Temporal elasticity as the most important tool for expressive performance of German late Romantic organ music

Yulia Draginda

Schulich School of Music
McGill University, Montreal, Quebec

May 2021

A paper submitted to McGill University in partial fulfillment of the requirements of the degree of
D. Mus Performance Studies

Contents

Abstract

Music expression is often regarded as an elusive phenomenon requiring "*a complicated and mysterious analysis*" (Matravers 2001, 225). However, several aspects of performance expression can be investigated using modern computational and musicological methods. In organ music, because the organist has only limited control over local timbre variations or note intensity, timing often becomes the main expressive parameter by which the performer conveys most of the instrument's musical expressivity. Thus, my project is dedicated to the investigation of timing as the most important tool to create an expressive performance.

The principal goal of my research is to introduce a single timing parameter, **temporal elasticity**, use it to describe the expressive impact quantitatively and find such values of this parameter for the German late Romantic organ music, which would be, on the one hand, stylistically appropriate, and on the other hand, emotionally comprehensible for the modern listener. The proposed temporal elasticity model's practical applications include but are not limited to the comparative performance analysis and computer simulation of expressive timing. To illustrate the relevant expressive tempo deviations, I performed a computer simulation of Hugo Riemann's (1884) hierarchical phrasing scheme. This pattern was applied to the equitemporal MIDI files of different Max Reger's organ works and compared against the recordings of professional organists, as well as Reger's own recording. An evaluation of this model was made through the regression analysis and listening tests (REB #20-06-021). According to my knowledge, this is the first study to examine the phrasing in late Romantic organ music; quantitative research on expressive timing has thus far focused either on the piano and other instruments (Todd 1985; Palmer 1996; Repp 1998; Gabrielsson 2003) or the Baroque organ music (Jerkert 2004; Gingras et al. 2010).

Résumé

L’expression musicale est souvent considérée comme un phénomène insaisissable nécessitant “une analyse complexe et mystérieuse” (Matravers 2001, 225). Cependant, plusieurs aspects de l’expression de la performance peuvent être étudiés à l’aide de méthodes modernes de calcul et de musicologie. En musique d’orgue, parce que l’organiste n’a qu’un contrôle limité sur les variations locales du timbre ou l’intensité des notes, le timing devient souvent le principal paramètre expressif par lequel l’interprète transmet la majeure partie de l’expressivité musicale de l’instrument. Ainsi, mon projet est dédié à l’investigation du timing comme l’outil le plus important pour créer une performance expressive. Le but principal de ma recherche est d’introduire un seul paramètre temporel, **l’élasticité temporelle**, l’utiliser pour décrire l’impact expressif quantitativement et trouver de telles valeurs de ce paramètre pour la musique d’orgue allemande tardive Romantique, qui serait, d’une part, stylistiquement approprié, et d’autre part, émotionnellement compréhensible pour l’auditeur moderne. Les applications pratiques du modèle d’élasticité temporelle proposé comprennent, sans toutefois s’y limiter, l’analyse comparative des performances et la simulation du timing expressif. Pour illustrer les écarts de tempo expressifs pertinents, j’ai réalisé une simulation du schéma hiérarchique d’Hugo Riemann (1884). Ce schéma a été appliqué aux fichiers MIDI équitemporals de différentes œuvres d’orgue de Max Reger et comparé aux enregistrements d’organistes professionnels, ainsi qu’à l’enregistrement de Reger lui-même. L’analyse de régression et les tests d’écoute (REB #20-06-021) ont permis d’évaluer ce modèle. Selon mes connaissances, il s’agit de la première étude à examiner le phrasage dans la musique d’orgue Romantique tardive; jusqu’à présent, la recherche quantitative sur le timing expressif s’est concentrée soit sur le piano et d’autres instruments (Todd 1985; Palmer 1996; Repp 1998; Gabrielsson 2003), soit sur la musique d’orgue Baroque (Jerkert 2004; Gingras et al. 2010).

Acknowledgements

I want to thank my doctoral supervisors, Prof. Hans-Ola Ericsson (organ performance) and Prof. Ichiro Fujinaga (music technology research), as well as my doctoral committee members: Chair Dr. Lena Weman and Dr. Jonathan Wild. Their guidance helped shape this project.

I want to thank my former organ teachers: Galina Semyonova, Prof. Alexei Semyonov, Prof. Ludger Lohmann, Prof. David Franke, Prof. Martin Sander, and Christian Lane. They gave me a solid background in organ playing and performance research.

I want to thank the FRQSC (Fonds de recherche du Québec—Société et culture) for the financial support of my doctoral studies.

I want to thank the CIRMMT (The Centre for Interdisciplinary Research in Music Media and Technology) for the Student Award granted to this research project.

I want to thank the Office of Graduate Studies for approving my mobility grant to do research at Max Reger Institute, Karlsruhe (Germany). While in Karlsruhe, I had a priceless opportunity to discuss my model with renowned researchers of Max Reger's work Dr. Jürgen Schaarwächter and Dr. Stefan König, whose help is highly appreciated.

I want to thank Dr. Adrian Foster for the professional recording of my final doctoral recital at The Church of Saint Andrew and Saint Paul in Montreal.

I want to thank my dear mother, Maria Nikolaeva, and my dear friends Vera Korotaeva, Néstor Nápoles, Lauma Akmene, Maria Gajraj, Dr. Jonathan Embry, Raphaël Ashby, Marina Tyurina, Sofya Iglitskaya, Anna Yurkova, Maxim Milkov, Irina Marchenko, and Greg Couture for being with me every step of the way.

A special thank you and my deepest appreciation go to Prof. Dr. Natalia Sander. Without her kind help and mentorship, my professional organ career would be over before it ever started.

I. Introduction

The organ is one of the most impressive classical music instruments, with "the longest history of all, with a repertory larger than that of any other instrument and with a magnificence beyond any other musical invention from the Greeks to the present day" (Williams 1980, 14). However, during its centuries-long history, the organ has been perceived in a very controversial way. Some performers and composers were inspired by its sound capacities; for example, W. A. Mozart wrote that "the organ always was, both in my eyes and ears, the king of all instruments" (Mozart 1866, 67). In contrast, others described it as a mechanical, "essentially non-expressive instrument," claiming that "it is wrong to play upon it in an emotional way" (Thalben-Ball 1950). The famous 20th-century composer 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, 46). The primary purpose of this work is to conduct interdisciplinary performance research based on both appropriate musicological methods and modern computational algorithms in order to reveal the expressive potential of the organ.

While listening to a live organ performance or a recording, one is listening to music played by humans and that, therefore, contains a human expression. W. Goebl wrote that "without this expressivity, the music would not attract people; it is an integral part of the music" (Goebl et al. 2008). In organ performance, the instrument's constraints do not allow refined continuous control over note intensity or local timbre variations; the possibilities of dynamic manipulations are not universal and strongly depend on the instrument in use. In the case of Romantic music, articulation—another commonly used expression tool—is usually defined by the composer or set to legato, so it also cannot be changed significantly. However, the expressive timing always remains available to the performer, and that is why it becomes one the most attainable tools to realize the organist's expressive intent. My research, thus, is dedicated to the investigation of temporal elasticity as the most important tool to create an expressive performance.

Quantitative research on expressive timing in music performance has so far primarily focused on the piano (Todd 1985; Palmer 1996; Repp 1992; Gabrielsson 2003), as well as some other instruments, such as the cello (Johnson 1999; Hong 2003), the guitar (Juslin 2000), the clarinet (Vines et al. 2006), the violin (Cheng and Chew 2009), and the harpsichord (Gingras et al. 2016). The development of MIDI (Musical Digital Instrument Interface) has greatly contributed to expressive performance research as well (Windsor and Clarke 1997; Repp 1998; Friberg et al. 2006; Cancino-Chacón et al. 2018). However, due to the nature of the instrument, the conclusions related to the piano or other instruments are not directly applicable to organ performances. Gabrielson and Juslin (1996) investigated expressive performance strategies for various instruments and showed that the variations in expressive timing were heavily dependent on the musical instrument and style.

Although organ music is a significant part of Western musical tradition, only very few empirical studies on organ performance have been published so far (Nielsen 1999; Gingras et al. 2015). However, they touched only the general problems in organ playing, such as practicing and performers' errors. The expressive organ performance was investigated by Jerkert (2004) and Gingras (2008, 2010); these studies, nonetheless, were focused exclusively on Baroque organ music. As far as I know, no one has ever performed quantitative research on expressive timing in German late Romantic music, which is the main focus of the present work. I hope that my research will cover this knowledge gap, with its fundamental aim to create a quantitative expressive timing model specifically for the organ and, more precisely, for the German late Romantic style.[1]

[1] Results of this research were presented at COBS 2020 (Draginda 2020), ISMIR 2020 (Draginda and Fujinaga 2020), and published in *Per Musi* Scholarly Music Journal (Draginda and Fujinaga 2021).

II. Literature overview

Despite the fact discussed above that the expressive timing models for other instruments cannot be applied to the organ in their original form, it is highly insightful to investigate prior work with a view to make it a reference point for the new model created in this project. This Chapter gives a short overview of expressive timing models and provides an initial motivation to propose the Riemannian theoretical concepts as the model basis for German late Romantic organ music.

II.1 Expressive timing models

By definition, "models are attempts at codifying hypotheses about expressive performance in terms of mathematical formulas or computer programs so that they can be evaluated in systematic and quantitative ways" (Cancino-Chacón et al. 2018, 1). In general, four different approaches to modelling expressive timing could be distinguished (Widmer and Göbl 2004):

- the synthesis-by-rule concept
- the complex mathematical simulation
- the data-driven approach relying on AI algorithms
- the predictive modelling based on musical structure.

Sundberg et al. (1983) first proposed the synthesis-by-rule concept, where the synthesized performance was created based on seven specific rules taken from the professional performance practice. These rules were later elaborated and transformed into the "analysis-by-synthesis," putting the professional musicians and researchers in a permanent feedback loop to find the best rules' parameters (Friberg et al. 2006). The *Phrase Arch rule* defined expressive timing in this ruleset: "a musical phrase is often performed with an arch-like shape applied to tempo and dynamics… The phrase is typically slow/soft in the beginning, fast/ loud in the middle and ends slow/soft, modelling a crescendo/accelerando and decrescendo/rallentando pattern” (Friberg et al. 2006, 149). Looking ahead, as it will be shown in the next Section II.2, this rule fully coincided with the Riemannian concept of joint expressive dynamic and phrasing (Riemann and Fuchs 1890, 13). For the organ, nevertheless, this might be done solely by the variations in timing (Riemann 1900, 91[2]).

The complex mathematical simulation of expressive performance was proposed in the ambitious “Mathematical Music Theory” by Guerino Mazzola (1990)—voluminous research incorporating music-theoretical, mathematical, philosophical, psychological and aesthetic areas of knowledge. Mazzola’s model consisted of two parts: an analytical part with several analysis tools assigning the specific *weights* to the notes, and a performance part, providing music generation based on the analysis made. This model was used by the RUBATO Composer System (Mazzola et al. 2008) to create an artificial expressive performance in different styles, but the expressive impact of this music has not been empirically evaluated. Mazzola’s idea of weighting has indirectly influenced this research project; however, instead of the note-by-note approach, the weighting of hierarchical levels was implemented for Romantic organ music (see Chapter III for details).

The data-driven approach implied challenging the computer to look for specific patterns in the data extracted from the professional human performance. An extensive overview of expressive models based on AI algorithms was given in Cancino-Chacón et al. (2018). Usually, the modelling was done at the note level, predicting the expressive timing from the score features and/or previously learned rules. For the acceptable level of accuracy, this approach required a huge corpus of performance data, which would be barely possible to collect in the case of German late

[2] Original text: “Dynamik und Agogic wirken beim Klavier und den meisten übrigen Instrumenten zusammen: die Orgel ist auf Agogik allein angewiesen” (Riemann 1900, 91).

Romantic organ music because of the small number of available consistent professional recordings. However, this methodology might be promising later in the future if the recordings' quantity significantly increases.

The attention to the expressive potential of the musical structure was first directed by Shaffer (1980), who investigated the recordings of concert pianists and discovered the hierarchical organization of the expressive timing. Clarke (1982) confirmed that the expressive characteristics of musical performance were related to the piece's structural characteristics and showed the correlation between rhythm and tempo. One of the earliest attempts to model the hierarchical structure of expressive timing computationally was made by Neil Todd (Todd 1985, 1989, 1992). He wrote: "a valid performance depends primarily on the perception and communication of the rhythmic life of the composition. That is, we must first discover the shape of the piece and then try to make it as clear as possible to our listeners." (Todd 1985, 40). It is essential to notice here that Hugo Riemann had the same vision of the correct (or "valid") performance (Riemann 1900, 90).

Todd's models were based on "*Generative Theory of Tonal Music*"— the fundamental theoretical framework developed by Lerdahl and Jackendoff (1983), which in turn inherited the core musicological principles from Schenkerian analysis. Lerdahl and Jackendoff combined Schenker's ideas with the laws of generative linguistics. They elaborated the following *Reduction Hypothesis*: "the listener attempts to organize all the events of a piece into a single coherent structure, such that they are heard in a hierarchy of relative importance" (Lerdahl and Jackendoff 1983, 106). The *Time Span Reduction*, representing a particular case of *Reduction Hypothesis*, defined that the piece of music could be presented in terms of articulated melodic groups organized in a tree-like hierarchy starting from the smallest metric unit and moving through larger levels.

Such groups were determined by specific properties, making them coherent: the full or partial cadences or other melodic movements towards the tonal stability. The degree of stability was varied according to the hierarchic level of its group.

Todd (1985, 1989) used the *Time Span Reduction* and modelled the tree-like hierarchy. He proposed a parabolic encoding function with six parameters to model the segments in the duration structure. He compared the model with the real performances and obtained visually similar curves for the human and algorithmic performances; however, he did not provide any quantitative evaluation data.

In 1992, Todd improved his model, and instead of one simply connected tree, used a set (or a "forest") of binary trees organized on several hierarchic levels (Todd 1992). In the same work, he introduced the strong connection between the timing and dynamics ("the faster, the louder, the slower, the softer") and suggested that expressive tempo changes were governed by analogy to physical motion. The result was awe-inspiring: the regression of the algorithmic model against the real performance data for Chopin's *Prelude* gave $R^2 = 0.74$, which was comparable for the values of the variance accounted for by a repeat human performance. Conic section curves, indeed, appeared to have a great potential in fitting tempo deviations for Romantic piano music. For example, Bruno Repp (1992) also found that local timing patterns followed a parabolic timing function in Schumann's *Träumerei*. In the current study, another member of the conic sections' family—elliptic curve—was proposed to model expressive timing in the German late Romantic music for organ (see Chapter III).

Windsor and Clarke (1997) applied Todd's model to the algorithmic analysis of Schubert's *G♭ major Impromptu* op. 9. Even despite the relatively low correlation between algorithmic and the professional pianist performance (see Appendix 1), they showed that the model, in general, was statistically significant. However, unlike Todd, they came to the fundamental conclusion that timing and dynamics were not connected by a definitive function (Windsor and Clarke 1997, 147). Interestingly, the same finding was confirmed by Repp (1999) by investigating the various piano performances of Chopin's *Etude* op. 10 №3. Up to the present moment, the model in Windsor and Clarke (1997) came closest to the model proposed in the current paper, which is why it was chosen as a reference point for the analytical evaluation in Chapter III.

The further advance in this field was the multi-level model (Widmer and Göbl 2004), which combined the predictive model (Windsor and Clarke 1997) with machine learning algorithms: the rule learning algorithm was used to learn a rule-based model of the local residual effects resulting after fitting parabolic approximation to a given tempo curve. An artificially expressive performance of Mozart's op. 11/1, created in accordance with this model, won a Second Prize in a Computer Performance Rendering Contest in Tokyo in 2002, where computer interpretations of classical music were rated by listeners (Widmer and Göbl 2004, 210). The proposed ensemble of different models was, in my personal opinion, the most up-and-coming for this task and yielded a computational model of expressive timing comprising both the global hierarchical structure of the music and the local musical context. This idea of a hybrid approach was the inspiration for the Chapter VI of this project.

In general, the hierarchical structure-based approach showed excellent potential for expressive timing modelling. However, it relied on Schenkerian harmonic analysis to define the phrasing boundaries, making it significantly less approachable for computational simulation, especially for the harmonically complex late Romantic music. Furthermore, the fundamental model (Todd 1992) contained a number of different parameters, which were hard to interpret and control. Consequently, it would be beneficial to create a similar model based on hierarchical music structure, which would be independent of the harmonic analysis and would have a single, easy interpretable model parameter. That is how the temporal elasticity concept was born.

The model presented in this study is also based on music structure and, therefore, benefited from the conceptual advantages of the predictive structural approach. However, it has only one raw-number parameter (defined as temporal elasticity) and relies on the Riemannian phrasing theory, which does not require any knowledge of music harmony for building a hierarchical phrasing scheme. A more detailed account of Riemannian phrasing principles is given in the following Section.

II.2 Riemannian phrasing theory and German late Romantic organ music

Hugo Riemann (1849–1919) was one of the most influential music theorists of his time. Riemann's ability to collect the contemporary trends of his time, abstract them into robust scientific theories and methodically describe them in his numerous books made his scholarship an appealing basis for mathematical modelling. "What is more, his systematic musical thought linked all the major aspects of the burgeoning discipline of musicology, from acoustics to aesthetics and history (and back again). For the first time, it seemed possible that musicology could indeed stand up to the objective, scientific scrutiny: Riemann helped define music as a stable, knowable entity – and, what is more, as an object worthy of scientific study" (Rehding 2003, 183).

Riemann's phrasing rules were described in his two fundamental studies (Riemann 1884, 1903), as well as in other books and articles (Riemann 1900, 1912; Riemann and Fuchs 1890). The earliest work (Riemann 1884) was especially interesting because of its reference to the questions of performance practice, while the last major study (Riemann 1903) was primarily focused on the more formal analysis. For Riemann, the phrasing aimed to underline the clarity of the musical structure and harmony (what he considered to be an essential part of the correct performance), contributed to the expression of “Life, Color, Warmth and Truth” (what was important for the touching, expressive performance); but the wrong phrasing might ruin the performance and make it grotesque.[3]

In order to methodize the phrasing, Riemann introduced the following signs:

1) Agogic accent (“*Der agogische Akzent*”) – the sign indicating the prolongation of the stressed note (and the shortening of the preceding note, respectively). In addition to the single accent (^), there were also double and triple accents denoting the greater degrees of prolongation. The agogic accent was vital for the German Romantic organ music, where it was barely possible to make an accentuation through the articulation (Laukvik 2006, 263).

2) Reading mark (“*Lesezeichen*”) – the sign marking the boundaries of the small motives (Riemann 1884, 9). There was a single (|) and the double sign for the motives of different degree of importance (||); this sign was “meant to counter the natural tendency of musicians to group notes according to barlines—this tendency certainly being a relic of a Baroque performance practice” (Lohmann 1995, 261).

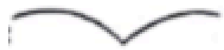
3) Different groups of slurs. In his writings, Riemann replaced the *legato* slur with the phrasing slur because *legato* was the default articulation at that time;[4] “the new office of the slur was to indicate the articulation of the musical thought (themes, periods, movements) into its natural divisions (phrases)” (Riemann and Fuchs 1890, 13). Riemann suggested that it would drastically

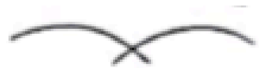
---

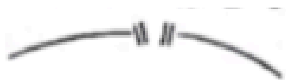
[3] Original text: “…die Phrasierung i) der Deutlichkeit dient, d.h. die Taktart, die motivische Gliederung und die Harmonie klarlegt, also ein unentbehrlicher Faktor des korrekten Vortrags ist; ferner, dass sie 2) dem Ausdruck erst Leben, Farbe, Wärme, Wahrheit gilt, sodass ein packender, ergreifender Vortrag ohne sie unmöglich ist. Endlich aber ist auch das Umgekehrte nicht außer Acht zu lassen, nämlich dass 3) falsche Agogik den Ausdruck ins Fratzenhafte verzerren, das Erhabene zum Lächerlichen machen muss (was kaum die falsche Dynamik in gleichem Maaßen vermag).” (Riemann 1900, 90).

[4] Original text: “legato ist die schlichte, die gewöhnliche Art der Tonverbindung, besonders für Töne, die zu einem Motive oder zu einer Phrase zusammen gehören.” (Riemann 1888, 8).

improve the quality of musical writing (Riemann 1884, 242).[5] The slurs might be grouped in the following way (Sievers 1967, 509):

- Connected slurs ("*Bogenanschluß*"): sign , indicating the immediate beginning of the following phrase after the end of the previous one;

- Crossed slurs ("*Bogenkreuzung*"): sign , indicating that the following phrase should start before the end of the previous one;

- Broken slurs ("*Abbrechende Bogen*"): sign , used when the first phrase did not reach its logical end; the second part after the break of the slur meant the so-called " negative part" (or the end) of the following second phrase.

- Stuttgart comma ("*Stuttgarter Komma*"): sign , prescribing the break in the middle of the phrase.

- Double slurs ("*Doppelbogen*"): sign , showing the outline of motives (bottom slurs) and long phrases (top slurs) at the same time.

In order to define the rhythmical and metrical structure, Riemann introduced three basic terms (Riemann 1903, 7–13):

1) Rhythmic quality ("*Rhythmische Qualität*"): the tempi between 60 MM and 120 MM were called "middle times" ("*mittlere Zeiten*") and considered to be "normal", whereas the tempi

[5] Original text: "Ich [Riemann] verband von Anfang an bei Abfassung dieses Buches den praktischen Zweck der Vorbereitung einer folgenschweren Verbesserung unserer Notenschrift, nämlich der Ersetzung der Legatobögen durch Phrasenbögen" (Riemann 1884, 242).

which were faster than 120 MM or slowly than 60 MM were brought into line with the "middle times" by summation or division respectively.

2) Metric quality ("*Metrische Qualität*") or different weights. According to Riemann, the music consisted of alternating the short light time ("*Nebenzeit*") with long heavy time ("*Hauptzeit*"); the light time always came first (e.g., at the upbeat).

3) Thematic motives ("*Thematische Motiven*") – the parts of melody, usually consisting of two or three notes and representing the smallest possible musical units of stand-alone expressive importance ("*eine kleinste Einheit von selbständiger Ausdrucksbedeutung,*" Riemann 1903, VIII). According to Riemann, the motives formed larger groups[6] in the symmetric hierarchical structure as shown in Figure 1 (in terms of tempo and with the down-beat start).

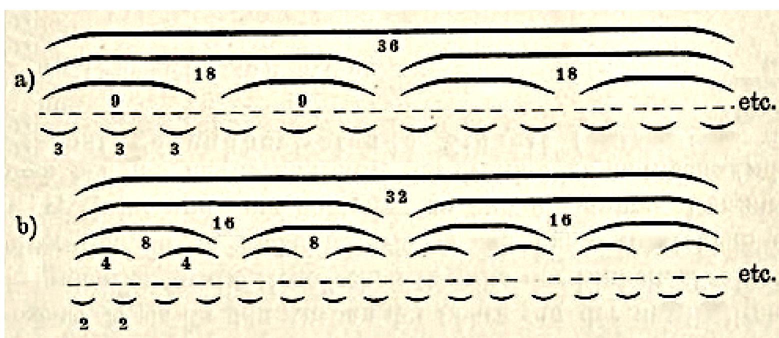


Figure 1. Riemann's hierarchical motivic scheme (Riemann 1884, 8).

In his later work, Riemann (1903, 197) "locked" the metrical structure in the strict 8-bars period with the metric quality principles applied first to notes (short light – long heavy), then to bars (light first bar – heavy second bar), then, similarly, to the two-bars groups, and finally to the 4-bars sentences ("*Vordersatz*" and "*Nachsatz*"). The most straightforward metric scheme for the 8-bars period for the 2/4 time signature is shown in Figure 2.

[6] Original text: "zwei (oder drei) enger aneinander geschlossene Töne, denen zwei (oder drei) andere gegenübertreten, wodurch größere Gruppen entstehen, die einander ebenso entsprechen und mit der wachsenden Größe eine deutlichere Sonderung erfordern" (Riemann 1884, 8)

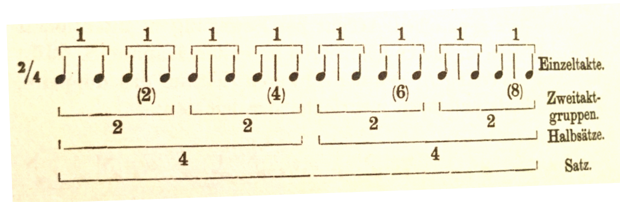


Figure 2. Metric scheme of the 8-bars period (Riemann 1903, 198).

The following weights (or metric qualities) represented the hierarchical organization of the bars:

- Degree 1 for the bars 1, 3, 5 and 7;
- Degree 2 for the bars 2 and 6;
- Degree 3 for the bar 4; and
- Degree 4 for the bar 8.

Strictly speaking, this system was not invented by Riemann; it was rather the "part of the standard themes in theoretical and critical writings on performance of this time" (Doğantan-Dack 2012, 14). Morgan (1978) showed that the symmetrical scheme existed already in the second half of the XVIIIth century and was reported by Johann Kirnberger, who, however, did not describe the accents in the larger rhythmical units. The first hierarchical accentuation was introduced by Gottfried Weber in 1824: "There is also a higher symmetry. Just as beats together form small groups, several groups can also appear bound together as beats of a larger group, of a larger or higher rhythm, a rhythm of higher order. One can even go further and place such a rhythm of higher order with a similar one, or third, so that these two or three together form a yet higher rhythm… The grouping of the larger rhythm is a more broadly conceived symmetry, which is, incidentally, exactly like that of the measures, only on a larger scale" (Morgan 1978, 437). Thus, according to Weber, the larger formal units resulted from an accumulation of smaller units through a process of addition into larger groups. Weber's assumptions were further developed by Moritz Hauptmann, who also conceived the meter as a symmetrical system and considered it beginning accented, but also introduced the complementary "metrically negative" series, starting from the non-accented (light) time.

In the most general case, the phrase for Riemann meant both changes in dynamics and timing: "all in all, the tone-force and rubato furnish the surest indication for determining the limits of slurs, so far as every phrase has but one climax, up to which it waxes in intensity, and after which to close it decreases, relaxing in tempo as well" (Riemann and Fuchs 1890, 13). Thus, it could be helpful for the composers to indicate the phrases not only by slurs but also by the *crescendo* and *diminuendo* signs.[7] However, Riemann especially mentioned that for the organ, the phrasing might be done only by the variations in agogic (Riemann 1884, 8; Riemann 1900, 91). Any increase of motion indicated a start, whereas the prolongation (of a tone) was anticipated as a moment of relaxation. Thus, the perception of motivic structure within the large phrase trailed the agogic/dynamic intermediate peaks of the small individual motives, while the bounds between the motives corresponded to their respective agogic/dynamic minima.

The concepts of symmetry and motivic development presented here built a fundamental methodological basis for my mathematical model of expressive timing. The appropriacy of the Riemannian concepts for the stylistically correct interpretation of the German late Romantic music was confirmed by many renowned organ performance researchers (Lohmann 1995; Laukvik 2006; Sander 2006; Szabó 2016). As it was shown in Lohmann (1995), Riemann's phrasing rules to some extent may be used for the performance of F. Liszt's *BACH*, more precisely, for the motivic shaping of the opening B-A-C-H motive. Laukvik (2006) also mentioned Liszt's *BACH* and proposed the "waves-like" agogic shaping for the 16th-notes passages, as well as showed the

[7] Original text: "Da die übliche Bogenbezeichnung nicht auf Abgrenzung der Phrasen berechnet ist, so helfen sich die Komponisten, um die weitere Ausdehnung der Phrasen anzudeuten, entweder durch die längeren Schattirungszeichen oder durch die Wortvorschriften crescendo und diminuendo" (Riemann 1884, 257).

possibility of the Riemannian dynamic/agogic shaping in organ *Choral Preludes* by J. Brahms and *1st Organ Sonata* by F. Mendelssohn. But, of course, most of the performance research on Riemann's principles was focused on the organ music of Max Reger.

Max Reger (1873–1916) has learned the Riemann's phrasing and metric rules already during his first piano lessons with Adalbert Lindner—the respective markings were found in his instructional piano scores starting from 1884 (Sievers 1967, 556). That is why I found the earlier motivic idea (Riemann 1884) to be more relevant to Reger's music than the later dogmatic metric scheme (Riemann 1903). This observation was later confirmed during my conversation with Dr. Jürgen Schaarwächter and Dr. Stefan König at the Max Reger Institute in Karlsruhe (Germany) in July 2019. From 1890 Riemann became Reger's primary composition and theory teacher, first at the Conservatory in Sondershausen, and later in Wiesbaden, where he became a "member" of an "international phrasing office" ("*Mitglied in einem ganz internationalen Phrasierungsbureau*") (König 2020, 5).

The influence of Riemann's phrasing on Reger's earliest works was quite evident. The performance-level markings made bei red ink ("*Vortragsebene,*" for example, see Figure 3) in his compositions written in the 1890s contained different Riemann's phrasing signs (König 2020, 8). According to König, Reger's most favourite phrasings signs were the crossed slurs, the upbeat slurs and particularly the agogic accent, which he found helpful for the "*fast understanding*" (König 2020, 9).[8] For example, the agogic accent was present in opp. 1–3, 6, and 14; the crossing slurs were used in opp. 3, 9, 48, and 49. Some other signs, such as crossed slurs or double slurs, were present in his homework transcriptions made under the direct supervision of Riemann (Sievers 1967, 539). The most well-known Reger's organ piece containing Riemann's indications

[8] Original text: "Ich [Reger] hätte dieses Zeichen ^ sehr gerne, da es sehr, sehr viel zum schnellen Verständnis nützt" (König 2020, 9).

was the *Choral prelude* "*Komm, süßer Tod*" WoO IV/3, where Riemannian upbeat slur and agogic accent appeared already in the first bar (see Figure 3):

Figure 3. Choral prelude "Komm, süßer Tod!" WoO IV/3 (Reger 1894).

The small star sign (*) near the agogic accent was leading to the footnote, where Reger clearly explained its meaning, as well as the implication of his crescendo and diminuendo signs: "The sign ^ denotes a slight lingering on the note or rest, over which it is placed; < and > have *dynamic* (Swell) and *agogic* meaning" (Reger 1894). Moreover, Reger repeated literally the same explanation in the comment to his *Choral Fantasia* "*Freu dich sehr, o meine Seele*" op. 30 (Reger 1899).[9] It was especially important for the interpretation of the phrases when the usage of expression box was for some (e.g., technical) reasons not possible—then the effect of increasing and decreasing of the motion might be done only agogically, in full accordance with the previously cited Riemannian point of view.

Sievers (1967, 551) also pointed out that the clear Riemannian 8-bars periodic structure could be found in many of Reger's works, for example, in the *Organ Sonata* №1 op. 33 (in the

[9] Original text: "Die < > beziehen sich auf den Gebrauch des Jalousieschwellers; doch kann man auch im Tempo bei < etwas *string*. u. bei > etwas *ritard*. (Tempo rubato)" (Reger 1898).

third part), *Organ Suite* op. 16 (fourth part), *Organ Trio* №4 op. 47, *Choral Preludes* op. 135a, as well as in some "*Marienlieder*" op. 61.

After 1900 the professional relationships between Reger and Riemann went cold. Reger tried to distance himself from his teacher and denounced Riemann's theory to be too conservative, tight, dogmatic, and useless for the good musician (Sievers 1967, 589; König 2020, 13). Some of the Riemann's phrasing signs (such as agogic accent or crossed slurs) Reger had never used in his later works (Sievers 1967, 540); other signs (such as slurring in general) were modified compared to the Riemann's initial meaning (König 2020, 12).

However, several performance researchers showed that some aspects of Riemann's theories, especially those being the common trends of the time, still had an influence on the mature Reger's works. For example, Lohmann wrote: "The traces of Riemann's thinking may easily be detected in Reger's works, even in those written after 1907, although these ideas, of course, are to a great extent a reflection of trends current at the time" (Lohmann 1995, 278). Lohmann, Laukvik and Sander clearly considered the Riemannian phrase shaping to be fully relevant for Reger's organ music and provided examples from opp. 40 and 135b (Laukvik 2006, 270–271), opp. 52/2, 59/9, and 60 (Lohmann 1995, 281–282), op. 73 (Sander 2006, 93). Sander (2006) outlined the influence of Riemann's teaching to the Reger's phrasing, also affirming the agogic/dynamic concept and describing the Riemannian hierarchical organization of the large phrases: "The large phrases are built from the short motives; furthermore, its movement is like waves: the culmination point of each following motive is stronger than the previous" (Sander 2006, 11). The importance of the Riemann's phrasing scholarship in Reger's work also emphasized Balázs Szabó, who especially mentioned in this context the ability of the Riemannian phrasing to visualize the motivic structure of the piece (Szabó 2016, 57).

With regard to his time, Riemann's aim was "to give a kind of system of coordinates; he did not want (but was often misunderstood as wanting!) to dictate rules to composers and performers" (Lohmann 1995, 264). After the presented study of the respective literature, this "system of coordinates" was chosen as the methodological foundation for the mathematical modelling of expressive timing in German late Romantic organ music.

III. Temporal elasticity model and its analytical evaluation

The Riemannian hierarchical phrasing scheme (see Figure 1) was the starting point for my mathematical model. I simulated the tempo arch at each level as the positive semi-ellipse, where the long axe was defined by Riemannian motivic length, and the short axe was proportional to the metronomic tempo. The equation for Riemann's phrasing arcs on each level took then the following form:

$$y_{ij} = \left(\sqrt{1 - \left(x - \frac{h_{ij}}{a_{ij}}\right)^2}\right) * b_{ij} + T, \tag{1}$$

where i denotes the number of the level; j, the sequence number of the semi-ellipse on the i^{th} level; a_{ij}, the long axe of the semi-ellipse, corresponding to the Riemannian motivic length; h_{ij}, x-coordinate of the semi-ellipse's center, corresponding to the middle point of each motive; T, starting metronomic tempo value (constant); b_{ij}, the short axe of the semi-ellipse, proportional to T:

$$b_{ij} = e_{ij} * T \tag{2}$$

The parameter e is defined as ***temporal elasticity****:* it shows the maximum of the model tempo deviation against the metronomic tempo for each level. If $e = e_0$ is the temporal elasticity for the global arch over the whole piece ($a_{ij} = a_0 = 0.5 * (length\ of\ the\ piece)$), then, in the symmetric case, temporal elasticities on subsequent levels are related to the e_0 through the same weight coefficient 1.5:

$$e_{ij} = e_i = 1.5 * e_0/N_i \, , \tag{3}$$

where N_i is an overall quantity of semi-ellipses on the i^{th} level. However, in real performance practice, absolute symmetry is rarely kept, and some irregularities may be possible. In the more general case, temporal elasticities take values:

$$e_{ij} = k_{ij} * e_0/N_{i,} \tag{4}$$

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

According to my research hypothesis, temporal elasticity e_0, a single raw-number model parameter, can carry the expressive information in organ performance in the German late Romantic style. This hypothesis was rooted in several psychological studies showing that the listeners were able to perceive some musical expression resulting from the performer's structural interpretation of the piece (Palmer 1996; Gabrielsson 2003). Because my mathematical model for the Riemannian phrasing scheme was based on the structural hierarchy, I suggested that the difference in values of the model's parameter, temporal elasticity, might result in a difference in the expressive impact. If the hypothesis is true, the goal of this research becomes determining such parameter values for e_0 that, on the one hand, would preserve the Riemannian idea of the built-in motivic symmetry, and on the other hand, would approximate the real performance data and therefore might be used in computer simulation of expressive timing.

III.1 Model for the duple meter

The model was evaluated analytically on Max Reger's *Choral Prelude* op. 135a/1. It is a textbook example of the Riemannian scheme with the time signature of 4/4 (Figure 1b): eight bars long, clear cadences in bars 2, 4, 6, 8 (see Appendix 2). The performance data was collected as the MIDI recording of the professional organist interpretation at the Casavant organ at The Church of Saint Andrew and Saint Paul in Montreal (Canada). Temporal information was extracted through the manual beat-mapping process in *Logic Pro X* and exported to *Matlab* for further processing. The local tempo at the 16^{th}-note level was calculated with *Matlab* as:

$$T_{(n)} = 60 * \frac{b_{onset(n+1)} - b_{onset(n)}}{t_{onset(n+1)} - t_{onset(n)}}, \quad (5)$$

where b_{onset} and t_{onset} are the onset time of note n in the score (in beats) and in the recording (in seconds), respectively. If there was no event at the 16^{th}-note level, the local tempo value was linearly interpolated. The model tempo curve Y involving the global arch and 4 subsequent levels was created in *Matlab* as:

$$Y = Y_0 + \sum_{i=1}^{4} \sum_{j=1}^{N_i} \left(\sqrt{1 - \left(x - \frac{h_{ij}}{a_{ij}} \right)^2} \right) * e_{ij} * T \quad (6)$$

with the following global set-up: starting tempo of the performance data $T=21$ *bpm;* total number of sixteenth notes in the piece $S_{16}=128$; total number of note onsets $n_{chuncks}=S_{16}+1=129$; Y_0, global arch obtained from the equation (1) with center $h_0 = S_{16}/2 = 64$ and long axe $a_0 = S_{16}/2 = 64$. For the subsequent levels, the quantities of ellipses N_i on the i^{th} level are N_1=2, N_2=4, N_3=8, N_4=16. The center of the j^{th} semi-ellipse on the i^{th} level is defined as $h_{ij} = m_k * h_0 / N_i$, where $m_k = (2*k+1)$, $k \in \mathbb{Z}$, $0 \leq k \leq N_i - 1$, and $a_{ij} = h_0 / N_i$ denotes the long axe of the j^{th} semi-ellipse on the i^{th} level.

III.1.1 Symmetric model and optimization procedure

An example symmetrical model curve (normalized to the mean tempo value of human performance data) with temporal elasticity values $e_0 = 0.5$ and $e_{ij} = 1.5 * e_0/N_i$ is shown in Figure 4.

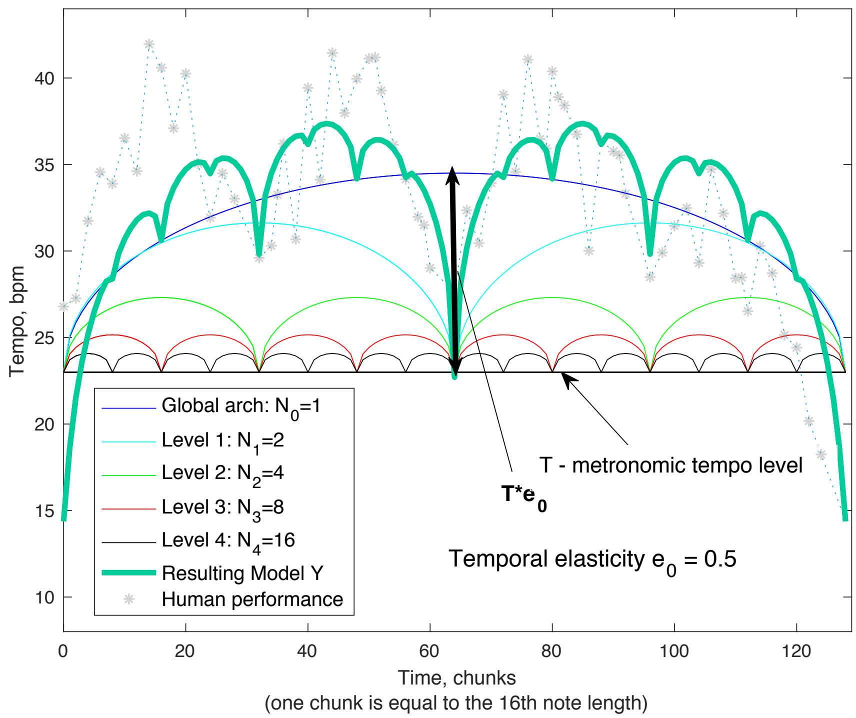


Figure 4. Mathematical model of Riemann's motivic scheme (Choral Prelude op. 135a/1).

Level 1 of this model corresponds here 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.

The straightforward regression of this symmetric model (3) against the human performance data gave $R^2 = 0.46, p < 0.0001$. It can be compared it with the results in Windsor and Clarke (1997), where the highest R^2 obtained for timing from the similar symmetric model was $R^2 = 0.299, p < 0.0001$ (see Appendix 1). However, this comparison is only notional, because the model in Windsor and Clarke (1997), despite the similar structure, has a different methodological background, as well as was applied to the early Romantic piano piece.

Then the *Matlab* built-in Nelder-Mead simplex algorithm (Lagarias et al. 1998) was used to evaluate the generic model with the varying values of e_{ij}. Temporal elasticities at the levels 1–4 were represented from (4) as $e_{ij} = k_{ij} * e_0/N_i$, and the coefficients k_{ij} together with the global value e_0 were set as parameters to optimize for the *Matlab* `fminsearch` function so to minimize the distance between the model curve Y and the performance data. The values $e_0 = 0.5$ and symmetrical coefficients $k_{ij} = 1.5$, were used as initial guess for the first simplex. The obtained curve (Figure 5) provided a highly significant coefficient of determination $R^2 = 0.83, p < 0.0001$ which might be comparable with the results in Todd (1992). The correlation between the optimized curve and human performance $R = 0.91$ is even higher, than the correlation between two professional organists' interpretation of this piece ($R = 0.89$, see Chapter V).

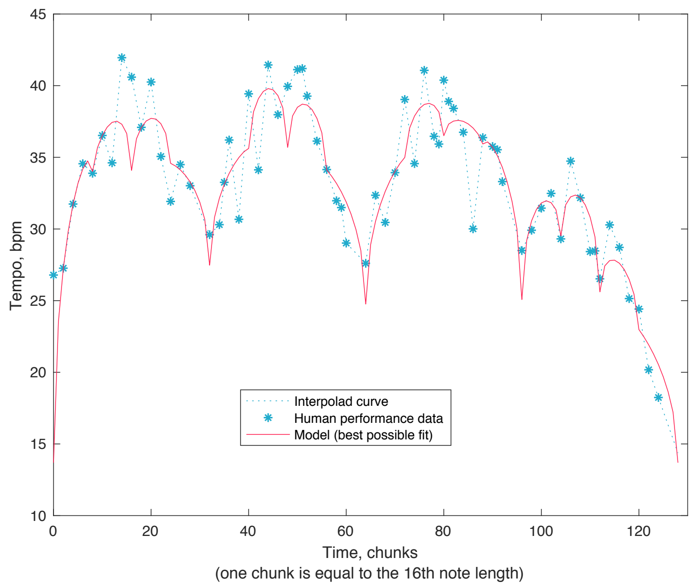


Figure 5. Model curve with optimized parameters obtained by Nelder-Mead simplex algorithm (duple-meter model).

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. I 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.

The optimized temporal elasticity was close to the initial guess: $e_0 = 0.53$, and the coefficients k_{ij} were distributed in the range from $3.6*10^{-8}$ to 4.97 with the mean over all levels $mean(k_{ij}) = 1.64$, which correlates well with the initial value for symmetric model $mean(k_{ij}) = k_{ij} = 1.50$ (see Figure 6).

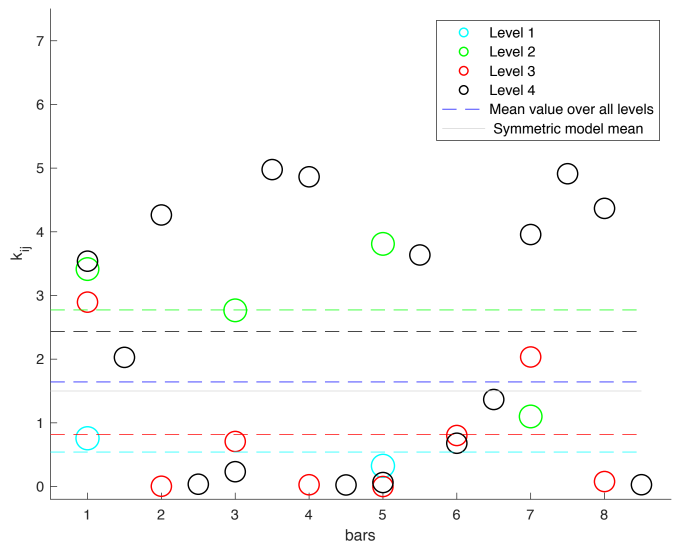


Figure 6. Optimized coefficients k_{ij} at the levels 1–4. Colored dashed lines represent the mean values for each level.

This obtained result contains three important factors to consider:

1) The levels 2 and 4 are more elastic than the levels 1 and 3. 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 marked by the respective phrasing slur endings (see Appendix 2).

2) The first coefficient at the level 2 $k_{21} = 3.34$, (see Figure 6, bar 1) is higher than the level mean value $mean(k_{ij}) = 2.7$. It is possible to notice that the first 'hump' of the optimized curve is higher than the 'hump' for the symmetric model (see Figures 4 and 5). It might be a result

of the initial performer's attempt to grab the audience's attention right at the beginning of the piece by substantial tempo increase.

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 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.

III.1.2 Improved model

Three aforementioned factors from the generic model analysis were used to create the appropriate modifications and improve the original dogmatic symmetric model.

Modification 1: The levels 2 and 4 were set to have more weight than the levels 1 and 3. The respective values of k_{ij} are shown in the Table 1. The mean value for the improved model was deliberately kept the same as for the initial symmetric model: $mean(k_{ij})_{symmetric} = mean(k_{ij})_{improved} = 1.50$.

Table 1. Coefficients k_{ij} for the improved symmetric model (duple meter).

Levels	*k_{ij}, average value (best fit model curve)*	*k_{ij}, improved model*
1	*0.54*	*0.5*
2	*2.7*	*2.5*
3	*0.71*	*0.5*
4	*2.43*	*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 performer's expression at the beginning of the piece. Figure 7 shows the improved model curve with two modifications.

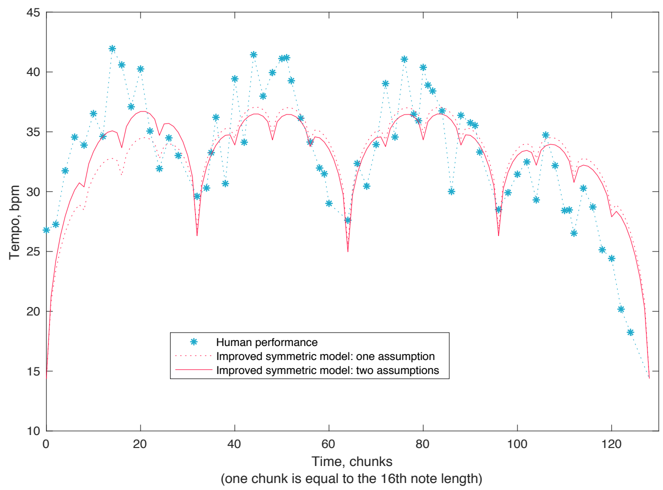


Figure 7. Improved model with one and two modifications (duple meter).

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 8.

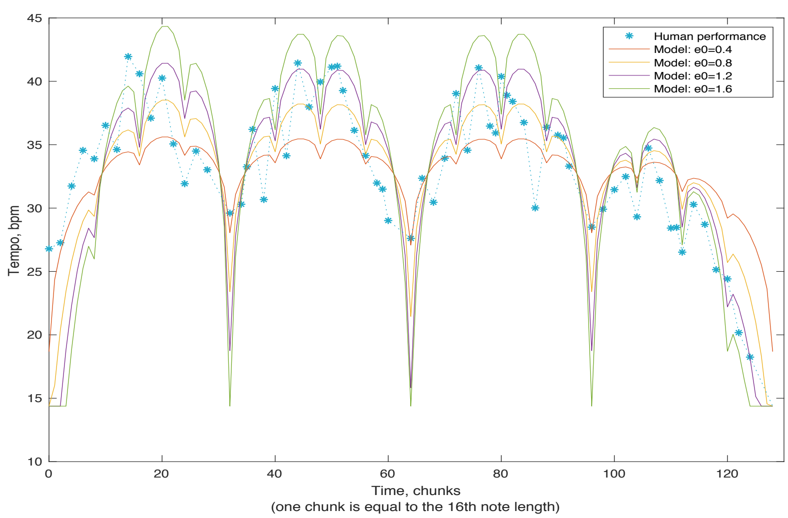


Figure 8. Improved model with three modifications (duple meter).

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} (4), and therefore ensure the Riemannian symmetry. The improved model was evaluated mathematically with the same algorithm as described in previous Section III.1.1, but this time only the value of e_0 was optimized. The summary of regression analysis depending on the modifications made is presented in the Table 2:

Table 2. Summary of regression analysis for improved symmetric model (duple meter).
All coefficients R^2 are significant with $p<0.0001$.

Number of modifications	*Optimal value of e_0*	*R^2*
No modifications ('pure' symmetric)	*0.53*	*0.46*
One (only unequal levels)	*0.56*	*0.5*
Two (unequal levels and k_{21} increase)	*0.56*	*0.61*
Three (unequal levels, k_{21} increase and boundary conditions)	*0.62*	*0.66*

It is revealed, that the improved model has a better performance, than the 'pure' symmetrical model. Specifically, introducing the levels' inequality and elasticity increase at the second level made a significant difference. It is a meaningful finding for the performance practice illustrating how essential are the first bars of the piece ("*well begun is half done*"). The boundary conditions allow the model to increase the R^2 as well; however, it is interesting to see, that when the elastic extension of the model curve becomes too high, the R^2 slowly decreases (see Figure 9). It shows, that after the certain values of e_0 the timing becomes over-expressive (described in performer's jargon as "*too much*") and might sound unconvincing for the listener.

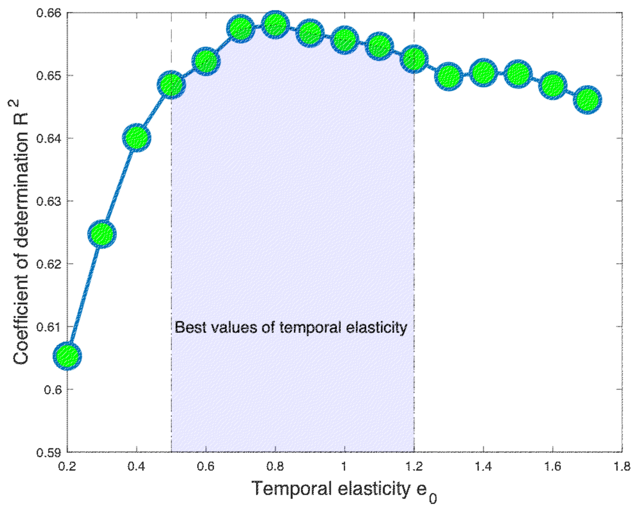


Figure 9. Coefficient of determination for improved duple-meter model depending on e_0.

This is an important avenue towards the computer simulation of expressive timing: the values of temporal elasticity within the interval with the highest values of R^2 might be musically the most convincing expressive strategies for the performer. The less elastic phrasing might be considered as mechanical, or non-expressive, while the hyper-elastic phrasing is tasteless or grotesque.

III.2 Model for the triple meter

A similar model was built for the triple-meter (Figure 1a) and analytically evaluated on Max Reger's op. 73 (bars 133–138, starting from the upbeat; see in Appendix 2). As can be seen from Figure 1, the triple-meter pieces have the same built-in symmetry as the duple-meter ones, where the number of semi-ellipses doubles while moving down from level 1. The main difference is in the number of semi-ellipses on the lowest, "motivic" level: whereas in the duple-meter case, it consists of two semi-ellipses and two-notes motives, in the triple-meter case, it has three semi-ellipses and three-notes motives.

In order to get the simulated symmetric phrasing for the selected excerpt, three "blocks" from Figure 1b were stuck together under the main arch with e_0. The model curve was obtained from Equation (6) with the respective numbers of ellipses $N_1=3$, $N_1=6$, $N_1=12$ and $N_1=24$, as well as with the following global parameters: $n_{chuncks}=216$, $h_0=108$, and $a_0 = 108$. Human performance data was collected in the same way as for the duple-meter model. The regression of the symmetric model with fixed elasticity values against the human performance data gave an $R^2 = 0.42, p < 0.0001$. Then the same optimization procedure was performed as described in previous Section III.1.1. Figure 10 shows the model curve after optimization.

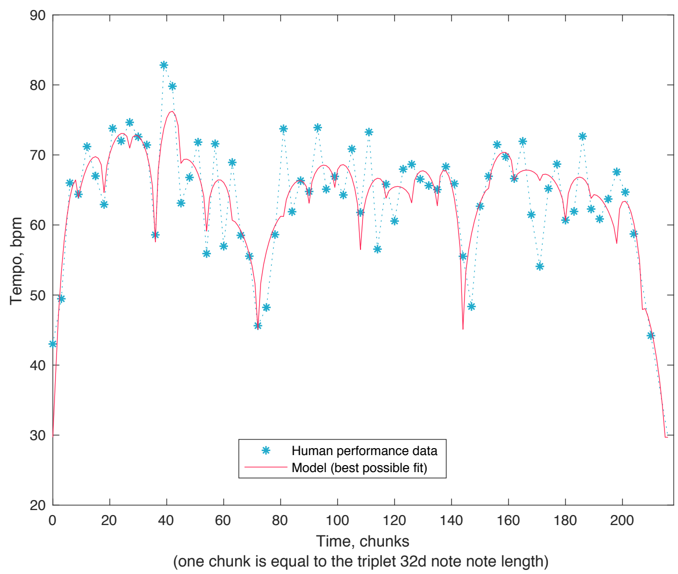


Figure 10. Model curve with optimized parameters obtained by Nelder-Mead simplex algorithm.

The optimized curve gave an $R^2 = 0.79, p < 0.0001$, demonstrating a relatively high fitting potential of the model. The value of optimal temporal elasticity was $e_0 = 1.05$, and the coefficients k were distributed in the range from $2.4*10^{-4}$ to 7.26 with the mean over all levels $mean(k_{ij}) = 1.68$, which, again, is not far from the initial value for symmetric model $mean(k_{ij}) = k_{ij} = 1.5$ (see Figure 11). The factors proposed in the previous Section for improved duple-meter model were valid for the triple model as well, but, in this case, with the different levels' strength. The "motivic" level 4 was the strongest, as for the duple-meter model; level 1, corresponding to the quasi-cadences in bars 134 and 136, and level 3, corresponding to the four short subsections in bars 134 and 135, marked by four slurs, were also relatively strong. It seems that the number of slurs plays an important role in the definition of the level's strength; this hypothesis will be further discussed in Chapter VI.

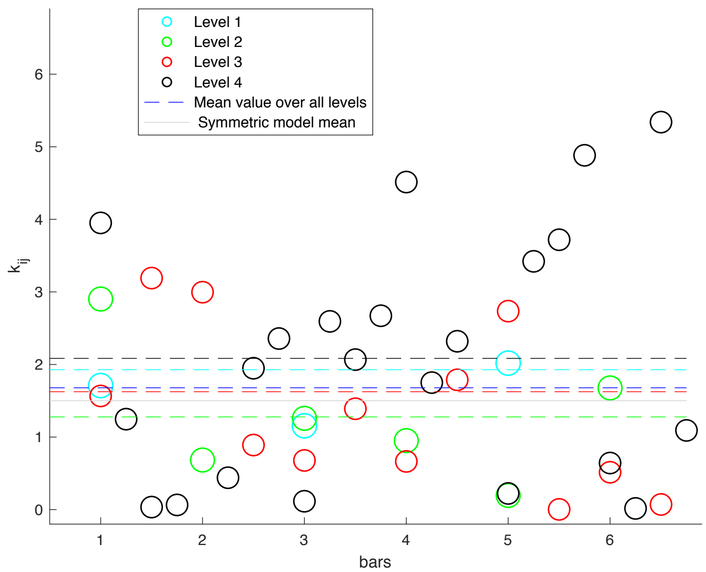


Figure 11. Optimized coefficients k_{ij} at the levels 1–4 (triple-meter model). Dashed lines represent the mean values for each level.

The improved model was created analogously to the duple-meter procedure, with unequal levels based on the values of k_{ij} presented in Table 3, as well as with increasing of the first coefficient on the second level and introducing of the boundary conditions.

Table 3. Coefficients k_{ij} for the improved symmetric model (triple meter).

Levels	*k_{ij}, average value (best fit model curve)*	*k_{ij}, improved model*
1	*1.72*	*1.5*
2	*1.23*	*1*
3	*1.62*	*1.5*
4	*2.08*	*2*

The introduction of the respective modifications led to the increasing of the coefficient of determination for the improved model up to $R^2 = 0.56, p < 0.0001$. Figure 12 shows the improved model with different values of temporal elasticity for the triple meter.

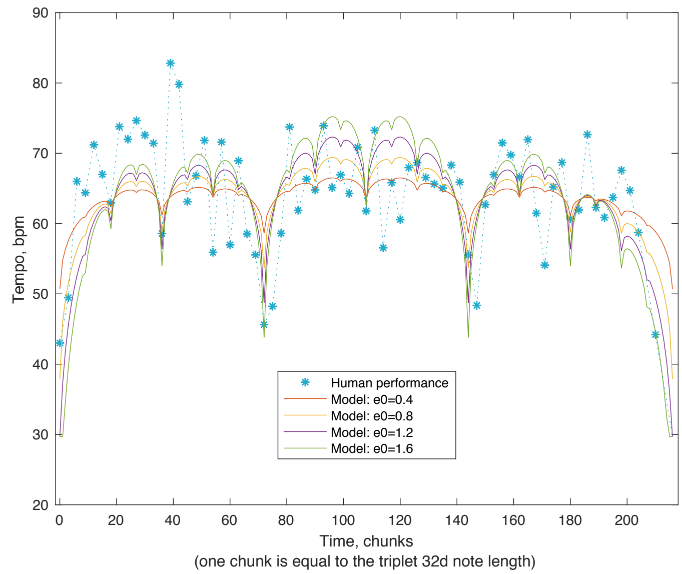


Figure 12. Improved model (triple meter).

As in the case of the duple-meter model, the coefficient of determination was slightly decreasing in the areas of low or high elasticities (see Figure 13). However, for the triple-meter model, this effect was not really prominent.

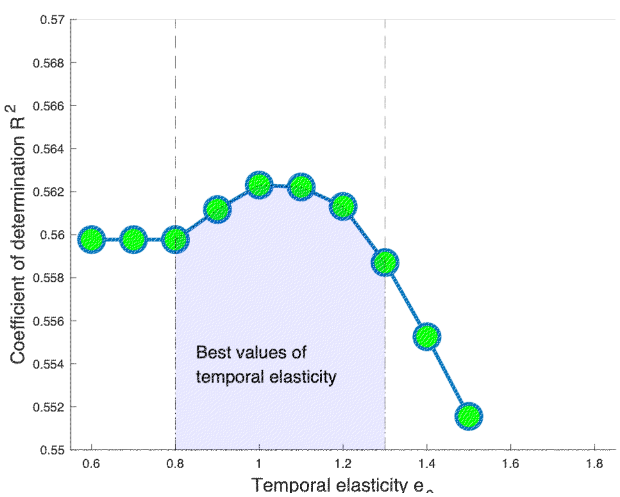


Figure 13. Coefficient of determination for improved triple-meter model depending on e_0.

In this Chapter, the Riemannian phrasing model was built and analyzed analytically. It was shown that the modifications based on the simple music score analysis allowed to improve the model's fitting capacity. Furthermore, the coefficient of determination of the improved model was proven to depend on the temporal elasticity, with the maximum values falling approximately into the elasticity interval [0.8, 1.2] (see Figure 9 and Figure 13). This interval is suggested to be the "safest" strategy for a performer, where the phrasing is expressive enough but is not "*too much*" or grotesque. The Chapter that follows moves on to consider the empirical evaluation of the proposed model through the listening tests.

IV. Evaluation of the temporal elasticity model through the listening tests

The listening experiment (REB file #20-06-021) was designed to empirically evaluate the proposed model by investigating the possible correlation between temporal elasticity values and listeners' perception of performance expression. The conducted experiment had two parts: the first part was introductory and intended to evaluate the overall model's naturalness; the second (main) part was entirely focused on evaluating the perceived expression.

IV.1 Methodology

Twenty-nine people identifying themselves as "organ students, working professionals, organ amateurs or enthusiasts" (referred further as "*Organists*"), and twenty-four persons with "little or no experience with organ or its repertoire" (referred further as "*Non-Organists*") voluntarily participated in the anonymous online survey. All participants gave their informed consent prior to the beginning of the experiment. The stimuli were different versions of Max Reger's pieces described in the previous Chapter III: *Choral Prelude* op. 135a/1 (duple-meter model) and the short excerpt from the op. 73 (bars 133–138, triple-meter model). One of the versions for each piece was a professional organ recording, and the other versions were MIDI files, which were modified to have the tempo curves simulated according to the mathematical model.

In order to achieve the experiment's purity, the following methodology was elaborated. First, the professional organ performance of the pieces was recorded in both MIDI and audio formats in The Church of Saint Andrew and Saint Paul in Montreal at the *Casavant* organ, equipped with the *Solid State* MIDI control system. The MIDI recordings were made using the *OrganAssist* software sequencer (OrganAssist 2021), which enabled to properly save all information about organ registrations and the swell boxes positions in *SysEx* messages. Then, the recorded MIDI files were imported into *LogiPro X* and accurately beat-mapped so to match the original score (this process was identical to the workflow described in Chapter III and used to extract the temporal information from the files). Next, the tempo curve was "flattened," and the file tempo was set to the constant (metronomic) value. At the last step, the tempo values obtained from the Matlab model (6) were programmatically applied to these "flat" MIDI files using the high-level C#/.NET framework *DryWetMidi* (2020). In each part of the experiment, the MIDI files were modelled with different values of temporal elasticity. All MIDI files were then played back at the same instrument with the same registration as the original human performance and professionally recorded from the listener's position so to capture the acoustics of the church. This approach allowed to eliminate any possible performance discrepancies in the files except the variations in timing.

The experiment's web interface was designed with the *Web Audio Evaluation Tool* (Jillings et al. 2015) and included the task description, volume control, audio files and play buttons. Before the experiment start, listeners were given the opportunity to check and adjust the sound level.

Part 1: In the first part, the stimuli were the human recordings and two models, symmetric and improved. The models were created with the closest to the human recording temporal elasticity values as discussed in the previous Chapter III ($e_0 = 0.6$ and $e_0 = 1.0$ for op. 135a/1 and op. 73, respectively). The exact task for the listeners was: "*Please, select the preferred interpretation by clicking on its letter (A, B or C).*" There was a designated page for each of two pieces, and the stimuli on the page were shown in the randomized order. The listeners were required to listen to all the music excerpts in their entirety and then select only one variant they liked the most.

Part 2: In the second part, the stimuli were the human recordings (the same as in Part 1), as well as the improved symmetric models with temporal elasticity values $e_0 = 0.4$, $e_0 = 0.8$, $e_0 = 1.2$ and $e_0 = 1.6$ (see Figures 8 and 12). The exact 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).*" The sliders were provided in the web interface for the response of each stimulus.

It is important to notice that this question was intentionally chosen in accordance with the results described by Bhatara et al. (2011), where a similar methodology was used for another experiment. In particular, Bhatara (2008) showed that the additional remark for the participants that it did not matter which particular emotion they felt while listening was proven to be important for the experimental results. The following ticks' descriptions were attached to the sliders so to help the participants: "*Not expressive, mechanical*" (at position 20), "*Moderately expressive*" (at position 50) and "*Extremely emotional*" (at position 80). The introducing of the word "*emotional*" at the last tick was also done intentionally: Bhatara (2008) showed that the replacement of the word “expressive” to the equivalent word “emotional” in the experiment question facilitated the decision-making process for the participants and led to the improvement in the experimental results. The advantage of these essential methodological aspects was taken into account in my experiment.

Just like in the first part, the listeners were required to listen to the excerpts in their entirety: this time, on each of the two pages, there were five different versions shown in random order. The listeners were obligated to use the full range of the slider’s scale and were allowed to listen to each version as many times as they needed so to make their final decision. After the experiment’s closure, all participants’ ratings were first divided by 100 and then imported to the *IBM SPSS* platform for detailed statistical analysis.

IV.2 Experimental results

Part 1. In the first part, the human performance was the most selected (preferred) version for op. 135a/1 and op.73 (referred further as “135H” and “73H”, respectively), followed by the improved symmetric model (“135M” and “73M”); the least preferred became the symmetric model (“135SM” and “73SM”). Out of all 53 participants, 25, 18, and 10 people selected the 135H, 135M, and 135SM versions, respectively; 23, 21, and 9 participants have chosen the respective versions 73H, 73M, and 73SM. The *Chi-square* goodness-of-fit statistical test revealed that the test answers distributions (25, 18, 10) and (23, 21, 9) were significantly different from the chance level[10]: $\chi^2\ (2,\ N = 53) = 6.38, p < 0.05$ for op.135a/1 and $\chi^2\ (2,\ N = 53) = 6.49, p < 0.05$ for op. 73.

[10] The chance-level null hypothesis assumes the values for each fragment are equal: 53/3, 53/3, 53/3.

The confidence intervals comparison method was chosen to analyze the individual differences between the versions of each piece. Each response in favor of the selected version was assigned the value 1, then the mean, the mean standard error, and the confidence intervals were calculated in *SPSS*. Table 4 shows the results' outcome. It is possible to notice that some confidence intervals on the 95% level fall into the category of *barely non-overlapping intervals* (Goldstein and Healy 1995). In this case, Afshartous and Preston (2010) proposed using an 83% confidence interval of the mean for each group, which reduced the type II error rate. Consequently, this approach was applied to the analysis of Part 1 experimental results.

The mean values and their respective confidence intervals are presented in Figure 14a (op. 135a/1) and Figure 14b (op.73). From the figure above, we can see that the human performance did not significantly differ from the improved model, which indirectly indicated the high naturalness of this model. However, the symmetric model got a significantly lower listeners' preference: its confidence interval did not overlap with the human performance and in the case of op. 73, neither with the human performance nor with the improved model.

Table 4. Experimental results for Part 1 (all participants).

Version	*N*	*Mean*	*SEM*	*CI-95%, Low*	*CI-95%, High*	*CI-83%, Low*	*CI-83%, High*
135H	*53*	*0.47*	*0.069*	*0.33*	*0.61*	*0.38*	*0.57*
135M	*53*	*0.34*	*0.066*	*0.21*	*0.47*	*0.25*	*0.43*
135SM	*53*	*0.19*	*0.054*	*0.08*	*0.30*	*0.11*	*0.26*
73H	*53*	*0.43*	*0.069*	*0.30*	*0.57*	*0.34*	*0.53*
73M	*53*	*0.40*	*0.068*	*0.26*	*0.53*	*0.30*	*0.49*
73SM	*53*	*0.17*	*0.052*	*0.07*	*0.27*	*0.1*	*0.24*

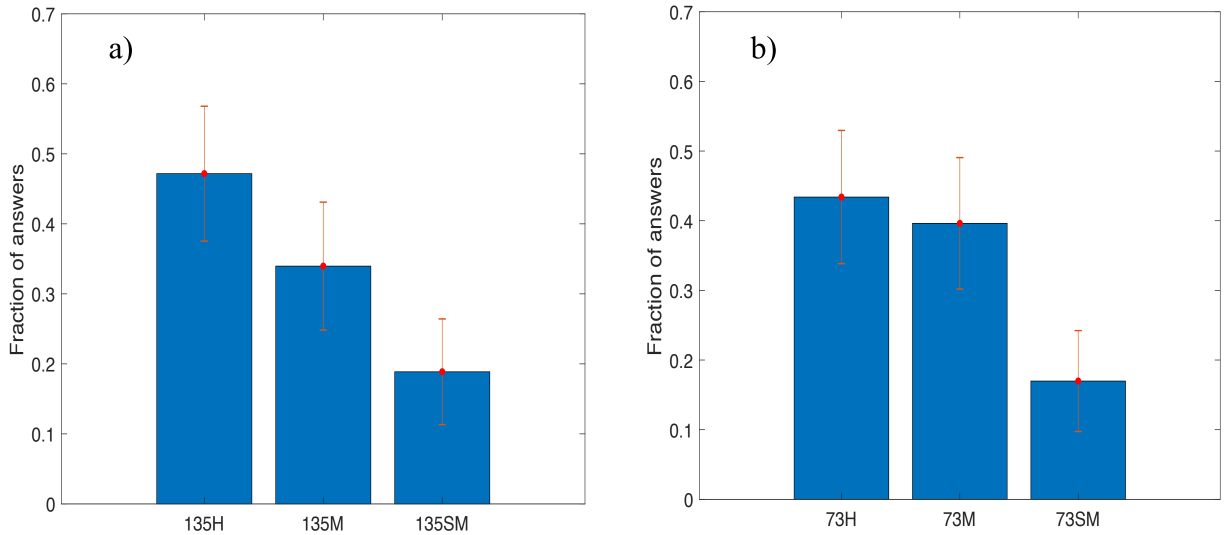


Figure 14. Results of the listening experiment, part 1 (all participants). a) - op. 135a/1, b) - op. 73.

This could serve as as empirical confirmation for the appropriacy of the factors theoretically predicted to improve the original symmetric model (see Chapter III).

Further analysis was performed to reveal the possible difference between the *Organists'* and *Non-Organists'* perceptions of the model (see full result tables in Appendix 3). Figures 15 and 16 provide an overview of the experimental results for both groups.

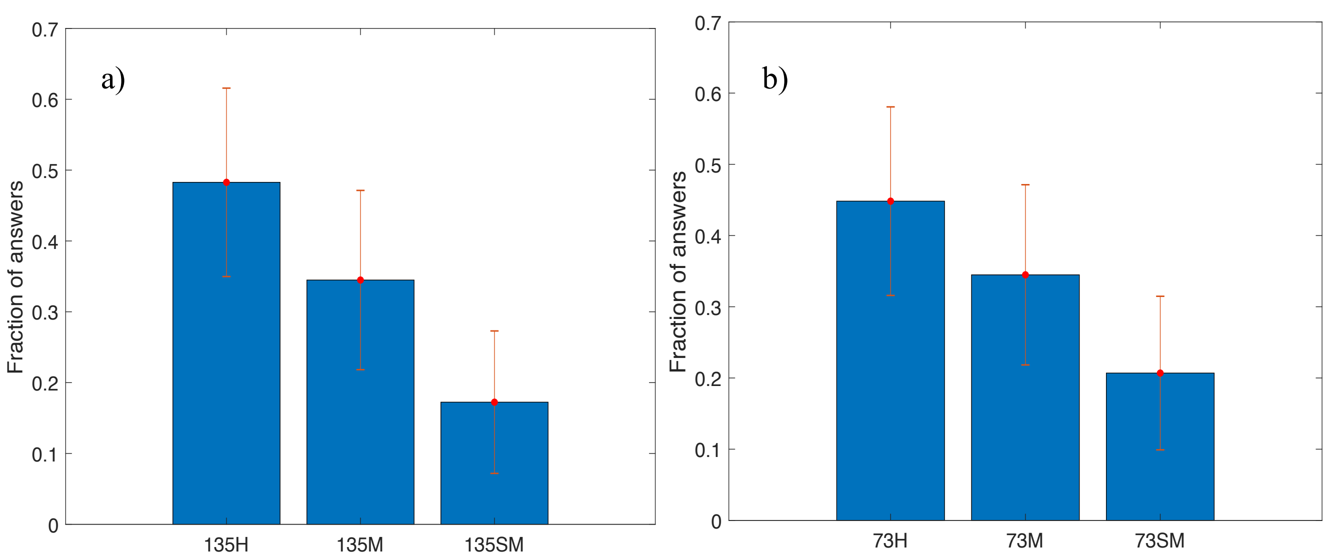


Figure 15. Results of the listening experiment, part 1 (Organists group). a) - op. 135a/1, b) - op. 73.

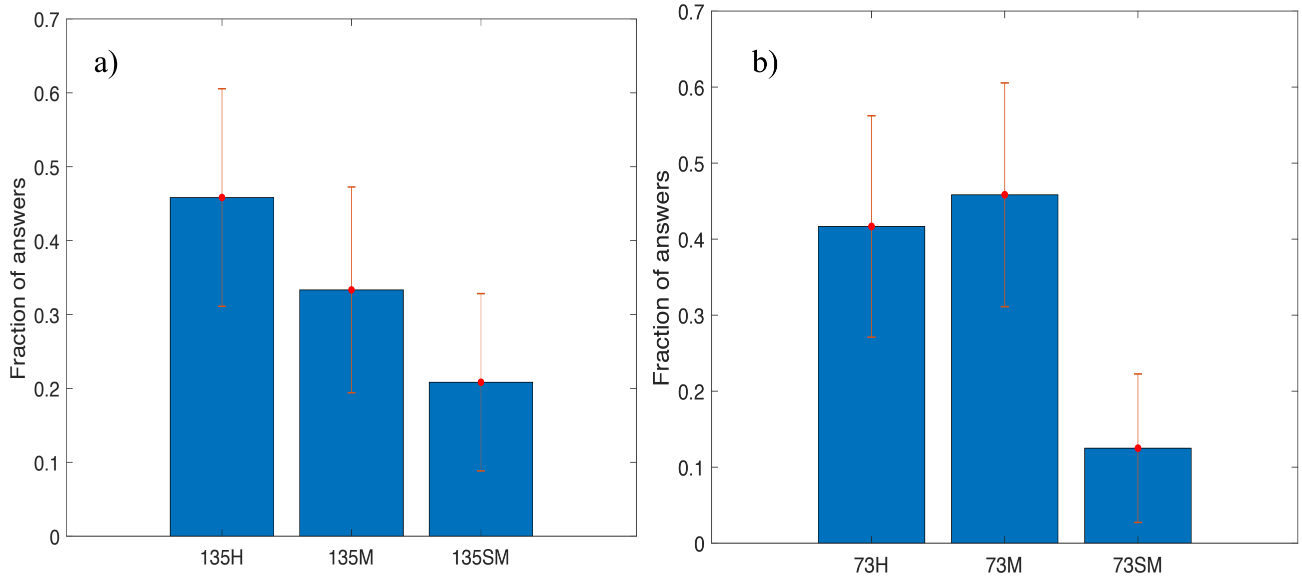


Figure 16. Results of the listening experiment, part 1 (Non-Organists group). a) - op. 135a/1, b) - op. 73

In general, both groups showed a similar trend in their preferences, favoring human performance and discarding the symmetric model. For the *Non-Organists* group, the mean rating for the improved model was even higher than the mean rating for human performance (0.46 and 0.42, respectively). However, this difference was not statistically significant because the confidence intervals vastly overlap.

Part 2. The following abbreviations were used for the stimuli in this part: human performance of op. 135a/1 and op. 73 was referred to as “135H”, and “73H” (same as in Part 1), the improved models with the temporal elasticities $e_0 = 0.4$, $e_0 = 0.8$, $e_0 = 1.2$ and $e_0 = 1.6$ were referred to as “135M0.4”, “135M0.8”, “135M1.2”, “135M1.6” for op 135a/1 and “73M0.4”, “73M0.8”, “73M1.2”, “73M1.6” for op. 73.

The grand mean of listeners’ ratings was $M = 0.54$ $(SE = 0.05)$ for all stimuli of op. 135a/1 and $M = 0.6$ $(SE = 0.05)$ for op.73, which demonstrated that responses were relatively well centered around the scale. Individual means (participants wise) ranged from 0.37 to 0.68 (SD=0.1) for op 135a/1 and from 0.47 to 0.76 (SD=0.1) for op. 73.

A repeated measures one-way ANOVA determined that the effect of temporal elasticity level in the four simulated performances on the listeners' expressivity ratings was significant for both op. 135a/1 ($F(3,156) = 23.96, p < 0.05$) and op. 73 ($F(2.5,132.6) = 25.9, p < 0.05$).[11] The effect size was large for both pieces: $\eta_p^2 = 0.31$ and $\eta_p^2 = 0.33$ (for op.135a/1 and op.73, respectively). In general, 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 least value $e_0 = 0.4$ (for op. 135a/1) or second-least value $e_0 = 0.8$ (op. 73) had the lowest rating.

The linear regression of the mean ratings for the model elasticity values was significant ($R^2 = 0.97, p = 0.013$) for op. 135a/1. For op. 73, only mean ratings in the elasticity range from $e_0 = 0.8$ to $e_0 = 1.6$ showed a significant linear trend ($R^2 = 0.99, p = 0.03$). Figures 17 and 18 show the mean listeners' ratings for the simulated performances, human performances and the significant linear trends.

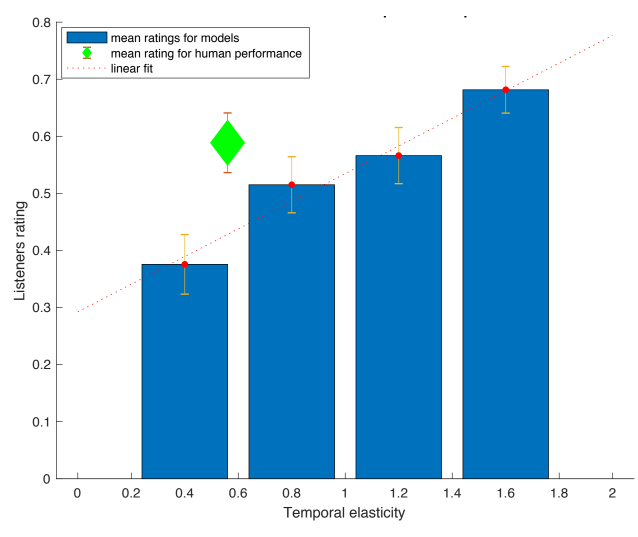


Figure 17. Average listeners' ratings for models with different values of temporal elasticity and human performance (op.135a/1).

[11] In this case, Mauchly's sphericity test was significant ($\chi 2$ (5) = 14.7, p = 0.012); that is why the Greenhouse-Geisser correction was used.

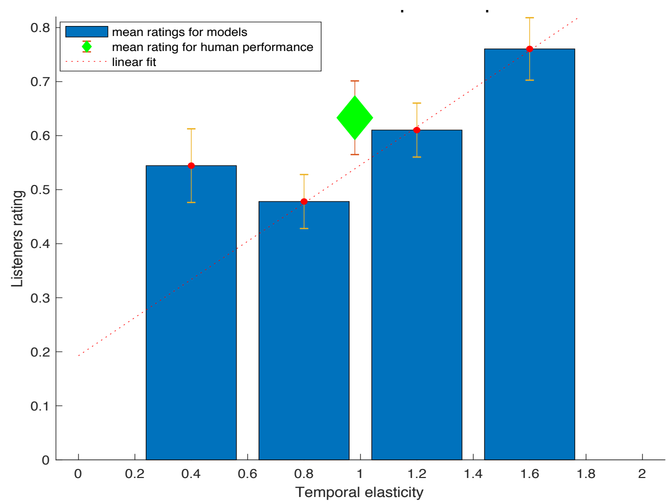


Figure 18. Average listeners' ratings for models with different values of temporal elasticity and human performance (op. 73).

In order to explore the difference between the ratings for individual models, as well as the rating for the human performance, the pairwise comparison with the Turkey HSD ("*Honestly Significant Difference*") post-hoc test was performed. Full test result tables are attached in Appendix 4. For the models of op. 135a/1, it was revealed that the difference between mean ratings for 135M0.4 and mean ratings for all other stimuli was significant ($p<0.01$), the ratings for 135M0.8 and 135M1.2 were different from the ratings for 135M0.4 and 135M1.6 ($p<0.05$), but did not significantly differ from each other. The listeners' expression perception for human performance (135H) was significantly different only from the 135M0.4; however, it was also noticeably apart from the 135M1.6, at the margin of statistical significance ($p<0.08$). For the models of op. 73, listeners' expressivity rating difference was significant for all pairs among 73M0.8, 73M1.2 and 73M1.6 ($p<0.01$). Surprisingly, the mean rating for the least elastic model 73M0.4 was higher than for the more elastic model 73M0.8 (0.54 and 0.48, respectively), but this difference was not statistically meaningful; among all stimuli, the rating for 73M0.4 was significantly different only from the rating for 73M1.6 ($p<0.001$). The human performance (73H) was rated significantly higher than 73M0.8 ($p<0.01$) and significantly lower than the most expressive version 73M1.6 (p<0.01), but was not significantly different from 73M1.2 and 73M0.4.

The same statistical analysis was then performed on the *Organists* and *Non-Organists* groups individually to investigate the possible differences in listeners' responses in greater detail. Table 2 presents a results overview.

As expected, the effect of the temporal elasticity on the listeners' rating of expressivity was significant in all four groups, and the effect size as measured by η_p^2 was large (see Table 5). The largest effect size was for the *Organists*' ratings of the op. 135a/1 expression level. For op. 73, the effect size was almost the same regardless of the organ experience. The slightest effect was achieved by *Non-Organists* while rating the op.135a/1 expression, which proved difficult.

Table 5. Part 2 results overview for Organists and Non-Organists.

Experience	*Piece*	*Effect of temporal elasticity*	*Effect size*	*Linear trend*[12]
Organists	*op. 135a/1*	$F(3, 84) = 21.57, p < 0.001$	$\eta_p^2 = 0.43$	$R^2 = 0.93, p = 0.03$
Organists	*op. 73*	$F(2.4, 66.3) = 14.3, p < 0.001$	$\eta_p^2 = 0.34$	$R^2 = 0.98, p = 0.05$
Non-Organists	*op. 135a/1*	$F(2.4, 56.5) = 5.36, p < 0.005$	$\eta_p^2 = 0.19$	$R^2 = 0.99, p = 0.005$
Non-Organists	*op. 73*	$F(2.6, 60.5) = 11.5, p < 0.001$	$\eta_p^2 = 0.33$	$R^2 = 0.99, p = 0.000$

The linear regression was significant in the same elasticity intervals as for all participants' ratings (in the case of *Organists* and op. 73—marginally). Figures 19 and 20 show the mean listeners' ratings and the respective linear trends for each group.

[12] Linear regression was made in the elasticity interval [0.4, 1.6] for op. 135a/1 and [0.8, 1.6] for op.73.

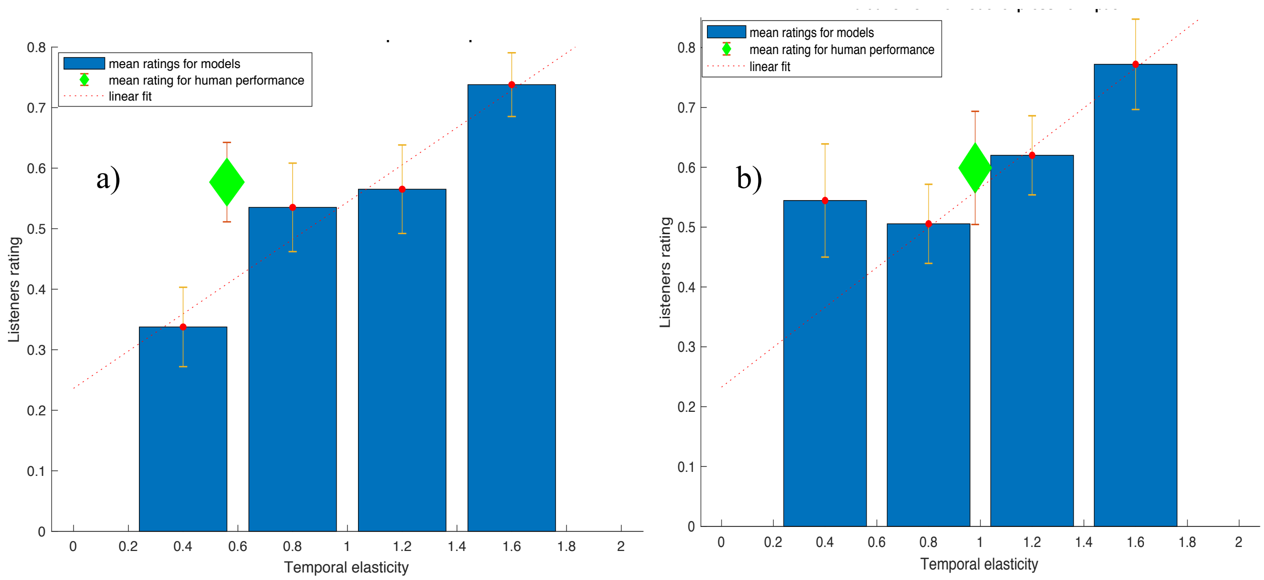


Figure 19. Results of the listening experiment, part 2 (Organists group). a) - op. 135a/1, b) - op. 73.

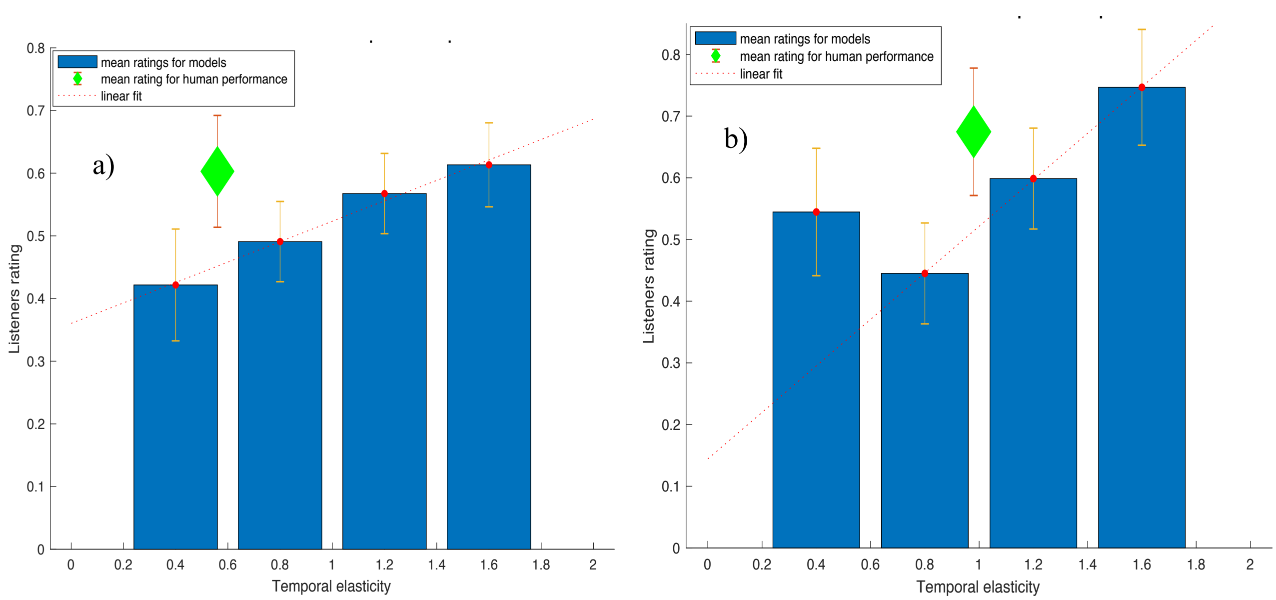


Figure 20. Results of the listening experiment, part 2 (Non-Organists group). a) - op. 135a/1, b) - op. 73.

A pairwise comparison of the listeners' mean ratings to the simulated and human performances was made with the Turkey HSD post-hoc test, in this case, individually for each group. The full results tables are attached in Appendices 5 (for *Organists*) and 6 (for *Non-Organists*). It can be observed that *Organists* were more coherent in their ratings than *Non-Organists,* providing a more distinctive evaluation of the expressive impact. In particular, among the modelled stimuli, the difference in means of *Non-Organists'* ratings was significant at the

0.01 level only for "edge" cases: 135M0.4 and 135M1.6, 73M0.4 and 73M1.6, 73M0.8 and 73M1.6, whereas the *Organists*' mean ratings were significantly different at the 0.05 level for 73M0.4 and 73M1.6, 73M0.8 and 73M1.6, 73M1.2 and 73M1.6, as well as for all model pairs except 135M0.8 and 135M1.2 in op. 135a/1. The human performance was significantly more expressive than the lowest-scored models 135M0.4 (with 135H, p<0.01) and 73M0.8 (with 73H, p<0.01) for *Non-Organists*. For *Organists*, the difference was also significant for 135M0.4 and 135H (p<0.01), however, for op. 73, only 73M1.6 and 73H were significantly distinguishable (p<0.01).

Despite the observed difference, the overall rating trends were very similar for both *Organists* and *Non-Organists*, following the same patterns as the joint ratings for all participants. The performed two-way repeated measures ANOVA with temporal elasticity as the within-subjects factor and the organ experience as the between-subjects factor revealed that the effect of organ experience on the mean listeners' ratings was not statistically significant neither for op. 135a/1 ($F(1,51) = 0.29, p = 0.59$) or for op. 73 ($F(1,51) = 0.73, p = 0.78$).

IV.3 Discussion

An empiric evaluation of the model revealed that the expressivity manifested by the temporal elasticity was clearly comprehensible for the modern listeners, both with or without prior organ experience. The significant linear trends showed that the most elastic models were rated as the most expressive, while the least (op. 135a/1) or second-least (op. 73) elastic models were rated as the least expressive.

Despite the clarity of the linear trends, the experimental results also allow a non-linear interpolation. Figure 21 shows the summary for the listeners' ratings for both pieces from Figures

17 and 18, but, for this time, with the hinge fit curves. Hinge approximation permits the definition of three areas, such that the elasticity values are significantly different between the areas, whereas remaining indistinguishable within them. The area I corresponds to the low, area II—to the medium and area III to the high expressivity rating (see Figure 21). It is interesting to notice that the medium area roughly coincides with the model's best elasticity range defined in the previous Chapter III (see Figures 9 and 13), which empirically supports the proposed suggestion of this range as the "safest" performer's expressive strategy.

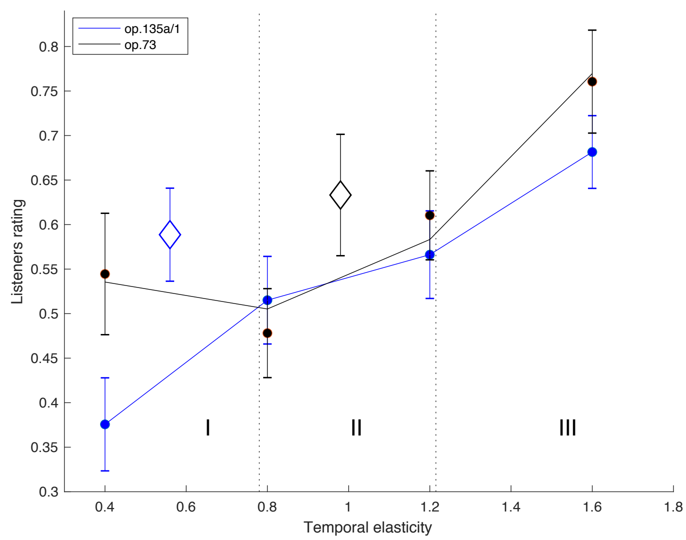


Figure 21. Hinge interpolation for the listeners' ratings for op. 135a/1 (blue) and op. 73 (black). I, II and III show the areas of low, medium and high expressivity.

Another interesting moment is the difference in the listeners' rating in the low elasticity range, which might be due to the local tempo variation at the beginning of the piece. More precisely, the change in elasticity value from $e_0 = 0.4$ to $e_0 = 0.8$ resulted in the noticeable tempo decrease for the opening notes (see Figures 8 and 12 in Chapter III), leading to the statistically significant increase of the listener's expression rating for the initially slow piece (op. 135a/1) and the slight decrease of the rating for the initially fast piece (op. 73), which was not significant, though. This accords well with the experimental results in Vieillard et al. (2012), where it was shown that the effect of expressiveness resulting from the slowing down the tempo was significant for the slow (*sad*) piece, whereas the observed positive correlation of higher tempo and higher expressiveness for the fast (*happy*) piece was not statistically reliable (Vieillard et al. 2012: Table 1). For the other versions, with $e_0 = 0.8, e_0 = 1.2$, and $e_0 = 1.6$, the tempo values for the first few notes were the same because of the boundary conditions, and for these values, the listeners' ratings followed a similar pattern for both pieces. This observation is important for the performance practice: it shows that in order to significantly increase the expressive impact, it is not enough just to play the fast piece faster (even with the technical proficiency); it is necessary to use stylistically appropriate phrasing patterns highlighting the musical structure—then it will be appreciated by the listeners and result in a significantly higher expressive impact.

V. Temporal elasticity and performance analysis

The temporal analysis of various audio recordings was performed in order to compare them with the model. The recordings in focus were made on different instruments (authentic and modern) and, in the case of the duple-meter model, performed by organists with diverse organ experience (organ amateurs and professional organists). An additional analysis featuring the history of performance practice strategies over the last decades was performed in the case of the triple-meter model, where only the professional recordings were taken into consideration (the recordings were generously provided by Max Reger Institute in Karlsruhe, especially for this research project).

V.1 Beat-tracking systems for audio organ recordings

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 appropriacy 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 church. The audio recording of this MIDI file was compared to the original MIDI track, and the original 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 8^{th}-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 16^{th} -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* (8^{th}-notes level);

3) TGBR-beats detected by *Tempogram Toolbox* and manually corrected by *BeatRoot* (8^{th}-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 8^{th}-note level. The summary of the evaluation analysis is given in Table 6.

Table 6. Evaluation of beat tracking systems.

Data set	*F-measure*	*Precision*	*Recall*
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 6 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 the tolerance window (see Figure 22). 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.

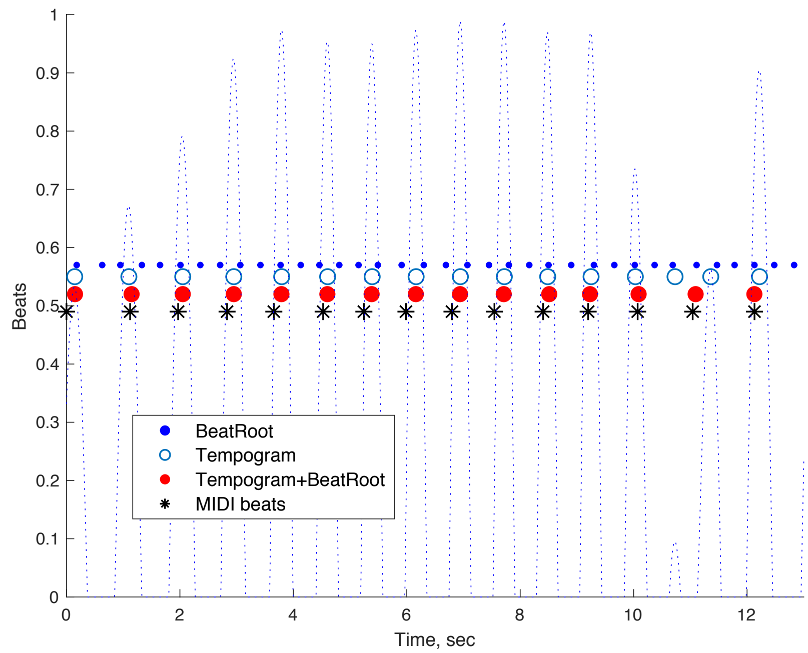


Figure 22. Beats detected by different beat tracking systems for the first 12 seconds of op. 135a/1.

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.

V.2 Comparative performance analysis for the duple- and triple-meter models

The following recordings of Max Reger's op.135a/1 (duple-meter model) 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 7 shows the outlines of the analysis (all coefficients R^2 were significant at the level $p<0.0001$).

Table 7. Regression analysis of 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*

According to the analysis, professional organists (Dupont, Lohmann, and Buttmann) tend to play more expressive than organ amateurs (Pirkl and Pals): the average values of temporal elasticity are $e_0 = 0.71$ (improved) and $e_0 = 0.39$ (symmetric) across the professional organists against the respective values $e_0 = 0.6$ and $e_0 = 0.36$ for 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 Chapter III. For both models, improved and symmetric, the average values for professionals $R^2 = 0.6$ and $R^2 = 0.5$ were slightly higher than the respective values for amateurs: $R^2 = 0.54$ and $R^2 = 0.47$ because the professional organists were more likely to be aware of advanced Riemannian phrasing principles (usually taught at the graduate school).

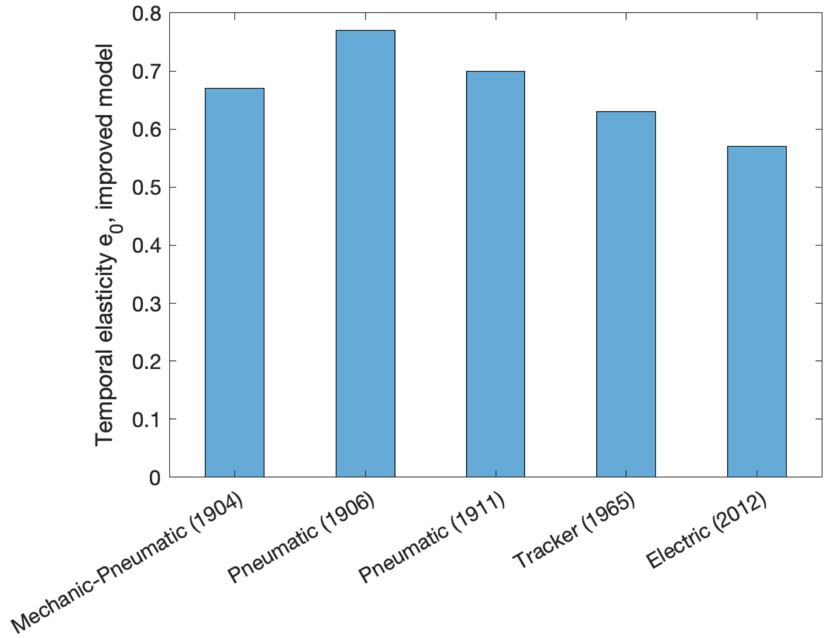


Figure 23. Values of temporal elasticity e_0 depending on the organ action.

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 23). 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 8.

Table 8. 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). 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 24). As shown in Figure 24, 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.

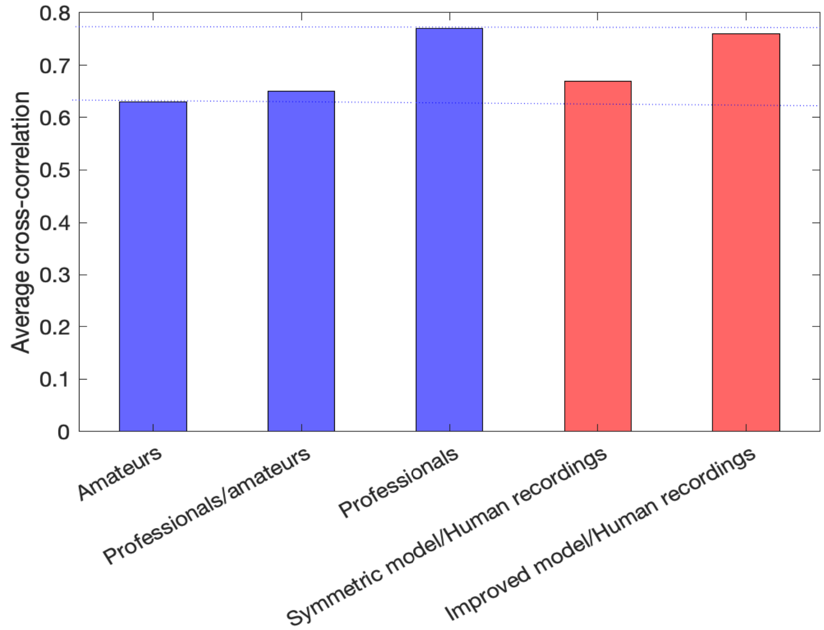


Figure 24. Average cross-correlation between human performances (blue bars) and between models and human performances (red bars).

For the triple-meter model (Max Reger's op. 73, bars 133-138), only the professional recordings were investigated:

1) Bernhard Buttmann / Sauer organ, Berlin Cathedral (Buttmann 2015, CD 2);

2) Bernhard Haas / Rieger organ, Konzerthaus Wien (Haas 1997);

3) Werner Jacob / Willi Peter Orgel, St. Nicolai Hamburg (Jacob 1973);

4) Rosalinde Haas / Albiez organ, Frankfurt (Haas 2011);

5) Martin Sander / Walcker organ, Riga Cathedral (Sander 1994);

6) Phillip Steinhaus / Harrison organ, Church of the Advent, Boston (Steinhaus 1956);

7) Heinz Wunderlich / Kemper-Orgel, Hauptkirche St. Jacobi, Hamburg (Wunderlich 1980).

All recordings were analyzed with *Tempogram Toolbox* and *BeatRoot*, then the data containing beat times at the 16th-notes level was imported to *Matlab* for processing. The local tempo was calculated from Equation (6) and compared to the symmetric and improved triple-meter models described in Chapter III. Table 10 shows the outlines of the analysis.

Table 9. Regression analysis of Max Reger's op. 73 (bars 133–138) audio recordings against the model.

Recording	*Optimal value of e_0, Improved model*	*R^2*	*Optimal value of e_0, Symmetric model*	*R^2*
Buttmann	*1.20*	*0.56*	*0.40*	*0.49*
Haas, B.	*0.96*	*0.58*	*0.30*	*0.49*
Jacob	*0.70*	*0.48*	*0.58*	*0.46*
Haas, R.	*0.65*	*0.45*	*0.40*	*0.42*
Sander	*1.52*	*0.56*	*1.10*	*0.51*
Steinhaus	*0.75*	*0.46*	*0.18*	*0.32*
Wunderlich	*0.85*	*0.46*	*0.32*	*0.34*

It is possible to notice that, similar to the duple-meter model, the most elastic recordings were made at the period instruments. The recording with the highest value of temporal elasticity $e_0 = 1.52$ was made by Martin Sander at the famous large Walcker organ in Riga Cathedral, which was built in 1883 and "represented an ideal instrument of the time when Reger was a student" (Sander 1994). Furthermore, this organ was mentioned in Hugo Riemann's organ manual (Riemann 1888, 143), in which he provided an up-to-date disposition and his recommendations about the registration for this instrument. Thus, it was not surprising to find that the recording on this instrument (Sander 1994) gave the highest coefficient of determination ($R^2 = 0.51$) against the symmetric model, which is indeed the most dogmatic mathematical implementation of Riemannian theory. The most "mechanical" (or the least elastic) performances were made on the relatively modern organs with mechanic action (such as Albiez organ in Frankfurt, built in 1983), or on the organs with electric action (Willi Peter organ in Hamburg, built in 1966, and the Harrison organ in Boston, 1936).

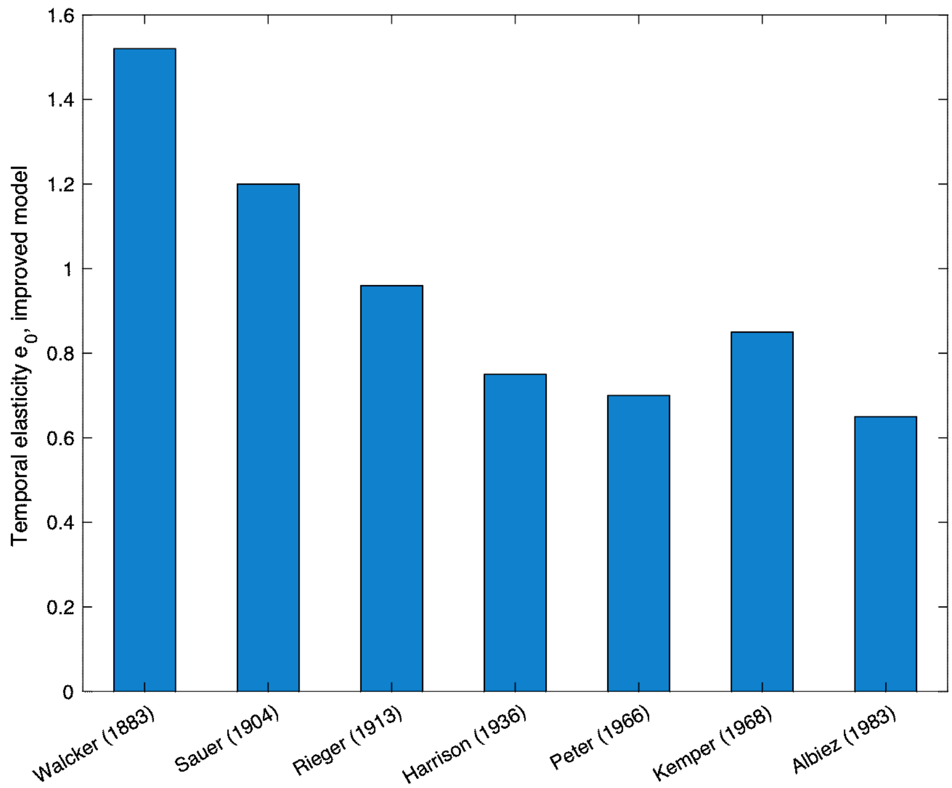


Figure 25. Values of temporal elasticity e_0 and the recordings' organs.

Figure 25 shows the temporal elasticity and the respective organ used for the investigated recordings. Remarkably, the recording with the relatively high elasticity at the tracker (Kemper organ, Hamburg, 1968) was made by Heinz Wunderlich, a well-renowned interpret of German late Romantic music and student of Max Reger's long-year friend Karl Straube. This is an important finding showing how the expressive performance of the late Romantic piece could be made even at the stylistically different instrument.

The temporal elasticity concept allowed also to explore the changes in performance practice strategies over the last decades. Figure 26 depicts the temporal elasticity values, as well as the mean recording's tempo values against the recordings' years. The professional human recording used for the analytical model evaluation in Chapter III was added to complete the picture.

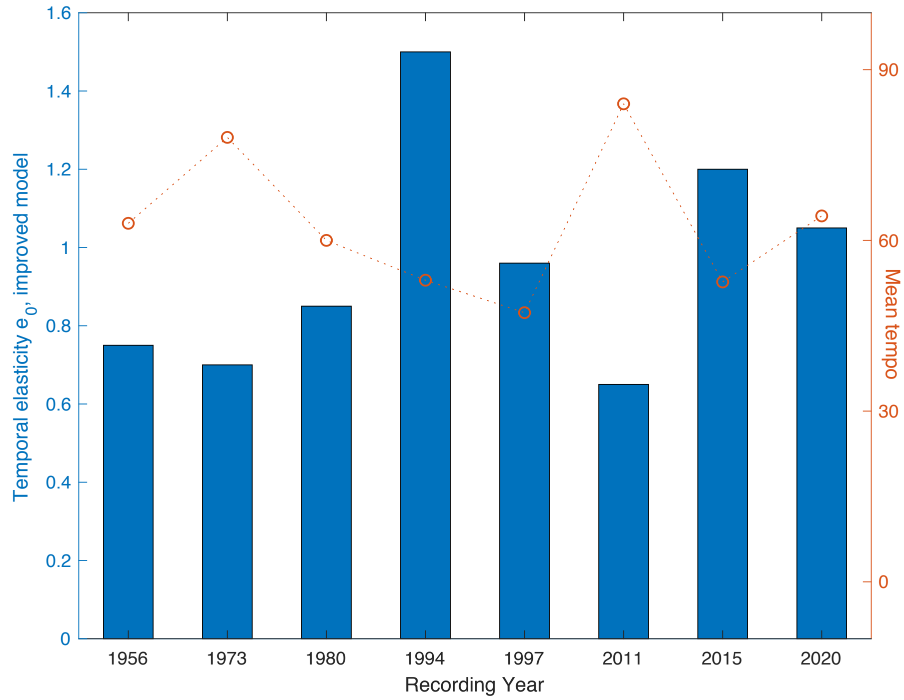


Figure 26. Temporal elasticity and mean tempo values depending on the recording year.

From the graph above, we can see that the recordings in the 50s–80s show less elasticity and higher mean tempo values, which might be due to the preceding *Orgelbewegung* influence. In contrast, the recordings made in the 90s, with higher elasticity values and slower tempi, might signal the gradual return to the more authentic performance practice with respect to the late Romantic style and Riemannian principles. The contemporary performance palette includes a variety of interpretations with (hopefully) prevailing of the more elastic ones. It was interesting to find that despite the relatively small sample size, the mean tempo and temporal elasticity showed a statistically significant negative correlation $R = -0.7, p < 0.05$: the slower was the recording's mean tempo, the higher was its expressive impact.

V.3 Performance analysis of Max Reger's recording

The temporal elasticity concept was used to investigate Max Reger's playing of his own music. The recording (Reger 1913) was made in 1913 in Freiburg im Breisgau on the newly built Welte Philharmonic organ (see disposition in Appendix 7). This two-manuals electro-pneumatic instrument was equipped with the reel mechanism and was initially conceived for the *Titanic*-like large *Britannic* ship; however, it was never installed there because of the beginning of World War I. Rumsey (2014) described Reger's recording process as follows: "He [Reger] played 16 works of his own music for Welte, arriving in Freiburg in style around the 28th May 1913… He sat down at the Welte's console and—seemingly with minimal preparation—started recording."

Figure 27. Max Reger playing the Freiburg recording organ in May 1913 (Rumsey 2014).

Two sections from Max Reger's *Basso Ostinato* op. 92/4 were analyzed here with the same steps as before for the op. 135a/1:

1) building a symmetric (obviously, duple-meter) model,

2) optimization procedure, and

3) defining the weights coefficients to create an improved model.

In the opening section (see Figure 28), the model was made from the second quarter note in bar 1 to the second quarter note in bar 9. The regression of Reger's performance data against the symmetric model gave $R^2 = 0.52, p < 0.0001$. Reger's temporal elasticity value was relatively low, only $e_0 = 0.18$, which might be an explanation why performers sometimes did not find Reger's own organ recordings really expressive and considered him being "not the greatest organist of his time" (The Classical Music Guide Forums 2018).

Figure 28. Basso Ostinato op. 92/4, bars 1–9 (Reger 1966).

Reger's differentiation of the phrasing levels was very logical and methodic: he clearly gave more weight to the Riemannian fundamental "motivic" level 4. Furthermore, level 3 (over the main repeated *Ostinato* pattern) and level 1 (approximately) marked by the two long slurs were also relatively strong. The weakest level in Reger's performance was level two (over 2-bars segments), which did not contain any marked phrasing events in the score.

Table 10. Coefficients k_{ij} for the Max Reger's performance (best fit curve) and improved model.

Levels	*k_{ij}, average value (best fit model curve)*	*k_{ij}, improved model*
1	*1.39*	*1.5*
2	*0.019*	*0.5*
3	*1.27*	*1.5*
4	*2.73*	*2.5*

Table 10 shows the mean values of weight coefficients for Reger's performance optimization and those chosen for the improved model. The improved model was built in accordance with the modifications described in Chapter III with the respective weight coefficients k_{ij} from Table 10; it gave a highly significant coefficient of determination $R^2 = 0.64, p < 0.0001$.

Another interesting finding is Reger's usage of slurring: Reger's tempo profile (Figure 29) shows that the long slurs were used to shape the phrase (bars 2–4 and bars 6–8). In contrast, the short slurs were used only for articulation, resulting in the long tempo plateau in the second part, corresponding to the sections with the short slurs only (See Figure 28). Thus, it might be possible to guess the slurring from the tempo profile and vice versa, using the long phrasing slurs to determine the phrase boundaries for the timing simulation purpose. This observation would be highly beneficial to the computer modelling algorithm proposed in Chapter VI.

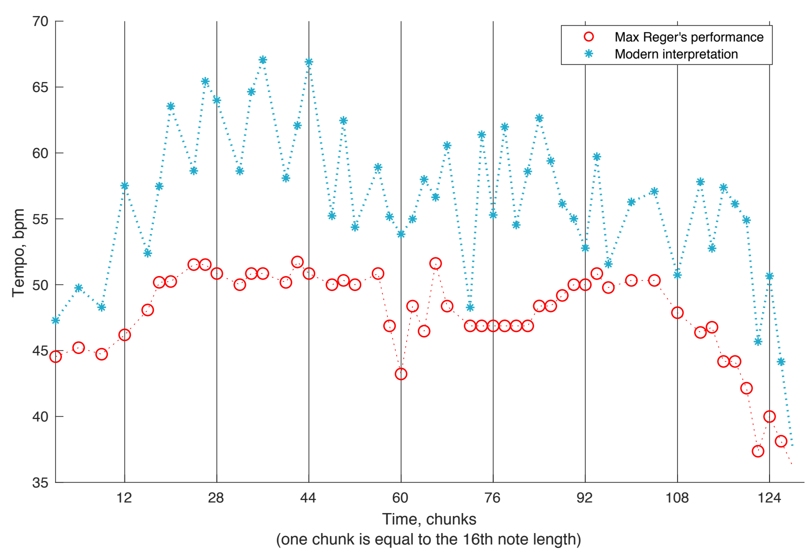


Figure 29. Tempo profiles for Max Reger's recording and modern recording (bars 1–9).

It was very insightful to compare the modern (*my own*) performance of the same section, recorded at the *Casavant* organ in The Church of Saint Andrew and Saint Paul in Montreal in October 2019. It could be observed that Reger's tempo was remarkably slower than the tempo of the modern performance (mean values 47.8 BPM and 56.6 BPM, respectively); however, this might be due to the technical inconsistencies while playing back the Reger's initial recording rolls (Rumsey 2005).

In general, the modern performance compared to Max Reger's was more elastic with $e_0 = 0.3$, as well as less determined by the model, neither symmetric ($R^2 = 0.5, p < 0.0001.$), nor improved ($R^2 = 0.61, p < 0.0001.$).The phrasing levels had a similar weight distribution, where the "motivic" level 4 was the most prominent, and level 2 was the weakest. Nevertheless, the modern performer preferred a short-scale phrasing at level 3 rather than beard the long-scale structure (level 1).

The analysis of another section of the piece (Figure 30, bars 26–34 with the model starting from the second quarter note) showed very similar results. The modern performance was less determined by the model than the Max Reger's one (for the improved model, $R^2 = 0.72$ and $R^2 = 0.6$, respectively), and the temporal elasticity value was slightly higher for modern performance.

Figure 30. Basso Ostinato op. 92/4, bars 26–34 (Reger 1966).

Remarkably, Reger's temporal elasticity increased over the piece and reached $e_0 = 0.25$ in this section; the modern performer's elasticity remained the same for both excerpts ($e_0 = 0.3$). As regards the levels' distribution, the "Riemannian" motivic level 4 was still the most prominent for both performers, and level 2 was the weakest. In this section, Reger continued to phrase methodically clear the *Ostinato* and thus made level 3 relatively prominent. Modern performer, in this excerpt, paid less attention to the shaping of the repeated *Ostinato* but succeeded in shaping the long 4-bars phrase instead.

Similar to the previous section, there was a visually noticeable correlation between the slurring in the music score (see Figure 30) and Max Reger's tempo profile: there was a slightly descending quasi-linear tempo area which corresponded to the section in bars 31–32 without long slurs (see Figure 31, chunks 88–104). In contrast, the sections under slurs in bars 26–30, 30–31 and 33–34 coincided with the arch-like local tempo shape (see Figure 31).

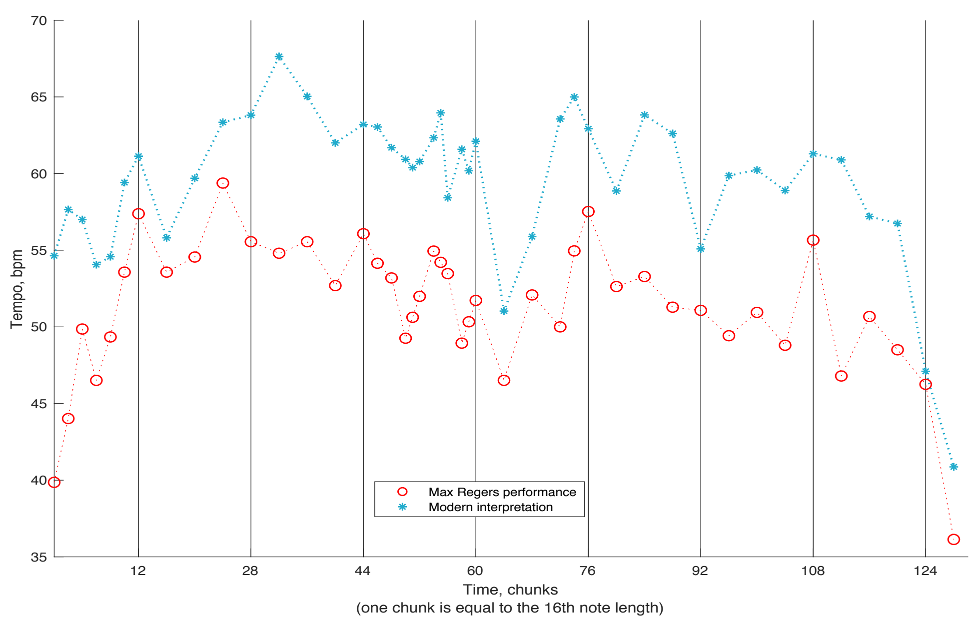


Figure 31. Tempo profiles for Max Reger's recording and modern recording (bars 26–34).

Reger's coefficient of determination for the dogmatic symmetric model $R^2 = 0.62$ ($p < 0.0001$) was exceptionally high. This might indirectly confirm the influence of Riemann's basic scholarship on Reger's performing strategies. However, the results of the first part of my listening experiment in Chapter IV showed that the interpretation simulated with the symmetric model timing profile was significantly less comprehensible for the modern listener.

VI. Temporal elasticity and computer modelling of expressive timing

The proposed temporal elasticity model opens an important avenue towards expressive timing simulation. The symmetric timing model could be easily calculated by just putting the elasticity values from (3) into equation (6), but, according to my listening experiment, the listeners clearly preferred the improved model rather than the symmetric one (see Table 4 and Figure 14 in Chapter IV). To create an improved model, the weight coefficients k_{ij} in (4) should be determined before using the equation (6). In the previous Chapters it was done by the optimization procedure from the real performance data and the manual score analysis. An Artificial Intelligent system can provide an alternative solution, which would be more suitable for the computer timing simulation.

VI.1 Automatic detection of the weight coefficients

To approach the automatic levels' weighting problem, I created a Demo classifier based on two Machine Learning algorithms.[13] The dataset contained 136 samples from the analyzed MIDI files of Max Reger's op. 73, op.92 and op.135a (whole pieces and piece segments), as well as excerpts from *Monologue* by Rheinberger op. 162. All files were recorded at The Church of Saint

[13] Demo notebook is available at https://draginda.org/#myprojects

Andrew and Saint Paul in Montreal and processed analogously to the workflow described in Chapter III for the op. 135a and op. 73 (for the duple and triple-meter models, respectively). The obtained levels weight coefficients were put into three classes:

1) Weak (if the mean level's weight coefficient was less than 0.75, yielding $k = 0.5$ for the improved model);

2) Middle (if the level's weight coefficient was more than 0.75 and less than 1.75, yielding $k = 1$ or $k = 1.5$ for the improved model);

3) Strong (if the level's weight coefficient was higher than 1.75, yielding $k = 2$ or $k = 2.5$ for the improved model).

The following elements in a music score, which, according to Riemannian studies, could impact the phrasing (see Section II.2), were selected as input features:

- Number of explicitly marked phrasing slurs in the score.
- Number of supporting *crescendo-diminuendo* pairs under slurs. Only paired hairpins fully enclosed under the slur were counted; incomplete or non-paired hairpins were ignored and considered as pure dynamic markings with no relation to the phrasing.
- Number of slurs ending with the *fermata* sign, signalizing the end of the choral sentence in the *Choral Preludes* or similar genres.
- Mean tempo of the piece indicated in the score in BPM or as one of the common tempo indications (assigned to one three groups: slow, medium, fast).
- Most frequent note duration, more known as *music pulse*.
- First dynamic mark in the score.
- Time signature of the piece (or piece's segment in the case of changing meter).
- Presence or absence of the final *ritardando* mark in the score.

- Stepwise dynamic changes at the beginning and the end of the level; multiple steps on the same level were summed up.
- Overall number of notes in the score (quantitative information about music texture).

Two other model-specific features were added to complete the set:

  - Number of semi-ellipses corresponding to this level in the model.
  - Temporal elasticity e_0 of the main arch.

Two algorithms were chosen to classify the level's weight based on these features: Polynomial multiclass Logistic Regression (PLR) and Random Forest (RF). The rationale for this choice was their ability to handle the non-linear data, and, in the case of the Random Forest, also its high interpretability.

PLR classifier performed best with the polynomial degree *2*, the *Newton-CG* solver and a strong regularization term $C = 0.001$. The achieved test accuracy was relatively poor (86%), with four misclassified level's weights from the test set. Errors analysis showed that the most problematic to recognize was the Middle weights' class with its three false negatives values. Table 11 shows the Precision, Recall, and F-score values for this model.

Table 11. PLR classifier: performance evaluation.

Level	*F-measure*	*Precision*	*Recall*
Weak	*0.89*	*0.80*	*1.00*
Middle	*0.86*	*0.92*	*0.80*
Strong	*0.84*	*0.80*	*0.89*

RF classifier with *35* ensembled *Decision Trees* and *Gini* splitting criterion reached 100% accuracy on the training set and achieved a remarkably high accuracy value on the test set, 96%. RF misclassified only one weight coefficient from the test set; noticeably, the same coefficient was wrongly predicted by the PLR as well. Presumably, there were some interpretation inconsistencies for this excerpt (Max Reger, op.135a/24, level 2); another cause could be the half-note pulse of the piece, unique for this dataset. Obviously, more data with unusual pulse values would be beneficial for this model. Table 12 provides the RF evaluation outline.

Table 12. RF classifier: performance evaluation.

Level	*F-measure*	*Precision*	*Recall*
Weak	*0.89*	*0.80*	*1.00*
Middle	*0.97*	*1.00*	*0.93*
Strong	*1.00*	*1.00*	*1.00*

It is noteworthy that the RF classifier achieved the perfect one for the Strong level's *F*-score because this level, mathematically, has the greatest effect on the resulting timing curve; the correct detection of its coefficients is essential for the model.

An insightful result was obtained from the RF features' analysis. It was revealed that the most important features for decision splits were the number of semi-ellipses (0.28), the number of slurs (0.14) and the number of notes (0.125). This was yet another confirmation for the suggestion in Chapter III that the slurring combined with the information about music structure and texture could serve as significant factors for the expressive phrasing.

VI.2 Algorithm proposal

The ability of an AI-based system to correctly predict the levels' strength gives an excellent opportunity to simulate the expressive timing computationally without using the optimization procedure on the human performance data. The starting point for the proposed algorithm would be the symbolic music score file in MusicXML or in a similar format, supporting the slurring and advanced dynamics notation. The expressive timing might be simulated then through the following steps:

- Detecting all phrasing slurs and assigning them a respective level (usually, the approximately four-bars long section under slur represents level 1, approximately two bars long –> level 2, approximately one bar long –>level 3). The slur must be at least one bar long to be considered as the phrasing slur; otherwise, it should be ignored as an articulation indication.
- Building a symmetric model; the remaining levels must be defined by appending the subsequent sections under the next slur or dividing the existing sections by two (for levels 1 and 2).
- Adding the "motivic" level 4 according to the piece's time signature (duple- or triple-meter model).
- Determining the weight coefficients for each level with the AI-based classifier (for example, RF classifier as proposed in Section VI.1) and computing the resulting tempo curve from the improved model Equation (6).

It is essential to notice here that the temporal elasticity of the global arch e_0 was also an input parameter for the RF classifier from Section VI.1. By creating a fully automatic system, e_0 should be chosen either from the proposed in Chapters III and IV "safest" area [0.8; 1.2] or slightly higher if the highly expressive performance is desired. After analyzing several case studies, my most generic recommendation for choosing e_0 would be picking up the values from the range [0.8–0.9] for the slow or medium tempi and (0.9–1.2] for the virtuosic piece in a fast tempo.

This proposed algorithm will work best for the short Romantic organ pieces with clear and consistent slurring. The resulted tempo curve might be applied to the equitemporal MIDI file of the piece, which can then be played back, as it was done in the listening experiment. The complete engineering implementation of this system is, of course, outside the scope of this performance research paper but is already reserved for future work.

I want to emphasize that the simulation of expressive performance could be beneficial if used as an investigation tool. As noticed by Crawford and Gibson (2009, 115), "the aim of modelling expressive performance should not be to replace or compete with a human in this domain, but to create models with which the complex nature of expressive performance can be better investigated and understood."

VII. Conclusion

In the present thesis, for the first time, the mathematical model of the Riemannian motivic scheme was created and evaluated, both analytically and empirically, through the listening tests. The proposed model is based on music structure and, therefore, benefits from the methodological advantages of the previous structural models found in the literature, such as high determination potential (Todd 1992) or clear correlation with the expressive impact (Palmer 1996). The advantage of my model against its predecessors is in its unlikability from the complex harmonic analysis and its controllability by a single, easy interpretable parameter. The model has a two-fold application: it can be used in performance analysis, as well as in computer simulation of expressive timing.

The model parameter **temporal elasticity** was proven to carry the expressive information of the organ performance in late Romantic style. The results of the listening experiment clearly showed that the most elastic models were rated as the most expressive, while the least or second-least elastic models were rated as the least expressive. The most reliable values of temporal elasticities for this model were found within the interval [0.8, 1.2].

This thesis has provided a deeper insight into the organ performance practice of the German late Romantic repertoire. In particular, it was found that the instrument type played an important role in the performance expression, and period organs allowed performers the more elastic phrasing. The obtained significant negative correlation between the tempo and temporal elasticity showed that the slower the recording's mean tempo, the higher its expressive impact.

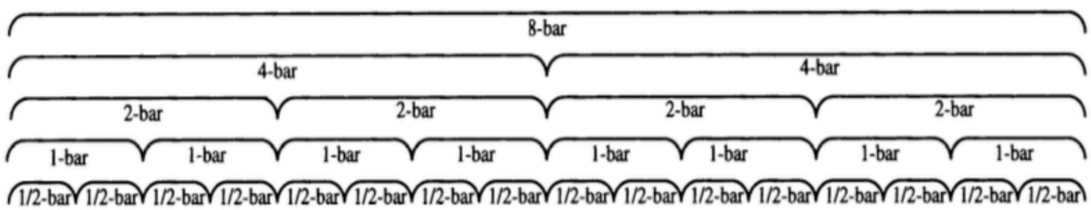
The analysis of Max Reger's performance undertaken here has extended the knowledge of authentic slurring in the late Romantic style. It was shown that, in accordance with Riemannian theory, slurs more than one-bar long were used to shape the phrase rather than being an articulation indication. The importance of slurring for the phrasing profile was also confirmed by the Random Forest features' analysis performed in the Chapter VI.

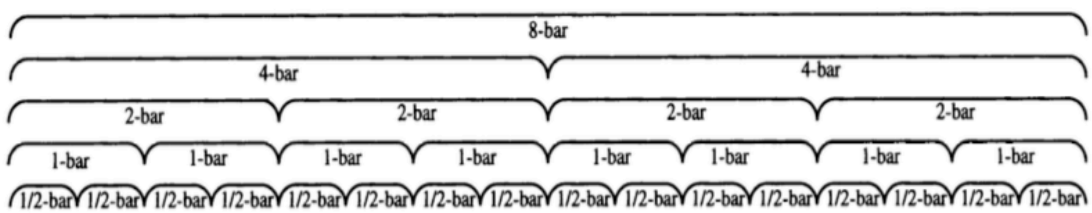
This work was based on a solid musicological foundation, Riemannian theory, which represented stylistically appropriate phrasing rules for the late Romantic organ music. The results of this work demonstrated that fulfilling these rules might be helpful for the modern performers to make their interpretations not only stylistically correct but also expressive and touching. I want to conclude my thesis with the following words of Ludger Lohmann (1995, 251): "Why should organists of today be interested in Hugo Riemann? As a composition teacher, Riemann influenced not only his student Max Reger—and thus one of the most important oeuvres in the organ repertoire, but, being one of the most renowned figures in European musical life around the turn of the century, also helped shape fundamental ideas about musical performance in a whole generation of musicians by his numerous writings, whose impact can still be felt today."

VIII. Appendices

Appendix 1

A schematic illustration of the hierarchical structural representation that the model used as an input structure (Windsor and Clarke 1997, 135)





Summary of Regression Analyses

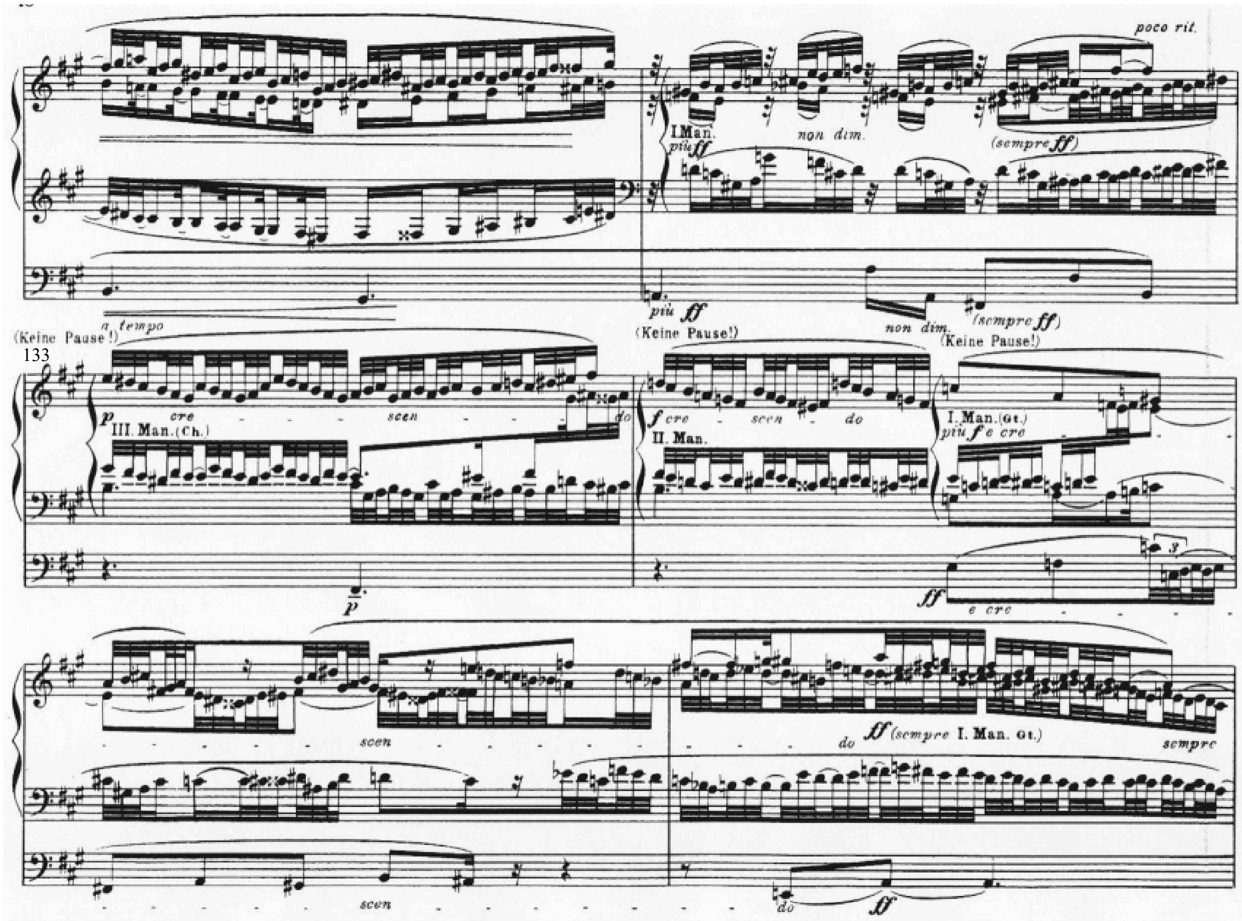
	Value of R^2	
Algorithmic Performance	**Timing**	**Dynamics**
AP (1, 1, 1, 1, 1)	.299	.086
AP (1, 1, 1, 1, 2)	.375	.054
AP (1, 1, 1, 2, 1)	.321	.041*
AP (1, 1, 2, 1, 1)	.235	.126
AP (1, 2, 1, 1, 1)	.258	.107
AP (2, 1, 1, 1, 1)	.217	.089

*Despite the low values of R^2, these regressions are all significant at the .0001 level: e.g. $R^2 = .041$, $df = 382$, $p = .000006701$.

Appendix 2

Music score used for the duple-meter model (Reger 1915).

Music score used for the triple-meter model (Reger 1904).



Appendix 3

Part 1 experimental results for *Organists* group

Version	*N*	*Mean*	*SEM*	*CI-95%, Low*	*CI-95%, High*	*CI-83%, Low*	*CI-83%, High*
135H	*29*	*0.48*	*0.094*	*0.29*	*0.68*	*0.35*	*0.62*
135M	*29*	*0.34*	*0.090*	*0.16*	*0.53*	*0.22*	*0.47*
135SM	*29*	*0.17*	*0.071*	*0.03*	*0.32*	*0.07*	*0.27*
73H	*29*	*0.45*	*0.094*	*0.26*	*0.64*	*0.32*	*0.58*
73M	*29*	*0.34*	*0.090*	*0.16*	*0.53*	*0.22*	*0.47*
73SM	*29*	*0.21*	*0.077*	*0.04*	*0.36*	*0.10*	*0.31*

Part 1 experimental results for *Non-Organists* group

Version	*N*	*Mean*	*SEM*	*CI-95%, Low*	*CI-95%, High*	*CI-83%, Low*	*CI-83%, High*
135H	*24*	*0.46*	*0.104*	*0.24*	*0.67*	*0.31*	*0.61*
135M	*24*	*0.33*	*0.098*	*0.13*	*0.54*	*0.19*	*0.47*
135SM	*24*	*0.21*	*0.085*	*0.03*	*0.38*	*0.09*	*0.33*
73H	*24*	*0.42*	*0.103*	*0.20*	*0.63*	*0.27*	*0.56*
73M	*24*	*0.46*	*0.104*	*0.24*	*0.67*	*0.31*	*0.61*
73SM	*24*	*0.13*	*0.069*	*0.00*	*0.27*	*0.03*	*0.22*

Appendix 4

Descriptive statistics and Turkey HSD post-hoc test results for all participants (op. 135a/1)

Dependent Variable: Mean rating, op. 135a/1

Elasticity	Mean	Std. Deviation	N
.4	.3757	.17845	53
.6**	.5887	.18953	53
.8	.5151	.14818	53
1.2	.5662	.18992	53
1.6	.6815	.21438	53
Total	.5454	.20972	265

Tukey HSD

(I) Elasticity	(J) Elasticity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
.4	.6**	-.2130*	.03600	.000	-.3119	-.1141
	.8	-.1394*	.03600	.001	-.2383	-.0405
	1.2	-.1906*	.03600	.000	-.2895	-.0917
	1.6	-.3058*	.03600	.000	-.4048	-.2069
.6**	.4	.2130*	.03600	.000	.1141	.3119
	.8	.0736	.03600	.248	-.0253	.1725
	1.2	.0225	.03600	.971	-.0764	.1214
	1.6	-.0928	.03600	.077	-.1917	.0061
.8	.4	.1394*	.03600	.001	.0405	.2383
	.6**	-.0736	.03600	.248	-.1725	.0253
	1.2	-.0511	.03600	.615	-.1500	.0478
	1.6	-.1664*	.03600	.000	-.2653	-.0675
1.2	.4	.1906*	.03600	.000	.0917	.2895
	.6**	-.0225	.03600	.971	-.1214	.0764
	.8	.0511	.03600	.615	-.0478	.1500
	1.6	-.1153*	.03600	.013	-.2142	-.0164
1.6	.4	.3058*	.03600	.000	.2069	.4048
	.6**	.0928	.03600	.077	-.0061	.1917
	.8	.1664*	.03600	.000	.0675	.2653
	1.2	.1153*	.03600	.013	.0164	.2142

Based on observed means.
The error term is Mean Square (Error) = .034.
*. The mean difference is significant at the .05 level.
**. Human performance (135H).

Descriptive statistics and Turkey HSD post-hoc test results for all participants (op. 73)

Dependent Variable: Mean rating, op. 73

Elasticity	Mean	Std. Deviation	N
.4	.5445	.18123	53
.8	.4781	.20973	53
1.0**	.6332	.24731	53
1.2	.6104	.15229	53
1.6	.7606	.13058	53
Total	.6054	.21000	265

Tukey HSD

(I) Elasticity	(J) Elasticity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.4	.8	.0664	.03668	.370	-.0343	.1672
	1.0**	-.0887	.03668	.114	-.1894	.0121
	1.2	-.0658	.03668	.378	-.1666	.0349
	1.6	-.2160*	.03668	.000	-.3168	-.1153
.8	.4	-.0664	.03668	.370	-.1672	.0343
	1.0**	-.1551*	.03668	.000	-.2558	-.0543
	1.2	-.1323*	.03668	.003	-.2330	-.0315
	1.6	-.2825*	.03668	.000	-.3832	-.1817
1.0**	.4	.0887	.03668	.114	-.0121	.1894
	.8	.1551*	.03668	.000	.0543	.2558
	1.2	.0228	.03668	.971	-.0779	.1236
	1.6	-.1274*	.03668	.005	-.2281	-.0266
1.2	.4	.0658	.03668	.378	-.0349	.1666
	.8	.1323*	.03668	.003	.0315	.2330
	1.0**	-.0228	.03668	.971	-.1236	.0779
	1.6	-.1502*	.03668	.001	-.2509	-.0494
1.6	.4	.2160*	.03668	.000	.1153	.3168
	.8	.2825*	.03668	.000	.1817	.3832
	1.0**	.1274*	.03668	.005	.0266	.2281
	1.2	.1502*	.03668	.001	.0494	.2509

Based on observed means.
The error term is Mean Square (Error) = .036.
*. The mean difference is significant at the .05 level.
**. Human performance (73H).

Appendix 5

Descriptive statistics and Turkey HSD post-hoc test results for *Organists* group (op. 135a/1)

Dependent Variable: Mean rating, op.135a/1 (Organists)

Elasticity	Mean	Std. Deviation	N
.4	.3376	.19222	29
.6**	.5769	.17244	29
.8	.5352	.13845	29
1.2	.5652	.19897	29
1.6	.7379	.20348	29
Total	.5506	.22108	145

Tukey HSD

(I) Elasticity	(J) Elasticity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
.4	.6**	-.2393*	.04797	.000	-.3719	-.1067
	.8	-.1976*	.04797	.001	-.3302	-.0650
	1.2	-.2276*	.04797	.000	-.3602	-.0950
	1.6	-.4003*	.04797	.000	-.5329	-.2678
.6**	.4	.2393*	.04797	.000	.1067	.3719
	.8	.0417	.04797	.907	-.0909	.1743
	1.2	.0117	.04797	.999	-.1209	.1443
	1.6	-.1610*	.04797	.009	-.2936	-.0285
.8	.4	.1976*	.04797	.001	.0650	.3302
	.6**	-.0417	.04797	.907	-.1743	.0909
	1.2	-.0300	.04797	.971	-.1626	.1026
	1.6	-.2028*	.04797	.000	-.3353	-.0702
1.2	.4	.2276*	.04797	.000	.0950	.3602
	.6**	-.0117	.04797	.999	-.1443	.1209
	.8	.0300	.04797	.971	-.1026	.1626
	1.6	-.1728*	.04797	.004	-.3053	-.0402
1.6	.4	.4003*	.04797	.000	.2678	.5329
	.6**	.1610*	.04797	.009	.0285	.2936
	.8	.2028*	.04797	.000	.0702	.3353
	1.2	.1728*	.04797	.004	.0402	.3053

Based on observed means.
The error term is Mean Square (Error) = .033.
*. The mean difference is significant at the .05 level.
**. Human performance (135H).

Descriptive statistics and Turkey HSD post-hoc test results for *Organists* group (op. 73)

Dependent Variable: Mean rating, op. 73 (Organists)

Elasticity	Mean	Std. Deviation	N
.4	.5445	.17373	29
.8	.5055	.19850	29
1.0**	.5990	.24857	29
1.2	.6200	.13638	29
1.6	.7721	.10874	29
Total	.6082	.19964	145

Tukey HSD

(I) Elasticity	(J) Elasticity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.4	.8	.0390	.04724	.923	-.0916	.1695
	1.0**	-.0545	.04724	.778	-.1850	.0761
	1.2	-.0755	.04724	.501	-.2061	.0550
	1.6	-.2276*	.04724	.000	-.3581	-.0970
.8	.4	-.0390	.04724	.923	-.1695	.0916
	1.0**	-.0934	.04724	.282	-.2240	.0371
	1.2	-.1145	.04724	.115	-.2450	.0161
	1.6	-.2666*	.04724	.000	-.3971	-.1360
1.0**	.4	.0545	.04724	.778	-.0761	.1850
	.8	.0934	.04724	.282	-.0371	.2240
	1.2	-.0210	.04724	.992	-.1516	.1095
	1.6	-.1731*	.04724	.003	-.3037	-.0425
1.2	.4	.0755	.04724	.501	-.0550	.2061
	.8	.1145	.04724	.115	-.0161	.2450
	1.0**	.0210	.04724	.992	-.1095	.1516
	1.6	-.1521*	.04724	.014	-.2826	-.0215
1.6	.4	.2276*	.04724	.000	.0970	.3581
	.8	.2666*	.04724	.000	.1360	.3971
	1.0**	.1731*	.04724	.003	.0425	.3037
	1.2	.1521*	.04724	.014	.0215	.2826

Based on observed means.
The error term is Mean Square (Error) = .032.
*. The mean difference is significant at the .05 level.
**. Human performance (73H).

Appendix 6

Descriptive statistics and Turkey HSD post-hoc test results for *Non-Organists* group (op. 135a/1)

Dependent Variable: Mean rating, op.135a/1 (Non-Organists)

Elasticity	Mean	Std. Deviation	N
.4	.4217	.15159	24
.6**	.6029	.21126	24
.8	.4908	.15869	24
1.2	.5675	.18262	24
1.6	.6133	.21128	24
Total	.5392	.19584	120

Tukey HSD

(I) Elasticity	(J) Elasticity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
.4	.6**	-.1813*	.05335	.008	-.3291	-.0334
	.8	-.0692	.05335	.694	-.2170	.0787
	1.2	-.1458	.05335	.055	-.2937	.0020
	1.6	-.1917*	.05335	.004	-.3395	-.0438
.6**	.4	.1813*	.05335	.008	.0334	.3291
	.8	.1121	.05335	.227	-.0358	.2599
	1.2	.0354	.05335	.964	-.1124	.1833
	1.6	-.0104	.05335	1.000	-.1583	.1374
.8	.4	.0692	.05335	.694	-.0787	.2170
	.6**	-.1121	.05335	.227	-.2599	.0358
	1.2	-.0767	.05335	.605	-.2245	.0712
	1.6	-.1225	.05335	.154	-.2704	.0254
1.2	.4	.1458	.05335	.055	-.0020	.2937
	.6**	-.0354	.05335	.964	-.1833	.1124
	.8	.0767	.05335	.605	-.0712	.2245
	1.6	-.0458	.05335	.911	-.1937	.1020
1.6	.4	.1917*	.05335	.004	.0438	.3395
	.6**	.0104	.05335	1.000	-.1374	.1583
	.8	.1225	.05335	.154	-.0254	.2704
	1.2	.0458	.05335	.911	-.1020	.1937

Based on observed means.
The error term is Mean Square (Error) = .034.
*. The mean difference is significant at the .05 level.
**. Human performance (135H).

Descriptive statistics and Turkey HSD post-hoc test results for *Non-Organists* group (op. 73)

Dependent Variable: Mean rating, op. 73 (Non-Organists)

Elasticity	Mean	Std. Deviation	N
.4	.5446	.19368	24
.8	.4450	.22224	24
1.0**	.6746	.24454	24
1.2	.5988	.17185	24
1.6	.7467	.15423	24
Total	.6019	.22267	120

Tukey HSD

(I) Elasticity	(J) Elasticity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.4	.8	.0996	.05774	.423	-.0604	.2596
	1.0**	-.1300	.05774	.169	-.2900	.0300
	1.2	-.0542	.05774	.881	-.2142	.1059
	1.6	-.2021*	.05774	.006	-.3621	-.0421
.8	.4	-.0996	.05774	.423	-.2596	.0604
	1.0**	-.2296*	.05774	.001	-.3896	-.0696
	1.2	-.1538	.05774	.066	-.3138	.0063
	1.6	-.3017*	.05774	.000	-.4617	-.1416
1.0**	.4	.1300	.05774	.169	-.0300	.2900
	.8	.2296*	.05774	.001	.0696	.3896
	1.2	.0758	.05774	.683	-.0842	.2359
	1.6	-.0721	.05774	.723	-.2321	.0879
1.2	.4	.0542	.05774	.881	-.1059	.2142
	.8	.1538	.05774	.066	-.0063	.3138
	1.0**	-.0758	.05774	.683	-.2359	.0842
	1.6	-.1479	.05774	.084	-.3079	.0121
1.6	.4	.2021*	.05774	.006	.0421	.3621
	.8	.3017*	.05774	.000	.1416	.4617
	1.0**	.0721	.05774	.723	-.0879	.2321
	1.2	.1479	.05774	.084	-.0121	.3079

Based on observed means.
The error term is Mean Square (Error) = .040.
*. The mean difference is significant at the .05 level.
**. Human performance (73H).

Appendix 7

Disposition of the Welte organ used for Max Reger's recording (Rumsey 2005).

Freiburg, Welte Company's premises
Pre-1913?

Manual	I	**Manual**	
Bordun	16	Wienerflöte	8
Principal	8	Bordun	8
Traversflöte	8	Viola	8
Gambe	8	Aeoline	8
Viol. d'orch.	8	Dolce	4
Vox coelestis	8	Quinte	$2^{2}/_{3}$
Flöte	4	Clarinette	16
Piccolo	2	Trompete	8
Sesquialter		Horn	8
Fagott	8	Oboe	8
Harfe		Vox Humana	8
Glocken			

The entire organ was enclosed in a single swell-box and an additional open-shut echo box houses the Vox Humana within this.

Pedal	
Violonbass	16
Subbass	16
Cello	8
Gedackt	8
Posaune	16

Compass: Manuals: C-a^3 (c4?); Pedals: C-f^1(?);
Accessories: Vox Humana Echo (opens/shuts Vox Humana's separate box lid); Tremolo

IX. Bibliography

Afshartous, David, and Richard A. Preston. 2010. "Confidence Intervals for Dependent Data: Equating Non-overlap with Statistical Significance." *Computational Statistics and Data Analysis* 54: 2296–2305.

Bhatara, Anjali. 2008. "Music as a Means of Investigating Perception of Emotion and Social Attribution in Typical Development and in Autism Spectrum Disorders." Ph.D. diss., McGill University.

Bhatara, Anjali, Tirovolas, Anna, Duan, Lilu, Levy, Bianca, and Daniel Levitin. 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, CD 2*. Oehms Classics OC854, compact disc.

Buttmann, Bernhard. 2015. *Max Reger: Complete Organ Works, Volume 4, CD 3*. Oehms Classics OC854, compact disc.

Cancino-Chacón, Carlos E., Maarten Grachten, Werner Göbl, and Gerhard Widmer. 2018. "Computational Models of Expressive Music Performance: A Comprehensive and Critical Review." *Frontiers in Digital Humanities* 5(25): 1–25.

Cheng, Eric, and Elaine Chew. 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): 325–228.

Clarke, Eric F. 1982. "Timing in the Performance of Erik Satie's 'Vexations'." *Acta Psychohgica* 50: 1–19.

Craft, Robert, and Igor Stravinsky. 1982. *Dialogues*. University of California Press.

Crawford, Tim, and Lorna Gibson. 2009. *Modern Methods for Musicology: Prospects, Proposals, and Realities (Digital Research in the Arts and Humanities).* Farnham: Ashgate.

Dixon, Simon. 2007. "Evaluation of the Audio Beat Tracking System Beatroot." *Journal of New Music Research* 36 (1): 39–50.

Doğantan-Dack, Mine. 2012. "Phrasing—the Very Life of Music: Performing the Music and Nineteenth-Century Performance Theory." *Theoretical and Critical Contexts in Nineteenth-Century Performance Practice* 9(1): 7–30.

Author's abstracts and publication:

Draginda, Yulia. 2020. "Temporal Elasticity: Demystifying Performance Expression in German Late Romantic Organ Music." 2d CIRMMT-OICRM-BRAMS Student Symposium, Montreal, Canada. Abstract: https://www.cirmmt.org/activities/general-assembly/previous-years/cobs-2020-program, last accessed May 25, 2021.

Draginda, Yulia, and Ichiro Fujinaga. 2020. "First Steps Towards Modelling Expressive Timing in German Late Romantic Organ Music." 21st International Society for Music Information Retrieval Conference, Late-Breaking-Demo Session. Abstract: https://program.ismir2020.net/static/lbd/ISMIR2020-LBD-426-abstract.pdf, last accessed May 25, 2021.

Draginda, Yulia, and Ichiro Fujinaga. 2021. "Life, Color, Warmth and Truth" : Expressive Timing in German Late Romantic Organ Music." *Per Musi*: *Computing Performed Music* 41.

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 086-087, compact disc.

Friberg, Anders, Roberto Bresin, and Johan Sundberg. 2006. "Overview of the KTH Rule System for Musical Performance." *Advances in Cognitive Psychology* 2(3): 145–161.

Fugatto. 2020. "The Link Organ of the Ev. Stadtkirche at Giengen-an-der-Brenz (Germany)." last accessed May 1, 2020. http://fugatto.free.fr/giengen_adb.html

Gabrielsson, Alf. 2003. "Music Performance Research at the Millennium." *Psychology of Music* 31(3): 221.

Gabrielsson, Alf, and Patrik Juslin. 1996. " Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience." *Psychology of Music* 24(1): 68–91.

Gingras, Bruno. 2008. *Expressive Strategies and Performer-listener Communication in Organ Performance*. Ph. D. diss., McGill University.

Gingras, Bruno, Stephen McAdams, and Peter Schubert. 2010. " The Performer as Analyst: a Case Study of J. S. Bach's 'Dorian' Fugue BWV 538." In *Music Theory and Interdisciplinarity–8th Congress of the Gesellschaft für Musiktheorie Graz,* Saarbrücken: Pfau-Verlag, 305–318.

Gingras, Bruno, Caroline Palmer, Peter Schubert, and Stephen McAdams. 2015. "Influence of Melodic Emphasis, Texture, Salience, and Performer Individuality on Performance Errors." *Psychology of Music* 44(4): 847–863.

Gingras, Bruno, Marcus Pearce, Meghan Goodchild, Roger Dean, Geraint Wiggins, and Stephen McAdams. 2016. "Linking Melodic Expectation to Expressive Performance Timing and Perceived Musical Tension." *Journal of Experimental Psychology: Human Perception and Performance* 42(4): 594–609.

Goldstein, Harvey, and Michael Healy. 1995. "The Graphical Presentation of a Collection of Means." *Journal of the Royal Statistical Society* 158(1): 175–177.

Göbl, Werner, Simon Dixon, Giovanni De Poli, Friberg, Anders, Bresin, Roberto, and Gerhard Widmer. 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.

Grosche, Peter, and Meinard Müller. 2011. "Extracting Predominant Local Pulse Information from Music Recordings." *IEEE Transactions on Audio, Speech and Language Processing* 19 (6): 1688–1701.

Haas, Bernhard. 1997. *Max Reger: Organ Works*. Audite 20.020, compact disc.

Haas, Rosalinde. 2011. *Reger: Complete Organ Works and Arrangements*. MDG 3151645, compact disc.

Hong, Ju-Lee. 2003. "Investigating Expressive Timing and Dynamics in Recorded Cello Performances." *Psychology of Music* 31(3): 340–352.

Jacob, Werner.1973. *Max Reger. Selected Organ Works*, *Volume 1*. MHS 1563/66, D3, LP.

Jerkert, Jesper. 2004. "Music Articulation in the Organ." In *Proceedings of Joint Baltic–Nordic Acoustics Meeting*, Mariehamn, Finland.

Jillings, Nicholas, Brecht De Man, David Moffat, and Joshua Reiss. 2015 "Web Audio Evaluation Tool: A Browser-Based Listening Test Environment." In *Proceedings of the 12th Sound and Music Computing Conference,* Maynooth.

Johnson, Christopher. 1999. "The Performance of Bach: Study of Rhythmic Timing by Skilled Musicians." *Bulletin of the Council for Research in Music Education* 141: 66–71.

Juslin, 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.

König, Stefan. 2020. "Max Regers Schreibprozess und seine Entwicklungen in der Wiesbadener Zeit (1890-1898)." *Methodology of Music Research* 12: 107–128.

Lagarias, Jeffrey, James A. Reeds, Margaret Wright, and Paul Wright. 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." In *Orgelschule zur historischen Aufführungspraxis—Teil 2: Orgel und Orgelspiel in der Romantik von Mendelssohn bis Reger und Widor,* 258–314. Stuttgart: Carus.

Lerdahl, Fred, and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music.* Cambridge: The MIT Press.

Lohmann, Ludger. 1995. "Hugo Riemann and the Development of Musical Performance Practice." In *Proceedings of the Göteborg International Organ Academy*, 1994, edited by Hans Davidsson and Sverker Jullander, 251–283. Department of Musicology: Göteborg University.

Lohmann, Ludger. 1998. *Max Reger: Organ Works, Volume 2*. Naxos 8.553927, compact disc.

Matravers, Derek. 2001. *Art and Emotion*. Oxford: Oxford University Press.

Mazzola, Guerino. 1990. *Geometrie der Töne: Elemente der Mathematischen Musiktheorie*. Basel: Birkhäuser.

Mazzola, Guerino, Gérand Milmeister, Karim Morsy, and Florian Thalmann. 2008. "Functors for Music: The Rubato Composer System." In *Transdisciplinary Digital Art. Sound, Vision and the New Screen. Communications in Computer and Information Science*, edited by Randi Adams, Steve Gibson, and Stefan Arisona, 238–254. Berlin: Springer–Verlag.

McKinney, Martin, Dirk Moelants, Matthew Davies, and Anssi Klapuri. 2007. "Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms." *Journal of New Music Research* 36(1): 1–16.

Milligan, Keziah, and Nicholas Bailey. 2015. "A Review of Software for Note Onset Detection." In *Proceedings of the 4th International Scientific Meeting for Sound and Musical Instrument Studies*, Tavira, Portugal.

Morgan, Robert P. 1978. "The Theory and Analysis of Tonal Rhythm." *The Musical Quarterly* 64(4): 435–473.

Mozart, Wolfgang Amadeus. 1866. *The Letters of Wolfgang Amadeus Mozart (1769–1791), Vol. 1*. New York: Hurd and Houghton.

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.

OrganAssist. 2021. Organ MIDI Made Simple. https://www.organassist.com/, last accessed April 16, 2021.

Palmer, Caroline. 1996. "On the Assignment of Structure in Music Performance." *Music Perception* 14: 23–56.

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 Choralvorspiele op. 135a - Teil 1 (Nr. 1–6)*. YouTube video, 9:07. Posted September 2, 2016. https://www.youtube.com/watch?v=KKyZ7fcWjbc

Reger, Max. 1894. *Vorspiel "Komm, süßer Tod!"* WoO IV/3. Autograph from the archive of Max-Reger Institut in Karlsruhe (MRI).

Reger, Max. 1899. "*Phantasie für Orgel über den Choral " Freu' dich sehr, o meine Seele.*" Leipzig: Jos. Aibl Verlag.

Reger, Max. 1904."*Variationen und Fuge über ein Originalthema.*" Leipzig: Lauterbach & Kuhn.

Reger, Max. 1913. "Organ Suite No. 2 in G Minor, Op. 92: IV. Basso Ostinato." Performed by Max Reger. *The Britannic Organ, Vol. 8., CD 1.* OEHMC Classics. OC847, compact disc.

Reger, Max. 1915. "*Dreißig kleine Choralvorspiele Kleine Choralvorspiele zu den Gebräuchlichsten Chorälen für Orgel.*" New York: Peters.

Reger, Max. 1966. "*Sämtliche Werke, Band 18.*" Wiesbaden: Breitkopf und Härtel.

Rehding, Alexander. 2003. *Hugo Riemann and the Birth of Modern Musical Thought (New Perspectives in Music History and Criticism).* Cambridge: Cambridge University Press.

Repp, Bruno. 1992. "Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann's "*Träumerei.*" *Journal of the Acoustical Society of America* 92: 2546–2568.

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.

Repp, Bruno. 1999. *Music, Mind, and Science. XVI. Individual Differences in the Expressive Shaping of a Musical Phrase: The Opening of Chopin's Etude in E major*. Edited by Suk Won Yi. Seoul National University Press.

Riemann, Hugo. 1884. *Musikalische Dynamik und Agogik, Lehrbuch der musikalischen Phrasierung auf Grund einer Revision der Lehre von der musikalischen Metrik und Rhythmik*. Hamburg: D. Rahter.

Riemann, Hugo. 1888. *Katechismus des Klavierspiels*. Leipzig: M. Hesse.

Riemann, Hugo. 1900. "Über Agogik." In *Präludien und Studien: gesammelte Aufsätze zur Ästhetik, Theorie und Geschichte der Musik, Band II,* 88–97. 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 Carl Fuchs. 1890. *Katechismus der Phrasierung*: *praktische Anleitung zum Phrasieren.* Leipzig: Hesse & Becker.

Rumsey, David. 2005. *Organists on a Roll—the Welte Organ's Mechanically-recorded Performances*. Göteborg: GOART.

Rumsey, David. 2014. Foreword in *The Britannic Organ, Vol. 8., CD 1*. OEHMC Classics. OC847, compact disc.

Sander, Martin. 1994. *Max Reger: Die Großen Orgelwerke Gespielt auf Historischen Originalinstrumenten, Volume 2.* Motette CD 12001, compact disc.

Sander, Natalia. 2006. *Max Reger's Organ Works in Aspect of its Biographical, Instrumental and Interpretational Questions*. Ph. D. diss., M. Glinka Nizhny Novgorod State Conservatory (In Russian).

Shaffer, Henry L. 1980. "Analysing Piano Performance. A Study of Concert Pianists." In *Tutorials in Motor Behavior*, edited by G. Stelmach and J. Requin, 443–455. Amsterdam: North-Holland Publishing Company.

Sievers, Gerd. 1967. "Hugo Riemanns Metrik und Phrasierung und Max Regers Werk." In *Die Grundlagen Hugo Riemanns bei Max Reger,* 534–575. Wiesbaden: Breitkopf & Härtel.

Steinhaus, Phillip. 1956. *Phillip Steinhaus Plays Johann Sebastian Bach and Max Reger***.** Boston Records, B 700, LP.

Sundberg, Johan, Anders Askenfelt, and Lars Frydén. 1983. "Musical Performance: A Synthesis-by-Rule Approach." *Computer Music Journal* 7(1): 37–43.

Szabó, Balázs. 2016. *Zur Orgelmusik Max Regers.* Bonn: Dr. Josef Butz Verlag.

Thalben-Ball, George. 1950. "*The President's Address*." The Musical Times 91(1292): 387–389.

The Classical Music Guide Forums. 2018. *A "quip" about composer Max Reger*. https://www.classicalmusicguide.com/viewtopic.php?t=48209#p485274, last accessed May 2021.

Todd, Neil. 1985. "A Model of Expressive Timing in Tonal Music." *Music Perception: An Interdisciplinary Journal* 3(1): 33–57.

Todd, Neil. 1989. "Towards a Cognitive Theory of Expression: The Performance and Perception of Rubato." *Contemporary Music Review* 4(1): 405–416.

Todd, Neil. 1992. " The Dynamics of Dynamics: A Model of Musical Expression." *Journal of The Acoustical Society of America* 91(6): 3540–50.

Vieillard, Sandrine, Roy, Mathieu, and Isabelle Peretz. 2012. "Expressiveness in Musical Emotions." *Psychological Research* 76: 641–653.

Vines, Bradley, Carol Krumhansl, and Daniel Levitin. 2006. "Cross-modal Interactions in the Perception of Musical Performance." *Cognition* 101(1): 80–113.

Widmer, Gerhard, and Werner Göbl. 2004. "Computational Models of Expressive Music Performance: The State of the Art." *Journal of New Music Research* 33(3): 203–216.

Williams, Peter. 1980. *A New History of the Organ from the Greeks to the Present Day.* Bloomington: Indiana University Press.

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.

Wunderlich, Heinz. 1980. *Max Reger. Orgelwerke, Volume 4.* Signum Records CLX0740, compact disc.