Synthetic and Pseudo-Synthetic Music Performances: An Evaluation

Tilo Hähnel and Axel Berndt

Dept. of Simulation and Graphics
Otto von Guericke University
Magdeburg, Germany
(+49-391) 67-1 21 89

{tilo, aberndt}@isg.cs.uni-magdeburg.de

ABSTRACT

Synthetic Baroque timing was evaluated by applying a newly developed concept of *macro* and *micro* timing. Subjects rated three different synthetic performances. The results showed clearly that modeled *macro* and *micro* timing had influenced human listeners' ratings in a positive direction. This paper further includes a study of human prejudice against synthetic performances. We let listeners believe they were rating a completely synthetic performance, which, in fact, was a recording of a human performance. This analysis in particular is of importance regarding the ranking of synthetic performances.

Keywords

Synthetic Performance, Timing, Evaluation

1. INTRODUCTION

In the last decades several performance systems were developed that shape musical expression automatically. Some are based on theoretical models [11], others are based on performance analyses [6,10]. The evaluation of these tools is often limited to a comparison of empirical data derived from human musicians and parameters of modeled performances. Because these tools are not invented to copy a certain individual characteristic, this procedure is not without difficulty. Furthermore, if performance parameters that influence tempo, articulation, and loudness are derived from mean values, then the extreme characteristics, which are important to a marked human performance, get lost. The result will be a flattened characteristic.

Consequently, it is to evaluate the effects of human like performances on listeners. This analysis by synthesis approach is nevertheless limited. One synthetic performance can be compared to another [8] but hardly to a real one. Often, stimuli are simple sequencer sounds or even artificial stimuli like sinus-tones. The conclusiveness of such results is always due to a comparison of different models. But to answer the question if and to what degree a synthetic performance is perceived as real, remains speculative.

Moreover, even if one were to judge synthetic performances of high quality, people may tend to look at those with pre-conceived notions.

This paper demonstrates an evaluation of performance features with a focus on timing parameters. Several timing phenomena, such as phrase arch playing [14,16] or the quadratic shape of final ritardandi [7] were discovered through an analysis of Classic or Romantic music and are therefore inadequate for other styles like Baroque music.

In the study we evaluated highly flexible mathematical models that shape expressive music performances in MIDI data. They were developed subsequently from a number of literature studies and human performance measurements of primarily Baroque music [3].

2. METHOD

2.1 Design

The study was made during the "Long Night of Science" at the Otto-von-Guericke University in Magdeburg in 2009. We tested 42 male and 24 female German participants of different age and with different experiences of Baroque music, as shown in Figure 1. All participants were confronted with three synthetic performances (*"flat"*, *"macro"* and *"micro"*) in a counter-balanced design, comprising the first six bars of Telemann's trumpet concert in D Major TWV 51:D7. These were presented as MIDI data by using the *MuSIG* Engine (see Section 2.2) and high quality samples from the *Vienna Symphonic Library* [15]. The *flat* version contains none of the three expressive features tempo, dynamics, or articulation. The *macro* version included a macro-timing, i.e. a phrase arch performance, similar to the Model of Windsor & Clarke [16], but with respect to the phrase structure of the trumpet concert. In this case the first larger phrase ends on the first beat in the third measure, where the second phrase already starts. This dovetailing function of single notes is typical in Baroque music. The second phrase contains

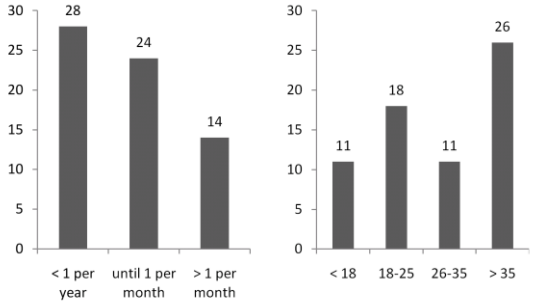

Figure 1: Left = Frequency of listening to Baroque Music. Right = Age of all participants in years.

SysMus10, Cambridge, UK, September 13-15, 2010

the measures three to five and terminates at the first beat in measure six. The *micro* version included metrical accentuations and prolongations of those notes that occurred on the beat as well as little ritardandi to mark the phrase boundaries.
The mean tempo of all three performances was the same (31.5 bpm), but the tempo of the *micro* performance differed from 20 bpm to 32 bpm and from 24 bpm to 34.8 bpm in the *macro* performance.

All participants rated *liveliness*, *expressiveness*, and their *overall impression* by selecting marks from 1 (very good) to 6 (very bad), which correspond to the marks A – F as given in British schools. The Figures include the corresponding letters. The significance of different ratings was computed in a Wilcoxon-test, which is a nonparametric test for two related samples.

Following Hypotheses were tested: It was assumed that both *micro* and *macro* performances were rated better than the *flat* performance in liveliness, expressiveness and overall impression. Since the *macro* performance is rather consistent with a so called "historically informed" performance, we assumed that the listeners' preferences regarding Baroque music might influence the estimation of the *macro* and *micro* performances. We also supposed that the difference in age might have an effect due to the increased experience of elder participants. In addition, all participants were asked to mark the fastest, slowest, and best performances.

In a second study the participants rated a "high-end synthetic performance". What was suggested to be the sounding result of cooperation between several universities that modeled ornamentation, room acoustics and every single instrument separately in a 3d space, was in truth a recording of an ensemble specialised in historically informed performance. In this task the participants were additionally asked to rate the *authenticity* of this *(pseudo-) synthetic* performance. Were the rank "A", the performance would be ranked as being as good as a real performance. Accordingly, any other rank would reflect the prejudice against synthetic performances. Regarding this it was important to ensure that the participants perceive the difference between the synthetic and *(pseudo-) synthetic* performance. Hence this performance was presented immediately after the first task, therefore we expected the real performance to be rated high.

2.2 Performance Synthesis

This section gives a conceptual introduction and overview of our performance synthesis system, the *MuSIG* engine. An in-depth description is provided in [2].

The musical raw data is given in the MIDI format. This flat version contains no tempo information, no dynamics, and all notes remain unarticulated. If tempo or dynamics information are nonetheless present, they are ignored. Instead, such performance information are provided by a separate XML file. Here, multiple performance styles can be defined.

One such performance style comprises all necessary information to render an expressive MIDI sequence. This includes tempo (macro timing), rubato (micro timing), information on the temporal precision and synchrony, dynamics, dynamic ranges for each part to scale the dynamics to what is actually possible on the instrument, schemes for metrical emphasis, articulation style definitions, and the actual articulation instructions.

All these performance features can be classified as *header* (or a priori) information and *temporally fixed* information. An articulation style, for instance, may define the articulation instructions available, hence, header information. But their actual application to articulate certain notes in the score is temporally fixed information. The latter are organised as sequences of performance instructions, called *maps*. Thus, we have tempo maps, rubato maps, dynamics maps, metrical emphasis maps, articulation maps, and so forth.

Furthermore, all performance information can be defined *globally* for all musical parts or *locally*, i.e. part-exclusive. A typical situation in music production is the following. All parts play synchronously, hence, they have one global tempo map. But they differ on the micro level. Each part has its own rubato map with individual instructions.

One further distinction of temporally fixed instructions has to be introduced, the discrimination of *point instructions* and *temporally extensive* instructions. The first class, i.e. point instructions, is defined only at discrete positions within the time domain. The articulation of a single note is an example of this; it applies only to one note at a particular score position. Formally, a point instruction I_i defines a date d_i and the information v_i that has to be applied to the musical material at that position $I_i=(d_i, v_i)$.

Temporally extensive instructions, by contrast, cover an interval greater than 0 in the time domain. They are basically defined as the quadruple $I_i=(d_i, v_{1,i}, v_{2,i}, shape_i)$ and describe a continuous value transition from $v_{1,i}$ to $v_{2,i}$ in the time frame $[d_i, d_{i+1})$ with the characteristic $shape_i$. An example from the dynamics domain: The dynamics instruction $I_0=(0, mf, f, \text{linear})$ defines an initial loudness level (*mezzoforte*) which transitions linearly to *forte* from date $d_0=0$ to date d_1 of the succeeding dynamics instruction I_1.

For the technical implementation of musical terms like piano, mezzoforte, forte, allegro, vivace, andante, legato, tenuto, accentuated, etc., mappings into numerical domains have to be defined. In the MIDI standard, loudness has to be mapped onto integer values in [0, 127]. Tempo instructions can be converted into values of beats per minute (bpm). Articulations change note parameters (duration, loudness, timbre, etc.) which can be expressed numerically. All these mappings can be freely defined in the header information—the loudness of forte, the tempo of allegro, and the description of articulations. Thus, the actual editing of the *v* parameters of the instructions is relatively intuitive and straight forward.

The *shape* term, however, is more complicated. The characteristics of dynamics transitions generally differ from those of tempo transitions. Even the shapes of metrical emphasis schemes feature unique characteristics that cannot be found in other classes of performance features. Each class has its own form for the shape term. The shape characteristics we have implemented are summarised in the following. As this is only a rough overview please refer to [3,4,9] for further details.

Timing

Tempo transitions (ritardando, accelerando) are traditionally modeled by quadratic functions. Our measurements of CD-productions, live recordings, and some specially prepared etudes could partly confirm this. Tempo transitions feature potential characteristics but differ with respect to the curvature. Very determined tempo changes feature a stronger curvature than the more neutral tempo changes which tend to the linear shape. Such differences could also be observed in different musical contexts. Ritardandi and accelerandi that accentuate a particular musical point (e.g., the final chord) are more determined than those just leading over to a different ongoing tempo.

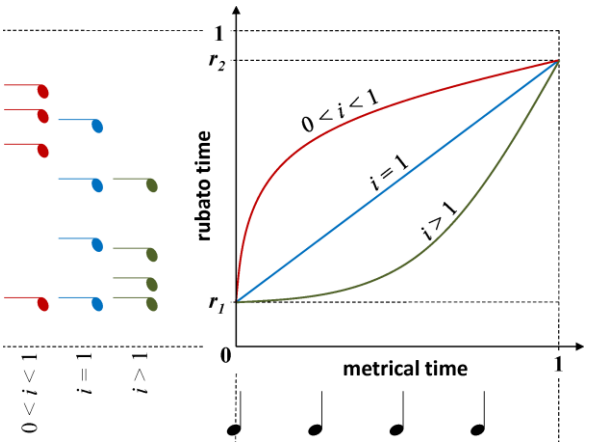

Figure 2: Rubato distortions. Parameter i is the exponent of the potential distortion.

Rubati are small self-compensating timing distortions, also modeled by potential functions in the unity square which represents the time frame to be distorted. Here, they map metrical score position onto rubato position (see Figure 2).

Random imprecision (normal distribution) and constant asynchrony can easily be added after computing the exact millisecond dates of the musical events.

Dynamics

Macro dynamics describes the overall loudness and loudness changes over time. This comprises crescendi and decrescendi. Both are modeled by cubic Bézier curves to create sigmoid characteristics. The straightness (linearity) and tendency (fast change at the beginning or end) of the loudness transition can be controlled by two parameters which are then converted into the four points of the control polygon. Thereby, neutral or more determined loudness changes can be made.

Upon the overall macro loudness micro deviations are added which reflect the metrical order of the musical piece, i.e. its time signature. Basically, the metrical emphasis scheme defines a sequence of emphases at certain points in the measure and transition characteristics (static or linear) in between them. The intensity of accentuation can be scaled, thus, the same emphasis scheme can be applied more unobtrusively or very markedly.

Articulation

Articulation is in part also an aspect of micro dynamics. However, the articulation of a musical note not only changes its loudness, but also its duration, timbre, envelope, and intonation. Loudness and duration changes are directly rendered into the corresponding MIDI events. For timbre, and envelope changes we switch between different instrumental sample sets of the *Vienna Symphonic Library*. These also include some deviations in tuning. Less subtle detuning necessitates additional work with the Pitch Wheel controller which has not yet been used.

After defining the effects of articulation instructions in the articulation style they are ready to be applied in the articulation map. Here, an articulation indicates the note to be articulated and its instruction. Even combinations of instructions, like an accentuated legato, are possible. Furthermore, multiple articulation styles can be created which implement the same instructions differently. Style switches in the articulation map allow changes between them.

Summary

A major design goal was the flexibility of all the formal models for timing, dynamics, and articulation. This tool kit allows a big variety of performance styles including most subtle nuances which makes the *MuSIG* engine a powerful tool to explore variations, for instance in the context of historically informed performance practices, and to explore their effect on the listener.

Table 1: Significance of rating differences between *flat*, *micro*, and *macro* performances.

Aspect	Difference between	p value
Overall impression	flat and macro	0.000
	flat and micro	0.000
	macro and micro	0.022
Expressiveness	flat and macro	0.000
	flat and micro	0.001
	macro and micro	0.011
Liveliness	flat and macro	0.000
	flat and micro	0.000
	macro and micro	0.230

However, even the best performance will be judged synthetic if the sound generation quality is insufficient. To get instrumental sounds of best quality we apply the *Vienna Symphonic Library*, a comprehensive sample library of orchestral instruments. To fully utilise its capabilities the *MuSIG* engine implements a separate playback mode that generates specialised controller messages for the related software sampler *Vienna Instruments*.

3. RESULTS

In general, differences regarding the participants' age and experience with Baroque music were insignificant.

The results of both evaluation procedures are shown as boxplots in Figure 3-6 and in Tables 1 and 2. The *flat* performance shows a wide spread distribution with a median rating of C. Both time-modeled versions were rated between B and C and estimated better than the *flat* version with a significance of p=0.001 or less. In the *micro* timing test the median grade of expressiveness was like the *flat* version a C, too, but the distribution shows a strong tendency towards B. The differences between the *macro* and *micro* versions were less significant but not insignificant at all, except the difference in liveliness, which was insignificant. During the test, some subjects stated that they could not hear any difference between the three versions or specify what the difference was. Others recognised a tempo difference in the *macro* version, the tempo of which was modeled more intensely.

The fewest subjects rated the *flat* performance as the best, whereas the most stated that the *macro* performance had been the best one. Regarding the tempo estimation the result was

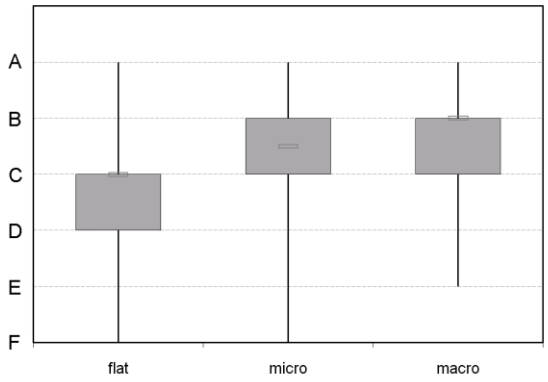

Figure 4: Liveliness ratings.

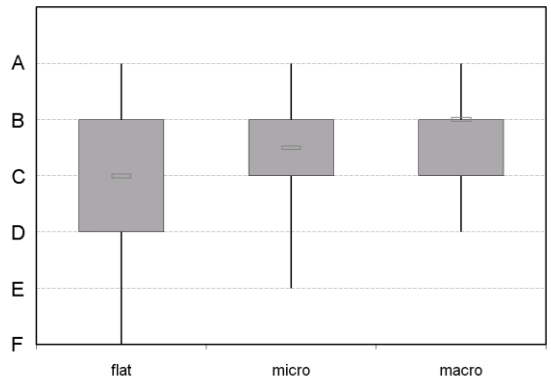

Figure 5: Overall impression ratings.

unclear. Although Figure 7 (right) shows that most participants either did not recognise a difference in mean tempo or perceived the macro version as the fastest, the differences turned out insignificant in a Chi-Square test (see Table 2). Similar results concerning the distribution of statements about the slowest performance were found. Here most participants—if not hearing no difference—suggested that the *flat* version had been the slowest. These differences are only significant at the $\alpha < 0.1$ level and at least not completely caused by chance.

The median rating of the *(pseudo-) synthetic* performance was B in all categories. Differences between categories were only significant between liveliness and overall impression with p=0.050, and liveliness and expressiveness with p=0.067, which is still a weak statistical attribute.

4. GENERAL DISCUSSION

Timing expression in music deals with subtle rubati and micro deviations along with large changes in tempo. Baroque Music in particular requires the former [3,12]. Even if many participants could not name the difference between *flat*, *macro* and *micro*, the subjective rating was better when timing was modeled. Moreover, the concept of micro and macro timing described in Section 2.2 turned out to improve the subjective impression of German listeners significantly.

Because the estimation of the *(pseudo-) synthetic* performance was made in an additional task, the median ranks of both studies cannot be compared directly. Any comparison must be made very carefully, of course, but the second study shows an ample indication that a median rate of "A" is hardly expectable for any synthetic performance. A reverse test-setup, in which participants believe that they are hearing a real performance but are confronted with a synthetic one, is still a very challenging project, for there is still much room for improvement in all facets from acoustics to performance features.

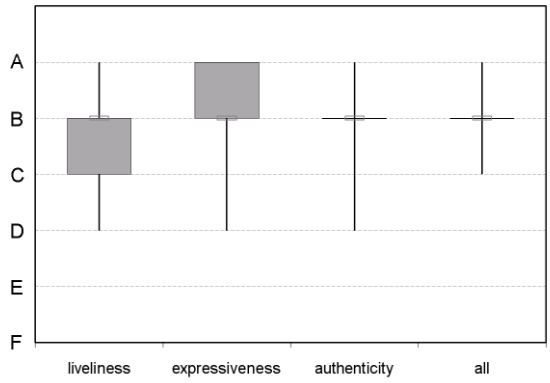

Figure 6: (Pseudo-) synthetic performance ratings.

Interestingly, the estimations about timing quality differed not with respect to the age and preference for Baroque Music. One explanation might be that the amount of experienced listeners was small. Another might be that although there were participants who have an affinity for Baroque Music, the study included neither expert musicians nor other experts in historically informed performances like musicologists. On the other hand crucial topics in historically informed performances are the mean tempo in general, instrumentation, the size of ensembles and articulation. Despite the importance of timing differences, timing itself is a rather subtle element of musical expression.

The loudness changes that paralleled the shape of the tempo structure and articulation decisions were very small. Had they not been added, the result would have sounded more un-balanced than in the *flat* performance, which was consequent in its flatness at least. Of course, future research should focus on all expressive features to the same extent. Then it will be easier to analyse the quality of a complete performance as well as the consequences of any manipulation of single features. However, against this background the results are still significant with respect to timing, because all other expressive features were only slightly adjusted.

Baroque Music is not "Phrase-Arch Music" in the sense of Romantic music [12]. Nevertheless, the adagio used in this study is more compatible to a Romantic interpretation than, for instance, the allegro movements. Since in the *macro* timing version the tempo differences were more obvious than in the *micro* performance, it was easier to notice that something was different. This might explain why the rating of the *macro* was better than that of the *micro* performance.

The tempo ranking between *flat*, *micro* and *macro* was hardly significant. However, the *flat* performance was perceived slower than the *micro* and *macro* performances. Though it is known that the tempo perception depends on expressive features like articulation and loudness [1], in this case a further explanation might be added: Both performances included a ritardando at the end of a phrase. Assuming that those ritardandi are perceived as normal and are therefore not very remarkable, the tempo estimation would focus less on those ritardandi. Consequently, the perceived mean tempo, even if ignoring a single ritardando, indeed increases. Since the *macro* timing version included more and larger ritardandi than the *micro* version, the former is quite likely to be perceived faster than the latter.

Today's synthetic performance systems still have many

Table 2: Chi-Square Test of differences between estimations of fastest, slowest, and best performance.

Estimation	Chi-Square	df	p
fastest	5.810	3	0.126
slowest	6.945	3	0.075
best	14.394	2	0.001

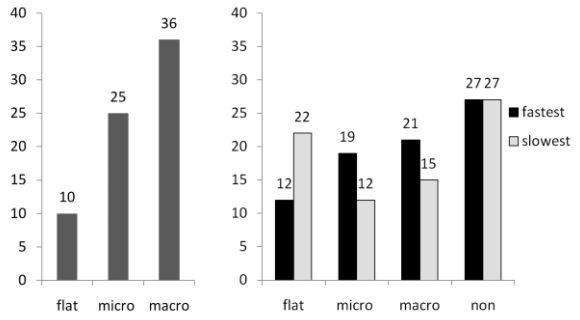

Figure 7: Left = Best performance. Right = Tempo estimation.

drawbacks. However, seen in the light of our observations, the subjective rating of listeners is additionally influenced by the fact that music is not adequate unless it is made by humans. Of course, many subjects were impressed by the quality of the *(pseudo-) synthetic* performance, but they still heard a difference between their imagination of a real performance and the real performance of which they believed it was synthetic.

5. REFERENCES

[1] W. Auhagen & V. Busch, The influence of articulation on listeners' regulation of performed tempo, in R. Kopiez & W. Auhagen, eds, *Controlling creative processes in music*, Peter Lang, Frankfurt a.M., 1998, 69-92.

[2] Berndt, A. Decentralizing Music, Its Performance, and Processing. In *Proc. of the Int. Computer Music Conf. (ICMC) 2010*. New York, Stony Brook, USA, June, 2010, 381-388.

[3] A. Berndt and T. Hähnel, Expressive Musical Timing. In *Proc. of the Audio Mostly 2009: 4th Conf. on Interaction with Sound*, Glasgow, Scotland, Sept., 2009, 9-16.

[4] A. Berndt and T. Hähnel, Modelling Musical Dynamics. In *Proc. of the Audio Mostly 2010: 5th Conf. on Interaction with Sound*, Piteå, Sweden, Sept., 2010.

[5] C. P. Bach, *Versuch über die wahre Art das Clavier zu spielen*. Bärenreiter, 1753-97. Faksimile-Reprint (1994) of Part 1 (Berlin, 1753 and Leipzig 1787) and Part 2 (Berlin, 1762 and Leipzig 1797).

[6] A. Friberg, R. Bresin, and J. Sundberg, Overview of the kth rule system for musical performance, *Advances in Cognitive Psychology*, Special Issue on Music Performance, vol. 2, no. 2-3, 2006, 145-161.

[7] A. Friberg and J. Sundberg, Does music performance allude to locomotion? a model of final ritardandi derived from measurements of stopping runners, *J. Acoust. Soc. Am.*, vol. 105, March 1999, 1469-1484.

[8] A. Gabrielsson, Interplay Between Analysis and Synthesis in Studies of Music Performance and Music Experience, *Music Perception* 3(1),1985, 59-86.

[9] T. Hähnel and A. Berndt, Expressive Articulation for Synthetic Music Performances. In *Proc. of the 2010 Conf. on New Interfaces for Musical Expression (NIME 2010)*, Sydney, Australia, June, 2010, 277-282.

[10] R. Kopiez & W. Auhagen, Preface, in R. Kopiez & W. Auhagen, eds, *Controlling creative processes in music*, Peter Lang, Frankfurt a.M., 1998, VII-IX.

[11] G. Mazzola, S. Göller, and S. Müller, *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance.* Zurich, Switzerland: Birkhäuser Verlag, 2002.

[12] [12] S. Pank, Der Fingersatz als ein bestimmender Faktor für Artikulation und Metrik beim Streichinstrumentenspiel, In *Michaelsteiner Konferenzberichte*, vol. 53, Michaelstein, 1998, 95-103.

[13] J. J. Quantz, *Versuch einer Anweisung die Flöte traversière zu spielen*. Berlin: Bärenreiter, 1752. Faksimilereprint (1997).

[14] N. P. Todd, The dynamics of dynamics: A model of musical expression, *The Journal of the Acoustical Society of America*, vol. 91, no. 6, 1992, 3540-3550.

[15] Vienna Symphonic Library GmbH. *Vienna Symphonic Library*. http://vsl.co.at/ (last visited: March, 2010).

[16] W. L. Windsor and E. F. Clarke, Expressive Timing and Dynamics in Real and Artificial Musical Performance: Using an Algorithm as an Analytical Tool, *Music Perception*, vol. 15, no. 2, 1997, 127-152.