



HAL
open science

Listener perception of changes in historically informed performance of solo baroque music due to room acoustics

Nolan Eley, Catherine Lavandier, Théodora Psychoyou, Brian F.G. Katz

► **To cite this version:**

Nolan Eley, Catherine Lavandier, Théodora Psychoyou, Brian F.G. Katz. Listener perception of changes in historically informed performance of solo baroque music due to room acoustics. *Acta Acustica*, 2024, 8 (6), pp.1-15. 10.1051/aacus/2023069 . hal-04432718

HAL Id: hal-04432718

<https://hal.science/hal-04432718>

Submitted on 3 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Listener perception of changes in historically informed performance of solo baroque music due to room acoustics

Nolan Eley^{1,2,*}, Catherine Lavandier¹, Théodora Psychoyou³, and Brian F.G. Katz²

¹CY Cergy Paris University, ETIS Laboratory ENSEA, CNRS, 2 Av. Adolphe Chauvin, 95300 Pontoise, France

²Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert UMR7190, 4 Place Jussieu, 75005 Paris, France

³Sorbonne Université, Institut de Recherche en Musicologie, Quai François Mauriac, 75013 Paris, France

Received 7 February 2023, Accepted 18 December 2023

Abstract – A listening test was performed to determine whether there were any perceived differences in the playing styles of musicians as a function of the room in which the music was performed. This listening test was part of a larger study aimed at investigating the impact of room acoustics on historically informed performance (HIP) of baroque music which included an objective analysis framework designed to capture dimensions of performance important to a historical baroque playing style. The test included both flute and viol examples and included the participation of 20 musically trained listeners with a background in HIP. The results of the test showed some significant differences in certain performance parameters for both instruments. In particular, the flute examples demonstrated differences that aligned with reports from the flutists about the strategies they used to adapt to the different acoustics, and there was agreement between listener ratings and previously extracted objective parameters. These findings suggest that room acoustics play a role in shaping musical performance, and that the resulting changes can be perceived by listeners.

Keywords: Room acoustics, Music perception, Performance analysis

1 Introduction

During a musical performance, musicians monitor and adjust their playing based on, among other things, aural feedback. This aural feedback is directly linked to the acoustical properties of the room in which the performance is taking place. While room acoustics have been known to influence the composition, performance, and perception of music for centuries [1], in recent decades there have been a growing number of controlled scientific investigations into the matter.

In this study, a listening test was performed in order to better understand, from the perception of a listener, the differences in performing style which may be due to different acoustic settings. This listening test was part of a broader effort intended to study the effect of acoustics on historically informed performance (HIP) of baroque music. While previous studies have examined the effect of acoustics on musical performance more generally, this study seeks to clarify whether the acoustics of historical rooms facilitates performances of music from the same era. In a previous study [2], 10 musicians (3 flutists, 3 theorbists, and 4 violists da gamba) specializing in historical baroque performance performed multiple compositions in two rooms, a baroque-era hall and a modern hall. This study will only make

use of the flute and viol recordings due to issues with the theorbo performances previously noted in [2]. In a follow-up study [3], a musicologically-informed objective analysis framework was developed to identify performance features important to a historically informed baroque playing style.

A subset of the audio recordings from those studies was used here to create a listening test in order to better understand the perceptibility of the previously measured changes in objective parameters. Additionally, this study was intended to shed light on the effectiveness of the objective analysis in capturing differences in performing style. Lastly, this study was also expected to provide more information on how the performance intentions of musicians are communicated to the listener.

1.1 Background

1.1.1 Historically informed performance

When preparing a musical performance, much effort is put into fulfilling the composer's intentions. However, in much music of the past, certain conventions were so well understood by performers that writing them in the score was deemed unnecessary ([4], p. 2). In addition, during the Baroque period, composers were often directly involved in performances as musicians, conductors, or both ([5], p. 9), reducing even further the need for detailed instructions in

*Corresponding author: nolan.eley@sorbonne-universite.fr

the score. The HIP movement is largely concerned with uncovering those details which, at one point, were implicit but have been gradually forgotten. In HIP, contrary to mainstream performance practice, musicians deliberately imbue their performance with stylistic tendencies from the era during which the composition was written. Many musicologists, by researching primary sources such as performance manuals and treatises, have created helpful references for adopting historical performance practice [4, 6, 7].

1.1.2 Baroque performance practice

It should be noted that the HIP movement is not immune to trends and that various styles, all claiming to be “historically informed” have appeared throughout the course of the last century. However, since around the 1980s, the view on what constitutes a HIP of baroque music has solidified somewhat. An extensive list of stylistic tendencies of the era have been well catalogued in [8–10] among others. Fabian and Schubert ([11], p. 39) summarized these performance attributes well, stating that a historically informed baroque performance style consists of:

... locally nuanced and clearly punctuated articulation, well defined metric groups and strongly projected/inflected rhythmic gestures, shallow and selectively used vibrato, and a general revelling in the characteristics of eighteenth-century instruments (*e.g.*, the uneven bow strokes, the variety of tonguing patterns and their effect on tone qualities).

Robert Donington described additional stylistic features of baroque playing style which he deemed as essential, stating, “[t]here are two basic characteristics of baroque sound which, under whatever conditions of performance, it is necessary to achieve: a transparent sonority, and an incisive articulation.” ([8], p. 167).

The term baroque expressiveness, coined in [12], was developed to describe expressive musical performances whose expressive characteristics embody the historically informed baroque performance style described above.

It is important to note that the characteristics of historical baroque performance discussed in this analysis are simplified for clarity. For example, while it remains broadly true that phrasing in HIP of baroque music tends to be articulated rather than continuous, it is still possible for continuous phrasing to be valid in certain contexts of historical baroque performance. The same holds true for all of the considered parameters. Furthermore, the implementation and perception of expressive devices in music can be strongly influenced by the composition and instrument [13], and this interaction has not been very well studied.

1.1.3 Acoustics–performance interaction

While there have been a number of studies examining the effect of acoustics on music performance, it is somewhat difficult to infer universal trends from these studies. This is partly due to the difference in methodology among them.

For example, some studies have used virtual acoustic environments [14–19], some have used real acoustic environments [20, 21], and others have used a mixed approach [22–25]. Furthermore, the variety of acoustical parameters under consideration in these studies is large, as is the range of methods analyzing the resulting musical performances.

One of the most commonly reported relationships in these studies is between tempo and reverberation time, partly because these are commonly considered musical and acoustical parameters. Even these findings, however, have not been very consistent across studies. For example, some studies seem to support the intuition that musicians tend to slow their tempo in reverberant environment [15, 16, 26], or at least partially support it [17]. For example, [24], showed that this relationship only held true to a certain extent. Firstly, this effect was only observed for compositions which were generally slow and was not observed in fast compositions. Secondly, this trend was only linear within a certain range and that slower tempos were also used in extremely short reverberation times, perhaps to compensate for the lack of acoustical decay. This nonlinear relationship between tempo and reverberation time was also found in [15]. Lastly, there are some studies that found little to no correlation between reverberation time and tempo [25, 27].

Another important and somewhat common finding in these studies is that adaptation strategies tend to be somewhat dependent on musical content and on individual musicians [15, 17, 24, 25, 28]. Because the musical content in this study (historical baroque music) is different from previous studies, it is unknown what kind of role the acoustics may be expected to play in musicians’ performance.

1.2 Rooms and musical performances

The data used for this study were acquired as part of a previous experiment investigating the impact of room acoustics on the HIP baroque music. The rooms and music performance recordings are described in detail in [2] but are summarized here. In the previous experiment, the musicians, all specializing in historically informed baroque performance practice, played in two halls, a historical baroque-era hall and a modern hall.

The historical hall is the Salon des Nobles from the Château de Versailles which hosted solo and small ensemble concerts during the Baroque era. The length and width of the room are approximately 9 m each while the height of the room is about 7 m, yielding an approximate volume of 564 m³. The modern hall is the amphitheater of the Cité de la Musique in Paris, built in the late 20th century. The amphitheater is a roughly fan-shaped, asymmetrical hall with an approximate volume of 1430 m³ and a seating capacity of approximately 250. It is suitable for solo and small ensemble performances and has controlled humidity to protect the historic musical instrument collection associated with the *Musée de la Musique*. Plan views of the halls can be seen in Figure 1, including the relevant source and receiver positions used for acoustic measurements. The measured reverberation time for these two rooms are shown

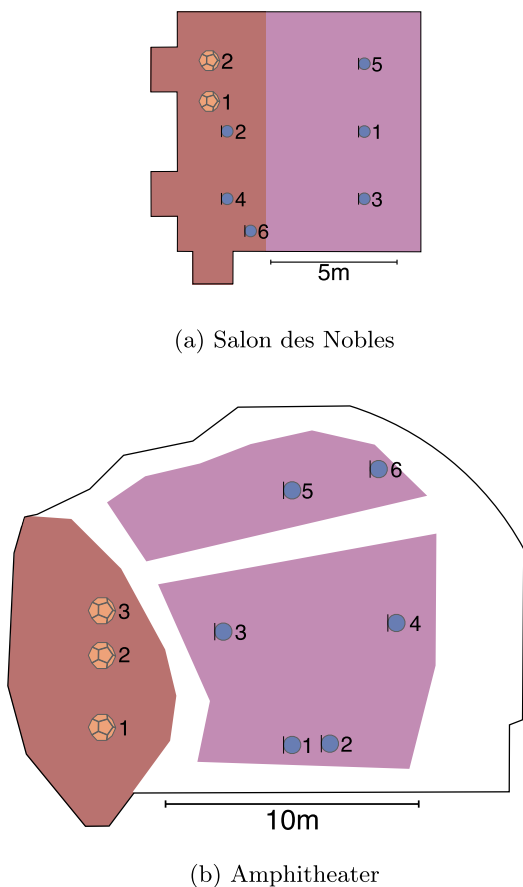


Figure 1. Plan view and map of source and receiver positions for the acoustic measurements taken in both rooms. Stage area shown in brown, while audience areas are shown in purple. (a) Salon des Nobles, (b) Amphitheater.

in Figure 2. Clarity measures (C_{80}) across source and receiver positions are shown for both halls in Table 1.

The different dimensions and acoustics of these two rooms are expected to provide a different performance experience for the musicians. The smaller nature of the historical room would be expected to offer a more immediate acoustic response. The longer reverberation time for the same room, at least in the mid-frequency region, will also provide additional energy return to the performers. This longer reverberation time in the mid-frequency region may also serve as needed acoustical support for baroque instruments which tend to be quieter than their modern counterparts ([10], pp. 151–152). In contrast, the flat reverberation time of the modern room can be expected to induce minimal coloration, retaining spectral neutrality, while the larger fan-shaped design and raked seating (not shown) would direct more early energy to the audience, with less late reverberation, compared to the smaller room.

The musicians performed several pieces in each setting, repeating each performance three times. The repertoire was chosen with the assistance of musicologists and was the same for each instrument class. Three couplets, composed by Marin Marais were chosen for the viol, three preludes by Jacques Hotteterre were selected for the flute.

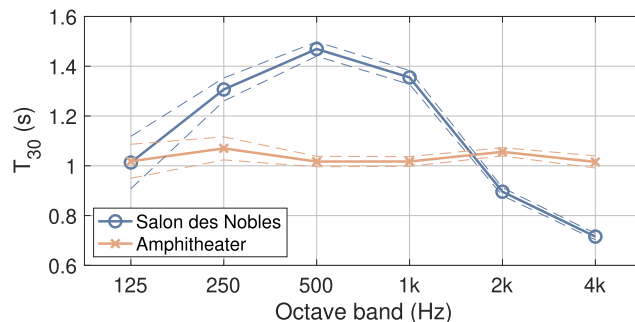


Figure 2. The measured reverberation time of the rooms used in this study. The dotted lines represent the standard deviation across all source/receiver combinations.

Table 1. Clarity values (C_{80} , averaged across the 500 Hz and 1 kHz octave bands) for each source and receiver combination.

| | R 1 | R 2 | R 3 | R 4 | R 5 | R 6 |
|------------------|------|------|------|-------|------|------|
| Salon des Nobles | | | | | | |
| S 1 | 0.95 | 6.73 | 2.48 | 1.90 | 1.66 | 0.88 |
| S 2 | 2.24 | 2.41 | 3.03 | −0.15 | 0.81 | 0.93 |
| Amphitheater | | | | | | |
| S 1 | 1.50 | 1.56 | 1.81 | 2.89 | 1.99 | 2.40 |
| S 2 | 1.74 | 1.71 | 1.67 | 2.03 | 1.64 | 1.59 |
| S 3 | 1.42 | 1.94 | 2.23 | 1.49 | 1.18 | 1.76 |

The time between each session was 10 days and the order of the halls was mixed. Performances were recorded with a cardioid microphone (AKG C414) positioned 1 away from, and directed towards, the instrument.

1.3 Questionnaire responses

After each session, the musicians responded to a number of questionnaires regarding their impression of the acoustics and its potential influence on their performance. A full reporting of the questionnaire results is out of the scope of this study. However, responses to two questions which are particularly relevant to this listening test are reported in Table 2.

While no universal adaptation strategy was reported, there are commonalities shared among some musicians. In the Salon des Nobles, two musicians (flutist 3 and violist 4) claimed to have adopted a slower tempo due to the acoustics. Violist 4 also claimed to play with a “lighter articulation,” and flutist 1 said they put effort into projecting upwards. In general, however, the musicians seemed to revel in the acoustics, claiming it was “suitable for the music” and “literally adapted to my instrument.” In the amphitheater, three musicians (flutists 1 and 2 and violist 4) mentioned the need to lengthen notes in order to compensate for the lack of acoustical decay of the hall. Several musicians (flutists 1 and 3 and violist 1) also mentioned the desire to project further or better support their own dynamics in response to the hall’s acoustics.

Table 2. Flutist and violist responses to the questionnaire.

| | How did this space impact your playing? How did it make you feel? | Were you conscious of any adjustments to your performance due to the space? To what would you attribute this? |
|------------------|--|---|
| Salon des Nobles | | |
| Flutist 1 | This space made me want to imprint my sound more as if to be able to extend it to every corner of the room. I like to be able to appreciate the ends of phrases, the ends of sounds. [I had the] impression of a full space. | Adjustment to the level of vertical space. More anchoring to the ground and more projection upwards. More effort to put my stamp on the sound for a greater immersion in the space. |
| Flutist 2 | Pleasant room. Suitable for the music. | No, but pleasure to hear the response of the room. |
| Flutist 3 | The acoustics seem very simple. The sound is alive, the response is very pleasant. The fact of playing “in situ” in the salon of the château is very pleasant, stimulating for the imagination. | I think I slowed down the tempo a bit to allow time to hear the sound resonate and bring the long notes to life. |
| Violist 1 | This space is made for playing the viol. It was an immense pleasure to play in this place, the acoustics are absolutely perfect, the tone reproduced to the highest degree – an unforgettable experience. | I let myself be carried away by the moment. |
| Violist 2 | Very pleasing | Not really. |
| Violist 3 | Versailles “home to” Marin Marais is the best acoustic/ framing for his music. What to say – inspiring. | Yes – more nuance as the room responds. |
| Violist 4 | The size of the room allows you to feel more free, perhaps less intimidating. The instrument being baroque, I feel better in an environment “literally adapted” to my instrument. | Once the reverb was heard/taken into account I was able to play more slowly in general with more straightforward or lighter articulations depending on the movements. The more I hear myself, the more I try to reign in certain details. |
| Amphitheater | | |
| Flutist 1 | My playing seemed to be very exposed where every detail could be heard. I also felt like I needed to project further, to hold on until the very end of the note values in order to reach the back of the room. | I think I accentuated the character of the pieces, sharper for <i>Animé</i> maybe slower of <i>Lentement</i> . I played very in front of me with horizontal projection & note values held to the very end. |
| Flutist 2 | Neutral. | Played longer because of the dryness of the room. |
| Flutist 3 | The space is large, sounds quite generous, it’s a fairly “neutral” room (in decor) so it’s quite easy to concentrate. There is a certain ease in finding yourself in a fairly traditional concert venue. | Feeling like I was in a big space led me to dig deeper into the dynamics for the 2nd and 3rd takes, to support the sound more, to make the phrases longer. |
| Violist 1 | I was a little bothered by the light. The color possibilities are quite huge and the dimensions of the room very pleasant. I was also a little tense, remembering a concert a long time ago in this same place on a Sunday morning at 11 am when I was very nervous! | Yes, as the resonance in the treble = almost zero, I tried to do the maximum without overdoing it I believe. |
| Violist 2 | This space had a very positive impact on my playing. The feeling was very good. | I didn’t need a conscious adjustment. |
| Violist 3 | Fairly neutral. Almost like a studio. | Not much resonance so [I was] aware of this and therefore adapt[ed] to these considerations. |
| Violist 4 | The size of the venue influenced the level of dynamics I used, and the reverb influences the tempi. I always need some adaptation time = the first few minutes of adjustment are difficult then ultimately I’m very comfortable. | I adjusted to the reverberation, adapted the tempi and the length of the notes as well as the dynamics, all this according to the response of the room. |

From these responses it is difficult to tell precisely which acoustical properties caused them to adjust their playing in the amphitheater, however, some comments indicate that a lack of reverberance, relative to the Salon des Nobles, may have been an important consideration. For example, flutist 2 claimed their adjustments were due to the “dryness of the room,” while violist 1 said the “resonance in the treble” was “almost zero,” and violist 3 mentioned there was “not much resonance” in the amphitheater.

1.4 Objective analysis of performance data

Objective analysis of the resulting audio recordings was performed and is described in detail in [3]. The analysis framework was developed in accordance with musicological principles to focus on areas which have been identified as important to communicating baroque expressiveness, notably phrasing, tone production, and vibrato [11].

Musicians have been found to communicate expressive phrasing primarily through manipulations of tempo and

loudness [29]. In order to capture phrasing, smoothed tempo and intensity (calculated by taking the frame-wise root mean square (RMS) of the audio signal) curves were computed of segments of different lengths (1, 2, 4, 8, and 16 bars). Features were then extracted from these segments including range, standard deviation, and coefficients of a 2nd order polynomial fit.

To capture tone production, mel-frequency cepstral coefficients (MFCC) were calculated on the audio signals, as well the harmonic and percussive components of these signals. The MFCCs of the original audio signal will be used in this study since they tended to be the most discriminative, compared to MFCCs from the harmonic or percussive components of the signal.

The vibrato features used include the rate, extent, and coefficients of a polynomial used to model the change in rate over the duration of the note.

1.5 Study objective

This listening test is a crucial component of a larger study investigating the relationship between the acoustics of historical halls and the performance of music from the same era. Prior to the test, objective measurements of recordings from performances in both a historical and modern hall revealed significant differences. Musicians also reported, through questionnaire responses, that they made intentional performance adjustments in response to the acoustics. However, previous research suggests that even large variations in acoustics may lead to only subtle changes in performance. Therefore, the primary aim of this listening test is to determine whether listeners can perceive these changes in performance style that were previously measured and reported.

The results of the test will provide insight into the effectiveness of the objective measures used to capture performance indications. Additionally, the results will indicate whether or not the musicians were able to effectively convey their intentions to the listeners through their performance.

The research question centers around determining whether a historically informed baroque playing style is facilitated by the acoustics of historical halls. Therefore, the parameters selected for this listening test, as well as for the previous objective analysis, were chosen for their relevance to the expression of a historical baroque style.

2 Methodology

2.1 Recording selection

Because the number of recordings from the original experiment was so large, a subset had to be selected for use in the listening test.

Because of the generally subtle observed differences in performance style between the two rooms, a random selection of these recordings would have likely resulted in a generally small effect size. As such, rather than selecting performances at random, the selections were curated to

include two types of performances: one which represented an average performance in that room for that composition and feature type (here referred to as “average” performances), and one which represented the performance in that room which was most different from the performances in the other room for that composition and feature type (here referred to as “extreme” performances). The intention was to measure the maximum conceivable effect of the room by restricting analysis to the extreme performances relative to more representative effects via average performances. The data was partitioned into smaller subsets of individual musicians and compositions. Within each subset and feature set, the Mahalanobis distance¹ was calculated between the cluster centroid of one performance’s group of features (either phrasing, tone production, or vibrato) and the distribution of another performance’s group of features. All performances within each subset were pairwise compared using this distance measure. The data for each composition was treated separately, and no inter-composition distances were calculated. Performances that represented the largest inter-room distance were chosen as the extreme examples while performances that represented the smallest intra-room distance were chosen as the average examples.

Only two of three compositions were chosen for the flute and viol, so as to reduce the total number of examples. These compositions were selected for their ability to highlight the desired features. The selection process resulted in some examples being selected twice, resulting in 15 examples for the flute and 21 examples for the viol. The duration of each recording ranged from approximately 30–60 s. All recordings within each instrument group were normalized to have the same RMS level in order to remove the influence of any potential differences in loudness from one recording to another.

While it is possible that this recording selection process may result in unrepresentative performances being included in the “extreme” examples, this was thought to be preferable to any kind of manual selection which would be subject to bias. A preliminary listening test revealed that certain viol examples noticeably diverged from others. Nonetheless, the extent to which these audible differences would be noticeable among a broader listener group remained uncertain. Consequently, this systematic selection approach was preferred. Furthermore, by restricting the analysis to “average” examples, one can avoid these potentially unrepresentative performances.

2.2 Design

Part of the purpose of this study was to shed light on the effectiveness of the objective analysis framework by comparing those results to perceptual ratings from listeners. In order to facilitate this, the three musical dimensions which made up the framework (phrasing, tone production, and vibrato) were the primary focus of this listening test. Listeners knowledgeable of baroque music were asked to rate the musical examples on a series of eight scales under the three broad categories of phrasing, tone production (including vibrato), and baroque expressiveness. The terms for

Table 3. Listening test rating categories. The right-side descriptions in each category are associated with a more historically-informed baroque playing style.

| Phrasing | |
|------------------------|-------------|
| Continuous | Articulated |
| Strict | Flexible |
| Mechanical | Varied |
| Tone production | |
| Forced/intense | Light |
| Muddy | Clear |
| Straight | Uneven |
| A lot of vibrato | No vibrato |
| Baroque expressiveness | |
| Not at all | Very |

each rating scale are listed in Table 3. Vibrato was only available during the viol examples. These terms were selected after a review of the literature examining listener perception of baroque performance [11, 31, 32]. In general, the right side column of descriptors in Table 3 correlate with performance characteristics which tend to be more appropriate to a historical baroque style. The following definition of baroque expressiveness was provided during the first few examples of the test: *a performance which adopts stylistic attributes which are characteristic of historically-informed baroque performance practice.*

The test was administered over headphones (Sennheiser HD650), in a quiet studio environment, through an application designed using MATLAB App Designer. The interface has a play/pause and stop button for the current audio example. Each rating scale was listed on a 7-point (from -3 to $+3$) continuous slider.

Participants were asked their age, level of general education (currently pursuing bachelor’s degree, currently pursuing master’s degree, or completed master’s degree), and level of familiarity with baroque HIP (not very familiar, familiar, very familiar).

The test was divided into three sections; a preliminary training section in which the participant was given two examples to judge in order to familiarize themselves with the interface, and one section for each instrument. The last two sections were presented in a randomized order, and, for each instrument’s section, the order of the audio examples was also randomized. Two audio examples were duplicated for both the flute and viol in order to have a measure of the repeatability of the participants’ ratings. The participants were allowed to revisit prior examples and adjust their answers if they desired. The average duration of the test was approximately 45 min.

2.3 Participants

Twenty participants took part in the test. All participants were required to have some formal musical training at the university level and at least some familiarity with historical baroque performance practice. Most participants were recruited from the population of music students at

the Clignancourt campus of the Sorbonne University. The test was available in both French and English so that participants could take it in the language with which they were most comfortable (17 chose French, while 3 chose English). All of the terminology was translated with the assistance of a French musicologist. The participants were compensated for the test, and the experiment protocol was approved by the Research Ethics Committee (Comité d’Éthique de la Recherche) at Sorbonne University (CER-2022-ELEY-EVAATest).

The average age of the participants was 24.8 (SD: 6.7). Eight participants self-reported some familiarity with baroque HIP, 11 participants claimed they were familiar with it, while one participant claimed they were very familiar with the subject. Nine participants were current bachelor’s students, six participants were current master’s students, and five participants had completed their master’s degree.

3 Results

Figures 3 and 4 show box plots of results for performances in each category, grouped by room. Participant responses were subject to t -tests for each rating category, with the responses for each room being treated as independent samples. These t -tests were applied to several data subsets: the “average” examples, the “extreme” examples, and all examples together. The null hypothesis for the t -test is that there is no difference between listener ratings based on the room. The resulting p -values are reported in Tables 4 and 5. A Cohen’s d accompanies p -values at the 5% significance level or greater, giving a measure of the effect size, normalized by the standard deviation ([33], pp. 262–264). A resulting positive d value indicated that performances were perceived as more baroque appropriate in that dimension in the Salon des Nobles, the smaller baroque-era room.

As previously mentioned, two duplicates were included in the examples for both instruments in order to examine the reproducibility of the participant responses. The average absolute difference between the ratings across all categories of these duplicates for the flute was 1.11 (on the 7-point scale) with a standard deviation of 0.54. For the viol, the average absolute difference was 1.06 with a standard deviation of 0.41.

3.1 Flute results

No significant differences were observed among the average flute performances according to the t -tests (see Tab. 4). However, among the extreme performances and among the combined set of performances, significant differences were found in every category except for the muddy–clear dimension of tone production and the baroque expressiveness rating.

Within the continuous–articulated dimension of phrasing, the performances in the Salon des Nobles were rated as significantly more continuous, while those in the amphitheater were rated as more articulated ($p = .034$, $d = -0.23$, for all examples). This is contrary to claims

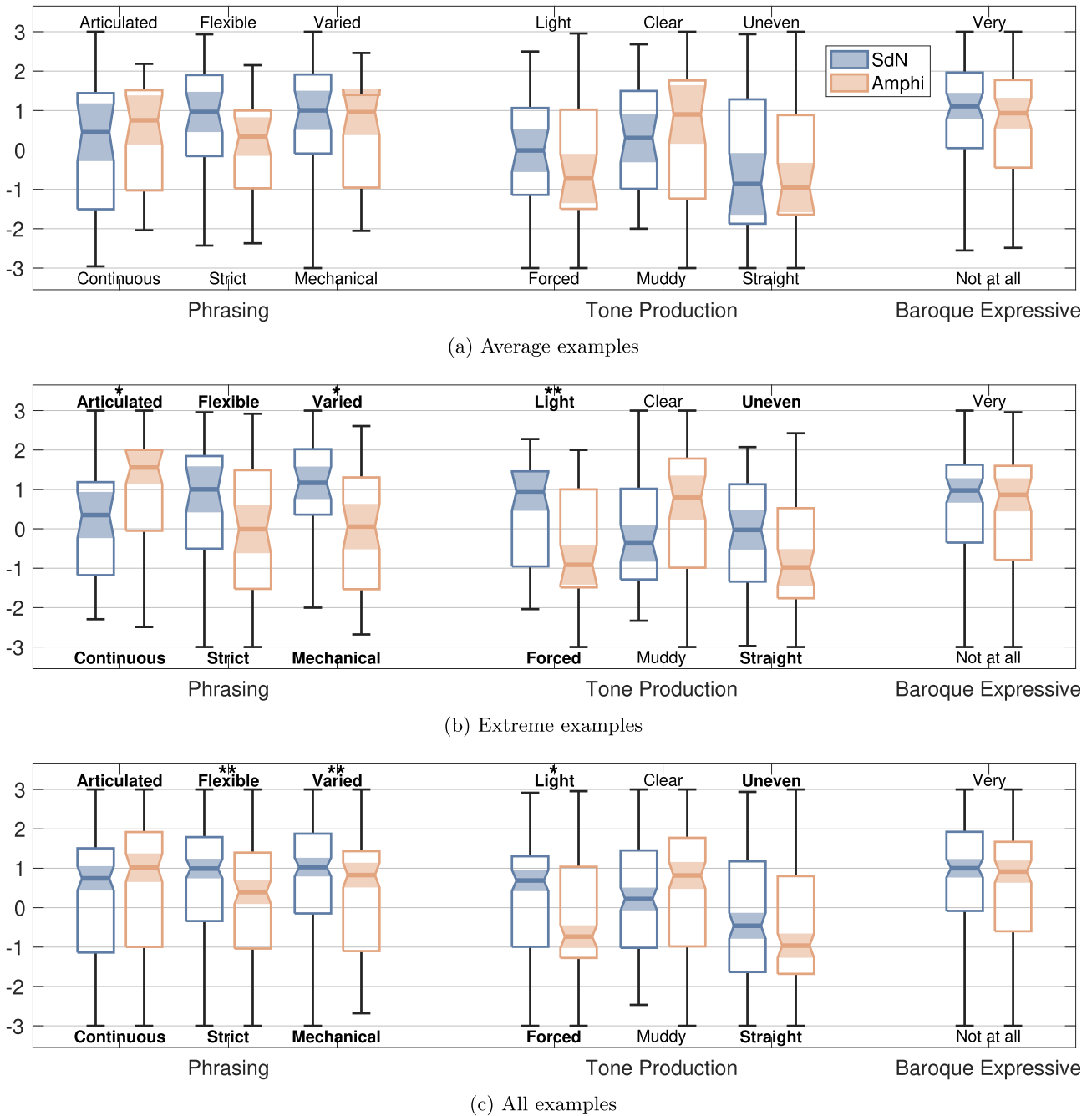
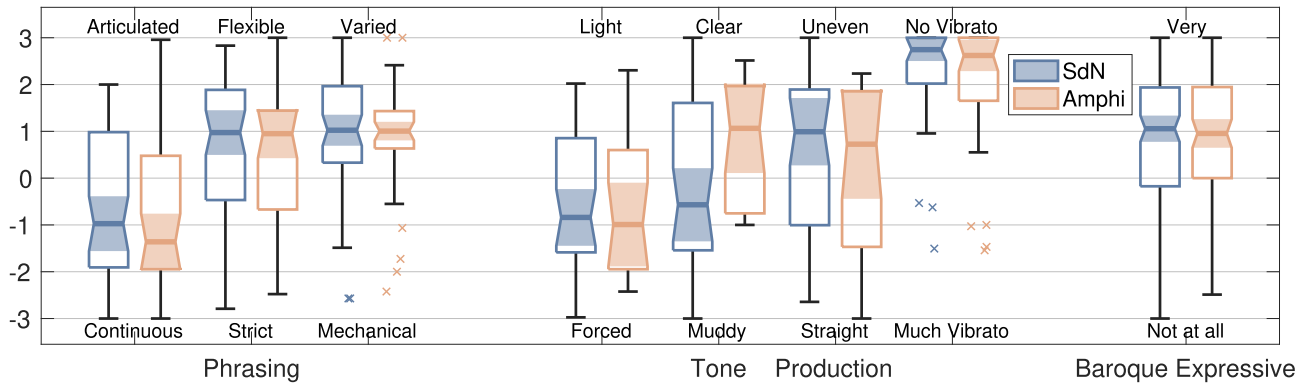


Figure 3. Box plots of listening test results for flute performances showing (a) average examples, (b) extreme examples, and (c) all examples. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima. Bold labels, asterisks (*) and double asterisks (**) indicate a p -values of $<.05$, $<.01$, and $<.001$, respectively.

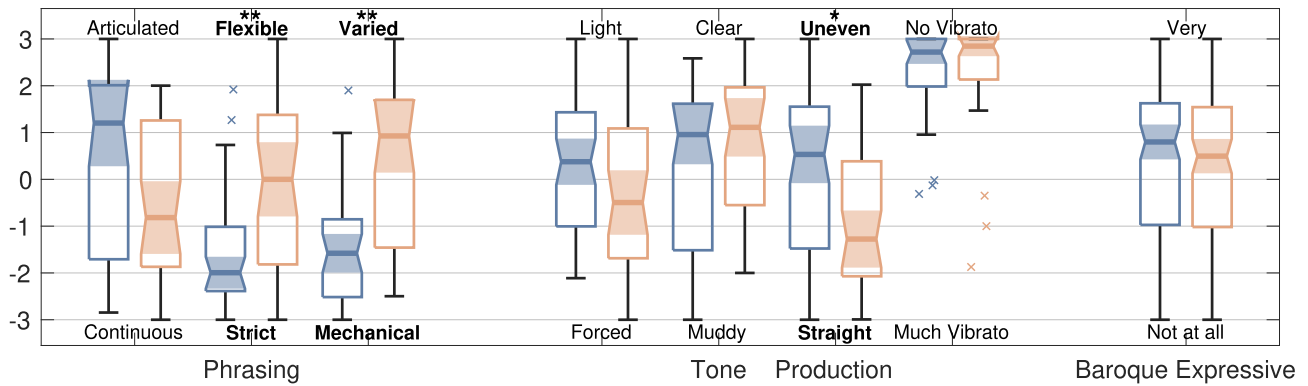
by some of the flutists in their questionnaire responses (see Sect. 1.3) that they intentionally tried to extend the duration of notes when playing in the amphitheater to compensate for the lack of acoustical decay in that hall, indicating that the flutists were perhaps not able to achieve their intended performance goals.

An area where there was agreement between the questionnaire responses and the listener ratings is the forced/intense–light dimension of tone production. Flutists

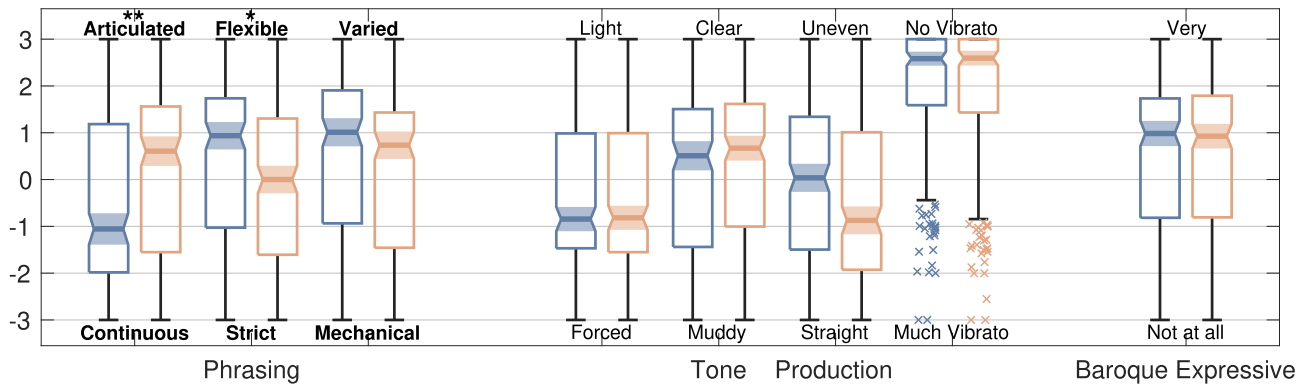
described the need to “project further” or “dig deeper into the dynamics” in the amphitheater, and one would expect this to result in a more “forced/intense” tone. This is exactly what was indicated by the participants in this listening test, with a high significance and moderate effect size ($p = .003$, $d = 0.32$, for all examples). This indicates that the intention to project more in the amphitheater was perceived by listeners. Furthermore, a light tone is associated with a historical baroque playing style meaning that, at least in this



(a) Average examples



(b) Extreme examples



(c) All examples

Figure 4. Box plots of listening test results for viol performances showing (a) average examples, (b) extreme examples, and (c) all examples. The thick line within the box represents the median, the box edges represent the upper and lower quartiles, and the whiskers represent the nonoutlier maxima and minima. Bold labels, asterisks (*) and double asterisks (**) indicate a p -values of $<.05$, $<.01$, and $<.001$, respectively.

dimension, the flutists were perceived as producing a more baroque-appropriate tone in the Salon des Nobles than in the amphitheater.

Highly significant differences and moderate effect sizes were found within the other two dimensions of phrasing; strict–flexible and mechanical–varied ($p < .001$, $d = 0.37$, for all examples in both dimensions). Listeners found the performances in the Salon des Nobles to be more flexible

and varied, while those in the amphitheater were judged to be more strict and mechanical. Flexible and varied phrasing is associated more with a historical baroque performing style, indicating that, among these dimensions, the flutists' performances were generally perceived as being more baroque appropriate in the Salon des Nobles.

The results also suggest that performances in the Salon des Nobles showed a tendency to be perceived as slightly

Table 4. Listening test results for flute examples showing p -values (and Cohen’s d where significant p -values were found) of t -tests of responses for performances in the two different rooms. Bold, single asterisk (*) and double asterisk (**) indicate significance at the $<.05$, $<.01$, and $<.001$, respectively.

| | Average | | Extreme | | All | |
|------------------------|---------|-----|-------------------|-------|-------------------|-------|
| | p | d | p | d | p | d |
| Phrasing | | | | | | |
| Continuous–Articulated | .552 | – | .003* | –0.62 | .034 | –0.23 |
| Strict–Flexible | .071 | – | .034 | 0.44 | <.001** | 0.37 |
| Mechanical–Varied | .512 | – | .004* | 0.60 | <.001** | 0.37 |
| Tone production | | | | | | |
| Forced–Light | .357 | – | <.001** | 0.67 | .003* | 0.32 |
| Muddy–Clear | .885 | – | .163 | – | .287 | – |
| Straight–Uneven | .518 | – | .016 | 0.45 | .017 | 0.26 |
| Baroque expressive | .242 | – | .160 | – | .081 | – |

Table 5. Listening test results for viol examples showing p -values (and Cohen’s d where significant p -values were found) of t -tests of responses for performances in the two different rooms. Bold, single asterisk (*) and double asterisk (**) indicate significance at the $<.05$, $<.01$, and $<.001$, respectively.

| | Average | | Extreme | | All | |
|------------------------|---------|-----|-------------------|-------|-------------------|-------|
| | p | d | p | d | p | d |
| Phrasing | | | | | | |
| Continuous–Articulated | .230 | – | .056 | – | <.001** | –0.34 |
| Strict–Flexible | .744 | – | <.001** | –0.89 | .004* | 0.27 |
| Mechanical–Varied | .867 | – | <.001** | –1.13 | .018 | 0.22 |
| Tone production | | | | | | |
| Forced–Light | .623 | – | .094 | – | .670 | – |
| Muddy–Clear | .054 | – | .076 | – | .163 | – |
| Straight–Uneven | .309 | – | .008* | 0.55 | .053 | – |
| Vibrato | .433 | – | .787 | – | .756 | – |
| Baroque expressive | .540 | – | .880 | – | .715 | – |

more baroque-expressive than those in the amphitheater ($p = .081$, for all examples). The overall ratings in this category indicated that all performances in both rooms were generally interpreted by listeners as being somewhat baroque expressive with mean ratings of 0.81 and 0.55 for the Salon des Nobles and amphitheater performances, respectively (on a scale from -3 , meaning not at all baroque expressive, to $+3$, meaning very baroque expressive).

3.2 Viol Results

As with the flute examples, no significant differences were observed among the average examples of the viol performances (see Tab. 5). Among the extreme examples, significant differences were found within the mechanical–varied and strict–flexible dimensions of phrasing and within the straight–uneven dimension of tone production. The results using all examples showed significant differences only in the three phrasing dimensions.

Among the extreme examples, the viol performances were judged to be significantly more strict ($p < .001$, $d = -0.89$) and mechanical ($p < .001$, $d = -1.13$) in the

Salon des Nobles and more flexible and varied in the amphitheater.

However, when looking at the results for all examples, the effect sizes are in the opposite direction for these two dimensions of phrasing. That is, the performances in the Salon des Nobles were rated as exhibiting slightly more flexible ($p = .004$, $d = 0.27$) and varied ($p = .018$, $d = 0.22$) phrasing, albeit with much smaller effect sizes. This indicates that, among these dimensions and all examples, these performances were perceived as being slightly more baroque appropriate in the Salon des Nobles. Within the remaining dimension in the phrasing category, continuous–articulated, the performances in the Salon des Nobles were rated as more continuous (and therefore *less* baroque appropriate in this dimension) than those in the amphitheater ($p < .001$, $d = -0.34$).

The only significant difference observed within the tone production category was among the straight–uneven dimension with the extreme examples ($p = .008$, $d = 0.55$). These ratings indicated that performances in the Salon des Nobles were rated as having more uneven tone production. When including all examples, this trend was still observed but was not statistically significant ($p = .053$).

As expected, based on the vibrato findings in [3], there was little to no perceived difference in vibrato usage among the performances in the two different rooms. The vibrato ratings indicated that there was very little perceived vibrato overall, aligning with the previously reported objective measures.

There appears to have been no difference in perceived baroque expressiveness as a function of the room within the viol performances, regardless of which subset of performances is examined ($p > .5$ for all performance subsets).

3.3 General discussion

For both instruments, no significant differences were observed when restricting the analysis to the “average” performances. These average performances were selected based on their proximity in a multivariate feature space to all other performances in the same room, making them the most representative performances in that room, for a specific feature and composition. The lack of significant differences found as a function of the room for these performances suggest that average performance adjustments are quite difficult for a listener to discern.

When restricting the analysis to the “extreme” performances only, or across all performances, significant differences were found for both instruments in several dimensions. The extreme performances were selected to be the most different from the average performance in the other room, for a specific feature and composition. This suggests that listeners are able to perceive performance changes on this scale.

Based on the questionnaire responses (see Sect. 1.3), one would have expected the flutists’ performances to be perceived as having more continuous phrasing and a more forced/intense tone in the amphitheater. While listeners did perceive a more forced tone in the amphitheater, they rated the phrasing there as more articulated. This suggests that the musicians were not successful in communicating all of their performance intentions to the listeners. The two primary performance intentions of the flutists in the amphitheater, according to questionnaire responses, were to increase their projection and to lengthen the notes to their full duration. However, an increased projection requires more effort and breath support, and would therefore likely render it more difficult to sustain notes to their maximum duration, meaning these two efforts were somewhat opposed to each other. Therefore, it is probable that only one of these intentions was achieved and communicated to the listener.

The flutists’ attempt to increase projection in the amphitheater may have resulted in their performances being perceived as more strict and mechanical in terms of phrasing. This focus on projection may have made it harder for them to focus on other aesthetic concerns, such as phrasing, which could have led to a more mechanical interpretation.

Significant differences were found in the phrasing category for both the extreme examples and all examples for the viol, but the effect sizes for these two data subsets were

in opposite directions. This suggests that the extreme examples selected to represent the phrasing features for the viol were perceived very differently from the other performances. An investigation into the viol examples revealed that this discrepancy may be partly due to a strikingly different interpretation of one of the compositions by one of the violists (violist 4) who represented the extreme phrasing examples for both rooms for this composition. This musician played the piece, which consisted only of chords, as single, abrupt strokes while all of the other musicians arpeggiated them. This violist performed the entire piece as single strokes in the Salon des Nobles, whereas in the amphitheater they arpeggiated the second half of the piece. This aligns with the results that show the extreme performances in the Salon des Nobles as exhibiting significantly more strict and mechanical phrasing. The recording selection process assumed that all interpretations were reasonable, and therefore the extreme examples would represent the maximum reasonable performance changes exhibited by musicians. However, the performances by this violist may have exceeded expected performance changes.

Figure 5 illustrates how distinct this violist’s performances were, compared to those of the other musicians, as indicated by objective features. It is notable that this violist reported having 3 years of professional experience, whereas the remaining three violists had self-reported 31–41 years of experience (mean: 37.3). It is possible that this difference in professional experience may have been partly responsible for this unconventional interpretation. While this performance stood out as being audibly different from the others, it was still included in favor of an automated approach which would be free of selection bias. Unfortunately, while this decision was taken to avoid selection bias, it may have had the unintended consequence imparting other biases on the viol results. In future experiments of a similar nature, particular attention should be paid to the selection of musicians. It is important to choose experienced professionals who can consistently convey their musical intentions, as opposed to less experienced individuals who may lack this level of control.

The lack of significant difference observed in baroque expressiveness between the two rooms does not necessarily indicate that listeners were not able to discriminate within this global parameter. As shown in [11], listeners discriminated along this broad dimension, however, in that study, the musical examples were chosen to represent a wide range of baroque playing styles and therefore a wide range of baroque expressiveness. In this study, all of the performances generally adopted the same baroque HIP style, so any differences within the baroque expressive parameter would be expected to be rather small.

Due to the fact that significant differences were found in many of the more narrowly-defined parameters within phrasing and tone production, most of which were in the direction of being more baroque expressive in the Salon des Nobles, one might expect a similar difference to be found in the baroque-expressive parameter. This was not the case, however, suggesting that it is easier for listeners to provide

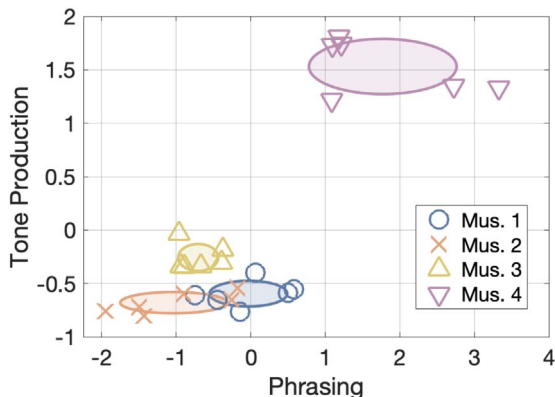


Figure 5. The distribution of objective data of violist performances of one composition in both rooms. The x -axis is the top principal component of the phrasing features while the y -axis is the top principal component of the tone production features. Individual data points as well as the standard deviation around their means are shown. The distributions show how different the interpretation of violist 4 was from the other musicians.

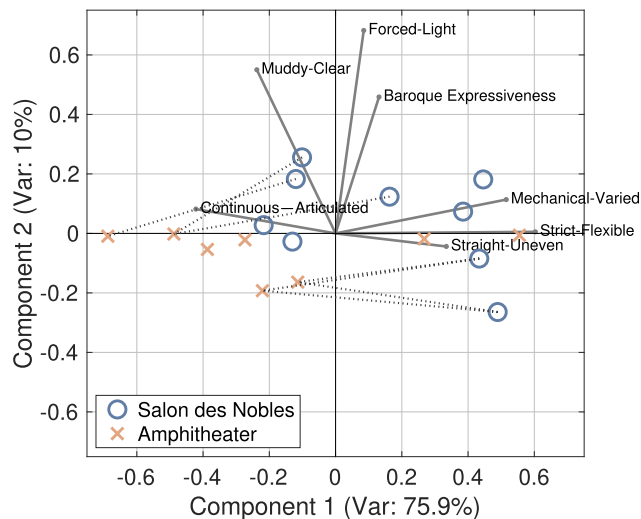
consensus among more narrowly-defined parameters than on global parameters such as baroque expressiveness.

3.4 Principal component analysis of ratings

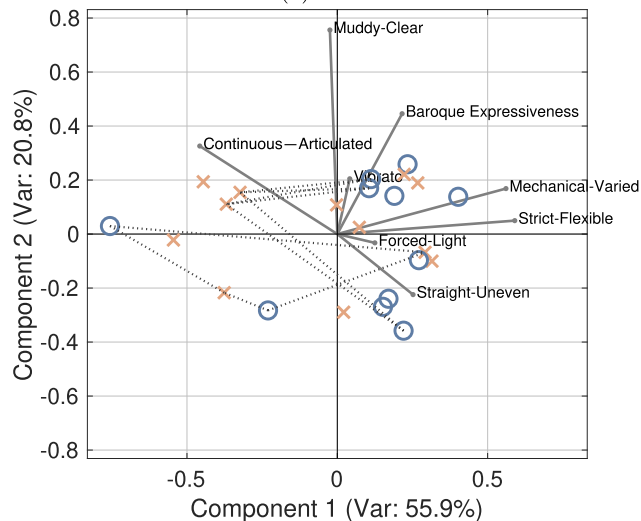
The results of the listening test were subject to principal component analysis (PCA) to explore what makes up the most salient perceptual dimensions. Figure 6 shows biplots of the resulting first two components of this analysis for both the flute and viol ratings, analyzed separately. Dotted lines are included to show connections between extreme examples of phrasing and tone production features of each composition between the two rooms. Connections among the extreme vibrato examples in the viol were left out to simplify the figure, since no difference was found within this dimension as a function of room. The first two components combined are responsible for 86.9% and 76.6% of the variance for the flute and viol examples, respectively.

The flute results (Fig. 6a) show a fairly clear separation between the rooms among the first component, which is mostly made up of the phrasing parameters, along with the straight-uneven dimension of tone production. The second dimension appears to consist of factors more related to tone production, including the forced-light and muddy-clear dimensions. The baroque-expressive dimension is made up somewhat of both dimensions, but is slightly more correlated with the second component. It is notable that the location of the baroque-expressive dimensions suggests that a performance is perceived as more baroque-expressive when it is judged to exhibit a lighter tone and more varied/flexible phrasing. This aligns well with musicological expectations.

The separation of classes (rooms) appears to be more distinct among the first component which is responsible for significantly more of the overall variance (75.9%) than the second component (10.0%). The third component was responsible for only 7.0% of the variance. Most of the vari-



(a) Flute



(b) Viol

Figure 6. Biplots of the first two principal components resulting from a PCA of the listening test responses of the (a) flute and (b) viol examples. Dotted lines show connections between extreme performances of the same compositions in different rooms (excluding vibrato).

ation of the extreme performances seems to be along the first component, as indicated by the dotted lines. This suggests that the first component is the primary dimension along which performances varied as a function of the room.

The viol results (Fig. 6b) show that the strict-flexible and mechanical-varied dimensions of phrasing are strongly correlated with the first component while the muddy-clear dimension of tone production is most strongly correlated with the second component. The baroque-expressive dimension seems to be made up of both components roughly equally. These results suggest that the viol performances were perceived as being more baroque-expressive when the performances were judged to have a clearer tone and more varied/flexible phrasing. This is also compatible with

musicological expectations. There does not appear to be a very clear separation of rooms among either of these components as there was within the flute results. The top two components contribute fairly significant variance each (55.9% and 20.8%), while the third component contributes only 12.2% of the overall variance. The variation of the extreme performances tends to be mixed along the two components, as indicated by the dotted lines. This suggests that there is no primary dimension along which the viol performances varied as a function of the room.

For both the flute and the viol, the strict–flexible and mechanical–varied dimensions appear to be the most active, being strongly correlated with the first principal component. This correlates with most significant differences being found in these dimensions. Furthermore, for both instruments, the mechanical–varied and strict–flexible dimensions of phrasing appear to be roughly opposed to the remaining phrasing dimension, continuous–articulated.

3.5 Comparison with objective ratings

The subjective responses from the listening test were compared with the objective measures described in Section 1.4. First, a subset of the objective features and subjective ratings were chosen (i.e., the objective phrasing features and the ratings of the phrasing dimensions from the listening test). Then, the median across all subjects for each piece and parameter was taken of the listening test ratings resulting in an $m \times n$ matrix where m is the number of performances and n is the number of rating categories. The mean of all objective features was taken for each piece resulting in an $m \times p$ matrix where p is the number of objective features. For both of these matrices and for each performance, m , the pairwise Euclidean distances to all other performances were calculated and then summed, resulting in an $m \times 1$ vector for both the subjective responses and the objective data. Each value in this vector represents the distance of that performance to all other performances for that specific metric (either subjective ratings or objective performance data). Lastly, these two vectors were used to calculate a Pearson correlation coefficient (r). This coefficient, r , serves as a measure of similarity for how a specific set of objective measures and subjective ratings differentiate between performances. Results from these comparisons are reported in Table 6.

In addition to comparisons with the subjective rating categories of phrasing, tone production, and baroque expressiveness, the top two principal components from the PCA performed in Section 3.4 were also included. These components were included because the most salient perceptual dimensions revealed by the PCA may not perfectly align with the two broad categories of phrasing and tone production. This offers another way to observe how the objective data aligns with the perceptual ratings of the listeners. Because very little difference in vibrato was found between the two rooms, according to both the objective measures and the subjective ratings, this parameter was left out of the following analysis.

Table 6. Pearson correlation coefficients (r) between the sum of pairwise distances of all flute and viol performances of listener rating data (columns) and objective performance data (rows). Bold labels, asterisks (*) and double asterisks (**) indicate a p -values of $<.05$, $<.01$, and $<.001$, respectively. BE = baroque-expressive.

| | Phrasing | Tone | BE | PC1 | PC2 |
|----------|--------------|---------------|-------|-------------|-------------|
| Flute | | | | | |
| Phrasing | 0.60 | 0.05 | −0.04 | 0.55 | −0.27 |
| Tone | −0.02 | 0.78** | 0.06 | 0.12 | 0.56 |
| Viol | | | | | |
| Phrasing | 0.18 | 0.42 | 0.23 | 0.28 | 0.19 |
| Tone | 0.54* | 0.34 | 0.21 | 0.43 | 0.09 |

The flute results in Table 6 show strong correlations between the objective measures and the listeners’ perception of the musical parameters they were intended to capture. The phrasing features show strong, significant correlations with the corresponding phrasing ratings while the tone production features show the same with their corresponding ratings. This indicates that these custom objective features are able to identify and isolate the perceptual dimensions they were designed to capture, at least for flute recordings. There is also some correlation between the phrasing features and the first principal component as well as the tone production features and the second principal component, adding further support that these are the most salient perceptual dimensions revealed by the questions posed. There are no significant correlations between the objective measures and the baroque-expressive ratings.

The results for the viol performances in Table 6 show that there is a significant correlation between the phrasing features and tone production ratings, and also the tone production features and phrasing ratings. A significant correlation was also found between the first principal component and the tone production ratings. One violist’s unconventional interpretation (previously discussed in Sect. 3.3) may have contributed to the unclear results in the viol performances. It is possible that these divergent performances were difficult to judge, rendering the viol results difficult to interpret in a generalizable way. An analysis of the average standard deviation of participant responses for each category across all examples did not show a significant difference between instruments, however. This indicates that listener ratings for the flute and the viol were similarly consistent and therefore, the most likely reason for these results is that the objective features were simply not very effective at capturing their intended expressive performance parameters for viol recordings. A direct comparison between the objective phrasing features and the subjective phrasing ratings for the flute and viol can be seen in Figure 7. The flute results show that as the range and standard deviation of the intensity curves increase, the phrasing is perceived as more continuous, flexible, and varied. However, for the viol, the opposite trend is true. In both cases, the features derived from tempo curves seem to be negatively correlated with

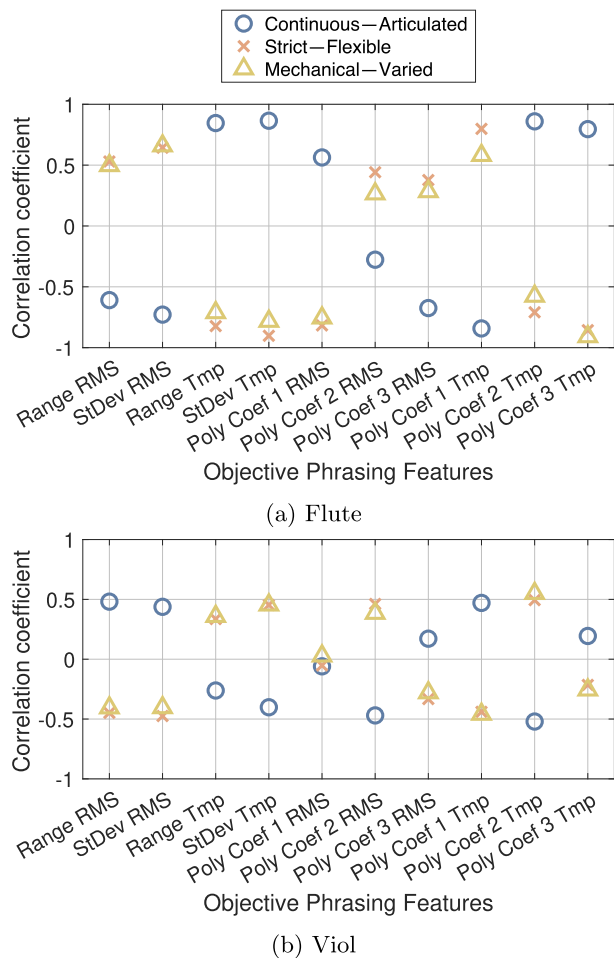


Figure 7. Correlation coefficients between subjective ratings in individual dimensions within the phrasing category and individual phrasing features from the baroque analysis framework of (a) flute and (b) viol performances. Tmp refers to the note-level tempo and RMS refers to the frame-wise RMS level. The objective phrasing features are described in Section 1.4.

those derived from intensity curves. In general, the correlations observed in the flute examples tend to be stronger than those observed in the viol examples.

4 Discussion

The purpose of this study was to gain insight into the perceived differences between performances in two separate rooms, and to determine if trained listeners' perceptions aligned with previously recorded objective differences. No significant differences were found when restricting the analysis to average performances. However, as previous studies have suggested that performance changes due to room acoustics are quite subtle, this outcome was not unexpected. When analyzing the extreme performances, or examples across all performances together, a number of significant differences were found in several performance dimensions for both instruments as a function of room.

The most significant differences with the largest effect sizes were found within the phrasing category for both instruments. Performances in the Salon des Nobles were rated as being more baroque appropriate in the strict-flexible and mechanical-varied dimensions. However, within the continuous-articulated dimension, the performances were rated as being less baroque appropriate in the Salon des Nobles.

Within the flute results, some findings were consistent with performance changes reported by musicians, such as the tone production being judged as more forced in the amphitheater. However, ratings in the continuous-articulated dimensions did not align with the intended performance changes reported by flutists.

There was fairly good agreement between the objective performance data and the listener ratings for the flute. There was a significant correlation between the proposed phrasing features and the listener ratings in the phrasing dimensions, while the same was true for the tone production features and their corresponding ratings. Additionally, the features showed almost no correlation with listener ratings in other categories. This is strong evidence to support the efficacy of these features to capture the expressive musical qualities that they were intended to capture, at least for flute performances.

The relationship between features and listener ratings was not as clear for viol performances. Tone production features correlated with ratings in the phrasing category and phrasing features correlated with ratings in the tone production category. These correlations, though significant, were not as strong as the correlations observed within the flute examples. These somewhat surprising results may be partially explained by an atypical interpretation of one of the violists, since, when these performances were removed, these correlations disappeared. Further analysis indicated that there was a similar consistency in responses for the two different instruments suggesting that the objective analysis was not very effective in capturing the expressive performance parameters of the violists.

This test provided evidence that there are some significantly different perceptual changes in performance style as a function of the room. Furthermore, listeners were able to perceive performance characteristics that were either reported by the performer, observed in objective parameters, or both. The most conclusive findings came from the flute examples, rather than the viol, suggesting that the previously used objective analysis framework does not apply equally well to all instruments. Future research could focus on improving the objective analysis to be more robust to different instruments. The study also provided further insight into how listeners perceive baroque expressiveness, finding that it tends to be correlated with certain dimensions of phrasing and tone production. However, the specific dimensions may be somewhat dependent on the instrument. While the study found some evidence to support that the acoustics of the baroque-era room facilitated the performance of historical baroque music, the variance in results between the two instruments, along with the small sample size, limit broader application of this conclusion.

Conflict of interest

Author declared no conflict of interest.

Data availability statement

Data are available on request from the authors.

Funding

This work was supported by the Paris Seine Graduate School Humanities, Creation, Heritage, Investissement d'Avenir ANR-17-EURE-0021 – Foundation for Cultural Heritage Science. This project takes place as part of http://www.sciences-patrimoine.org/projet/evaa_ver/EVAA_Ver:Expérience_Virtuelle_en_Acoustique_Archéologique. Additional funding provided by the European Union's Joint Programming Initiative on Cultural Heritage project PHE (The Past Has Ears, <http://phe.pasthasears.euphe.pasthasears.eu>, 20-JPIC-0002-FS).

Acknowledgments

We would also like to thank the research center of the Château de Versailles and the Musée de la Musique of the Philharmonie de Paris, as well as the musicians who participated in the study.

References

1. K. Schiltz: Church and chamber: the influence of acoustics on musical composition and performance. *Early Music* 31, 1 (2003) 64–80.
2. N. Eley, C. Lavandier, T. Psychoyou, M. Jossic, B.F.G. Katz: Performance analysis of solo baroque music played in a period and modern hall, in Proceedings of the 16th French Acoustics Congress, Marseille, France, 11–15 April, 2022, 1–6.
3. N. Eley, T. Psychoyou, C. Lavandier, B.F.G. Katz: A custom feature set for analyzing historically informed baroque performances, in Proceedings of the 24th International Congress on Acoustics, Gyeongju, October 24–28, Acoustical Society of Korea, Seoul, South Korea, 2022, 1–89.
4. C. Lawson, R. Stowell: The historical performance of music: an introduction. *Cambridge handbooks to the historical performance of music*. Cambridge University Press, Cambridge, UK, 2003.
5. J. Rink: *Musical performance: a guide to understanding*. Cambridge University Press, Cambridge; New York, 2002.
6. R. Donington: *The interpretation of early music*. Faber and Faber, London, UK, 1963.
7. J. Butt: *Playing with history: the historical approach to musical performance*. Cambridge University Press, Cambridge, UK, 2002.
8. R. Donington: *Baroque music style and performance: a handbook*. W.W. Norton & Company, 1982.
9. R. Donington: *A performer's guide to Baroque music*. Charles Scribner's Sons, New York, NY, 1973.
10. B. Haynes: *The end of early music: a period performer's history of music*. Oxford University Press, New York, NY, 2007.
11. D. Fabian, E. Schubert: Baroque expressiveness and stylishness in three recordings of the D minor Sarabanda for solo violin by J.S. Bach. *Music Performance Research* 3 (2009) 36–56.
12. D. Fabian, E. Schubert: Is there only one way of being expressive in musical performance? Lessons from listeners' reactions to performances of J. S. Bach's music, in: C.J. Stevens, D.K. Burnham, G. McPherson, E. Schubert, J. Renwick, Eds. Proceedings of the 7th International Conference on Music Perception and Cognition. Causal Productions [for] AMPS, Adelaide, South Australia, 2002, 112–115.
13. D. Fabian, E. Schubert, R. Pulley: A Baroque Träumerei: the performance and perception of two violin renditions. *Musiology Australia* 32 (2010) 27–44.
14. K. Kato, K. Ueno, K. Kawai: Musicians' adjustment of performance to room acoustics, part III: understanding the variations in musical expressions. *Journal of the Acoustical Society of America* 123 (2008) 3610.
15. K. Kato, K. Ueno, K. Kawai: Effect of room acoustics on musicians' performance. part II: audio analysis of the variations in performed sound signals, *Acta Acustica united with Acustica* 101 (2015) 743–759.
16. T. Fischinger, K. Frieler, J. Louhivuori: Influence of virtual room acoustics on choir singing. *Psychomusicology: Music, Mind, and Brain* 25, 3 (2015) 208–218.
17. S. Amengual Gari, M. Kob, T. Lokki: Analysis of trumpet performance adjustments due to room acoustics, in: Proceedings of the International Symposium on Room Acoustics (ISRA 2019), Amsterdam, 15–17 September, 2019, 65–73.
18. N. Eley, S. Mullins, P. Stitt, B.F.G. Katz: Virtual Notre-Dame: preliminary results of real-time auralization with choir members, in: 2023 International Conference on Immersive and 3D Audio (I3DA), Bologna, Italy, September 8–10, 2021, 1–6. Video available at <https://youtu.be/g6fwv8FzjS4>.
19. S. Mullins, V. Le Page, J. De Muyenke, E.K. Canfield-Dafilou, F. Billiet, B.F.G. Katz: Preliminary report on the effect of room acoustics on choral performance in Notre-Dame and its pre-Gothic predecessor. 150 (2021) A258.
20. W. Chiang, S.-T. Chen, C.-T. Huang: Subjective assessment of stage acoustics for solo and chamber music performances. *Acta Acustica United with Acustica* 89, 5 (2003) 848–856.
21. P. Luizard, J. Steffens, S. Weinzierl: Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions? *Journal of the Acoustical Society of America* 147 (2020) EL132–EL137.
22. A.H. Marshall, D. Gottlob, H. Alrutz: Acoustical conditions preferred for ensemble. *Journal of the Acoustical Society of America* 64 (1978) 1437–1442.
23. K. Kawai, K. Kato, K. Ueno, T. Sakuma: Experiment on adjustment of piano performance to room acoustics: Analysis of performance coded into MIDI data, in: Proceedings of the International Symposium on Room Acoustics (ISRA 2013), Toronto, 9–11 June, 2013, 1–6.
24. Z. Schärer Kalkandjiev, S. Weinzierl: Playing slow in reverberant rooms – examination of a common concept based on empirical data, in: A. Mayer, V. Chatzioannou, W. Goebel, Ed. Proceedings of the Third Vienna Talk on Music Acoustics, 16–19 September, Institute of Music Acoustics, (Wiener Klangstil), University of Music and Performing Arts Vienna, Vienna, 2015, pp. 215–219.
25. P. Luizard, E. Brauer, S. Weinzierl: Singing in physical and virtual environments: how performers adapt to room acoustical conditions, in: Proceedings of the AES International Conference on Immersive and Interactive Audio, 27–29 March, York, UK.
26. S. Bolzinger, J. Risset: A preliminary study on the influence of room acoustics on piano performance. *Journal de Physique IV* 2 (1992) C1-93–C1-96.
27. S. Bolzinger, O. Warusfel, E. Kahle: Study of the influence of room acoustics on piano performance. *Journal de Physique Colloque* 4 (1994) 617–620. <https://doi.org/10.1051/jp4:19945132>.

28. M. Kob, S. Amengual Gari, Z. Schärer Kalkandjiev: Room effect on musicians' performance, in: J. Blauert, J. Braasch, Eds. *The technology of binaural understanding*, Springer International Publishing, 2020, pp. 223–249.
29. C. Palmer: Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance* 15, 12 (1989) 331–346.
30. P.C. Mahalanobis: On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* 2 (1936) 49–55.
31. E. Schubert, D. Fabian: A taxonomy of listeners' judgments of expressiveness in music performance, in: D. Fabian, R. Timmers, E. Schubert, Eds. *Expressiveness in music performance: Empirical approaches across styles and cultures*, Oxford University Press, 2014, pp. 283–303.
32. E. Schubert, D. Fabian: The dimensions of baroque music performance: a semantic differential study. *Psychology of Music* 34 (2006) 573–587.
33. R.S. Witte, J.S. Witte: *Statistics*, 11th edn., Wiley, Hoboken, NJ, 2017.

Cite this article as: Eley N. Lavandier C. Psychoyou T. & Katz BFG. 2024. Listener perception of changes in historically informed performance of solo baroque music due to room acoustics. *Acta Acustica*, 8, 6.