Adaptation of singers to physical and virtual room acoustics

Paul Luizard, Jochen Steffens, Stefan Weinzierl

To cite this version:

Paul Luizard, Jochen Steffens, Stefan Weinzierl. Adaptation of singers to physical and virtual room acoustics. International Symposium on Room Acoustics, Sep 2019, Amsterdam, Netherlands. hal-02356057

HAL Id: hal-02356057
https://hal.archives-ouvertes.fr/hal-02356057

Submitted on 8 Nov 2019

ISRA 2019
AMSTERDAM
PROCEEDINGS of the
International Symposium on Room Acoustics
15 to 17 September 2019 in Amsterdam, Netherlands

Adaptation of singers to physical and virtual room acoustics

Paul LUIZARD; Jochen STEFFENS; Stefan WEINZIERL

Audio Communication Group, Technische Universität Berlin

ABSTRACT

As observed with instrumentalist musicians, singers are expected to react to variations of the acoustics of the venues where they perform by adapting their voice production. To which extent do these changes happen and how are they related to specific variations of room acoustic conditions? And does it make a difference whether they are physically present in the room or whether the room is simulated electro-acoustically? These questions were addressed by recording two musical solo pieces sung by four singers in eight physical acoustical environments. In addition to close-microphone recordings, binaural room impulse response datasets were measured at the position of the singer in order to reproduce the acoustical space through dynamic binaural synthesis. Room acoustical simulations were performed corresponding to the physical rooms and measurement configurations. The experiment was then replicated in an anechoic chamber where the singers would hear themselves in the various measured and simulated virtual spaces. The performances were analysed through automatic musical feature extraction and statistically related to the room acoustical parameters of each venue by means of mixed regression models. Results revealed statistically significant, but highly individual adaptation strategies, both in physical and virtual environments.

Keywords: Singing voice, Adaptation, Virtual acoustics

1. INTRODUCTION

How do singers vary their musical performances from one room to another? To which extent are the changes due to room acoustics or other factors? The present study is focused at the strategies of adaptation to room acoustical conditions used by singers, and aims at investigating their behaviour while performing in physical or virtual acoustical spaces. As previously observed with instrumentalist musicians, singers are expected to react to variations of the acoustics of the venues where they perform by adapting their vocal production.

Previous research on musical adaptation to room acoustics mostly involved instrumentalist musicians. In a recent study (1), recordings of a solo cellist playing the same musical program during concerts in seven different halls were analysed and compared to acoustical measurements of these venues. A statistical analysis of this data showed that more than 50% of the variance across different performance features could be explained by the acoustical properties of the halls. Following a similar methodology, the same authors conducted another study (2) where 12 solo instrumentalist musicians played six musical instruments in 11 virtual environments. The musical features, such as tempo and loudness, extracted from the recordings were compared to room acoustical parameters calculated according to ISO-3382-1 (3) from simulated room impulse responses. Here, a much lower amount of variance in musical performance (8%) could be explained by room acoustical features. Moreover, *individual* patterns of adaptation were observed, indicating that musicians developed their own approaches to play a given musical piece in different acoustical conditions. It should be noted that the different musicians played different musical excerpts, although each musician played the same repertoire in each virtual room, like the cellist in the previously mentioned study. The observed difference of variance explained by room acoustics could be due to the fact that certain musicians are less sensitive to the influence of room acoustics, and also that playing in virtual acoustics is a rather different experience than playing in physical rooms.

Simulation techniques from virtual acoustics were used in another study (4) to render different room acoustics through a multichannel loudspeaker array in an anechoic chamber where solo instrumentalists were recorded. Results of this study showed that variations in several performance features corresponded to questionnaire reports made by the musicians after the recordings where they were asked to evaluate their performance in a qualitative manner (i.e. lower or larger tempo, degree of staccato, or sound intensity), suggesting that the changes such as emphasised *staccato* or longer pauses between notes in longer reverberation, were made consciously (5).

Research involving singers and their adaptation to room acoustics mainly focused on choir performance. A pioneer study (6) analysed recordings of a choir of four singers in a semi-anechoic chamber, exposed to discrete early reflections combined with reverberated sound. It was shown that strong early reflections were appreciated if they did not arrive more than 40 ms after the direct sound. More recent research (7) investigated the impact of early reflections added to the natural reverberation of a church on small choirs of two to four singers. Results showed that singing in fast tempo benefits from strong early reflections while slow tempo singing is not significantly affected, indicating that an influence of room acoustics on musical performance might depend on the character of the musical piece performed. This question was previously addressed in the case of organ players (8) for whom the musical content had a key influence regarding their adaptation to various virtual room acoustics. In particular, multiple breaks between musical phrases within a piece were longer in higher reverberation levels, yielding lower tempo than pieces with few breaks. Another important question for musicians is the ability of hearing themselves among other sound sources. A previous study (9) suggested the self-to-other ratio which quantifies the difference of sound energy between the voice of a singer and the environment, and showed that room acoustics influenced both the sound power and the long-term averaged spectrum of choir recordings.

This literature review shows that little research has been conducted on solo singers regarding various aspect of room acoustics that might influence their performance. In addition, the potential difference of singing behaviour in physical and virtual rooms has not been clearly investigated so far, which is interesting for both methodological and psychological reasons. To which extent do changes in musical performance happen and how are they related to specific variations of room acoustical conditions? And does it make a difference whether the singers are physically present in the room or whether the room is simulated electro-acoustically? The present study has addressed these questions by running an experiment with singers who would perform several musical pieces in various physical and virtual rooms. By analysing the recordings of singing performances together with the room acoustical measurements in terms of audio features for the room and the musical performance, the degree of correlation between the musical performance and the acoustics of the room could be estimated.

2. EXPERIMENT

2.1 Measurements in physical rooms

The rooms where the experiment took place are depicted in Fig. 1. The rooms were chosen so that the variation of their acoustics was large, ranging from a recording studio, concert halls of various sizes, up to a church. In each venue, room acoustical measurements were conducted according to ISO3382-1 (3), using a dodecahedral sound source (Norsonic Nor276 and amplifier Nor280) generating logarithmic sine sweeps and an omnidirectional class-1 microphone (NTi Audio M2230) at several positions within the audience as well as on stage at the position of the singer. The interaural cross-correlation (IACC) was measured on stage by means of a dummy head (Neumann KU100) and a small broadband speaker (Fostex 6301B) with a directivity close to the human voice, and positioned just before the mouth of the head. The dummy head and the speaker were placed on a rotating plate to record binaural room impulse responses for a range of head orientations from -45° to +45° in steps of 5°, with 0° being the orientation of the singer facing the middle of the room.

The four singers who participated in this experiment were students with more than 10 years of musical training, specialised in *opera* and *lieder* singing styles. They were recorded in each room by means of a close-field professional microphone (DPA 4060). They all had the same initial position and orientation towards the end of the audience area and were allowed to slightly move to ensure ecological validity. They were asked to sing three excerpts of musical pieces that they were used to perform, so that the musical content adds as little variability as possible on the performances. Each singer was asked to choose two musical pieces, one fast and one slow (Tab. 1). Due to their different vocal ranges, it was impossible to ask the singers to perform the same lyrical pieces. However, to record a piece common to all singers, they were asked to sing “Happy birthday” in a lyrical manner and in their own vocal register. As a whole, the recordings occurred over a time period of six weeks.

Recording studio Chamber music hall Middle size concert hall

Cabaret theater Philharmonie Berlin Church

Figure 1 – The six venues where the singers were recorded.

2.2 Simulation of virtual acoustical spaces

The experiment was further replicated under virtual conditions in an anechoic chamber a few days after the last recording session in the physical rooms. This methodology was applied to let the singers perform pieces in the eight acoustical spaces within a reduced time (here two hours including pauses), thereby avoiding certain biases related to their psychological and physical state which might vary from one day to another. In addition, this method allows for eliminating influences of the respective visual environment in which the singers perform, and for focusing on the auditory modality only.

In our study, the virtual rooms were simulated by means of convolving impulse responses without direct sound in real-time with the singing voice captured by the same nearfield microphone that the singers wore in the physical rooms, as can be seen in Fig. 2. This task was performed by the SoundScape Renderer (SSR) software (10). The calibration of the simulated impulse responses was adjusted according to the method proposed in previous research (11), involving the simulation of an anechoic chamber to obtain the level of the isolated direct sound. The measured BRIRs were then calibrated to match the simulated direct sound.

2.2.1 Virtual rooms based on measurements

The headphones (AKG K701) used to reproduce the room response to the singers were equalised by the inverse filter of their transfer function. Although these headphones were open, a little amount of direct sound, which remained identical for all singers and rooms, was added to the feedback signal to compensate for the slight acoustic insulation. A head tracker (Polhemus Patriot) was used to select the interpolated room impulse responses according to the head orientation in the horizontal plane.

Table 1 – Singers and musical pieces chosen for the experiment. The singers all had at least 10 years of musical practice.

Voice type	**Age**	**Chosen musical pieces**	**Pace**
Soprano	23	W.A. Mozart – Ach ich liebte (Aria from *Entführung aus dem Serail*)	Fast
		G. Puccini – Aria from *Manon Lescaut*	Slow
Mezzo soprano	27	J.S. Bach – Esurientes implevit bonis (Aria from *Magnificat*)	Fast
		F. Schubert – Nur wer die Sehnsucht kennt	Slow
Tenor	21	F. Schubert – Ganymed	Fast
		A. Cesti – Intorno all idol mio	Slow
Baritone	26	G.F. Händel – Honor and arms (Aria from *Samson*)	Fast
		G. Fauré – Au cimetière	Slow

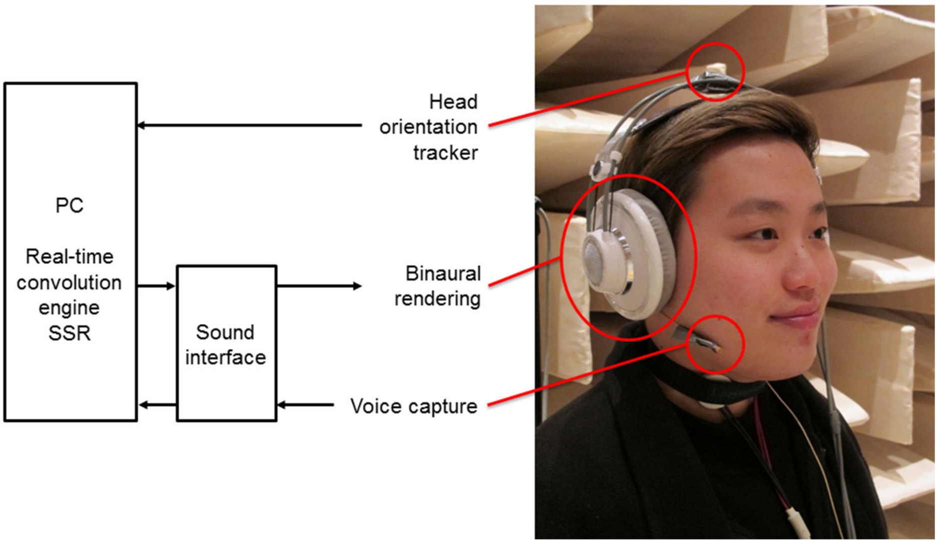

Figure 2 – Schematic of the real-time rendering system for binaural synthesis.

In a first phase, the binaural room impulse responses measured *in situ* with the artificial head and the speaker together positioned on the rotative plate were used. Hence the singers would hear the response of the rooms through the ears of the artificial head, dynamically oriented in the virtual room.

2.2.2 Virtual rooms based on simulations

In a second phase, a set of binaural room impulse responses was generated by means of ray-tracing simulations using the RAVEN software (12). 3D models were built according to geometric measurements of the rooms where the measurements had been performed. These acoustical models were calibrated according to the target reverberation times (T_{30}) measured in the audience and on stage of each physical room. The calibration was realised through an iterative process by tuning the absorption coefficients over third octave frequency bands of the "plaster" material, whose actual absorption can vary much depending on its thickness and installation, and which represented the largest surfaces in most room models. All other materials (wood, glass, brick, curtains, seats) were kept at their default absorption values in the RAVEN database. After the calibration phase, simulations were run with a specific sound directivity for the source, from measurements of a soprano singer in a spherical microphone array (13), and the head related transfer functions (HRTF) of the FABIAN head and torso simulator (14) for the receiver. Both source and receiver were placed on stage at the same position as the sound source and artificial head during the measurements. Each source and receiver orientation yielded simulated binaural room impulse responses that would be used later to reproduce, in the anechoic chamber, the experiment which was first run in the physical rooms.

2.3 Feature Extraction

Room acoustical parameters were then calculated out of the measured impulse responses and musical features were extracted from the singing recordings. The room acoustical parameters derived from the impulse responses measured on stage, 1 m away from the source in its horizontal plane, are shown in Tab. 2. They were chosen for their perceptual relevance according to the literature, where EDT was found to be a good predictor of the perceived reverberance (15), ST_{early} (or ST1) as a measure of the early sound energy on stage (16), G_{late} as a relevant estimation of the subjective level of late sound (17), $IACC_{late}$ as a measure of the listener envelopment (18), and BR as an estimator of the perceived tonal colour of the room (19).

The musical features were extracted in two steps. First, low-level features were calculated, accounting for both the recorded audio signal and the MIDI score of the musical piece, at the single notes and bar scales. This calculation requires the detection of each single note, which is a difficult task with singing voice signals due to the relative slow attack of each note, as compared for example to percussive instruments for which the onset detection algorithms yield much better results (20). Hence the onsets had to be verified and potentially manually modified. The second step involves another set of musical features which are interpretable in terms of musical expression of a performance. They were estimated in a previous study (21) involving musical experts who verbally characterised musical phrases. These verbal attributes were statistically related to the low-level features through linear regressions, yielding the possibility to quantify a musical recording by means of musically meaningful features. In the present study, eight musical features were used, possibly grouped into three musical dimensions, as proposed in Tab. 3. The four singers sung two musical pieces in eight different rooms related to three environments (physical, virtual based on measurements and on simulations), yielding 72 recordings of each piece composed of 100 notes on average. Hence about 15,000 values for each of the eight musical features were estimated in this experiment.

Table 2 – Room acoustical parameters of the venues where the experiment took place, averaged over the 500 Hz and the 1 kHz octave bands, except for $IACC_{late}$: 125 Hz to 4 kHz octave bands, according to ISO3382-1 (3). Abbreviations w/ & w/o stand for "with" & "without".

Room name	**Volume (m^3)**	**$IACC_{late}$**	**EDT (s)**	**ST_{early} (dB)**	**G_{late} (dB)**	**BR**
Recording studio (w/o banners)	420	0.73	0.7	-5.5	12.2	1.18
Recording studio (w/ banners)	420	0.74	0.5	-7.1	8.2	1.55
Kammersaal (chamber music)	590	0.74	1.0	-4.9	13.8	1.17
Cabaret theater Distel	1700	0.76	0.5	-9.3	8.5	1.23
Joseph Joachim Saal (w/o banners)	3660	0.77	1.8	-11.1	10.8	0.94
Joseph Joachim Saal (w/ banners)	3660	0.77	1.3	-11.4	8.9	1.21
St Eduard Church	9360	0.54	4.3	-13.9	14.0	0.97
Philharmonie Berlin	22000	0.74	1.2	-17.6	3.8	0.88
Relative differences (*)	-	11%	16%	18%	14%	8%

(*) Relative differences of room acoustic parameters between measured and simulated impulse responses, averaged across all rooms. The simulated absorption was calibrated using T30 (4% of resulting relative difference).

2.2.4 Statistical analysis

The design of the experiment (i.e. recording several pieces performed by different singers in a number of rooms of various types) led to a hierarchical data structure including repeated measurements / levels for each singer, room, and type of environment. To analyse our data, we computed several linear mixed models (22) where both fixed effects (here the room acoustical parameters that are expected to influence the singers) and random effects (the respective levels) are considered together. Before the analysis, all features were z-transformed (i.e. normalised by their mean and standard deviation values over the eight rooms) in order to compare their respective predictive contributions by means of their regression estimates *b*.

Table 3 – Musical features and associated musical performance dimensions.

Musical features	Tempo	Agogic	Loudness	Dynamic range	Hardness	Brightness	Fullness	Bandwidth
Dimensions	Time		Dynamics		Timbre			

3. RESULTS

First, intercept-only models were computed to estimate the proportion of variance that was explained by the different levels of the data (i.e. Interclass Correlation Coefficient [ICC]), for each musical feature. Results revealed that, averaged across all musical features, the variance in musical features was mostly explained by the pieces (ICC_{pieces}= 28%) and the singers ($ICC_{singers}$= 18%), while the different rooms and environment types had a much smaller influence (ICCs < 1%).

In the next step, to investigate the influence of the room acoustical features on the singing performance, several mixed linear models predicting the different musical features were computed including the "singer" and "piece" levels as random effect. Results of these models revealed no significant relationship between the musical features and the room acoustical parameters. Accordingly, the variance explained by the room acoustical parameters, indicated by the marginal R^2, was very low, with a maximum of 2%.

As no general effect of the room acoustics on the musical performance was found, we analysed the data in a more detailed manner which resulted in building mixed models for each singer and each musical feature, with the room acoustical parameters as fixed effects and the musical pieces as random effects. These models yielded a number of significant relations which are depicted in Fig. 3. The figure illustrated that the Hardness feature, which is one of the timbre descriptors, showed the largest amount of significant relations with room acoustical features while the Agogic feature representing the variation of tempo at bar level showed no significant relation at all. Considering the room acoustical parameters, it can be seen that BR is involved in the largest amount of significant relations, while EDT is the least involved.

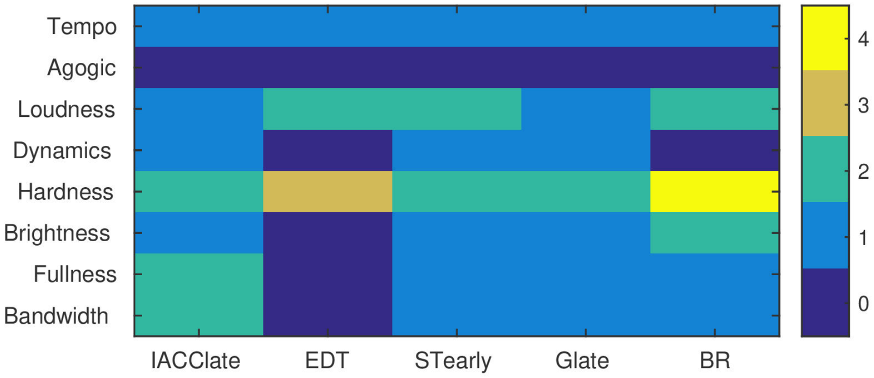

Figure 3 – Number of significant ($p<0.05$) relations between the musical features and the room acoustical parameters.

Looking at the $R^2_{marginal}$ of the respective models, it becomes obvious that the explained variance of the singer-specific models can explain much more variance than the general models. More concretely, the singer-specific models related to the physical rooms explained 15% of the variance on average by the room acoustical parameters, whereas the models related to the virtual rooms based on measurements and simulations obtained explained 7% and 11% of the variance, respectively.

Focusing on each significant relation between the room acoustical parameters and the musical features, as illustrated in Fig. 4, where the sign of the slopes, either positive or negative, provides additional information on the way singers adapted to the rooms. Fig. 4 indicates that a different number of significant relations was observed for each singer. It further suggests that, in case of Singers 1 and 2, only a few musical features were statistically related to room acoustical features. In contrast, Singers 3 and 4 showed many significant relations differing in nature and distribution. For example, Singer 3 adapted mostly timbre aspects and varied the loudness (i.e. Dynamics) of their voice according to the rooms while Singer 4 modulated Tempo, Loudness, and timbre features across rooms. In addition, we observed that Singer 3 presented significant relations almost exclusively in physical rooms while it happened in virtual rooms based on measured impulse responses for Singer 4, except for the Tempo feature which was related to physical rooms.

Furthermore, we analysed the sign of these slopes across singers. This analysis revealed that for Singers 1, 3, and 4, the musical features varied in the same direction, as opposed to Singer 2. It can also be observed that, for most relations, $IACC_{late}$, EDT, and ST_{early} present a common variation, while G_{late} and BR followed the opposite direction.

4. DISCUSSION

With respect to the central hypothesis of the present study, the results demonstrate that room acoustical conditions can explain a part of the variations in singing on the level of each singer within the nested data. However, rather than a common pattern of reaction, each singer developed a personal strategy of adaptation to various acoustical conditions of the environment. The extent of adaption is comparable to what was observed in a previous study involving several instrumental musicians (2). In contrast, a previous study involving only one musician who played the same repertoire in various rooms (1) yielded much higher musical variance explained by room acoustics. Hence the musical content might have a higher influence on the results than what the mixed models revealed, as previously showed for instrumentalist musicians (8).

The direct comparison of the explained variance in the singer-specific models across physical ($R^2_{marginal}$ = 15%) and virtual room based on acoustical measurements ($R^2_{marginal}$ = 7%) and simulations ($R^2_{marginal}$ = 11%) suggests that physical rooms evoke stronger adaptions of the vocal performance than the virtual rooms. This might be due to the full multisensory and ecologically valid experience offered by the real physical rooms as opposed to the purely acoustical experience in the virtual environments which did not offer any visual cues related to the simulated rooms, although a recent study (23) stated that visual cues do not affect the perception of very different reverberated sound fields. These discrepancies might also be due to slight acoustical differences between the physical rooms and their auralisation, essentially due to the non-individualised HRTFs.

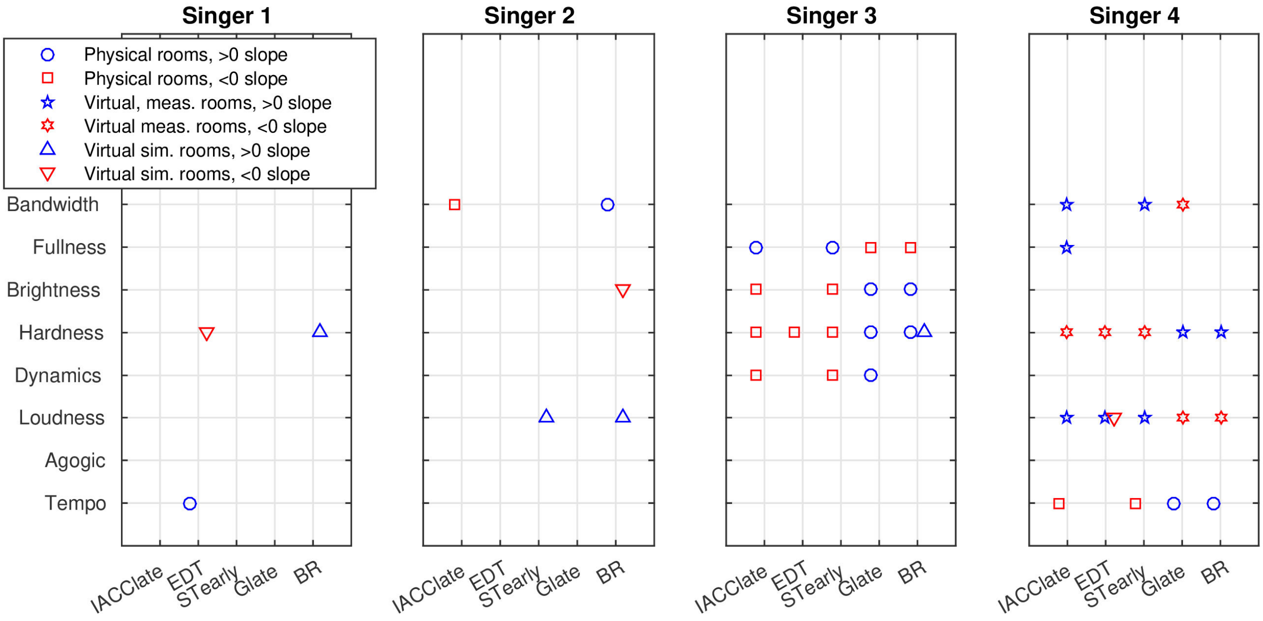

Figure 4 – Significant ($p<0.05$) linear relations between the room acoustical parameters and the musical features over all tested rooms in the framework of linear mixed models. Blue and red markers indicate positive and negative regression slopes respectively.

However, while three of the four singers would show more adaption in the physical rooms, one singer showed considerably more significant relations between the room acoustical conditions and his way of singing in the virtual condition. Also the willingness to get musically involved in the conditions of a virtual acoustic environment thus seems to be individually different. In general, however, the results tend to confirm the observations of two previous studies (1,2) where the musical variance could be explained by room acoustics to a much higher extent for a musician playing in physical rooms than for musicians in virtual rooms.

5. CONCLUSIONS

Four singers sung two musical pieces in eight rooms which were presented to the musicians in three different ways. The general hypothesis of this study was that the singers would adapt their way of singing to the room acoustical conditions of different rooms. To assess such relations, linear mixed models were used, yielding the proportion of variance in the singing recordings explained by room acoustics together with information regarding the specific relations between musical features and room acoustical parameters.

Results showed that no common pattern of adaptation could be observed across singers. Instead, the singers used strongly individual approaches in adapting their vocal performance to room acoustics. Timbre features such as Hardness were most often involved in the adaptation process, as well as features related to sound level. Besides, no specific room acoustical parameter was found to be responsible for the observed changes in singing. Instead, all of the selected parameters in this study seem to play a role, although the pattern was again different for the different singers. The adaptation of singers to room acoustics was observed to be slightly stronger in physical than in virtual rooms, probably due to ecologically fully valid environment including all sensory modalities, contrary to the virtual rooms where only sound rendering was enabled. However, the difference of explained variance of singing performance in each type of environment is not large (15% in physical rooms, 7% and 11% in virtual rooms based on measurements and simulations respectively), which suggests that dynamic binaural rendering is a suitable approach to study the behaviour of musical performers in virtual acoustic environments.

ACKNOWLEDGEMENTS

This work was partially funded by the A. von Humboldt Foundation. The authors would like to thank the singers for their participation in this study, the room managers for their collaboration, E. Brauer for helping on recording the singers, and C. Böhm for his help in running the real-time rendering system.

REFERENCES

1. Schärer Kalkandjiev Z, Weinzierl S. The Influence of Room Acoustics on Solo Music Performance: An Empirical Case Study. Acta Acust United Acust. 2013;99.
2. Schärer Kalkandjiev Z, Weinzierl S. The Influence of Room Acoustics on Solo Music Performance: An Experimental Study. Psychomusicology Music Mind Brain. 2015;25(3).
3. ISO 3382-1. Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces. 2009.
4. Ueno K, Kato K, Kawai K. Effect of room acoustics on musicians' performance. Part I: Experimental investigation with a conceptual model. Acta Acust United Acust. 2010;96(3):505–515.
5. Kato K, Ueno K, Kawai K. Effect of Room Acoustics on Musicians' Performance. Part II: Audio Analysis of the Variations in Performed Sound Signals. Acta Acust United Acust. 2015;101:743–59.
6. Marshall AH, Meyer J. The Directivity and Auditory Impressions of Singers. Acustica. 1985;58:130–9.
7. Noson D, Sato S, Sakai H, Ando Y. Singer responses to sound fields with a simulated reflection. J Sound Vib. 2000;232(1):39–51.
8. Amengual Gari S. Investigations on the Influence of Acoustics on Live Music Performance using Virtual Acoustic Methods. HfM Detmold; 2017.
9. Ternström S. Choir acoustics: an overview of scientific research published to date. Int J Res Choral Sing. 2003;1(1):3–12.
10. Geier M, Ahrens J, Spors S. The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In: In 124th AES Convention in Amsterdam, The Netherlands, May 17–20. 2008.
11. Schärer Kalkandjiev Z. The Influence of Room Acoustics on Solo Music Performances. An Empirical Investigation. Technische Universität Berlin; 2015.
12. Schröder D, Vorländer M. RAVEN: A real-time framework for the auralization of interactive virtual environments. In: Forum Acusticum. Aalborg Denmark; 2011. p. 1541–1546.
13. Weinzierl S, Vorländer M, Behler G, Brinkmann F, von Coler H, Detzner E, et al. A Database of anechoic microphone array measurements of musical instruments. 2017;
14. Brinkmann F, Lindau A, Weinzierl S, Müller-Trapet M, Opdam R, Vorländer M. A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations. J Audio Eng Soc. 2017;65(10):841–848.
15. Barron M. Subjective study of British symphony concert halls. Acta Acust United Acust. 1988;66(1):1–14.
16. Gade AC. Investigations of musician's room acoustic conditions in concert halls. II: Field experiments and synthesis of results. Acustica. 1989;89:249–62.
17. Dammerud JJ, Barron M, Kahle E. Objective Assessment of Acoustic Conditions for Symphony Orchestras. Build Acoust. 2011;18(3):207–19.
18. Beranek LL. Concert Halls and Opera Houses: Music, Acoustics and Architecture, 2nd ed. Springer-Verlag, New York, 667 p.; 2004.
19. Beranek LL. Music, acoustics, and architecture. Wiley, New York; 1962.
20. Böck S, Widmer G. Maximum filter vibrato suppression for onset detection. In: Proc of the 16th Int Conf on Digital Audio Effects (DAFx) Maynooth, Ireland. 2013.
21. Weinzierl S, Maempel H-J. Zur Erklärbarkeit der Qualitäten musikalischer Interpretationen durch akustische Signalmaße [On the explicability of qualities of musical interpretation by means of acoustic signal measures]. Loesch H von, Weinzierl S, editors. Schott, Mainz et al.; 2011.
22. McLean RA, Sanders WL, Stroup WW. A unified approach to mixed linear models. Am Stat. 1991;45(1):54–64.
23. Schutte M, Ewert SD, Wiegrebe L. The percept of reverberation is not affected by visual room impression in virtual environments. J Acoust Soc Am. 2019;145(3):EL229–EL235.