Works in Audio and Music Technology

edited by Axel Berndt

TUDpress
2015

Impressum / Bibliografische Informationen

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

ISBN 978-3-95908-021-7

Verlag der Wissenschaften GmbH
Bergstr. 70 | D-01069 Dresden
Tel.: +49 (351) 47 96 97 20 | Fax: +49 (351) 47 96 08 19
http://www.tudpress.de

Typeset by editor.
Printed in Germany.

Contributors

Nadia Al-Kassab
Leipzig, Germany

Axel Berndt
Detmold, Germany

Tilo Hähnel
Weimar, Germany

Robert Harald Lorenz
Dresden, Germany

Maxim Müller
Dresden, Germany

Felix Schönfeld
Dresden, Germany

Hendrik Schuster
Dresden, Germany

Reviewers

Georg Essl
University of Michigan, Ann Arbor, MI USA

Mark Grimshaw
Aalborg University, Aalborg, Denmark

Daniel Hug
University of Applied Sciences and Arts Northwestern Switzerland, Brugg-Windisch, Switzerland
Zurich University of the Arts, Zurich, Switzerland

Stefania Serafin
Aalborg University Copenhagen, Denmark

Proofreaders

Anne-Katrin Helmuth and Antje Schuster
Dresden, Germany

Contents

Robert Harald Lorenz, Hendrik Schuster

Nadia Al-Kassab, Axel Berndt

Axel Berndt

Chapter 4

Formalizing Expressive Music Performance Phenomena

Abstract What is the curve shape of a ritardando? How loud is forte and how does a decrescendo transition from forte into piano? What effect does a staccato articulation have on a tone? These are typical questions arising during analyzing and creating expressive music performances on the computer. We addressed these questions in a research project, performed detailed analyses of human performances and developed a series of mathematical models to emulate the characteristics that we observed. This chapter summarizes our analysis approach and compiles the mathematical models that we derived from the analyses. This comprizes a variety of macro and micro phenomena from the three domains timing, dynamics and articulation. A flexible model design makes it a versatile toolbox, esp. for music production, applicable to a variety of different musical styles.

4.1 Introduction

The expressive performance of music transforms a musical raw material (symbolic formats like music notation, MusicXML or MIDI) into a sounding output or an equivalent representation (i.e., audio data). A direct "one-to-one" mapping can be experienced by generating output directly from notated information where each note is precisely performed with its standardized pitch, duration, loudness, and timing. Though this may already be an expressive performance as it transforms the raw material into audio output, but in comparison to human musicians' performances this particular mapping appears to be a statistically insignificant special case.

A central aim of performance research in musicology as well as in the field of computer music is the identification and characterization of expressive performances. Another central aim is the investigation of typical applications in their application in specific musical contexts. The latter plays an intrinsic role for computer-generated performances. We denominate this *performance planning*, as it models the decision process that maps a musical situation to a combination of performance features, just like the musician who analyzes a musical piece. The musician then decides how he is going to play it, writes down an abstract performance plan by means of tempo, dynamics and articulation instructions. The rendering of a performance plan into an audible output or an equivalent representation is called *performance synthesis*. The goal of performance synthesis is to find formal models that adequately describe the performance phenomena. Typical questions are: How, precisely, is a ritardando played? What effect does a certain articulation have on a note? This is the toolkit for the work of performance planning.

This chapter is concerned with the latter field, performance synthesis, and summarizes respective results of an interdisciplinary research project of computer scientists and musicologists, who investigated the phenomena of expressive performance and developed mathematical models for their description. The technical implementation of these models has applications in music production, interactive media scoring, and interactive psychoacoustic studies. The models can also serve in performance analysis as they help abstracting from measurements (such as timing and loudness envelope measurements) and capturing meaningful features. Thereby we discovered, for instance, differences in the performance of continuous loudness transitions (crescendo, decrescendo) depending on the musical context. The following section gives an introduction to the performance features and their analysis—a comprehensive presentation is given by Hähnel (2013). Based on the findings, we detail and discuss the developed models, give remarks on their implementation and point out future directions of this research. Parts of this work have been previously presented by Berndt & Hähnel (2009, 2010), Hähnel & Berndt (2010), and Berndt (2011*a*,*b*).

4.2 Performance Features and Analyses

A musical score denotes a musical idea (or, which is more problematic, an idea of someone making a transcription). The performer analyzing the score derives a new musical idea, which can differ completely from the composer's

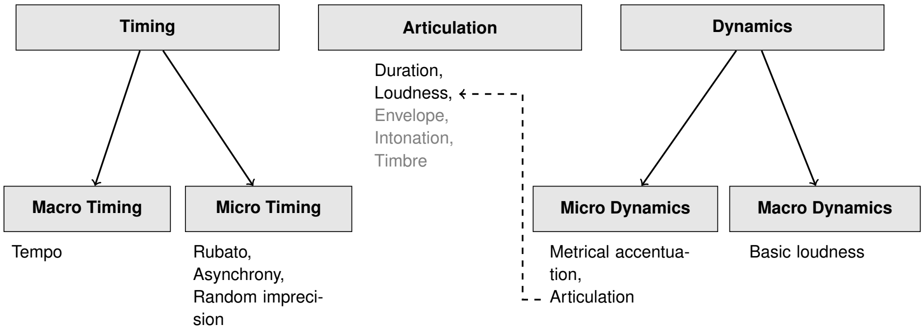

Figure 4.1: An overview of the performance feature classes described in this chapter. Envelope, intonation, and timbre are not included, thus, colored light gray.

idea. Expression in music performance aims at delivering the performer's idea of a musical score to the listener, who finally derives a new individual idea based on music perception and listening analysis. According to the modified lens model of Juslin (2009), music consists of acoustical cues, with which composer and performer want to achieve a certain kind of impression on the listener. The composer uses structural elements, like mode, pitch, melodic progression, and rhythm (which are in a strict sense rhythm classes, as Bengtsson (1974) wrote). The performer applies complex transformations, consciously and subconsciously. These transformations introduce a variety of details into the performance mainly with respect to the "three pillars" of expressive performance: timing, dynamics, and articulation. When developing models for phenomena in these domains it is not enough to know, e.g., that a ritardando slows down the tempo, that a crescendo raises the loudness, and that the time signature is somehow accentuated. We need to know precisely how these features are shaped. Which characteristics are appropriate and which degrees of freedom do they have? The succeeding section gives a brief introduction of the feature spaces discussed in this chapter, see Figure 4.1 for an overview. Following this, we detail our analysis methodology including the related work from which we then derive our modelling approach.

4.2.1 Feature Spaces

Tempo is the most deeply explored class of timing features. Tempo defines the musical meter in the format "number of symbolic time units per physical time unit", usually "beats per minute" (bpm). The tempo can remain con-

create highly polyphonic performance structures. A typical situation in a solo concert would be modelled as follows. Global maps are defined for the accompaniment, whereas, for the solo instrument which has its own more detailed musical structure demanding a more detailed performance, local maps are created.

Our performance engine further implements techniques to transition seamlessly between different performances. This allows us to combine them and create new performances interactively during realtime playback and opens up for live situations, like interactive media scoring (Berndt 2009, Berndt, Dachselt & Groh 2012), music-psychological studies involving a certain amount of interactivity, and analysis-by-synthesis studies (see chapter 5 for an example).

4.7 General Discussion & Future Directions

During the development of the models, we paid particular attention to make them easy to customize and easy to combine. The models support a big variety of characteristics covering those observed in the measurements of human musicians as well as atypical and extreme degeneracies, e.g. very strong micro timing distortions that hardly occur in human performances. Most models allow the definition of new instances, like accentuation schemes, dynamics instructions, and articulations. This flexibility opens up the whole bandwidth of performance styles for music production and pays off in study applications, especially when the study should not only cover typical characteristics.

In order to properly combine different models it is mandatory to prevent mutual interference. A micro timing distortion (rubato) affects local tempo fluctuations, but the average tempo remains unchanged. The same applies to metrical accentuation, articulation, and basic loudness. An articulation may change the loudness of a note, but the basic loudness and metrical accentuation are preserved. Our model design enables a change of one feature without affecting the other feature, even if they are interweaving on the same domain.

For the evaluation of which extent our models are able to describe plausible human performances (in Western music culture), we carried out a listener study with 46 participants. 23 of them had an expert background in historically informed music performance. The study consisted of two runs. In the first run, the listeners had to decide whether the musical samples they heard were real (human) or synthetic (computer) performances and grade it (1: human performance, 2: probably human, 3: probably synthetic, 4: synthetic). In the second run, all samples were synthetic performances and the listeners had to decide whether these were copies of human performances (1: yes, 2: probably yes, 3: probably no, 4: no).

All samples were Baroque pieces, mostly minuets. The 1:1 copies were detailed manual reproductions of real performances that correspond with them regarding timing (similar inter-onset intervals), dynamics, and articulation (transcribed from close-listening analyses and readjusted with respect to the used instrumental sample set). The plausible versions were replications of these performances using our models without random imprecision. The articulation-only performances excluded the work with timing and dynamics features, and kept only the constant asynchrony and articulations of the plausible versions. The extreme versions use the same characteristics as the plausible versions, but set extreme shapes (exaggerated rubati and articulations etc.). In the inverted versions the tempo, macro dynamics, rubato, and metrical accentuation characteristics are inverse to the 1:1 copies. Both, the extreme and inverted versions were untypical performances serving as control stimuli.

All versions worked with high-quality samples from the *Vienna Symphonic Library* (Vienna Symphonic Library GmbH 2015) and applied reverberation and stereo panning based on the human recordings. As the differences in sound generation and acoustics, nonetheless, most likely biased the results of the first part of the study, we excluded the human recordings in the second part. The 1:1 copies served as reference performances.

Table 4.4 shows a summary of the listener's ratings. A significance analysis of the rating differences reveals that, in the first test, the human recordings were rated significantly better than all other versions by both, expert and non-expert listeners. The mean rankings of the plausible (model-based) versions and the articulation-only versions were not significantly different from the 1:1 copies indicating that an elaborate articulation strongly contributes to the impression of plausibility. Significantly different ratings occurred between the plausible versions and the control group of extreme and inverse versions in both tests—the control groups were rated significantly worse. The models prove to lie in level with the reference versions. This suggests that the models are able to plausibly emulate typical characteristics of human performances.

The surprisingly good ratings of the articulation-only versions also have to be seen in the stylistic context of our stimuli, i.e. all Baroque music. Flat macro features and an intense work with micro features is close to the performance concepts of music of this style, confers the study of Shaffer (1981). With music

Type of performance	**23 non-expert ratings**		**23 expert ratings**	
	mean value	standard dev.	mean value	standard dev.
“ The current recording is real.”				
human recordings	1.70	0.470	1.35	0.487
1:1 copies	2.61	0.656	3.65	0.573
plausible models	2.52	0.665	3.35	0.775
inverted & extreme versions	2.35	0.775	3.26	0.915
“This synthetic version is a copy of a real performance.”				
1:1 copies	2.22	0.751	2.44	0.609
inverted versions	2.76	0.705	2.96	0.767
plausible models	2.26	0.619	2.48	0.846
extreme versions	2.94	0.460	3.11	0.656
articulation only	2.22	0.736	1.87	0.757

Table 4.4: Listeners’ ratings in the study.

from the Romantic period, we would expect to gain a rather significant difference to the 1:1 copies.

The models may already cover a big variety of performance features, but in the measurements we still observe a lot of details that cannot be captured yet. Random processes might be used to emulate what cannot even be derived from systematic reasons. There are also a lot of performance aspects not covered yet, such as fingering and bowing (Al Kasimi, Nichols & Raphael 2007, Hall & O'Donnell 2009), or the communication within ensemble play. The interaction between the players functions like a homeostatic system. Little timing variances of one player affect neighboring players, who gradually adapt to it. The player himself also tries to restore synchrony with the ensemble. Similar effects exist for dynamics and articulation. Their investigation and formalization still stands out and would be a valuable extension.

Although our models introduce a certain level of abstraction and offer a fairly intuitive parameterization, the creation of expressive performances is, nonetheless, a laborious process. The models constitute a comprehensive toolbox. However, the creation of a performance plan is still up to the user. This work can be supported by algorithmic methods, not necessarily like a fully autonomous virtual musician, but as music production tools which create initial performances that can be refined by human musicians. Building such a framework around the models described here is a promising future perspective of this work.

4.8 Summary

The expressive performance of music is not a simple one-to-one transformation of notated information into sounding output. Forms of printed music are known to be ambiguous and incomplete, which gives freedom to musicians when designing their own interpretations. However, studies have shown that musical culture has established certain design patterns. We traced the establishment of these patterns in musicological and theoretical literature (mainly on Western music), and found clues about their characteristics. This was complemented by analyses of audio recordings, including professional CD productions, life recordings at an international competition, and carefully prepared exercises. We gained valuable insights, not only into the basic characteristics and shapes of the phenomena, but also into their mutability. Based on these insights, we were able to develop a series of dedicated mathematical models. We see these models as a versatile and flexible toolkit for the creation of expressive performances with—

but not limited to—human-like impression. Potential applications can be found in music production, expressive performance generation, realtime adaptive music performance, analysis-by-synthesis studies, and as a musicological tool to emulate historically informed performance practices.

Acknowledgement The author would like to express thanks to all musicians who participated in the recordings, to all participants of the studies, to the team of the "Zentrum für Telemann-Pflege und -Forschung Magdeburg", in particular Carsten Lange, for the cooperation, inspiring discussions, and the chance to make recordings at the *5th International Telemann Competition for Historical Woodwinds* in 2009. My sincere thanks are also dedicated to Tilo Hähnel who cooperated with me in this project, conducted the studies and analyses, and, thereby, provided the invaluable basis for the iterative revision process in the model development.

Bibliography

Al Kasimi, A., Nichols, E. & Raphael, C. (2007), A Simple Algorithm for Automatic Generation of Polyphonic Piano Fingerings, *in* '10th Int. Society for Music Information Retrieval Conf. (ISMIR)', Austrian Computer Society (OCG), Vienna Univerity of Technology, Vienna, Austria, pp. 131–132.

Bello, J., Duxbury, C., Davies, M. & Sandler, M. (2004), 'On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain', *IEEE Signal Processing Letters* **11**(6), 553–556.

Bengtsson, I. (1974), Empirische Rhythmusforschung in Uppsala, *in* 'Hamburger Jahrbuch für Musikwissenschaft', Lang, Frankfurt a.M., pp. 195–220.

Berndt, A. (2009), Musical Nonlinearity in Interactive Narrative Environments, *in* G. Scavone, V. Verfaille & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 355–358.

Berndt, A. (2011*a*), Musical Tempo Curves, *in* M. Clarke, M. Adkins & B. Isaacs, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, University of Huddersfield, Huddersfield, UK, pp. 118–121.

Berndt, A. (2011*b*), *Musik für interaktive Medien: Arrangement- und Interpretationstechniken*, Verlag Dr. Hut, Munich, Germany.

Berndt, A., Dachselt, R. & Groh, R. (2012), A Survey of Variation Techniques for Repetitive Games Music, *in* 'Audio Mostly 2012: 6th Conf. on Interaction with Sound—Sound and Future Vision', Ionian Academy, Interactive Institute/Sonic Studio Piteå, ACM SIGCHI, Corfu, Greece, pp. 61–67.

Berndt, A. & Hähnel, T. (2009), Expressive Musical Timing, *in* 'Audio Mostly 2009: 4th Conf. on Interaction with Sound—Sound and Emotion', Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå, Glasgow, Scotland, pp. 9–16.

Berndt, A. & Hähnel, T. (2010), Modelling Musical Dynamics, *in* 'Audio Mostly 2010: 5th Conf. on Interaction with Sound—Sound and Design', Interactive Institute/Sonic Studio Piteå, ACM, Piteå, Sweden, pp. 134–141.

Bernhard, C. (1667/1999), Von der Singekunst oder -manier, *in* J. Müller-Blattau, ed., 'Die Kompositionslehre Heinrich Schützens in der Fassung seines Schülers Christoph Bernhard', 3rd edn, Bärenreiter. original publication 1667.

Bresin, R. (2001), Articulation Rules for Automatic Music Performance, *in* R. Dannenberg, ed., 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, La Habana, Cuba, pp. 294–297.

Clarke, E. F. (1999), Rhythm and Timing in Music, *in* D. Deutsch, ed., 'The Psychology of Music', 2nd edn, Academic Press, Elsevier, San Diego, CA, USA, chapter 13, pp. 473–500.

Dannenberg, R. B. (2006), The Interpretation of MIDI Velocity, *in* 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Tulane University, New Orleans, USA, pp. 193–196.

Dannenberg, R. B., Pellerin, H. & Derenyi, I. (1998), A Study of Trumpet Envelopes, *in* 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, San Francisco, USA, pp. 57–61.

de Poli, G. (2004), 'Methodologies for Expressiveness Modelling of and for Music Performance', *Journal of New Music Research* **33**(3), 189–202.

Dixon, S. (2000), Extraction of Musical Performance Parameters from Audio Data, *in* 'Proc. IEEE Pacific-Rim Conf. on Multimedia', IEEE Circuits and Systems Society, IEEE Signal Processing Society, and The Signal & Multimedia Processing Lab, Sydney, Australia, pp. 42–45.

Dixon, S. (2006), Onset Detection Revisited, *in* 'Proc. of the 9th Conf. on Digital Audio Effects (DAFx'06)', CIRMMT, Schulich School of Music/McGill University, Montreal, Canada, pp. 133–137.

Friberg, A. (1995), 'Matching the rule parameters of PHRASE ARCH to performances of "Träumerei": a preliminary study', *Quarterly Progress and Status Report* **36**(2–3), 63–70.

Friberg, A., Bresin, R. & Sundberg, J. (2006), 'Overview of the KTH rule system for musical performance', *Advances in Cognitive Psychology, Special Issue on Music Performance* **2**(2-3), 145–161.

Friberg, A. & Sundberg, J. (1999), 'Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners', *The Journal of the Acoustical Society of America* **105**(3), 1469–1484.

Fyk, J. (1995), *Melodic Intonation, Psychoacoustics, and the Violin*, Organon, Zielona Góra, Poland.

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

Gabrielsson, A., Bengtsson, I. & Gabrielsson, B. (1983), 'Performance of musical rhythm in 3/4 and 6/8 meter', *Scandinavian Journal of Psychology* **24**(1), 193–213.

Gatty, R. (1912), 'Tempo Rubato', *The Musical Times* **53**(829), 160–162.

Goebl, W. (2001), 'Melody lead in piano performance: Expressive device or artifact?', *Journal of the Acoustical Society of America* **110**(1), 563–572.

Goebl, W. (2011), Die ungleichzeitige Gleichzeitigkeit des Spiels: Tempo rubato in Magaloff's Chopin und andere Asynchronizitäten, *in* H. von Loesch & S. Weinzierl, eds, 'Gemessene Interpretation: Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen', TU Berlin, Schott Music, Mainz, Germany, chapter 10, pp. 145–156.

Hall, C. V. & O'Donnell, J. T. (2009), Using Regular Expressions to Express Bowing Patterns for String Players, *in* G. Scavone, V. Verfaille & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 37–40.

Hefling, S. E. (1993), *Rhythmic Alteration in Seventeenth- and Eighteenth-Century Music. Notes Inégales and Overdotting*, Schirmer Books, New York, NY.

Hähnel, T. (2013), *Baroque performance—a research study on characteristic parameters of 18th Century music*, Vol. 2 of *Studies in Cognitive Musicology*, Electronic Publishing Osnabrück, Osnabrück, Germany.

Hähnel, T. & Berndt, A. (2010), Expressive Articulation for Synthetic Music Performances, *in* 'Proc. of New Interfaces for Musical Expression (NIME) 2010', University of Technology Sydney, Sydney, Australia, pp. 277–282.

Honing, H. (2003), 'The Final Retard: On Music, Motion, and Kinematic Models', *Computer Music Journal* **27**(3), 66–72.

Hudson, R. (1994), *Stolen Time, The History of Tempo Rubato*, Oxford University Press.

Jerkert, J. (2003), Measurements and models of musical articulation, Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

Juslin, P. N. (2009), Emotion in music performance, *in* S. Hallam, I. Cross & M. Thaut, eds, 'The Oxford Handbook of Music Psychology', Oxford University Press, Oxford, chapter 35, pp. 377–389.

Juslin, P. N., Friberg, A. & Bresin, R. (2002), 'Toward a computational model of expression in music performance: the GERM model', *Musicae Scientiea* **5**(1), 63–122.

Madison, G. (2000), On the nature of variability in isochronous serial interval production, *in* P. Desain & L. Windsor, eds, 'Rhythm Perception and Production', Studies on New Music Research, Swets & Zeitlinger, Lisse, The Netherlands, pp. 95–113.

Maempel, H.-J. (2011), Musikaufnahmen als Datenquelle der Interpretationsanalyse, *in* H. von Loesch & S. Weinzierl, eds, 'Gemessene Interpretation: Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen', Schott, Mainz, Germany, chapter 11, pp. 157–171.

Maestre, E. & Gómez, E. (2005), Automatic Characterization of Dynamics and Articulation of Expressive Monophonic Recordings, *in* 'Proc. of the 118th Audio Engineering Society Convention', Audio Engineering Society (AES), Barcelona, Spain.

Parncutt, R. (2003), Accents and expression in piano performance, *in* K. W. Niemöller, ed., 'Perspektiven und Methoden einer Systematischen Musikwissenschaft', Peter Lang, Frankfurt/Main, Germany, pp. 163–185. Festschrift Fricke.

Quantz, J. J. (1752), *Versuch einer Anweisung, die Flöte traversière zu spielen*, Bärenreiter. Reprint (1997), H. Augsbach.

Ramirez, R., Hazan, A. & Maestre, E. (2005), Intra-Note Features Prediction Model for Jazz Saxophone Performance, *in* 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Universitat Pompeu Fabra, Barcelona, Spain.

Rasch, R. A. (1979), 'Synchronisation in Performed Ensemble Music', *Acustica* **43**, 121–131.

Repp, B. H. (1992), 'Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann's "Träumerei"', *Journal of the Acoustical Society of America* **92**(5), 2546–2568.

Riemann, H. (1884), *Musikalische Dynamik und Agogik: Lehrbuch der musikalischen Phrasirung auf Grund einer Revision der Lehre von der musikalischen Metrik und Rhythmik*, D. Rahter, Hamburg, Germany.

Shaffer, H. (1981), 'Performances of Chopin, Bach and Bartok: Studies in Motor Programming', *Cognitive Psychology* **13**, 326–376.

Sundberg, J. & Verrillo, V. (1980), 'On the Anatomy of the Retard: A Study of Timing in Music', *The Journal of the Acoustical Society of America* **68**(3), 772–779.

Todd, N. P. M. (1995), 'The Kinematics of Musical Expression', *The Journal of the Acoustical Society of America* **97**(3), 1940–1949.

Vienna Symphonic Library GmbH (2015), 'Vienna Symphonic Library', https://vsl.co.at/de [last visited: May 2015]. Sample library of orchestral instruments.

Yamada, M. (1998), Temporal fluctuation in musical performances: Fluctuations caused by the limitation of performers' controllability and by artistic expression, *in* S. W. Yi, ed., 'Proc. of the 5th Int. Conf. on Music Perception and Cognition', Western Music Research Institute, Seoul National University, Seoul, South Korea, pp. 353–358.