



HAL
open science

Using virtual acoustics and electroglottography to study the adaptation of singing voice production

Paul Luizard, Nathalie Henrich Bernardoni, Christoph Böhm

► To cite this version:

Paul Luizard, Nathalie Henrich Bernardoni, Christoph Böhm. Using virtual acoustics and electroglottography to study the adaptation of singing voice production. Forum Acusticum 2020, Dec 2020, Lyon (virtual), France. 10.48465/fa.2020.0303 . hal-03080847

HAL Id: hal-03080847

<https://hal.science/hal-03080847>

Submitted on 24 Jun 2021

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.

USING VIRTUAL ACOUSTICS AND ELECTROGLOTTOGRAPHY TO STUDY THE ADAPTATION OF SINGING VOICE PRODUCTION

Paul Luizard¹

Nathalie Henrich Bernardoni²

Christoph Bhm¹

¹ Audio Communication Group, Technische Universität Berlin,
Einsteinufer 17c, Berlin, D-10587, Germany.

² Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

paul.luizard@campus.tu-berlin.de, nathalie.henrich@gipsa-lab.fr, c.boehm@tu-berlin.de

ABSTRACT

Previous research has showed that singers and instrumentalist musicians tend to adapt their sound production to the acoustics of the venue in which they perform. Studying such an adaptation process is not trivial since many parameters are to be taken into account. For example, various performances can differ due to aspects of the halls other than acoustic, the interaction with the audience, or the physical and psychological state of the performer. A specific methodology has been used to study how singers vary their sound production when they perform in different acoustical spaces within a short period of time. Virtual concert halls were proposed to four singers who could hear themselves in reverberant sound fields by means of a real-time convolution process and dynamic binaural synthesis. The signals captured by a microphone in near-field and by an electroglottography unit were analysed and yielded several parameters related to voice production, namely vocal intensity, fundamental frequency, and glottal contact quotient. These voice-quality parameters were compared to room acoustical parameters from the eight different rooms proposed to the singers. Results showed that the singers reacted in specific manners by varying their voice quality and being sensitive to different room acoustical parameters.

1. INTRODUCTION

Musical adaptation across rooms was investigated essentially in terms of changes produced by instrumentalist musicians while playing a given musical program in various venues. Such an experiment was conducted by recording a cellist in seven concert halls [1]. Comparing musical features calculated from the recordings with room acoustic parameters showed that more than 50% of the variance in music performance could be explained by room acoustics variations.

Following a similar methodology, adaptation of singing voice was investigated by recording four singers in eight rooms [2]. It has been shown that singers would adapt in rather individual manners, involving specific musical feature related to their sound production, and following specific room acoustic parameters across rooms.

Other studies [3, 4] placed musicians in an anechoic

chamber and presented them with synthetic reverberated sound fields while they played. It has been shown that the differences induced by the change of room acoustics could be heard during a listening test and the musicians reported that they modified their performance in accordance with the room variations. A correlation analysis with the recordings showed that the adaptation process could be made consciously.

Other research involved musicians in virtual environments [5, 6], by means of computer simulations of the different tested venues. Several musical features (e.g. sound level, timbre, dynamics, tempo) could be correlated to room acoustic parameters (reverberation time, sound intensity of the room response).

Virtual venues were also used to investigate singing performance [7, 8] by means of high-level musical features which carried a musical meaning, such as tempo, loudness, sound brightness or fullness. A few relations between room acoustic parameters and musical features were also identified but the latter were exclusively based on audio signals.

Previous investigations on voice production involved electroglottography (EGG). This measurement technique is used to record glottal-contact cycles and to derive relevant information related to voice production [9], such as glottal contact [10], identification of laryngeal mechanisms [11], and vocal effort assessment [12], which is related to vocal intensity [13].

This overview of recent research has shown that singers musical performance can be influenced by the acoustics of physical or virtual rooms in which they perform. Virtual acoustics was used for instrumentalists. Is such sound environment really suited to investigations with singers? Would information be gained by using virtual environments in conjunction with recording devices other than microphones?

This study combines audio recordings in virtual acoustics and electroglottographic voice assessment to explore whether voice production parameters would be correlated to room acoustics parameters of the venues where solo singers perform.

2. MEASUREMENTS

Measurements were conducted in an anechoic chamber at the technical university in Berlin, Germany. Virtual environments were synthesised so that the singers could perform in eight different rooms within two hours, which is not possible in the physical world due to the distance between rooms. This short time of experiment allows for the singers to remain in a similar physical and psychological state during the whole experiment, contrary to the case of recording sessions in the different rooms that would be several days apart. Hence this experimental protocol tends to reduce the number of factors possibly involved in musical performance variation in different rooms.

2.1 Singers and musical pieces

The singers who took part in this experiment had more than 10 years of musical practice and at least 4 hours of daily vocal training. They were asked to sing two excerpts of musical pieces that they were familiar with, so that the variations between performances would not be due to a lack of knowledge of the pieces. Details about the singers and the musical pieces are presented in Tab. 1.

2.2 Voice assessment

The vocal performances of each singer were recorded to quantify potential variations occurring in the different rooms. Audio signals were recorded in near field by means of a lavalier microphone (DPA 4060) placed 3 cm away from the edge of the singer mouth, as illustrated in Fig. 1. Synchronised measurements of glottal activity was achieved by means of an electroglottograph (EGG) unit (Glottal Enterprises EG2) that is composed of two electrodes between which the voltage of a low-intensity electrical current is proportional to vocal-folds contact area. All signals were captured by an audio interface (RME Fireface UFX) and a personal computer.

The EGG signal was used to calculate the Contact Quotient (CQ) which represents the amount of time during which vocal folds are in contact with each other over the duration of a vibratory cycle. DECOM algorithm [10] was used. The EGG signal also served for calculation of fundamental frequency (F_0) by means of YIN algorithm [14], allowing to observe pitch accuracy and amount of vibrato. The third voice parameter under investigation was the sound pressure level, referred to as *Intensity*, calculated from the audio signal according to the standard *fast* integration time (125 ms).

2.3 Virtual environments

The virtual rooms were based on acoustic measurements of actual rooms. Binaural room impulse responses were acquired by means of an artificial head (Neumann KU100) placed on stage, at the position of a singer. A directional loudspeaker (Fostex 6301B) having a directivity close to the human voice was placed just before the mouth of the head. They were together placed on a rotating plate so

that impulse responses could be measured at various angles in the horizontal plane, spanning $\pm 45^\circ$ with 5° step. Acoustical characteristics of the rooms, calculated from omnidirectional room impulse responses according to ISO 3382 [15] were measured at the same location on stage. The results can be seen in a previous paper [2]. It should be noted that the acoustics of these rooms were very different, ranging from a dry recording studio, a theatre and concert halls of various sizes, up to a church.

The experiment with the musicians was conducted in a laboratory environment in the anechoic room of the TU Berlin. This environment offers the possibility to add an artificial diffuse sound field to direct sound. An open headphone served to render the artificially reverberated voice. A compensation filter was used to equalise the headphones and minimise their influence on the signal chain. With open headphones, the singers could perceive the direct sound of their own voice unaffected but extended by the virtual sound field containing only the diffuse field information. Therefore the direct sound component was removed from the measured impulse responses. The rendering of the sound field was carried out with the Sound Scape Renderer (SSR) [16]. The SSR performed a convolution of the live recorded signal with the measured binaural room impulse responses (BRIRs) in real-time. The rotations of the singers' heads were captured with a tracker and sent to the SSR in real time, while the impulse responses were dynamically exchanged according to the current head orientation. The system latency was measured and a value of 52 ms was found, which is below the value recommended by Lindau [17]. This dynamic rendering supports the perception of the three-dimensional sound field and directional reflections, and improves the naturalness of the vocal performance.

2.4 Statistical analysis

In the present experiment, four singers performed two musical pieces in eight rooms, yielding hierarchical levels of variance to the data. Hence an adapted framework of statistical analysis is the linear mixed-effects models (LMM) [18]. This considers dependent variables, here the voice production parameters (*Intensity*, *Contact quotient*, F_0) which can be partly explained by linear combinations of independent variables, or fixed effects, here the five selected room acoustic parameters. Random effects, i.e. the levels of hierarchical data (pieces, rooms, and singers if not considered individually), are also accounted for in each model. To ensure that the dependent variables would not present collinearity, which would bias the results, cross-correlation of the room acoustic parameters across all rooms was calculated, and the obtained maximal value of 0.67 was considered low enough to run the LMM analyses.

The first statistical models included all singers, musical pieces, and rooms, yielding, for each voice production parameter, a coefficient of determination R_{marginal}^2 (further termed R^2). The latter represents the proportion of variance of each voice production parameter explained by the

Table 1. Singers and musical pieces involved in the experiment.

Voice type	Age	Musical pieces	Pace
Soprano	23	W.A. Mozart - Ach ich liebe (Aria from <i>Entführung aus dem Serail</i>)	Fast
		G. Puccini - Aria from <i>Manon Lescaut</i>	Slow
Mezzo-soprano	27	J.S. Bach - Esurientes implevit bonis (Aria from <i>Magnificat</i>)	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 <i>Samson</i>)	Fast
		G. Fauré - Au cimetière	Slow

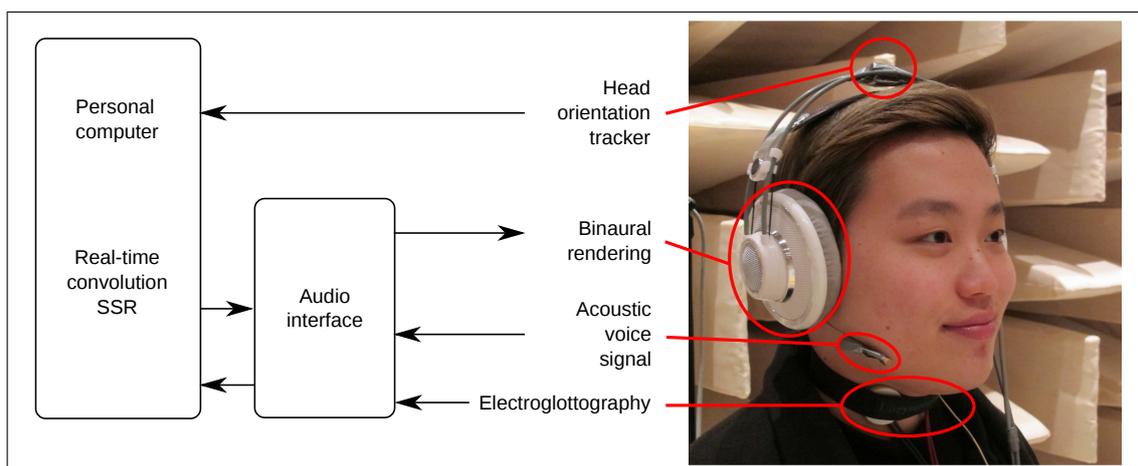


Figure 1. System for real-time rendering in virtual rooms.

room acoustic parameters. The obtained values were very low (mean $R^2 < 1\%$ and maximal $R^2 = 1.3\%$). Therefore, another series of statistical models was run, accounting for each singer in an individual manner. The results are exposed in the following.

3. RESULTS

Figure 2 presents the R^2 values obtained by the models run for each independent singer. The mean values over singers are 20% (Contact Quotient), 7% (Intensity), and 2% (F_0). The peak is obtained by the Tenor whose Contact Quotient reached $R^2 = 43\%$ across rooms and musical pieces. This parameter is related to the highest R^2 for the tenor and mezzo-soprano singers, whereas the soprano and baritone singers presented another pattern and obtained their highest R^2 for Intensity. The pitch accuracy F_0 presented $R^2 < 10\%$ for all singers.

Beyond these global statistics, it is also possible to focus on individual significant relationships between each voice parameter and the room acoustic parameters across all rooms. The term significant refers to p-values inferior to 5% for the linear relation model of a given voice parameter and one of the room acoustic parameters across rooms within the LMM framework. These significant relationships are presented in Fig. 3. It can be seen that the Soprano produced the highest number of voice parameters that are correlated to room acoustic parameters, just before

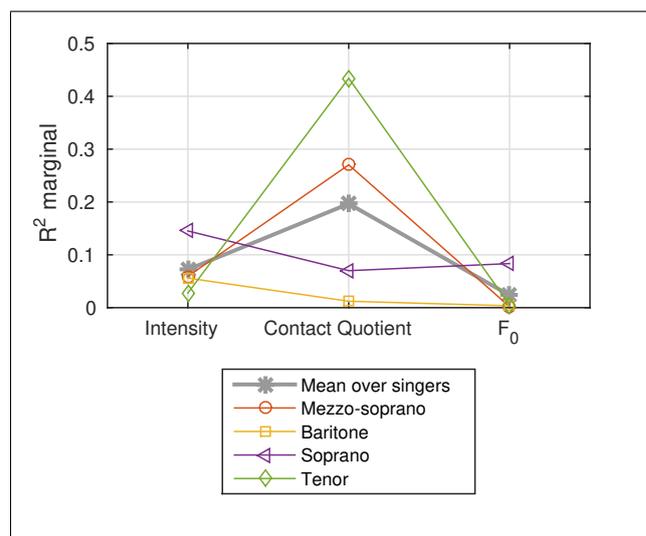


Figure 2. Coefficients of determination R^2 from the Linear Mixed Models for each independent singer.

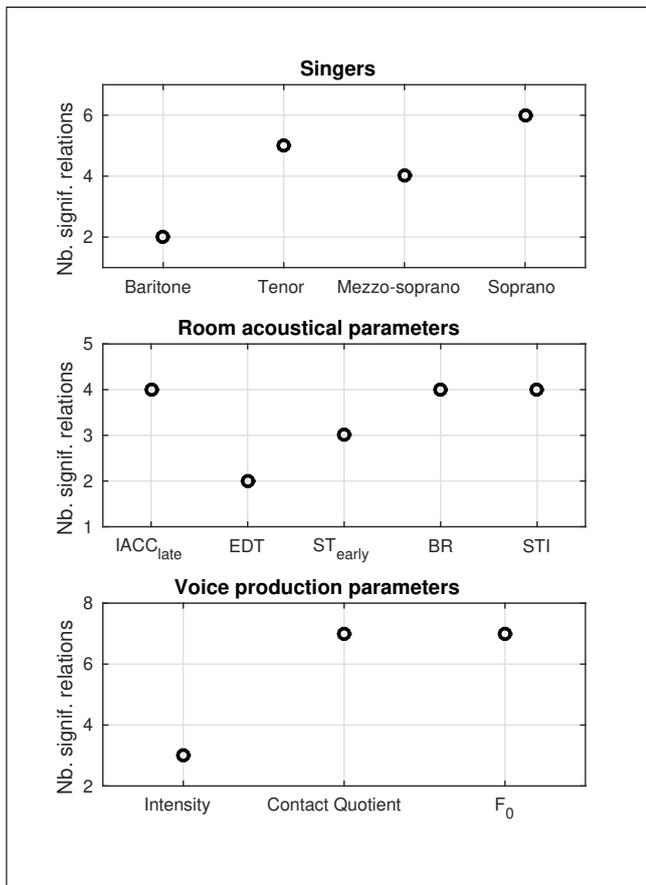


Figure 3. Number of significant relationships between the room acoustical parameters and the voice production parameters for each singer (top), each room parameter (middle), and each voice parameter (bottom).

the Tenor and Mezzo-soprano. The room parameters most involved in significant relationships were the late interaural cross-correlation (IACC_{late}), the bass ratio (BR), and the speech transmission index (STI), respectively related to the perceptual dimensions *spatiality*, *tonal colour*, and *speech clarity* according to previous research [19]. The early stage support (ST_{early}) and early decay time (EDT), related to the perceived loudness and reverberation, were involved to a lesser extent in significant relationships. Surprisingly, although F₀ obtained rather low R² scores, this voice parameter was involved in as many significant relationships as Contact Quotient. The Intensity parameter was much less involved.

4. DISCUSSION

The statistical models including all singers together yielded R² values inferior to 1% while models considering individual singers yielded average R² across singers between 2% and 20%. Hence, this analysis showed that singers behave differently from each other regarding adaptation to the acoustics of a series of venues. This result based on voice production parameters confirmed the previous findings on musical features of singing voice [2].

Moreover, looking at R² values of the independent

singers showed that, depending on the considered singer, large differences could be observed, both in terms of absolute difference for each voice parameter, and in terms of pattern. For example, the Mezzo-soprano and Tenor presented a maximal R² value for Contact Quotient, whereas the Soprano and Baritone maximised R² for Intensity. Also the number of significant relationships between voice and room parameters varied strongly from one singer to another. These observations detailed the various adaptation mechanisms that the singers developed.

The fact that the Contact Quotient received the highest value on average over singers showed that the laryngeal behaviour is impacted by variations of room acoustics from one room to another, as part of a more general singing adaptation process.

The room acoustic parameters most involved in significant relationships were IACC_{late}, BR, and STI, while EDT and ST_{early} were less concerned. According to a previous study that established links between room acoustic parameters and associated perceptual dimensions, this would mean that the singers used temporal cues to a lesser extent while they focused on spatiality, tonal colour, and speech clarity or intelligibility.

It can be remarked that R² for Intensity was higher than R² for F₀ while F₀ was involved in more than twice as many significant relations with single room parameters than Intensity. This would denote a larger deviation of F₀ across the room acoustic parameters while Intensity is more evenly distributed over room parameters, i.e. correlations with room parameters are larger on average for Intensity but presents strong peak values for F₀ which would be associated to specific room parameters.

5. CONCLUSION

This study investigated how singers adapt to various room acoustic conditions. Both audio and electroglottographic signals were used to assess voice production parameters in each room. The rooms were virtual, designed from binaural impulse responses measured in physical rooms, allowing to render in an anechoic chamber the virtual spaces through binaural synthesis.

This experiment confirmed that virtual acoustics is an effective tool to study the behaviour of performers in various acoustic conditions. This framework allows to reduce certain bias of a perceptual experiment in the sense that a performer could be recorded in various environments within a short time.

In complement of audio signals, electroglottography has brought additional information and turned out to be a functional tool to study singer adaptation patterns in this experiment.

More generally, the fact that singers adapt to various concert venues in a rather individual manner was confirmed. Such an adaptation process involved voice parameters related to both the produced sound and specific laryngeal mechanisms assessed by means of the EGG device.

6. ACKNOWLEDGEMENTS

The authors would like to thank the singers for their participation, the room managers for their collaboration, and E. Brauer for his support in the recording sessions. This work was supported by a grant of the Alexander von Humboldt Foundation.

7. REFERENCES

- [1] Z. Schärer Kalkandjiev and S. Weinzierl, “The influence of room acoustics on solo music performance: An empirical case study,” *Acta Acustica united with Acustica*, vol. 99, 2013.
- [2] P. Luizard, J. Steffens, and S. Weinzierl, “Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions?,” *Journal of the Acoustical Society of America*, vol. 147, no. 2, pp. EL132–EL137, 2020.
- [3] K. Ueno, K. Kato, and K. Kawai, “Effect of room acoustics on musicians’ performance. Part I: Experimental investigation with a conceptual model,” *Acta Acustica united with Acustica*, vol. 96, no. 3, pp. 505–515, 2010.
- [4] K. Kato, K. Ueno, and 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*, vol. 101, pp. 743–759, 2015.
- [5] Z. Schärer Kalkandjiev and S. Weinzierl, “The influence of room acoustics on solo music performance: An experimental study,” *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 3, 2015.
- [6] S. A. Gari, M. Kob, and T. Lokki, “Analysis of trumpet performance adjustments due to room acoustics,” in *Proc. International Symposium on Room Acoustics, Amsterdam, The Netherlands, 15-17 September, 2019*.
- [7] P. Luizard, J. Steffens, and S. Weinzierl, “Adaptation of singers to physical and virtual room acoustics,” in *Proc. International Symposium on Room Acoustics, Amsterdam, The Netherlands, 15-17 September, 2019*.
- [8] P. Luizard, E. Brauer, and S. Weinzierl, “Singing in physical and virtual environments: how performers adapt to room acoustical conditions,” in *Proc. Conference of the Audio Engineering Society on Immersive and interactive audio, York, United Kingdom, 27-29 March*, Audio Engineering Society, 2019.
- [9] D. Childers and A. Krishnamurthy, “A critical review of electroglottography,” *Critical reviews in biomedical engineering*, vol. 12, no. 2, pp. 131–161, 1985.
- [10] N. Henrich, C. dAlessandro, B. Doval, and M. Castellengo, “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” *Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–1332, 2004.
- [11] B. Roubeau, N. Henrich, and M. Castellengo, “Laryngeal vibratory mechanisms: The notion of vocal register revisited,” *Journal of voice*, vol. 23, no. 4, pp. 425–438, 2009.
- [12] D. Huang, F. Minifie, H. Kasuya, and S. Lin, “Measures of vocal function during changes in vocal effort level,” *Journal of Voice*, vol. 9, no. 4, pp. 429–438, 1995.
- [13] N. Henrich, C. dAlessandro, B. Doval, and M. Castellengo, “Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency,” *Journal of the Acoustical Society of America*, vol. 117, no. 3, pp. 1417–1430, 2005.
- [14] A. De Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [15] ISO3382-1, “Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces,” 2009.
- [16] M. Geier, J. Ahrens, and S. Spors, “The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods,” in *In 124th AES Convention in Amsterdam, The Netherlands, May 1720, 2008*.
- [17] A. Lindau, “The perception of system latency in dynamic binaural synthesis,” in *German Annual Conference on Acoustics (DAGA)*, vol. 35, (Rotterdam), pp. 1063–1066, 2009.
- [18] J. Hox, M. Moerbeek, and R. Van de Schoot, *Multilevel analysis: Techniques and applications*. Routledge, 2017.
- [19] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Disentangling preference ratings of concert hall acoustics using subjective sensory profiles,” *Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3148–3161, 2012.