na Música de Órgão Romântica Tardia Alemã

Resumo: Este artigo propõe um modelo matemático para o tempo expressivo na música de órgão romântica tardia alemã, baseado nos princípios do fraseado de Riemann. O modelo é construído seguindo o esquema de hierarquia motivica simétrica e, em seguida, aplicado ao trabalho para órgão de Max Reger. Foi demonstrado que o conteúdo expressivo da música é preservado no modelo, isso por meio de um parâmetro denominado elasticidade temporal. Junto com as avaliações analíticas e empíricas convencionais, o modelo também foi avaliado quanto à sua naturalidade. A correlação entre o modelo e a performance musical humana é comparável à correlação entre diferentes performances profissionais da obra.

Palavras-chave: Hugo Riemann; Recuperação de informação musical; Tempo expressivo; Performance musical de órgão.



“Life, Color, Warmth and Truth”:¹

Expressive timing in German late Romantic organ music

Yulia Draginda, McGill University/CIRMMT, yulia.draginda@mail.mcgill.ca

Ichiro Fujinaga, McGill University/CIRMMT, ichiro.fujinaga@mcgill.ca

1. Introduction

During its centuries-long history, the organ has been perceived in a very controversial way. Some composers were impressed by its sound capacities (for example, W. A. Mozart (1866) crowned it as “the king of all instruments”), while others blamed it for being too mechanical and inexpressive. For instance, Igor Stravinsky said: “I dislike the organ’s legato sostenuto, and its blur of octaves, as well as the fact that the monster never breathes” (Craft and Stravinsky 1982, p 46). This paper proposes an interdisciplinary approach to quantitative organ performance analysis and aims to reveal the instrument’s expressive potential.

Every performance, live or recorded, contains specific human expression conveyed by the musician to the listeners. W. Göbl wrote that “without this expressivity, the music would not attract people; it is an integral part of the music” (Göbl *et al.* 2008). In organ performance, the instrument’s constraints do not allow the refined continuous control over note intensity or local variations of timbre. Nonetheless, timing remains always available, and that is why it becomes one of the most important tools to realize the organist’s expressive intent.

Although organ music is a significant part of Western musical tradition, very few empirical studies on organ performance have been published so far (Nielsen 1999; Jerkert 2004; Gingras 2008). However, no one, as far as we know, has performed quantitative research on German late Romantic organ music, which is the main focus of the present work. A more detailed account of the expression specificity for this style is given in the following section.

2. Quantitative research on expressive timing

Quantitative research on expressive timing in music performance has so far primarily focused on the piano (Palmer 2005; Gabrielsson 2003), as well as some other instruments, such as the violin (Cheng and Chew 2009), the clarinet (Vines *et al.* 2006), the cello (Hong 2003), the guitar (Juslin 2000), and the harpsichord (Gingras *et al.* 2016), while the organ has generally been ignored. The development of MIDI technology has contributed significantly to piano performance research as well (Windsor and Clarke 1997; Repp 1998). However, due to the nature of the instrument, the conclusions related to the piano are not directly applicable to organ performances. The main difference is in the way one produces sound: for organs, the air is moved through pitched pipes, while for pianos, strings are struck with felted hammers. The dynamic changes on the

1. [Editor Note: This paper was made using a L^AT_EX template and has some adequation facing the docx template.]

organ are by default discrete and usually controlled by “pulling” stops (sets of pipes of different shapes that can produce different timbres), while the dynamic variety of the piano is continuous and controlled directly by varying the pressure on the key. That is why dynamics, as a concept managed by fingertips on piano keys, do not translate to the organ. In addition, the organ is a unique instrument producing a sound that does not decay; an organ tone will hold indefinitely as long as a key is pressed and the instrument is powered on. All these particular organ features make a significant difference to the expressive elements in organ performance compared to the piano and other instruments.

The initial motivation for this study was rooted in two papers (Gingras *et al.* 2010; Gingras and McAdams 2011), which are the only two empirical studies to have been published so far on expressive pipe organ performance using MIDI. But the focus in both studies was on Baroque music and J. S. Bach; that is why the observations made are stylistically inappropriate for Romantic music. The main difference is in the use of articulation as an expressive tool: while in Baroque music different types of articulation are allowed, in the Romantic style the articulation is prescribed as an absolute legato or is unambiguously defined by the composer (Laukvik 2010). In this paper, we present a new analytical concept according to the specific properties of the instrument, as well as to the stylistic attributes of German late Romantic organ music.

3. Mathematical model for Riemannian phrasing

The phrasing theory of Hugo Riemann (1849–1919), a renowned German musicologist of the XIXth century, was chosen as the basis for our model. Several studies on music performance have shown the importance of Riemannian principles for the stylistically correct German late Romantic music interpretation (Lohmann 1995; Laukvik 2006; Sander 2006; Szabó 2016), whereas Riemann’s scientific approach to expressivity noticed by Rehding (2003) is particularly beneficial for computer simulation. Furthermore, Riemann was the primary composition professor of Max Reger, whose work is the subject of this research. Riemann’s phrasing theory was described in detail in two central works (Riemann 1884, 1903), as well as was briefly touched in other books and articles (Riemann 1900, 1912; Riemann and Fuchs 1890). For this study, because of its direct reference to performance expression questions, the earliest work (Riemann 1884) was chosen as a model basis.

According to Riemann, the main goal of expressive phrasing is to underline the hierarchical motivic structure, which he considers essential to the correct music score interpretation (Riemann 1900, p 90). The phrasing at each level must have the arch-like timing shape: starting from slow, followed by an *accelerando*, then returning to the original tempo. This concept corresponds to the Phrase Arch rule in the KTH rule system (Friberg *et al.* 2006). Riemann shows that the arch-like shape is built on different hierarchical levels (see Figure 1) that is very similar to the hierarchical tree-based model proposed by Todd (1985). The main difference with the latter is in the underlying principles of hierarchical organization: while Riemann’s scheme is based on the numbers of notes in the motives, Todd’s timing model is relying on *Generative Theory of Tonal Music* (Lerdahl and Jackendoff 1983), which inherits the core musicological principles from Schenkerian harmonic analysis. Consequently, the Riemannian approach is more approachable for computational simulation, especially for the harmonically complex late Romantic music.

(Riemann 1903, p VIII) described the short 2- or 3-notes motives as “the smallest possible musical units of stand-alone expressive importance”, which form the large-scale symmetry as shown in Figure 1. We propose to model the tempo arch at each level as the positive semi-ellipse, where the long axe is defined by Riemannian length, and the short axe is proportional to the metronomic tempo. The coefficient of proportionality e is defined as **temporal elasticity**: it shows the maximum of the model tempo deviation against the metronomic

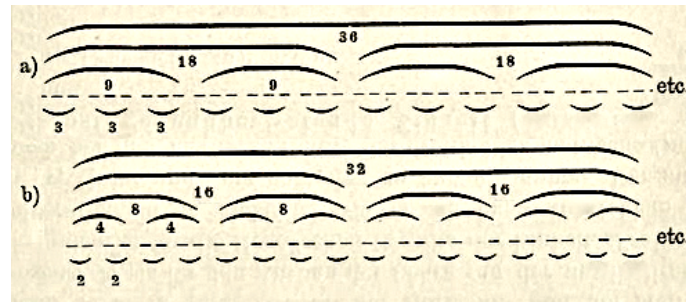


Figure 1: Riemann's motivic scheme (Riemann 1884, p 8).

tempo for each level. According to our research hypothesis, temporal elasticity is capable of carrying the expressive information in organ performance in the German late Romantic style. The following part of this paper describes in greater detail the formal mathematical model of the temporal elasticity concept and its empiric evaluation through the listening test.

3.1. Symmetric model

In the most general symmetric case, the temporal elasticities e_{ij} do not vary within the level and are uniformly related to the global coefficient e_0 for the arch over the whole piece:

$$e_{ij} = e_i = 1.5 * \frac{e_0}{N_i}, \quad (1)$$

where i , is the number of the level; j , sequence number of the ellipse on the i^{th} level, and N_i , quantity of ellipses on the i^{th} level. The model contains the sum of all arcs on all levels. For example, the model curve Y for one global arch and four subsequent levels will be:

$$Y = \left(\sqrt{1 - \left(x - \frac{h_0}{a_0} \right)^2} \right) * e_0 * T + \sum_{i=1}^4 \sum_{j=1}^{N_i} \left(\sqrt{1 - \left(x - \frac{h_{ij}}{a_{ij}} \right)^2} \right) * e_{ij} * T + T \quad (2)$$

where T is the constant metronomic tempo, h and a are the parameters of each ellipse (the ellipse's center and the short axe, respectively). For the duple meters (Figure 1, case b), the number of ellipses doubles when moving from the main arch to the subsequent levels. In our case, the quantities of ellipses N_i on the i^{th} level are $N_1 = 2$, $N_2 = 4$, $N_3 = 8$ and $N_4 = 16$.

Max Reger's *Choral Prelude* op. 135a/1 "Ach bleib mit deiner Gnade" was chosen for the analytical evaluation of this model. It is a textbook example of the Riemannian symmetry with the time signature 4/4: eight bars long, clear cadences in bars number 2, 4, 6 and 8. Level 1 in our model corresponds here to the cadences in bars 4 and 8; level 2 shows the cadences in bars 2, 4, 6 and 8; level 3 highlights the four quarter-note segments endings, and level 4 corresponds to the smallest microstructure of two quarter-note motives (see Figure 2)

The professional organ performance of this piece was recorded in MIDI format at the large 4-manuals *Casa-vant* organ in The Church of Saint Andrew and Saint Paul in Montreal (Canada). Local temporal information on

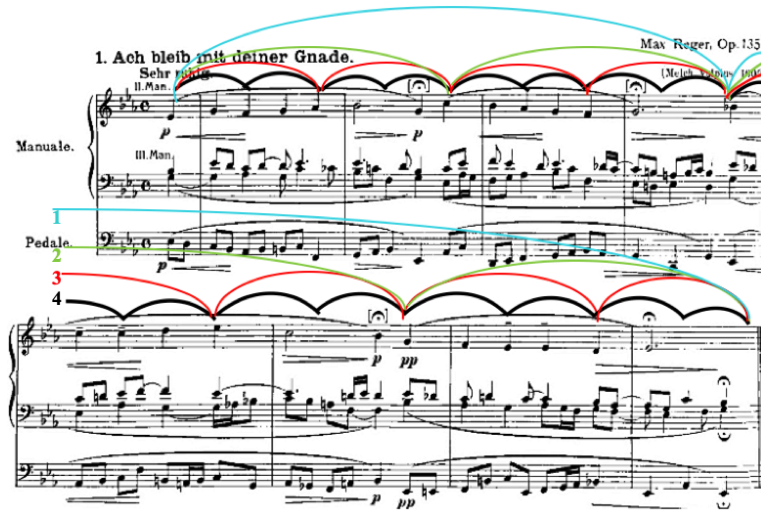


Figure 2: Max Reger, *Choral Prelude op. 135a/1* (with marked phrasing levels 1-4).

the sixteenth-note level was extracted manually in *Logic Pro X* through beat-mapping, exported to *Matlab*² and compared against the model curve Y obtained from the equation (2). Figure 3 shows the model levels and the resulting curve Y with the value of temporal elasticity $e_0 = 0.5$.

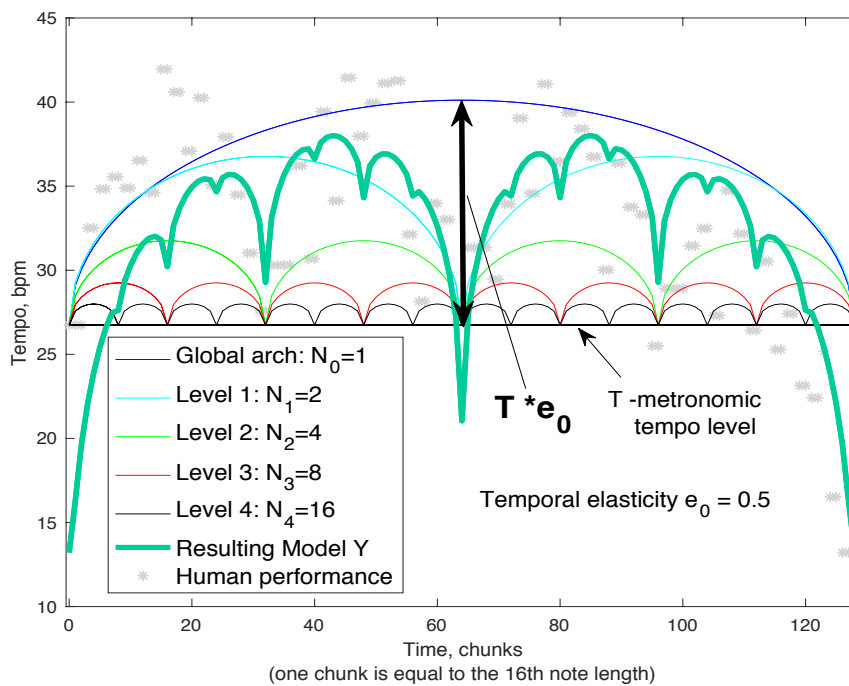


Figure 3: Mathematical model of Riemann’s motivic scheme for Max Reger *Choral Prelude op. 135a/1*.

The straightforward regression of the symmetric model with fixed elasticity values against the real human performance data gave $R^2 = 0.43$ ($p < 0.001$). It could be hypothetically compared with the results in (Windsor and Clarke 1997), where the highest R^2 obtained for timing from the similar symmetric model was 0.299. However, this comparison is speculative because the model described by Windsor and Clarke (1997) was applied to the early Romantic piece for piano and had a different methodological background.

2. If there was no event at the sixteenth-note level, the local tempo value was interpolated and set to the preceding notes local tempo.

3.2. Improved symmetric model

In real performance practice, the absolute symmetry as described by equation (1) is rarely kept, and some irregularities may be possible. In the more general case, temporal elasticities e_{ij} take values:

$$e_{ij} = k_{ij} * \frac{e_0}{N_i}, \quad (3)$$

where the weights coefficients k_{ij} may differ both within the specific level and over all levels.

To build an improved model, which takes into account the varying weight coefficients, the optimization using the *Nelder-Mead* simplex algorithm (Lagarias *et al.* 1998) was performed. The values of e_0 and k_{ij} were set as optimization parameters for the *Matlab fminsearch* function in order to minimize the distance between the model Y and the human performance data. Symmetric values $e_0 = 0.5$ and $k_{ij} = 1.5$ were used as the initialization guess for the first simplex. The obtained model provided a highly significant coefficient of determination $R^2 = 0.816$ ($p < 0.001$), which might be comparable with the values of the variance accounted for by a repeat human performance (Todd 1992). The correlation between the optimized curve and human performance $R = 0.9$ is even higher than the correlation between two professional organists' interpretation of this piece ($R = 0.89$, see Table 5 in Section 4).

But despite the high value of R^2 , the weights obtained through the optimization process cannot be directly used for the model simulation because they contain information about both relevant (performer's expressive intent) and irrelevant (e.g., related to the technical issues) tempo deviations. We undertook a detailed analysis of the weights' distribution so to determine the most prominent trends and map them to the relevant score features or historic performance practice principles.

Figure 4 shows the values for k_{ij} , as well as the mean values for the levels 1–4. The optimized temporal elasticity was close to the initial guess: $e_0 = 0.43$, and the coefficients k_{ij} were distributed in the range from $2 * 10^{-4}$ to 7.5 with the mean over all levels $mean(k_{ij}) = 1.53$, which correlates well with the initial value for the symmetric model $mean(k_{ij}) = k_{ij} = 1.5$ (see Figure 4).

This obtained result contains three important factors to consider:

1) Levels 1 and 3 are less elastic than levels 2 and 4. This coincides well with the “motivic paradigm” of the Riemannian model, with the small motives being its cornerstone, as well as to the brief score analysis of the piece showing that the cadences in bars 2, 4, 6 and 8 are all equally strong (and serve as breezing breakpoints between choral phrases) while the cadence in the 4th bar does not have any additional weight.

2) The mean value of the coefficients on the second level is $mean(k_{ij}) = 2.7$. But the first coefficient k_{21} , corresponding to the tempo arch over bar 1, is significantly higher than the level mean value, $k_{21} = 4.9$ (see Figure 4). This might result from the initial performer's attempt to grab the audience's attention right at the beginning of the piece by a substantial tempo increase. Interestingly, this corresponds to the observations made already back in 1902 by Charles Sears (1902), who investigated the rhythmical patterns in organ performance of the church hymns.

3) Neither the symmetric nor the generic model curve did not reach the maximal (fastest) tempo of human performance. That is understandable because with the further increase in temporal elasticity, the “tails” of

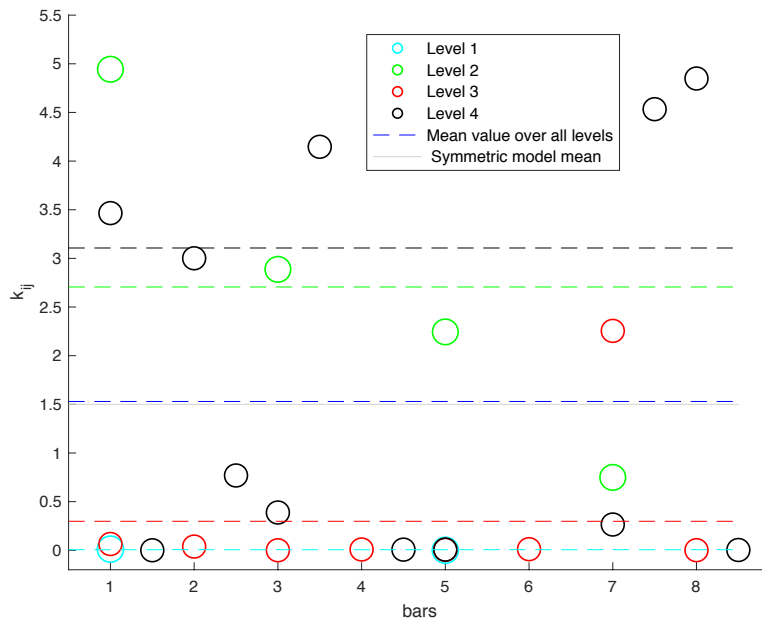


Figure 4: Optimized coefficients k_{ij} at the levels 1–4. Colored dashed lines represent the mean values for each level.

the model curve at the beginning and at the end cross the zero-level and thus output the sense-less negative tempo values. Hence, introducing appropriate boundary conditions for a start- and end- tempo might improve the model performance and make it more elastic.

These three observations from the generic model analysis were used to create the appropriate modifications for the improved symmetric model:

Modification 1: Levels 2 and 4 were set to have more weight than levels 1 and 3. The respective values of k_{ij} are shown in Table 1. The mean value for the improved model was deliberately kept the same as for the initial symmetric model: $mean(k_{ij})_{symmetric} = mean(k_{ij})_{improved} = 1.5$.

Table 1: Coefficients k_{ij} for the improved symmetric model.

Levels	k_{ij} , average value (best fit model curve)	k_{ij} , improved symmetric model
1	0.0062	0.5
2	2.7061	2.5
3	0.2981	0.5
4	3.1068	2.5

Modification 2: The first coefficient at the second level was increased up to $k_{21}=3.5$ (so to give the same elasticity as for the next four bars) for emulation of the performer’s expression at the beginning of the piece.

Modification 3: The model tempo curve was defined to be above or equal to the performance tempo minimum value (final *ritardando*) to prevent it from getting negative values while increasing the temporal elasticity. The improved model with different elasticity values is shown in Figure 5.

Henceforth, in the improved model, only the global arch elasticity e_0 can vary and thus encompass the variations in expressive timing. The other elasticity coefficients e_{ij} are fixed by being tied to the respective level values of k_{ij} (Table 1), and therefore ensure the Riemannian symmetry.

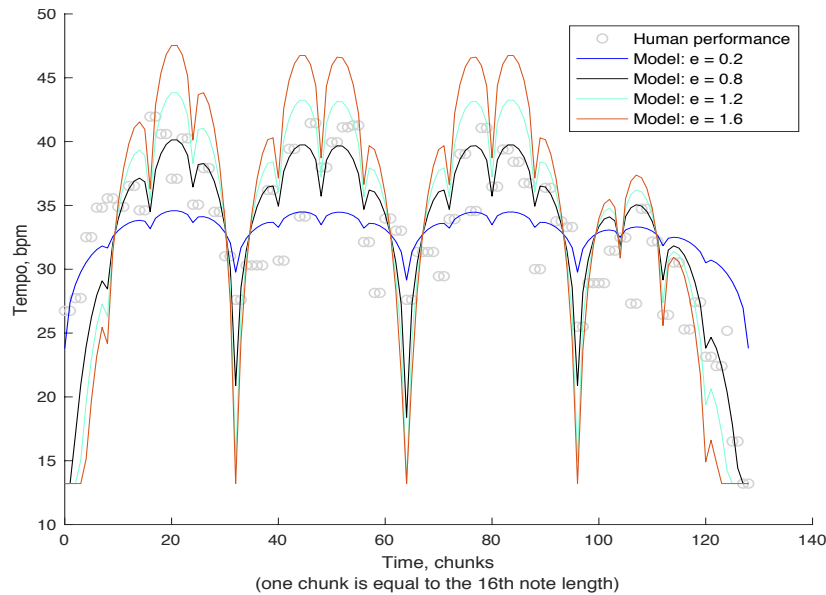


Figure 5: Improved symmetric model with three modifications.

The improved model was evaluated mathematically with the same algorithm as described above, but this time only the value of e_0 was optimized. The summary of regression analysis depending on the modifications made is presented in Table 2 (all coefficients R^2 were significant at the level $p < 0.001$).

Table 2: Summary of regression analysis for the improved symmetric model.

Number of modifications	Optimal value of e_0	R^2
No modifications (“pure” symmetric)	0.43	0.43
One (only unequal levels)	0.415	0.47
Two (unequal levels and k_{21} increase)	0.55	0.62
Three (unequal levels, k_{21} increase and boundary conditions)	0.62	0.63

It is revealed that the improved model has better performance than the “pure” symmetrical model. Specifically, introducing the elasticity boost at the second level makes a significant difference. It is a meaningful finding for the performance practice illustrating how essential are the first bars of the piece. The boundary conditions allow the model to increase the R^2 as well; however, when the elastic extension of the model curve becomes too high, the R^2 slowly decreases.

4. Comparative temporal analysis

The temporal analysis of five different audio recordings (Dupont 2012; Lohmann 1998; Buttman 2015; Pirkl 2016; Pals 2019) was performed in order to compare it with the model. The recordings in focus were made on different organs (authentic and modern) and performed by organists with diverse playing experience (organ amateurs and professional organists).

4.1. Evaluation of beat tracking systems

Beat tracking analysis for organ audio recordings is an extremely challenging task. Because of very reverberant acoustic conditions for each given recording, it is not easy to use an automatic procedure for onset detection

or tempo estimation. Furthermore, there is no standard organ: every organ has its particular stops disposition; the action might be mechanical, pneumatic, electric or electro-pneumatic, as well as different organs do not respond in the same way to a given touch. The developed polyphonic texture of the piece makes it hard to determine the beat “event”, when the active melody entrance might be confused with the down-beat, or the beats corresponding to the soft “over legato” articulation might be missing. Thus far, only one scientific research in beat tracking for the organ audio recordings was done (Jerkert 2004). But it was focused on the analysis of only terse music excerpts from fugues by J. S. Bach, and the beat-tracking procedure was done manually through visual inspection of the spectrogram. That is why a prior evaluation step was made here in order to confirm the appropriateness of selected automated and semi-automated beat tracking systems for German late Romantic organ music.

In order to determine the most appropriate procedure for beat detection, an existing MIDI file (used in the previous section) was played back on a real organ and professionally recorded from the listener’s position in the hall so to capture the acoustics of the church. The audio recording of this MIDI was compared to the MIDI track, and MIDI onset times were considered as the “ground truth”. According to Milligan and Bailey (2015), the onset detection algorithms, focused on periodicity, are more appropriate for the analysis of instrumental music with unclear soft onsets than the energy-based approaches for the onset detection procedure, commonly used for music with percussive sounds.

Consequently, the initial onset detection was done with Tempogram Toolbox (Grosche and Müller 2011), based on the general beat-tracking assumption that the beats must occur in a periodic fashion, at least within a certain time window. This algorithm succeeded in detecting the majority of onsets at the 8th-note level, with some confusions in the fragments with significant local agogical changes. These short fragments were refined manually using the BeatRoot system (Dixon 2007), displaying the musical data and beats in a graphical interface. BeatRoot was also used to add the onset times on the 16th-notes level where such note events were present.

Three following sets of data were exported to *Matlab* and evaluated against the “ground truth” in terms of Precision, Recall and F-measure:

- 1) BR–beats detected initially by BeatRoot (without Tempogram clicks);
- 2) TG–beats detected by Tempogram Toolbox (8th-notes level);
- 3) TGBR–beats detected by Tempogram Toolbox and manually corrected by BeatRoot (8th-notes level).

The tolerance window was set as proposed in McKinney *et al.* (2007) to the one-fifth of the average “ground truth” inter-onset interval at the 8th-note level. The summary of the evaluation analysis is given in Table 3.

Table 3: Evaluation of beat tracking systems.

<i>Data set</i>	<i>F-measure</i>	<i>Precision</i>	<i>Recall</i>
BR	0.54	0.37	1
TG	0.84	0.79	0.89
TGBR	0.98	0.98	0.98

The brief analysis of Table 3 shows that BeatRoot has a high Recall value (without false negatives detections) but a very low Precision, which is not acceptable for the current study. Tempogram Toolbox has a higher Precision value (fewer false positives detections) than BeatRoot but still contains some detections outside

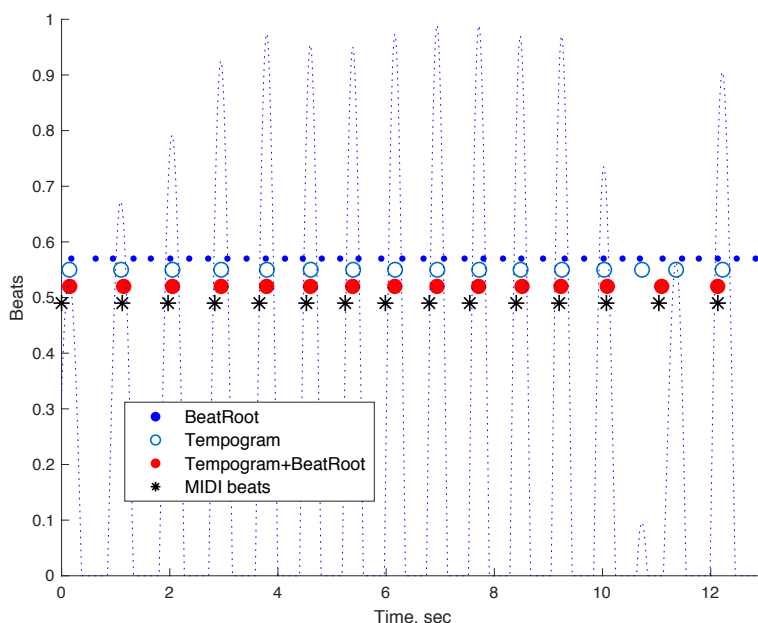


Figure 6: Beats detected by different beat tracking systems for the first 12 seconds of op. 135a/1 by Max Reger.

the tolerance window (see Figure 6). The manual correction with BeatRoot helped to fix this issue, providing the Precision equal to 0.98 with only one inaccurate detection for the whole piece.

Thus, the combination of the Tempogram Toolbox with BeatRoot was proven to be an effective beat tracking system for extracting beats from organ audio files.

4.2. Temporal analysis

The following recordings were taken from *Naxos* music library or downloaded from license-free *Youtube* videos:

- 1) Jean-Baptist Dupont / Walcker organ (1904), professional organist (Dupont 2012);
- 2) Ludger Lohmann / Link organ (1906), professional organist (Lohmann 1998);
- 3) Bernhard Buttmann / restored Koulen organ (1911), professional organist (Buttmann 2015);
- 4) Thorsten Pirkl / Kreienbrink organ (1965), organ amateur (Pirkl 2016);
- 5) Herman Pals / Hauptwerk/Sonarte organ (2012), organ amateur (Pals 2019).

All recordings were then analyzed with Tempogram Toolbox and BeatRoot. The data containing beat times at the 8th- and 16th-notes level (where present) was imported to *Matlab* for further processing. The local tempo was calculated at the 16th-notes level and compared to both symmetric and improved symmetric models. Table 4 shows the outlines of the analysis (all coefficients R^2 were significant at the level $p < 0.001$).

According to the analysis, professional organists (Dupont, Lohmann, Buttmann) tend to play more expressive than organ amateurs (Pirkl and Pals): the average values of temporal elasticity are $e_0=0.71$ (improved) $e_0=0.39$ (symmetric) across the professional organists against the respective values $e_0=0.6$ and $e_0=0.36$ for

Table 4: Regression analysis of Max Reger op.135a/1 audio recordings against the model.

Recording	Optimal value of e_0 , improved model	R^2	Optimal value of e_0 , symmetric model	R^2
Dupont	0.67	0.64	0.44	0.52
Lohmann	0.77	0.55	0.35	0.46
Buttmann	0.7	0.62	0.38	0.53
Pirkl	0.63	0.46	0.39	0.43
Pals	0.57	0.63	0.33	0.51

the amateurs. In all five cases, the improved model gave better values of R^2 than the symmetric model, thus confirming the appropriateness of modifications suggested in section 3.2.

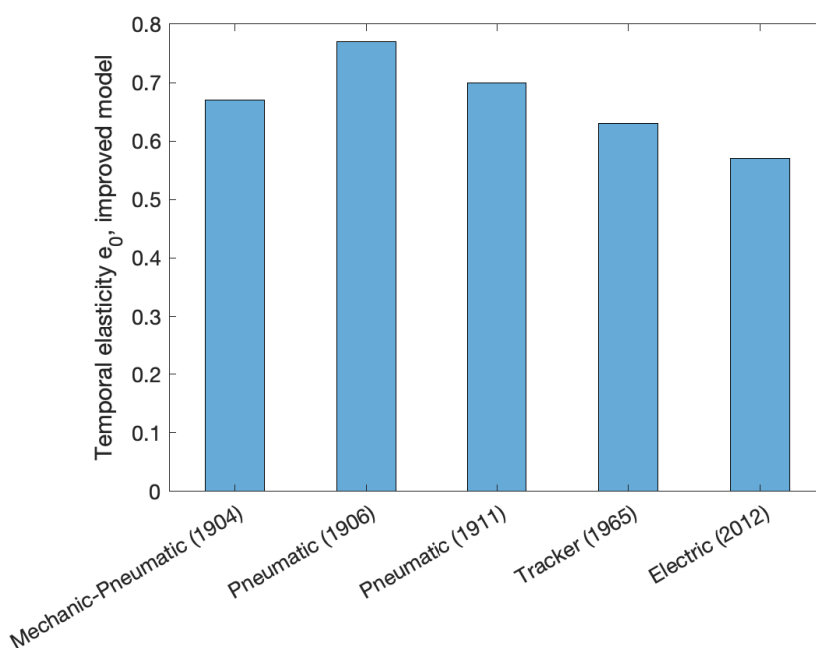


Figure 7: Values of temporal elasticity e_0 depending on the organ action.

For both models, improved and symmetric, the average values for professionals $R^2=0.6$ and $R^2=0.5$ are slightly higher than the respective values for amateurs: $R^2=0.54$ and $R^2=0.47$ because the professional organists are more likely to be aware of advanced Riemannian phrasing principles (usually taught at the graduate school).

It is also interesting to notice that the highest values for temporal elasticity were obtained for the recordings made on the period instruments with pneumatic action (see Figure 7). The most elastic performance with $e_0=0.77$ (Lohmann 1998) was recorded at the Link organ (1906) of the Evangelische Stadtkirche in Giengen an der Brenz, which is “one of the best-preserved instruments of the Reger period, and as such, ideally suited to the realization of the Bavarian composer’s music” (Fugatto 2020). Therefore, it is possible to conclude that the instrument’s type plays an important role in the performance expression, and historic organs allow performers the more elastic phrasing.

In order to see the overall model performance in terms of its naturalness, the cross-correlation analysis was made across all audio recordings, MIDI recording and two computed models: the symmetric model with $e_0=0.38$ (SM) and the improved symmetric model with $e_0=0.7$ (IM). The analysis results are presented in Table 5.

Table 5: Correlation matrix for audio recordings.

R	Dupont	Lohmann	Buttmann	Pirkl	Pals	MIDI	SM	IM
Dupont	1	0.78	0.81	0.71	0.78	0.73	0.71	0.8
Lohmann	0.78	1	0.89	0.67	0.79	0.71	0.67	0.74
Buttmann	0.81	0.89	1	0.69	0.86	0.72	0.73	0.79
Pirkl	0.71	0.67	0.69	1	0.63	0.49	0.66	0.68
Pals	0.78	0.79	0.86	0.63	1	0.73	0.64	0.79
MIDI	0.73	0.71	0.72	0.49	0.73	1	0.65	0.8
SM	0.71	0.67	0.73	0.66	0.64	0.65	1	0.92
IM	0.8	0.74	0.79	0.68	0.79	0.8	0.92	1

As expected, the highest correlation among the non-simulated performances was between two professional interpretations of the piece at the organ with the same kind of action (pneumatic): $R=0.89$ (Lohmann/Buttmann).

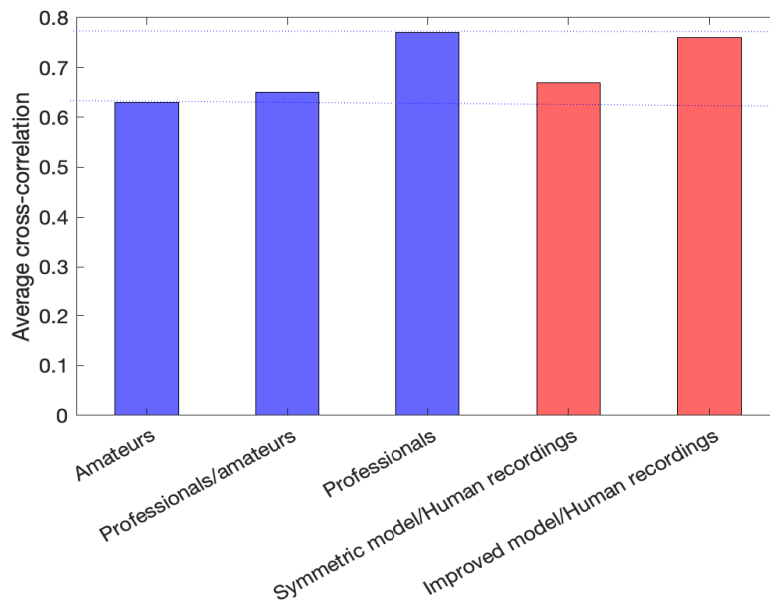


Figure 8: Average cross-correlation between human performances (blue bars) and between models and human performances (red bars).

The average correlation across the distinct professional recordings (including MIDI) $R=0.77$ is higher than the correlation between two amateurs $R=0.63$ and average cross-correlation between amateurs and professionals $R=0.75$. The average correlation across all recordings is $R=0.67$ and $R=0.76$ for the symmetric and the improved models, respectively (Figure 8). As shown in Figure 8, both correlation coefficients for symmetric and improved symmetric models fall within the interval $[0.63, 0.77]$; hence the model performance in terms of its naturalness outperforms the average amateur and is close to the professional human interpretation.

5. Empirical evaluation of the model

Several studies show that listeners can perceive the expression resulting from the performer’s structural interpretation of the piece (Palmer 1996; Gabrielsson 2003). Because our Riemannian model relies on the structure hierarchy, it was suggested that the difference in values of the temporal elasticity might result in a different expressive impact. We designed a small listening test to evaluate our model empirically.

5.1. Methodology

The listening experiment was conducted online with the 13 volunteers, where 9 participants identified themselves as “organ students, working professionals, organ amateurs or enthusiasts”, and 4 had “little or no experience with organ or its repertoire”. The stimuli were different versions of Max Reger’s op. 135a/1. One version was recorded by a professional human performer, and the other four versions were simulated to contain the artificial expressive timing according to the improved symmetric model with the elasticities values $e_0=0.4$, $e_0=0.8$, $e_0=1.2$, $e_0=1.6$ (see Figure 5). The model tempo curves were modelled in *Matlab* using the equation (2), then obtained tempo values were applied programmatically to the equitemporal MIDI file using the C#/.NET framework DryWetMidi (2020). Model files were played back and recorded from the listener’s position at the same instrument and with the same timbre (same organ stops) as the original human recording. This procedure allowed us to eliminate all other possible performance discrepancies except the timing variations.

The web interface presented to the listeners was designed with the *Web Audio Evaluation Tool* (Jillings *et al.* 2015) and included the volume control, audio files, play buttons, as well as sliders with the range 1–100 for the response of each stimulus. The question to the listeners was: “Despite the type of emotions felt, please evaluate how expressive each performance is at the scale 1-100 (by moving the slider)”; this question was chosen in accordance with the results in Bhatara *et al.* (2011), where a similar methodology was used. To facilitate the process for the participants, the additional descriptions were attached to the slider: “Not expressive, mechanical” (at position 20), “Moderately expressive” (at position 50) and “Extremely emotional” (at position 80). Users were asked to use the full range of the slider scale and were allowed to listen to each version as many times as they want so to make the final decision. For each user, the order of stimuli shown was randomized.

5.2. Results

Participants’ answers were divided by 100 and imported to *SPSS* for statistical analysis. The grand mean of listeners’ ratings was 0.52 (SE=0.05), which demonstrated that responses were centred around the scale. Individual means (participants wise) ranged from 0.32 to 0.72 (SD=0.09). A repeated measures one-way ANOVA with a Greenhouse-Geisser correction determined that the effect of temporal elasticity level was significant ($F(2.71, 32.4) = 5.45, p < 0.05$). As expected, the model version with the highest value of temporal elasticity $e_0=1.6$ was rated as the most expressive, and the model with the lowest $e_0=0.4$ had the lowest rating; the human performance, corresponding approximately to the $e_0=0.62$ (Table 2), was rated close to the most elastic model. The linear regression of the mean ratings for the model elasticity values was significant at the 0.05 level ($R^2=0.96$). Figure 9 shows the mean listeners’ ratings and the linear trend.

The post-hoc pairwise comparisons with the Turkey HSD test have determined that the difference between ratings for $e_0=1.6$ and $e_0=0.4$ was significant ($p < 0.01$), as well as the difference between rating for the human performance and the least elastic model $e_0=0.4$ ($p < 0.01$). At the margin of statistical significance was the difference between the ratings for $e_0=0.4$ and $e_0=1.2$ ($p < 0.065$); all other ratings did not significantly differ from each other.

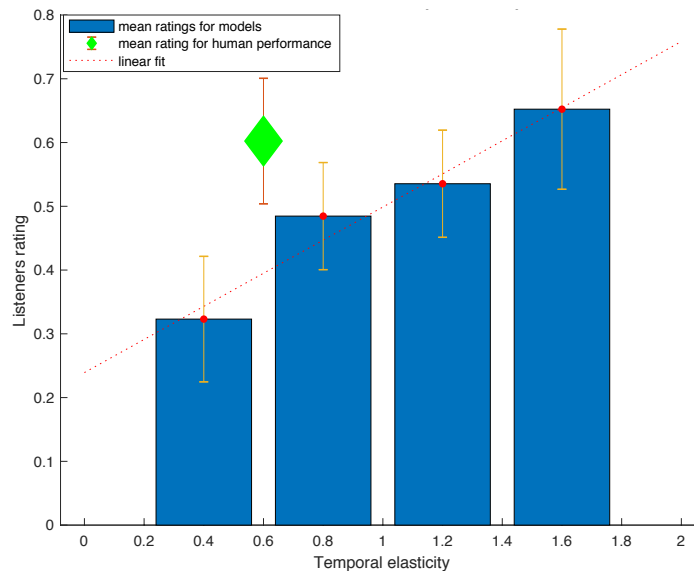


Figure 9: Average listeners' ratings for models with different values of temporal elasticity and human performance.

6. Discussion

The results of both analytical and empirical evaluation of our model for Riemannian phrasing show that the temporal elasticity is indeed a very effective tool to carry expression. The relatively high (as compared to the other structural timing models) value of the coefficient of determination, especially for the improved symmetric model, demonstrates that the model variance explains up to 63% of the overall timing variance in human performance. According to the comparative temporal analysis, the model is closer to the professional organists' than to the amateurs' interpretations, which indirectly confirms its stylistic correctness. It is especially remarkable that the correlation model/human performance is comparable with the correlation between distinct professional human interpretations.

An empiric evaluation of the model reveals that the expressivity manifested by the temporal elasticity is comprehensible for the modern listeners, both with or without prior organ experience. The least elastic model was rated as the less expressive or mechanical, followed by the linear upward trend up to the most elastic variant. However, further investigation is needed to explore the correlation between the finer differences in temporal elasticity and listeners' responses.

To further our research, we plan to generalize the proposed model to other Romantic pieces with various structures because the fundamental aspects of Riemannian phrasing theory allow creating models for any arbitrary combination of initial motives provided the general symmetry. The majority of German late Romantic organ pieces do not follow the strict 4+4 bars periodic pattern. However, exclusively for this style, it is possible to determine the phrase boundaries denoted by the phrasing slurs in the musical score and build upon them the symmetric hierarchical model using the equation (2). A major impediment to this process is the first modification for the improved model, which implies the inequality of the weights for different levels. The elaboration of the AI-based algorithm for the automated weighting in our model is reserved for future work.

7. Conclusion

Key music concepts defined by Hugo Riemann, one of the most influential theorists of his time, are still of great interest today. In this paper, a mathematical model for Riemannian phrasing was introduced and evaluated, both analytically and empirically, against professional human interpretation. The proposed model has a considerable two-fold implication: it can be used in performance phrasing analysis, as well as in computer simulation of expressive timing for German late Romantic organ music.

Hugo Riemann concludes: “The well-shaped phrasing contributes to the clarity of the motivic and harmonic structure so that it becomes an essential factor for the correct interpretation of the music score; furthermore, the touching, emotional performance would be impossible without the expressive phrasing, which evokes the musical expression of Life, Color, Warmth and Truth” (Riemann 1900, p 90)³.

8. Acknowledgement

This research has been supported by the Fonds de recherche du Québec – Société et culture (FRQSC).

9. References

- Bhatara, Anjali, Anna Tirovolas, Lili Duan, Levy, Bianca, and Levitin, Daniel. 2011. *Perception of Emotional Expression in Musical Performance*. *Journal of Experimental Psychology*, 37(3):921–934.
- Buttmann, Bernhard. 2015. *Max Reger: Complete Organ Works, Volume 4*. Oehms Classics OC854, compact disc.
- Cheng, Eric and Chew, Elaine. 2009. *Quantitative Analysis of Phrasing Strategies in Expressive Performance: Computational Methods and Analysis of Performances of Unaccompanied Bach for Solo Violin*. *Journal of New Music Research*, 37(4):325–338.
- Craft, Robert and Stravinsky, Igor. 1982. *Dialogues*. University of California Press.
- Dixon, Simon. 2007. *Evaluation of the Audio Beat Tracking System Beatroot*. *Journal of New Music Research*, 36(1):39–50.
- DryWetMidi. 2020. *C#/.NET library for high-level processing of standard MIDI files*. <https://github.com/melanchall/drywetmidi>, last accessed October 27, 2020.
- Dupont, Jean-Baptist. 2012. *Max Reger: Intégrale de l'œuvre pour Orgue, Volume 1*. Hortus O86-087, compact disc.
- Friberg, Anders, Bresin, Roberto, and Sundberg, Johan. 2006. *Overview of the KTH Rule System for Musical Performance*. *Advances in cognitive psychology*, 2(2–3):145–161.
- Fugatto. 2020. *The Link Organ of the Ev. Stadtkirche at Giengen-an-der-Brenz (Germany)*. http://fugatto.free.fr/giengen_adb.html, last accessed October 21, 2020.
- Gabrielsson, Alf. 2003. *Music Performance Research at the Millennium*. *Psychology of Music*, 31(3):221–272.

3. Translated from German by Yu. D.

- Gingras, Bruno. 2008. *Expressive strategies and performer-listener communication in organ performance*. McGill University, Montreal, Canada.
- Gingras, Bruno and McAdams, Stephen. 2011. *Improved score-performance matching using both structural and temporal information from MIDI recordings*. *Journal of New Music Research*, 40:43–57.
- Gingras, Bruno, McAdams, Stephen, and Schubert, Peter. 2010. *The performer as analyst: a case study of JS Bach's "Dorian" Fugue (BWV 538)*. *Music Theory and Interdisciplinarity—8th Congress of the Gesellschaft für Musiktheorie Graz*, Saarbrücken: Pfau-Verlag.
- Gingras, Bruno, Pearce, Marcus, Goodchild, Meghan, Dean, Roger, and McAdams, Geraint Wiggins Stephen. 2016. *Linking melodic expectation to expressive performance timing and perceived musical tension*. *Journal of Experimental Psychology: Human Perception and Performance*, 42(4):594–609.
- Grosche, Peter and Müller, Meinard. 2011. *Extracting Predominant Local Pulse Information from Music Recordings*. *IEEE Transactions on Audio, Speech and Language Processing*, 19(6):1688–1701.
- Göbl, Werner, Simon Dixon, Giovanni De Poli, Friberg, Anders, Bresin, Roberto, and Widmer, Gerhard. 2008. *Sense in Expressive Music Performance: Data Acquisition, Computational Studies, and Models*. In *Sound to Sense – Sense to Sound: A State of the Art in Sound and Music Computing*, Berlin:Logos.
- Hong, Ju-Lee. 2003. *Investigating Expressive Timing and Dynamics in Recorded Cello Performances*. *Psychology of Music*, 31(3):340–352.
- Jerkert, Jesper. 2004. *Music Articulation in the Organ*. *Proceedings of Joint Baltic-Nordic Acoustics Meeting*, Mariehamn, Finland.
- Jillings, Nicholas, Man, Brecht De, Moffat, David, and Reiss, Joshua. 2015. *Web Audio Evaluation Tool: A Browser-Based Listening Test Environment*. *Proceedings of the 12th Sound and Music Computing Conference*, Maynooth.
- Justin, Patrik. 2000. *Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception*. *Journal of Experimental Psychology: Human Perception and Performance*, 26:1797–1813.
- Lagarias, Jeffrey, Reeds, James A., Wright, Margaret, and Wright, Paul. 1998. *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions*. *SIAM Journal of Optimization*, 9(1):112–147.
- Laukvik, Jon. 2006. *Aspekte Musikalischer Interpretation in der Spätromantik*. *Orgelschule zur historischen Aufführungspraxis – Teil 2: Orgel und Orgelspiel in der Romantik von Mendelssohn bis Reger und Widor*, Stuttgart: Carus.
- Laukvik, Jon. 2010. *Historical Performance Practice in Organ Playing, Part 2. The Romantic Era*. Stuttgart: Carus.
- Lerdahl, Fred and Jackendoff, Ray. 1983. *A Generative Theory of Tonal Music*. Cambridge: The MIT Press.
- Lohmann, Ludger. 1995. *Hugo Riemann and the Development of Musical Performance Practice*. *Proceedings of the Göteborg International Organ Academy*, Department of Musicology: Göteborg University.
- Lohmann, Ludger. 1998. *Max Reger: Organ Works, Volume 2*. Naxos 8.553927, compact disc.
- McKinney, Martin, Moelants, Dirk, Davies, Matthew, and Klapuri, Anssi. 2007. *Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms*. *Journal of New Music Research*, 36(1):1–16.

- Milligan, Keziah and Bailey, Nicholas. 2015. *A Review of Software for Note Onset Detection*. *Proceedings of the 4th International Scientific Meeting for Sound and Musical Instrument Studies*, Tavira, Portugal.
- Mozart, Wolfgang Amadeus. 1866. *The Letters of Wolfgang Amadeus Mozart (1769-1791)*, Vol. 1. New York and Philadelphia.
- Nielsen, Siw. 1999. *Regulation of Learning Strategies during Practice: A Case Study of a Single Church Organ Student Preparing a Particular Work for a Concert Performance*. *Psychology of Music*, 27:218–229.
- Palmer, Caroline. 1996. *On the Assignment of Structure in Music Performance*. *Music Perception*, 14:23–56.
- Palmer, Caroline. 2005. *Timing in skilled piano performance*. *The Journal of the Acoustical Society of America*, 85:66.
- Pals, Herman. 2019. *Max Reger (1873–1916), 30 kleine Choralvorspiele, No 1–10, Opus 135a*. YouTube video, 15:47. Posted February 15, 2019. <https://www.youtube.com/watch?v=p34MHTksixU>.
- Pirkl, Thorsten. 2016. *Max Reger (1873–1916), 30 kleine Choralvorspiele, No 1–10, Opus 135a*. YouTube video, 9:07. Posted September 2, 2016. <https://www.youtube.com/watch?v=p34MHTksixU>.
- Rehding, Alexander. 2003. *Hugo Riemann and the Birth of Modern Musical Thought (New Perspectives in Music History and Criticism)*. Cambridge University Press.
- Repp, Bruno. 1998. *A Microcosm of Musical Expression. I. Quantitative Analysis of Pianists' Timing in the Initial Measures of Chopin's Etude in E major*. *The Journal of the Acoustical Society of America*, 104(2):1085–1100.
- Riemann, Hugo. 1884. *Musikalische Dynamik und Agogik, Lehrbuch der musikalischen Phrasierung auf Grund einer Revision der Lehre von der musikalischen Metrik und Rhythmik*. Hamburg, Germany.
- Riemann, Hugo. 1900. *Präludien und Studien: gesammelte Aufsätze zur Ästhetik, Theorie und Geschichte der Musik, Band II*. Leipzig: Hermann Seemann.
- Riemann, Hugo. 1903. *System der Musikalischen Rhythmik und Metrik*. Leipzig: Breitkopf & Härtel.
- Riemann, Hugo. 1912. *Vademecum der Phrasierung*. Leipzig: M. Hesse.
- Riemann, Hugo and Fuchs, Carl. 1890. *Katechismus der Phrasierung: praktische Anleitung zum Phrasieren*. Leipzig: Hesse & Becker.
- Sander, Natalia. 2006. *Max Reger's Organ Works in Aspect of its Biographical, Instrumental and Interpretational questions*. Ph.D. Dissertation. Nizhny Novgorod: M. Glinka Nizhny Novgorod State Conservatory press (In Russian).
- Szabó, Balázs. 2016. *Zur Orgelmusik Max Regers*. Bonn: Dr. Josef Butz Verlag.
- Todd, Neil. 1985. *A Model of Expressive Timing in Tonal Music*. *Music Perception: An Interdisciplinary Journal*, 3(1):33–57.
- Todd, Neil. 1992. *The Dynamics of Dynamics: A Model of Musical Expression*. *Journal of The Acoustical Society of America*, 91(6):3540–3550.
- Vines, Bradley, Krumhansl, Carol, and Levitin, Daniel. 2006. *Cross-modal interactions in the perception of musical performance*. *Cognition*, 101(1):80–113.
- Windsor, Luke and Clarke, Eric. 1997. *Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool*. *Music Perception*, 15(2):127–152.