

## Sound power and timbre as cues for the dynamic strength of orchestral instruments

Stefan Weinzierl, Steffen Lepa, Frank Schultz, Erik Detzner, Henrik von Coler, and Gottfried Behler

Citation: *The Journal of the Acoustical Society of America* **144**, 1347 (2018); doi: 10.1121/1.5053113

View online: <https://doi.org/10.1121/1.5053113>

View Table of Contents: <https://asa.scitation.org/toc/jas/144/3>

Published by the [Acoustical Society of America](#)

---

### ARTICLES YOU MAY BE INTERESTED IN

[A measuring instrument for the auditory perception of rooms: The Room Acoustical Quality Inventory \(RAQI\)](#)  
The Journal of the Acoustical Society of America **144**, 1245 (2018); <https://doi.org/10.1121/1.5051453>

[Assimilation of mobile phone measurements for noise mapping of a neighborhood](#)  
The Journal of the Acoustical Society of America **144**, 1279 (2018); <https://doi.org/10.1121/1.5052173>

[Structural contributions to phantom partial generation in the piano](#)  
The Journal of the Acoustical Society of America **144**, 1564 (2018); <https://doi.org/10.1121/1.5055225>

[Prediction of three articulatory categories in vocal sound imitations using models for auditory receptive fields](#)  
The Journal of the Acoustical Society of America **144**, 1467 (2018); <https://doi.org/10.1121/1.5052438>

[Lombard effect, ambient noise, and willingness to spend time and money in a restaurant](#)  
The Journal of the Acoustical Society of America **144**, EL209 (2018); <https://doi.org/10.1121/1.5055018>

[Exact solution for the acoustical impulse response of a line source above an absorbing plane](#)  
The Journal of the Acoustical Society of America **144**, 1539 (2018); <https://doi.org/10.1121/1.5054301>

---



CAPTURE WHAT'S POSSIBLE  
WITH OUR NEW PUBLISHING ACADEMY RESOURCES

Learn more 



# Sound power and timbre as cues for the dynamic strength of orchestral instruments

Stefan Weinzierl,<sup>1,a)</sup> Steffen Lepa,<sup>1</sup> Frank Schultz,<sup>1</sup> Erik Detzner,<sup>1</sup> Henrik von Coler,<sup>1</sup> and Gottfried Behler<sup>2</sup>

<sup>1</sup>Audio Communication Group, TU Berlin, Einsteinufer 17c, D-10587 Berlin, Germany

<sup>2</sup>Institute for Technical Acoustics, RWTH Aachen, Kopernikusstr. 5, D-52074 Aachen, Germany

(Received 2 March 2018; revised 19 August 2018; accepted 20 August 2018; published online 12 September 2018)

In a series of measurements, the sound power of 40 musical instruments, including all standard modern orchestral instruments, as well as some of their historic precursors from the classical and the baroque epoch, was determined using the enveloping surface method with a 32-channel spherical microphone array according to ISO 3745. Single notes were recorded at the extremes of the dynamic range (*pp* and *ff*) over the entire pitch range. In a subsequent audio content analysis, audio features were determined for all 3482 single notes using the timbre toolbox. In order to analyze the relative contributions of timbre- and amplitude-related properties to the expression of musical dynamics in different instruments, Bayesian linear discriminant analysis and generalized linear mixed modelling were employed to determine those audio features discriminating best between extremes of dynamics both within and across instruments. The results from these measurements and statistical analyses thus deliver a comprehensive picture of the acoustical manifestation of “musical dynamics” with respect to sound power and timbre for all standard orchestral instruments. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/1.5053113>

[AM]

Pages: 1347–1355

## I. INTRODUCTION

The sound power provides elementary information about the strength and dynamic range that can be produced by individual musical instruments. These data are important, for example, in predicting the sound impact in musical performance venues as a result of source power, stage design and auditorium acoustics. In musical performance studies, the sound power, in combination with other acoustical features of the source signal, can be considered an acoustical manifestation of the expressive potential of each instrument. For the study of musical performance practice, it is of fundamental interest to what extent the sound power and the spectral properties of musical instruments have changed as a result of the historical development of their design and how this might affect, for instance, the overall balance of orchestral groups. Future applications of this knowledge will arise through the implementation of virtual acoustic environments, where an appropriate calibration of acoustic scenes will only be able to be reached based on knowledge of the sound power and the directivity of each individual source.

When dealing with the “dynamics” of music or musical instruments, one should be aware of the fact that in a *musical* context, dynamics is used in terms of the intended or perceived sound strength, i.e., an absolute value indicated in the score by marks usually ranging from *pianissimo* (*pp*) to *fortissimo* (*ff*), whereas in a *technical* context, dynamics is normally used to reference the available amplitude range which can, for example, be given by the ratio of maximum to

minimum amplitudes available in a certain channel of communication. In order to avoid confusion, we will use the terms “dynamic strength” in a musical context, and “dynamic range” for the technical domain.

There are different methods to determine the sound power of musical instruments. In principle, the radiated sound power can be numerically simulated if a complete model of all constitutive parts of the instrument and their coupling is available (Chaigne *et al.*, 2004). However, the resulting acoustical efficiency of the system has to be referenced to a normalized excitation force rather than a human force with its complex interaction between instrument and musician. The same is true for sequential measurements of sound intensity (Lai and Burgess, 1990; García-Mayén and Santillán, 2011), from which the sound power can be determined according to ISO 9614-2 (1996), but which require a reproducible excitation. Hence, for an ecologically valid measurement with professional musicians, there remain the classical approaches for “single-shot” sound power measurements, i.e., the reverberation chamber method and the enveloping surface method according to ISO 3741 (2010) and ISO 3745 (2012).

Since a reliable determination of the sound power of acoustic instruments depending on the intended dynamic strength and the pitch of the notes played thus requires quite a large experimental effort, only limited data are available so far. Earlier studies on the power and the dynamic range of musical instruments mostly relied on comparative measurements of sound pressure values (Sivian *et al.*, 1931). The first comprehensive series of direct measurements of sound power for all standard orchestral instruments according to

<sup>a)</sup>Electronic mail: stefan.weinzierl@tu-berlin.de

the reverberation chamber method was performed by Meyer and Angster (1983). The data were later combined with earlier measurements of the sound pressure and sound intensity of musical instruments (Clarke and Luce, 1965; Burghauer and Spelda, 1971), which were transformed to sound power values based on assumptions about the acoustical conditions of the room, the recording distance, and the directivity of the sound source (Meyer, 1990). The results are given in the classic reference book *Acoustics and the Performance of Music* (Meyer, 2009). They include the recording of scales over two octaves and of selected single notes played at *pp* and *ff* in order to quantify the dynamic range of all standard orchestral instruments. In order to specify a single value for the dynamic range, Meyer selected the *pp* of the softest and the *ff* of the loudest note.

The acoustical expression as well as the perception of dynamic strength of musical instruments is, however, only partly related to their absolute sound power. This was already demonstrated by experiments where listeners were able to identify the intended dynamic strength produced by musicians, largely independently of the actual sound level (Nakamura, 1987). Accordingly, there must be other *perceptual cues* that encode a musician's expression of dynamic strength. By recording instrumental sounds at different pitches and intended dynamic strengths, and analysing the influence of the factors pitch, timbre, and loudness on the perceived musical dynamics in a full factorial design, Fabiani and Friberg (2011) could show that loudness and timbre have a similar impact on the *perceived* dynamic strength, while pitch seems to exert only a comparatively minor influence. With a limited sample of only five musical instruments, however, these authors were not able to investigate which features of the acoustical signal actually provided the expressive cues of dynamic strength. Meyer (1993, p. 204 and 2009, p. 35ff.) suggested using the decreasing difference in level between the strongest partials and those with a frequency of about 3000 Hz as an indicator for dynamic strength, without analyzing the validity of this hypothesis systematically. Hence, apart from a descriptive analysis of the sound power and the timbral properties of all standard orchestral instruments, the present study will analyse for which specific acoustical cues the dynamic strength, as expressed by professional musicians, becomes manifest.

As an empirical basis for these analyses, the study generated a comprehensive database of musical instrument recordings using the enveloping surface method. For 40 musical instruments, including all standard orchestral instruments of the classical and early romantic period, and different historical construction methods, single notes were recorded at *pp* and *ff* over the complete instrumental range in semitone distance, and scales over two octaves were also recorded. We then analysed the sound power for each instrument, each pitch, and both dynamic levels. With respect to the possible contribution of timbral properties to the expression of dynamic strength and to sound differences between epochs, we used the recorded signals to calculate all audio features available in the timbre toolbox (Peeters *et al.*, 2011). Based on a Bayesian linear discriminant analysis

(LDA, controlling for sound power and pitch), we selected those features that discriminated best between recordings of different dynamic. We then used a general linear mixed model analysis (GLMM) in order to estimate the relative predictive value of sound power and identified spectral features for explaining the intended dynamic strength.

## II. METHODS

### A. Measurement setup and calibration

The sound power measurements were performed using the enveloping surface method according to ISO 3745 (2012), using a quasi-spherical microphone array with a radius of approximately  $r = 2.1$  m, and 32 Sennheiser KE4-211-2 electret microphones with a nearly uniform frequency response from 20 Hz to 20 kHz [cf. Fig. 1(c)], located on the faces of a truncated icosahedron (soccer ball shape). The microphones were held in a framework by 90 lightweight but robust fiberglass rods. The entire setup can be seen in Fig. 2. The requirements defined by ISO 3745 (2012) regarding the measurement conditions for precision method 1 were met for all but a few measurements. With a free volume of  $V = 1070$  m<sup>3</sup> the fully anechoic chamber at TU Berlin has a lower limiting frequency of  $f = 63$  Hz. None of the musical instruments recorded exhibited a characteristic dimension of the sound radiating parts of  $d_0 > r/2 = 1.05$  m. The criterion  $r \geq \lambda/4$  is violated only for a few notes with a pitch below E1, corresponding to a fundamental frequency of 41 Hz at a tuning frequency of 440 Hz for A4. This applies to the lowest notes of the contrabassoon, the bass trombone, and the tuba. The recommended number of microphones was raised from 20 to 32 units in order to allow for a simultaneous acquisition of the directivity in higher spatial resolution (Shabtai *et al.*, 2017).

The frequency responses of the 32 microphones were equalized individually in order to compensate for nonuniformities of the microphones as well as for the influence of the pole structure holding the microphone array [Fig. 1(a)]. The individual sensitivities of all microphones were measured by means of a substitution measurement. A loudspeaker with a broadband frequency response over the range from 50 Hz to 20 kHz was used to produce a sine sweep signal, and a reference microphone (B&K 1/4 in. type 4939) was used to measure the sound pressure created at a distance of 1 m. All 32 microphones of the sphere were subsequently placed at the position of the reference microphone, and the measurement was repeated with the same signal. The result was a set of the microphone transfer functions, derived by complex spectral division of the microphone measurement by the reference measurement [Fig. 1(c)].

To estimate the influence of the pole structure, a reciprocal BEM simulation was performed. The geometry of the microphone in the mounting situation, with either five or six sticks originating from each node, was simulated with a point source at the opening of the microphone membrane, allowing us to calculate the transfer path from any point in space to the microphone. The microphone and a part of either the five-bar or six-bar node were modeled as a compact and rigid body placed at the microphone array's radius.

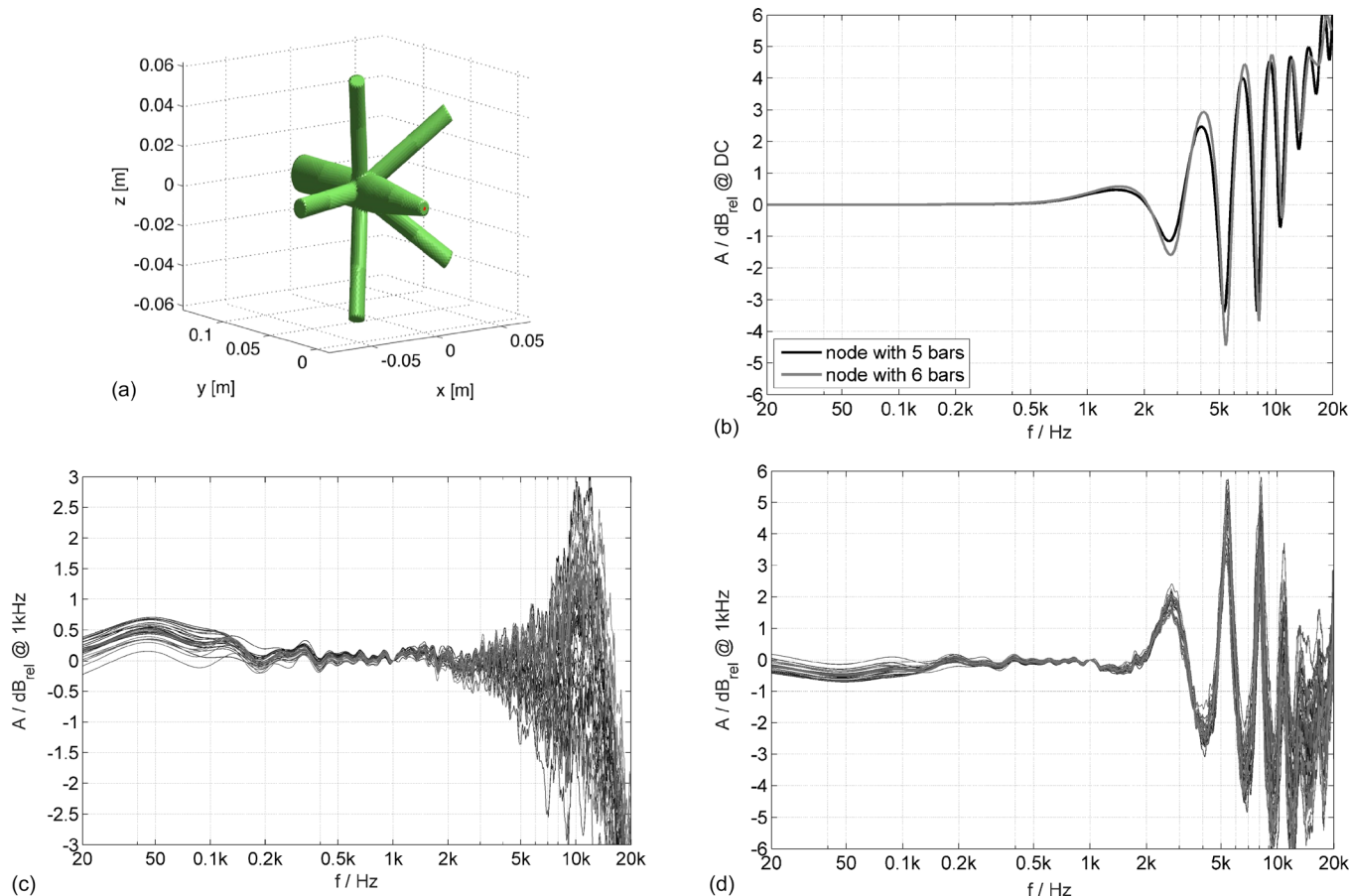


FIG. 1. (Color online) Compensation of the frequency responses of the spherical microphone array. (a) Mesh used for a boundary-element-method simulation of the influence of the pole structure holding the microphone array, with either five or six sticks originating from each node. (b) Averaged frequency responses of the diffraction patterns caused by the five-bar or six-bar node used in the surrounding spherical microphone array. (c) Frequency responses of the 32 individual microphones resulting from a substitution measurement. (d) Resulting compensation filters for the individual microphones. Additional bandpass weighting not shown here.

Assuming that most musical instruments are extended sound sources, and that sound therefore arrives at the microphone from different angles, centered around the frontal incidence ( $0^\circ$ ) pointing at the center of the microphone array, the acoustic transfer functions were simulated for different

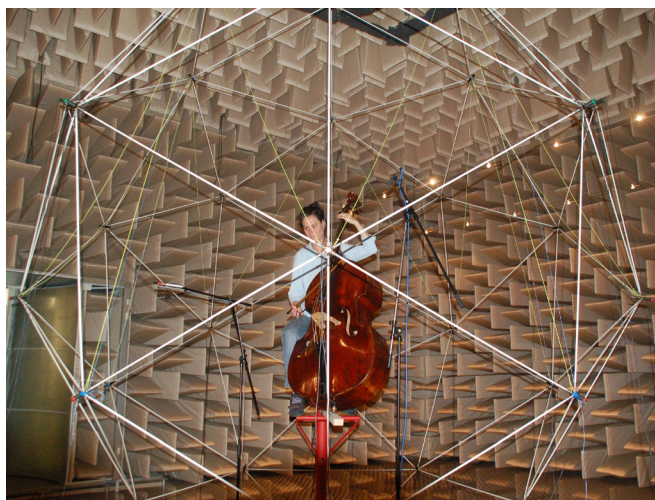


FIG. 2. (Color online) Spherical 32-channel microphone array surrounding a musician in the anechoic chamber of TU Berlin.

positions within a sphere with radius 1 m from the origin of the (reciprocal) source. A weighted average transfer function was then calculated with weights  $w_i = 1 - (\Delta r_i)^2$ , based on the distance  $\Delta r_i$  between the specific position  $i$  and the origin. Since any attempt to *measure* the transfer functions accordingly would have been affected by the nonuniform frequency response and the non-ideal directivity of the measurement loudspeaker, the simulation was considered to be a more reliable approach.

As can be seen in Fig. 1(b), the regular structure of the pole construction causes a comb filter-like ripple of the frequency response for frequencies above 1 kHz. The ripple is slightly larger for the six-bar node. Depending on the mounting position of each microphone, the measured transfer function  $H_{\text{mic}}$  was multiplied with either the five-bar or six-bar node transfer function  $H_{\text{BEM5,6}}$ . The resulting transfer function,

$$H = H_{\text{mic}} \cdot H_{\text{BEM5,6}}, \quad (1)$$

was inverted while preserving the phase, thus yielding the raw compensation filter  $H_{\text{inv}}$  in the frequency domain. After subsequent inverse FFT, all 32 impulse responses  $h_{\text{inv}}$  were windowed around their individual peak, using a Dolph–Chebyshev window with 140 dB stopband attenuation and 8193 taps and a subsequent rectangular window

with 4097 taps. The resulting compensation filters are shown in Fig. 1(d). Except for some of the lowest notes with a fundamental of  $f < 60$  Hz, a minimum-phase bandpass filter (63 Hz, 20 kHz, with fourth order Butterworth slopes) was additionally applied by default to suppress low- and high-frequency noise.

Four 8-channel RME OctaMic microphone preamplifiers and A/D converters connected to an audio workstation were used in order to record the microphone signals with 24 bit resolution at a sampling frequency of  $f_s = 44.1$  kHz. A calibration process was performed each time the gain factor of the measurement chain was changed. To measure all individual 32 gain factors, a sine sweep signal generated in MATLAB was fed to all the 32 input ports simultaneously, and the impulse response of the entire measurement chain was captured. The gain values were changed during the recording, taking into consideration the loudness of each instrument, to ensure that neither overload nor low modulation of the inputs would occur. After the calibration of the electrical measurement chain (microphone input), a pistonphone calibrator (B&K 4230, 94 dB @ 1 kHz) was used with the most accessible microphone in the sphere to obtain the absolute sensitivity of the measurement setup. These transfer functions were used to normalize the recordings of each individual microphone.

## B. Recording, musical instruments, musicians

All instruments of a typical Beethovenian orchestra (violin, viola, violoncello, double bass, flute, oboe, clarinet, bassoon, French horn, trumpet, trombone) were recorded both in their modern form and with instruments typical for the period around 1800 (some originals, some copies). Some popular orchestral instruments without an older historical predecessor (tenor saxophone, alto saxophone, bass clarinet, contra-bassoon, tuba) were also recorded; for some instruments, also a baroque precursor of the modern instrument was measured, such as a baroque bassoon, or a baroque transverse flute as a precursor of the classical keyed flute and the modern Boehm concert flute. Finally, a modern guitar, a modern harp, and a soprano singer were recorded. The modern instruments were played by members of the Deutsches Sinfonieorchester Berlin (<https://www.dso-berlin.de/>) and other professional orchestras in Berlin, and the historical instruments were all played by members of the Akademie für Alte Musik (<http://akamus.de/>), one of the most renowned ensembles for historically informed performance practice in Germany. The modern instruments were tuned to 443 Hz and the classical instruments to 430 Hz, the assumed tuning for an orchestra of the Viennese classical period; most baroque instruments were tuned to 415 Hz. Details of the recorded instruments, such as the maker as well as the strings, bows, mouthpieces, etc. can be found in the documentation of the database of all recorded tones, which is accessible online (Weinzierl *et al.*, 2017).

An adjustable chair was used in order to place the musical instrument as close as possible to the geometrical center of the array, and the musicians were asked to perform in a playing position that remained as constant as possible. Each musician

was asked to play single notes in *ff* (instruction: “play as loud as possible without sounding unpleasant”) and in *pp* (instruction: “play as soft as possible without allowing the sound to break up”) in semitone steps over the entire pitch range required in the standard orchestral repertoire. The musicians were asked to play without vibrato for approximately 3 s per note, which was considered to be sufficient for the steady-state analysis of each note. Of three notes played for each pitch and each dynamic level, the softest or loudest and at the same time musically convincing version was selected manually.

## C. Sound power analysis

For the sound power analysis, the stationary parts of all single note recordings were selected manually using a  $-3$  dB criterion for the beginning and the end of the stationary phase. In the case of all examined instruments, this resulted in durations between 200 and 4400 ms. The sound pressure  $p$  was averaged within the steady sound boundaries for each microphone position as

$$L_p = 10 \log_{10} \left( \frac{\frac{1}{N} \sum_{n=1}^N p[n]^2}{p_0^2} \right), \quad (2)$$

where  $N$  corresponds to the number of samples in the stationary phase and  $p_0 = 2 \times 10^{-5}$  Pa.

The resulting individual microphone pressure levels were averaged over the spherical enveloping surface as

$$\bar{L}_p = 10 \log_{10} \left( \frac{1}{M} \sum_{m=1}^M 10^{0.1L_{p,m}} \right) [\text{dB}], \quad (3)$$

where  $M = 32$ , thus yielding a sound power level of

$$L_W = \bar{L}_p + 10 \log_{10} \left( \frac{S_1}{S_0} \right) [\text{dB}], \quad (4)$$

with  $S_1 = 54.63 \text{ m}^2$  and  $S_0 = 1 \text{ m}^2$ .

To obtain perceptually meaningful values for the transient sounds (plucked guitar, harp), the sound pressures  $p[n]$  in Eq. (2) were subject to time-weighted filtering (“fast”) according to IEC 61672-1 (2013) prior to averaging:

$$L_\tau(t) = 20 \log_{10} \left\{ \left[ \left( \frac{1}{\tau} \right) \int_{-\infty}^t p^2(\xi) e^{-(t-\xi)/\tau} d\xi \right]^{1/2} / p_0 \right\} \times [\text{dB}], \quad (5)$$

with  $\tau$  as the time of the exponential function for time weighting F (fast,  $\tau = 0.125$  s) and  $\xi$  used for the integration from  $-\infty$  to the observation time  $t$ .

Finally, the ISO 3745 (2012) correction factors  $C_1 = -0.17$  dB and  $C_2 = -0.13$  dB were applied, considering the meteorological conditions inside the anechoic chamber with temperature  $\theta = 17^\circ \text{C}$ , static pressure  $p_s = 101.3$  kPa, and a relative humidity of 60%. The correction factor  $C_3$  was ignored, with values  $< 0.1$  dB for the most relevant part of the spectrum with  $f \leq 5$  kHz.

## D. Dynamic range indicators

The sound power values were calculated for each of the 3482 notes recorded, as described above. For string instruments, the values varied from note to note within a typical range of  $\pm 6$  dB, whereas for most wind instruments, there was a systematic increase with pitch, as illustrated in Fig. 3. To quantify the dynamic range of an instrument, we indicated the highest value for the *ff* ( $L_{W\_ff\_max}$ ) and the lowest value for the *pp* ( $L_{W\_pp\_min}$ ), following the procedure of Meyer (2009). From a musical point of view, however, these values are of limited practical relevance. This is first because the maximum and minimum values belong to very contrasting pitch regions, and the ranges for one specific pitch are typically much narrower. For the flute, for example, we obtain a dynamic range of 28 dB by contrasting the softest *pp* with the loudest *ff*, whereas the dynamic range is hardly more than 6 dB for one specific pitch over most of the tonal range (Fig. 3). The second reason is that the extreme values are often reached in pitch regions that are hardly used in the musical repertoire. Taking again the example of the flute, the highest sound power values are reached for the notes above Bb6, which are never used in the symphonies of Mozart, Haydn, and Beethoven (cf. Fig. 3 and Quiring and Weinzierl, 2016b).

In order to determine a more musically relevant value, indicating the actual contribution of an instrument to the orchestral balance, we have calculated a weighted average of the *pp* and *ff* values over pitch, using a typical distribution of pitch in the classical repertoire. This distribution was derived from symphonies no. 1–9 of L. v. Beethoven for each individual instrument, based on an analysis of the authors (Quiring and Weinzierl, 2016a), which is available online (Quiring and Weinzierl, 2016b). Beethoven's symphonies belong to the most popular orchestral works. In the Repertoire Reports of the League of American Orchestras

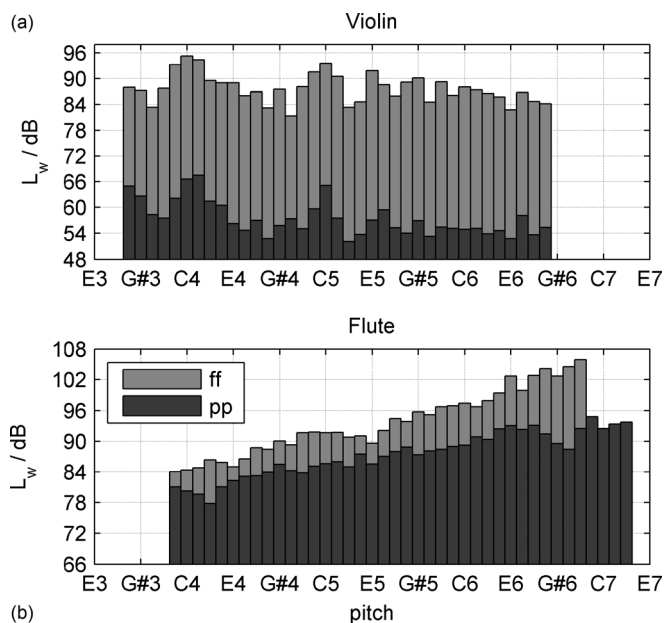


FIG. 3. Sound power levels for a modern violin (a) and for a flute (b) over pitch.

2002–2013, no composer appears more often than L. v. Beethoven (League of American Orchestras, 2018), and with about 593 000 individual notes, the sample seems sufficiently large to give a representative picture of how the different instruments are actually used in the classical-romantic orchestral repertoire. The weighted average values for the sound power in *ff* ( $L_{W\_ff\_av}$ ) and in *pp* ( $L_{W\_pp\_av}$ ) were thus calculated using the frequencies by which each pitch appears in the symphonies of L. v. Beethoven as weights.

## E. Timbral features

All audio data in the set was recorded at a sampling frequency of 44.1 kHz in  $M=32$  channels from the spherical microphone array. For further processing, only one of the 32 channels was used for the calculation of audio features per instrument. Calculating a sum of the channels was not considered to avoid comb filter effects. Instead, we selected the channel which most often exhibited the highest root-mean-square (RMS) signal level of the 32 channels over all notes played by each instrument as the principal channel, i.e., as the principal direction of sound radiation.

For this channel, we extracted audio features using the timbre toolbox (TTB, Peeters *et al.*, 2011). The toolbox is divided into global descriptors, referring to the temporal energy envelope, and time-varying descriptors, which extract spectral features using a sliding-window approach. Time-varying features were calculated as trajectories with a Hamming window of 23.2 ms duration and a hop size of 5.8 ms, as defined by the TTB. Two statistical single-value descriptors across time, namely, the median and the interquartile range (IQR) were obtained for each feature trajectory from each recording.

The not-so-common use of tristimulus features (Pollard *et al.*, 1982) was tested, drawing on the TTB implementation, as well as on a custom implementation of the same formulae, in order to increase robustness. For this, a sliding-window analysis with a window size of 9.29 ms and a hop size of 4.64 ms was applied for the partial tracking. The YIN algorithm (de Cheveigné *et al.*, 2002) was used for estimating the fundamental frequency  $f_0$  of each window. The  $f_0$  boundaries were set to 20 and 4000 Hz, since the highest pitch in the data lies at 2793.83 Hz (ISO pitch F7) and the lowest pitch has a fundamental frequency of 21.82 Hz (ISO pitch F0). The FFT was calculated with an additional zero padding to a length of  $2^{13}$  samples. For each window, the parameters of the first 30 partials were measured using quadratic interpolation (Smith and Serra, 2005). Median and IQR across time windows were subsequently calculated for each tristimulus feature recording.

## F. Statistical analysis

Initially, the available information on *pitch*, *sound power* and all 141 TTB features of the 3482 audio recordings (1764 *pp*-recordings and 1718 *ff*-recordings) were z-standardized to reduce possible later problems with scaling, multi-collinearity and comparative interpretation. New categorical variables were created to code intended *dynamic strength* (*pp* vs *ff*), *instrument* (see Table I for instrument

list), *instrument group* (*brass, string, woodwind, plucked strings, and voice*) and *epoch* (*classical vs modern*).

Two stepwise LDA were performed as data-mining procedures to identify the best informationally nonredundant predictors for *dynamic strength* contained within the dataset. While the first analysis included (and thereby controlled for) *sound power, pitch*, and all available spectral features, the second LDA left out the *sound power* variable to simulate a scenario without any loudness information. During the analyses, the overall *Wilks Lambda* coefficient was employed as the primary variable inclusion/exclusion criterion. Feature selection was stopped when either no significant decrease in *Wilks Lambda* was achievable or when tolerance values for single predictors fell below 0.1, thereby signaling an intolerable degree of multi-collinearity within the chosen predictor set.

In order to estimate the relative predictive value of *sound power* and the spectral features identified in the LDA, GLMM analyses (Skronald and Rabe-Hesketh, 2004) with robust maximum likelihood estimation were performed. In both models (GLMM 1, GLMM 2), *dynamic strength* was implemented as the binominal dependent, employing a logistic link function. Furthermore, both models estimated random intercepts for *instrument* clusters and thereby accommodated for instrument-specific dynamic and spectral ranges, but did not contain fixed intercepts due to z-standardization. In GLMM 1, *pitch, sound power* and the timbral features identified by the first LDA were introduced stepwise as fixed predictors, with *pitch* acting as a control variable. The GLMM 2 was realized in a similar fashion, drawing on spectral features identified in the second LDA, but here *sound power* was left out to simulate a scenario without loudness information. For each modeling step in both models, cumulative and incremental marginal and conditional  $R^2$  (Nakagawa and Schielzeth, 2013), as well as a likelihood-ratio-test of model improvement, were calculated.

### III. RESULTS

#### A. Sound power and dynamic range of orchestral instruments

The results of the sound power measurements are shown in Table I. They include the minimum and maximum values  $L_{W\_pp\_min}$  and  $L_{W\_ff\_max}$  reached for *pp* and *ff* over the entire pitch range, as well as the weighted averages for *pp* and *ff*, based on the pitch distribution of each instrument in the classical-romantic orchestra repertoire (see Sec. II D).

The dynamic range, derived from the difference between the sound power in *pp* and in *ff*, is remarkably different for the various instruments. It ranges from a minimum of 18–22 dB for the double reed instruments (oboe, bassoon, contrabassoon, dulcian) to a maximum of 57 dB for the clarinet. When taking the distribution of pitch into account, i.e., how the instruments are actually used in the orchestral repertoire, the averaged values range from 9 to 15 dB for the double reed instruments to 33 dB for the clarinet.

TABLE I. Sound power levels for 40 musical instruments, determined for single notes played at *pp* (“as soft as possible”) and *ff* (“as loud as possible”) over the entire chromatic range of each instrument. The level  $L_{W\_pp\_min}$  shows the minimum value, and  $L_{W\_ff\_max}$  shows the maximum value reached. The levels  $L_{W\_pp\_av}$  and  $L_{W\_ff\_av}$  show the average of the *pp* and *ff* values for the entire tonal range of each instrument, with the pitch distribution within the symphonies of L. v. Beethoven used as weights. These values could only be calculated for the modern and classical instruments that appear in these symphonies. The sound power values for each individual note (pitch) is available in the electronically published database of all recorded notes, as well as details of the recorded instruments, such as the maker as well as the strings, bows, mouthpieces, etc. (Weinzierl et al., 2017).

Instrument	$L_{W\_ff\_max}$	$L_{W\_pp\_min}$	$L_{W\_ff\_av}$	$L_{W\_pp\_av}$
Violin				
Classical	95	56	90	62
Modern	95	52	91	56
Viola				
Classical	94	57	91	62
Modern	97	53	93	62
Violoncello				
Classical	102	57	96	62
Modern	97	63	93	70
Double bass				
Classical	100	66	96	73
Modern	100	56	96	70
Flute				
Baroque transverse	101	69		
Keyed flute (classical)	106	68	101	84
Modern	105	77	101	89
Oboe				
Romantic	99	78	94	83
Classical	100	81	97	86
Modern	101	80	99	84
Cor anglais	101	79		
Clarinet				
Basset horn (F)	102	59		
Classical (Bb)	105	57	97	65
Modern (Bb)	110	53	102	69
Bass Clarinet (Bb)	102	65		
Bassoon				
Baroque	98	77		
Classical	101	82	99	86
Modern	104	82	101	86
Contrabassoon	98	80	94	85
Dulcian	98	77		
French horn				
Natural horn (A)	111	74	107	86
Double horn (F/Bb)	114	71	112	85
Trumpet				
Natural trumpet (D)	107	81	104	83
Modern (C)	112	74	107	83
Trombone				
Alto (Eb)	104	67		
Tenor (classical, C)	106	79	106	78
Tenor (modern, Bb/F)	113	68	112	82
Bass (classical, F)	109	78		
Bass (modern, Bb/F/G)	112	72		
Tuba	122	71		

TABLE I. (Continued)

Instrument	LW <sub>ff_max</sub>	LW <sub>pp_min</sub>	LW <sub>ff_av</sub>	LW <sub>pp_av</sub>
Saxophone				
Alto	111	82		
Tenor	113	82		
Timpani				
Hand crank	108	60		
Pedal	108	58		
Harp				
Guitar	91	54		
	88	59		

## B. The contribution of sound power and timbre to the expression of dynamic strength

The stepwise LDA 1 (incorporating *sound power*) was able to identify *sound power*, *spectral skewness* (ERBfft, median), and *decrease slope* as the best significant and non-redundant predictors for intended *dynamic strength*, resulting in the correct classification of 92% of the recordings. Stepwise LDA 2 (without incorporating *sound power*) was able to identify *spectral skewness* (ERBfft, median), *spectral flatness* (STFTmag, median), and *attack slope* as the best nonredundant predictors for intended *dynamic strength*, resulting in correct classification of 85% of cases.

The GLMM 1 employing *pitch*, *sound power* and the spectral features identified in LDA 1 was able to achieve a marginal  $R^2$  of 80% and a conditional  $R^2$  of 96%. Inspection of incremental  $R^2$  gains implies that *sound power* is able to explain 69% of dynamic strength and the timbre feature *spectral skewness* is able to explain an additional 9%. When accommodating for the different dynamic ranges of instruments with the help of random intercepts, however, *sound power* is able to explain 96% of *dynamic strength* alone, with only minor additional gains through spectral features (see Table II).

The GLMM 2 employing *pitch* as control and the spectral features identified in LDA 2 was able to achieve a marginal  $R^2$  of 72% and a cumulative  $R^2$  of 89%. Inspection of incremental  $R^2$  gains implies that *spectral skewness* is able to explain 35% of *dynamic strength* and *spectral flatness* an additional 29%. When accommodating for the different spectral ranges of instruments with the help of random intercepts, however, *spectral skewness* is able to explain 48% of *dynamic*

*strength* alone with 38% additional gains in predictive power with the help of *spectral flatness* (see Table III).

Scatterplots (Fig. 4) illustrate the interplay of the predictors identified by both model variants in discriminating between instrumental recordings of differing dynamic strengths. Table IV demonstrates the intercorrelations of *sound power*, *pitch*, and the spectral features used in the final models.

## IV. DISCUSSION

The current investigation presents a comprehensive dataset of sound power measurements for 40 musical instruments, including all standard orchestral instruments. With professional musicians instructed to play as softly and as loudly as possible, and covering the whole chromatic range of the individual instruments, these values describe the physical potential of each instrument with respect to the production of sound within the aesthetical limitations of musical practice. At the lower end of the dynamic range, when the tone can only just be steadily produced (*pp*), the sound power levels range from 53 dB for the violin to 82 dB for the bassoon and saxophone (tenor and alto). At the upper end of the dynamic range, where the tone can still be produced in an aesthetically acceptable manner (*ff*), these values range from 88 dB for the guitar up to 122 dB for the tuba. The dynamic ranges, determined by the difference between the minimum *pp* level and the maximum *ff* level, lie between 18 dB for the contrabassoon and 57 dB for the clarinet.

Since these extreme values are often only reached for certain notes (pitches), which sometimes lie outside the standard pitch range used in the orchestral repertoire, they bear only limited relation to musical practice. Earlier studies tried to address this by measuring not only single tones but scales or specific musical excerpts (Meyer, 2009). Since the resulting values, however, depend on the selected excerpt and the chosen register of the instrument and are thus not very reproducible, we chose another approach by calculating a weighted average of the individual notes and using the distribution of pitch of each instrument in the symphonies nos. 1–9 of L. v. Beethoven as weights. These distributions are publicly available (Quiring and Weinzierl, 2016b), so they can also be used for future investigations. Using these weighted averages to determine the mean dynamic range of each instrument gives values ranging from 9 dB for the contrabassoon to 33 dB for the clarinet.

TABLE II. Results of a generalized linear mixed model (GLMM 1, binomial target with logit-link), predicting dynamic strength by pitch, sound power and timbre features. Marginal  $R^2$  values provide the estimated explained variance in dynamic strength (as cumulative sum and incremental contribution of each predictor) when considering fixed effects only. Conditional  $R^2$  values provide the estimated explained variance in dynamic strength when also taking into account instrument-specific dynamic strength thresholds in terms of estimated random intercepts. The BIC and Deviance are information theoretical measures of the overall model fit when a predictor is included. The F and p values verify the significance of the model, and the Sign shows whether the predictor is positively or negatively correlated with dynamic strength.

Predictor	F	Sign	p (Wald)	Deviance	BIC	$\Sigma$ marg. $R^2$	$\Delta$ marg. $R^2$	$\Sigma$ cond. $R^2$	$\Delta$ cond. $R^2$
Pitch	51.4	–	0.023	8888	8896	0%	0%	0%	0%
Sound power	5.2	+	<0.001	29466	29474	69%	69%	96%	96%
Spectral skewness (ERBfft, median)	178.7	–	<0.001	19237	19244	77%	9%	96%	0%
Decrease slope	17.2	–	<0.001	21312	21320	80%	3%	96%	0%



TABLE III. Results of a generalized linear mixed model (GLMM 2, binomial target with logit-link), predicting dynamic strength by pitch and timbre features only. For the statistical measures see Table II.

Predictor	F	Sign	p (Wald)	Deviance	BIC	$\Sigma$ marg. $R^2$	$\Delta$ marg. $R^2$	$\Sigma$ cond. $R^2$	$\Delta$ cond. $R^2$
Pitch	8.1	–	0.005	13966	13974	0%	0%	0%	0%
Spectral skewness (ERBfft, median)	65.9	–	<0.001	15605	15613	35%	35%	48%	48%
Spectral flatness (STFTmag, median)	111.2	–	<0.001	25369	25377	64%	29%	86%	38%
Attack slope	20.6	+	<0.001	23785	23793	72%	9%	89%	3%

Since the measurements were conducted with only one musical instrument and one performer per instrument, they can, of course, not be straightforwardly generalized. There are certainly differences between individual instruments and the individual performers playing them. An indication of these person- and instrument-related individual differences might be given by comparing the results with previous results of Meyer (1990). For the 12 instruments measured in Meyer’s study, the values for the sound power at *ff* lie within  $\pm 4$  dB of our values, with a mean absolute difference of 2 dB, except for the tuba, for which Meyer’s value is 10 dB lower than ours. The values for the sound power at *pp* lie within  $\pm 10$  dB of our values, with a mean absolute difference of 6.8 dB. The *ff* values are thus quite reproducible, whereas the values for *pp* seem to depend much more on the instrument as well as the perception and technical abilities of the individual performer.

Based on an extraction of timbral features (Peeters *et al.*, 2011) for each of the 3482 recorded notes, we have attempted to quantify the relative contribution of sound power and timbre to the expression of dynamic strength. The results of a generalized linear mixed model analysis can be interpreted from the perspective of a hypothetical listener drawing on this information. If this listener had musical experience (knowing the dynamic potential of individual musical instruments) and room acoustical experience (being able to estimate the sound power of a musical instrument in a reverberant sound field), virtually no additional cues would be necessary to identify the tone of a musical instrument being played at *pp* or *ff*. If the individual properties of the musical instruments are not known, the reliability decreases

considerably, as can be seen by comparing the estimated marginal  $R^2$  with the conditional  $R^2$  (69% vs 96%), i.e., by comparing a model for *all* musical instruments (marginal  $R^2$ ) with a model, where the dynamic thresholds are allowed to vary between the instruments (conditional  $R^2$ ). In such situations, spectral properties can be used as additional cues to compensate for the loss of information. The most informative feature in this context is *spectral skewness*, with a left-skewed spectral shape indicating high dynamic strength, i.e., with the mode of the spectral distribution shifted towards higher partials. This cue, however, has to be weighted by the pitch of the tone in question, due to the general correlation between pitch and spectral skewness in most instruments.

We then considered a hypothetical situation where for some reason, no sound power information is available at all. This could happen for example when listening to audio recordings of instrumental music at arbitrary volume, or when the influence of the room and the source-receiver distance cannot be reliably estimated to extrapolate from sound pressure to sound power. As it turns out, even in such scenarios listeners are still quite reliably able to identify the intended dynamic strength by combining several dimensions of timbral information. This is again the *spectral skewness* of the tone, combined with *spectral flatness* and *attack slope*, again weighted by the pitch of the played note. Low *spectral flatness* provides a valuable cue for high dynamic strength, because the amplitude difference between the partials and the instrumental noise floor generated by wind or bow noise increases (and the flatness decreases) with dynamic strength,

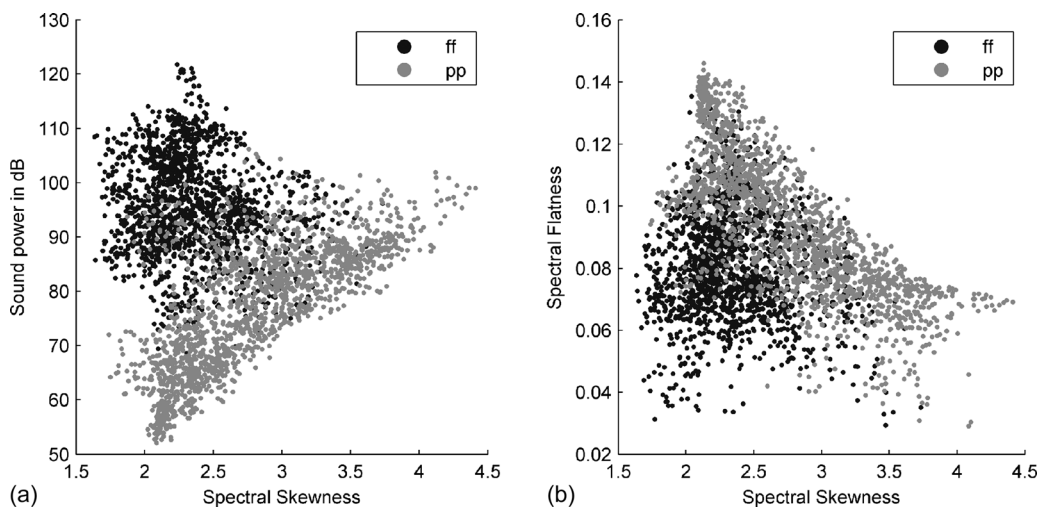


FIG. 4. (a) Sound power and spectral skewness (ERBfft, median) as predictors of dynamic strength. (b) Without sound power, spectral skewness (ERBfft, median) and spectral flatness (STFTmag, median) are the best predictors of dynamic strength.

TABLE IV. Spearman Correlation matrix of sound power, pitch, and spectral features used in final models.

	1	2	3	4	5	6
1 Pitch	1	0.03	-0.49	0.15	0.19	0.10
2 Sound power in dB	0.03	1	-0.14	-0.53	0.18	0.40
3 Spectral skewness (ERBfft, median)	-0.49	-0.14	1	-0.31	-0.13	0.00
4 Spectral flatness (STFTmag, median)	0.15	-0.53	-0.31	1	-0.24	-0.10
5 Decrease slope	0.19	0.18	-0.13	-0.24	1	0.00
6 Attack slope	0.10	0.40	0.00	-0.10	0.00	1

and so does the slope of the attack of the tone. With a combination of these timbral features, a level of determinancy of 72% can be reached with an instrument-unspecific model (marginal  $R_2$ ), and 89% with an instrument-specific model (conditional  $R_2$ ).

Taken together, the present results indicate the acoustical features on which listeners can draw in order to identify the intended dynamic strength when listening to classical, instrumental music. Even when sound power is difficult to estimate in the concert situation and even more when listening to recorded music, timbre-related temporal (*attack slope*) and spectral (*spectral skewness*, *spectral flatness*) features can be used to fill the information gap, and to still decode the dynamic expression in the acoustical signal almost reliably.

## ACKNOWLEDGMENTS

This investigation was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG FOR 1557, DFG WE 4057/3-2). The authors would like to thank Johannes Krämer, Alexander Lindau, and Martin Pollow for support in the measurements described in this paper. We thank the 40 professional musicians for participation in the study, and the editor and two anonymous reviewers for valuable comments on the text.

Burghauer, J., and Spelda, A. (1971). *Akustische Grundlagen des Orchestrierens (Acoustic Fundamentals of Orchestration)* (Gustav Bosse Verlag, Regensburg, Germany).

Chaigne, A., Derveaux, G., and Balanant, N. (2004). "Sound power and efficiency in stringed instruments," in *Proc. 18th International Congress on Acoustics*, Kyoto, Japan, pp. 2123–2126.

Clarke, M., and Luce, D. (1965). "Intensities of orchestral instrument scales played at prescribed dynamic markings," *J. Audio Eng. Soc.* **13**(2), 151–157.

de Cheveigné, A., and Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.* **111**(4), 1917–1930.

Fabiani, M., and Friberg, A. (2011). "Influence of pitch, loudness, and timbre on the perception of instrument dynamics," *J. Acoust. Soc. Am.* **130**(4), EL193–EL199.

García-Mayén, H., and Santillán, A. (2011). "The effect of the coupling between the top plate and the fingerboard on the acoustic power radiated by a classical guitar," *J. Acoust. Soc. Am.* **129**(3), 1153–1156.

IEC 61672-1 (2013). "Electroacoustics—Sound level meters—Part 1: Specifications" (International Electrotechnical Commission, Geneva, Switzerland).

ISO 3741 (2010). "Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Precision methods for reverberation test rooms" (International Organization for Standardization, Geneva, Switzerland).

ISO 3745 (2012). "Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Precision methods for anechoic rooms and hemi-anechoic rooms" (International Organization for Standardization, Geneva, Switzerland).

ISO 9614-2 (1996). "Acoustics—Determination of sound power levels of noise sources using sound intensity—Part 2: Measurement by scanning" (International Organization for Standardization, Geneva, Switzerland).

Lai, J. C. S., and Burgess, M. A. (1990). "Radiation efficiency of acoustic guitars," *J. Acoust. Soc. Am.* **88**(3), 1222–1227.

League of American Orchestras (2013). *Orchestra repertoire reports (ORR) 2002-2013*, <http://www.americanorchestras.org/> (Last viewed June 1, 2018).

Meyer, J. (1990). "Zur Dynamik und Schalleistung von Orchesterinstrumenten" ("On the dynamics and sound power of orchestral instruments"), *Acustica* **71**, 277–286.

Meyer, J. (1993). "The sound of the orchestra," *J. Audio Eng. Soc.* **41**(4), 203–213.

Meyer, J. (2009). *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers* (Springer Science+Business Media, New York).

Meyer, J., and Angster, J. (1983). "Sound power measurement of musical instruments," in *Proc. 11. International Congress on Acoustics*, Paris, p. 321.

Nakagawa, S., and Schielzeth, H. (2013). "A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models," *Methods Ecol. Evol.* **4**(2), 133–142.

Nakamura, T. (1987). "The communication of dynamics between musicians and listeners through musical performance," *Atten. Percept. Psychol.* **41**, 525–533.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., and McAdams, S. (2011). "The timbre toolbox: Extracting audio descriptors from musical signals," *J. Acoust. Soc. Am.* **130**(5), 2902–2916.

Pollard, H. F., and Jansson, E. V. (1982). "A tristimulus method for the specification of musical timbre," *Acta Acust. Acust.* **51**(3), 162–171.

Quiring, R., and Weinzierl, S. (2016a). "Tonhöhenverteilungen im klassischen Orchesterrepertoire" ("Pitch distributions in classical orchestra repertoire"), *Fortschritte der Akustik, DAGA Aachen*, pp. 1486–1489.

Quiring, R., and Weinzierl, S. (2016b). "Pitch distributions for individual instruments in the symphonies No. 1-9 of L.v. Beethoven," <http://dx.doi.org/10.14279/depositonce-5040>.

Shabtai, N. R., Behler, G., Vorländer, M., and Weinzierl, S. (2017). "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *J. Acoust. Soc. Am.* **141**(2), 1246–1256.

Sivian, L. J., Dunn, H. K., and White, S. D. (1931). "Absolute amplitudes and spectra of certain musical instruments and orchestras," *J. Acoust. Soc. Am.* **2**(3), 330–371.

Skrondal, A., and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modelling. Multilevel, Longitudinal, and Structural Equation Models* (Chapman and Hall, London).

Smith, J. O., and Serra, X. (2005). "Parshl: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," <https://ccrma.stanford.edu/STANM/stanms/stanm43/stanm43.pdf> (Last viewed June 1, 2018).

Weinzierl, S., Vorländer, M., Behler, G., Brinkmann, F., von Coler, H., Detzner, E., Krämer, J., Lindau, A., Pollow, M., Schulz, F., and Shabtai, N. R. (2017). "A Database of anechoic microphone array measurements of musical instruments," <http://dx.doi.org/10.14279/depositonce-5861.2>.