Predicting subglottal pressure in female singers based on electroglottographic and acoustic metrics

Helena Engström[1], Marcin Włodarczak[1], Sten Ternström[2]
[1]Department of Linguistics, Stockholm University, Sweden
[2]Speech, Music and Hearing, KTH Royal Institute of Technology, Sweden
helena.engstrom@ling.su.se, wlodarczak@ling.su.se, stern@kth.se

Abstract

The estimation of subglottal pressure (P_s) is a difficult task in experimental phonetics, voice research and clinical treatment of voice disorders due to the strict and invasive methods needed for obtaining qualitative measurements of P_s. This paper investigates the feasibility of using electroglottographic (EGG) and acoustic metrics to indirectly predict P_s using backwards stepwise regression modeling (BSR) and random forest regression analysis (RF) with a subject-specific leave-one-out (LOO) cross validation based on seven female singers. The results show an average root-mean-square-error (RMSE) of 2.89 cm H_2O from target using BSR and 3.24 cm H_2O from target using RF for P_s prediction. In addition, sound pressure level (SPL), fundamental frequency (f_o) measured in semitones, quotient of contact by integration (Q_{ci}) and the normalized peak EGG derivative ($dEGG_{max}$) were the most important predictors. However, the prediction may be affected by the small sized data. Based on these findings, electroglottographic metrics may serve as an option for exploring indirect estimation of P_s in future research.

Introduction

One of the most fundamental components in human voice production is subglottal pressure (P_s). As explained by McKenna et al. (2017), subglottal pressure is the build-up of air beneath the vocal folds that sets the vocal folds into vibratory motion once it is released and thus causes phonation. This makes subglottal pressure essential for controlling voice onset and offset, fundamental frequency (f_o) and sound pressure level (SPL) (Lin et al., 2019). It has been shown that subglottal pressure is instrumental in the treatment of voice disorders (McKenna et al., 2017; Lin et al., 2019; Fryd et al., 2016) since many voice difficulties stem from glottal inefficiencies that cause abnormal subglottal pressure regulation (McKenna et al., 2017). Unfortunately, subglottal pressure estimations require invasive interventions such as using tracheal needles and esophageal balloons, or strict elicitation of /p/-occlusions followed by sustained vowels at steady pitch which equates temporarily subglottal pressure with oral pressure (Lin et al., 2019). However, although this method is less invasive, it does not allow subglottal pressure estimations during conversational speech or lyrical singing, making it less informative for vocal therapy.

Recent attempts in voice research have investigated the feasibility of using indirect methods for obtaining subglottal pressure measurements by using mini accelerometers attached to the neck that measure neck surface vibration during phonation (Fryd et al., 2016; Lin et al., 2019; McKenna et al., 2017). Findings by McKenna et al. (2017) have shown that the magnitude of the accelerometer signal is significantly related to subglottal pressure, although the relationship was found to be speaker-specific. Lin et al. (2019) found that incorporating acoustic metrics such as cepstral peak prominence (CPP) may aid in the prediction of subglottal pressure in non-modal phonation. However, considering that McKenna et al. (2017) found that individual variation among the participants determined whether the magnitude of the accelerometer signal could relate to subglottal pressure, it could potentially be more suitable to perform subject-specific cross validations.

Similarly to using an accelerometer, electroglottography (EGG) is a non-invasive method for obtaining glottal information using electrodes placed at the neck to record vocal fold contact (Rothenberg, 1992). Given the previous findings by McKenna et al. (2017), Lin et al. (2019) and Fryd et al. (2016), electroglottographic signals could potentially also be used for the prediction of subglottal pressure.

Following the approach of Lin et al. (2019), this study attempts to use previously collected voice map data consisting of EGG metrics and acoustic metrics for the prediction of subglottal pressure using backwards stepwise regression modeling (BSR) to assess the prediction of P_s in female singers. In addition, a random forest regression analysis (RF) using decision trees for prediction was done to compare with the BSR approach and to investigate the hierarchical importance of the chosen metrics for prediction of subglottal pressure.

This investigation is based on these research questions: 1) Can subglottal pressure be predicted using EGG and acoustic metrics based on seven female singers? 2) Which metrics among the EGG metrics and the acoustic metrics contribute the most to the prediction of subglottal pressure based on seven female singers?

Material and methods

Data

The data was provided from a previous study researching the effects of lung volume on the electroglottographic signal by Ternström et al. (2020) consisting of eight female singers. Seven participants were chosen for this study since the mean subglottal pressure measurements of one participant were faulty due to a malfunctioning signal during the recording. The participants were instructed to sing *Frère Jacques* using the syllable /pa/ at high and low lung volumes and at different modes: *forte*, *mezzoforte* and *piano*. The electroglottographic signal had been recorded using a dual-channel electroglottograph (EG2, Glottal Enterprises) and the acoustic data was recorded using an AKG C520 microphone.

Intraoral pressure (P_{IO}) was recorded for obtaining mean subglottal pressure data during /p/-occlusions performed by the participant while holding a tube at the corner of the mouth. The tube had an inner diameter of 3 mm and was connected to a pressure transducer (PG-100E, Glottal Enterprises). For this study, only four participants had calibrated recordings of P_{IO} for expressing

subglottal pressure in units of cm H_2O. Therefore, the raw P_{IO} values (being signed 16-bit integers ranging from *2^(-15) = -32768 to 2^(+15)-1 = +32767*) were converted to express subglottal pressure in units of cm H_2O based on the existing calibrated data that was available. Based on the calibrated data, the value of 14 000 P_{IO} was found to be within a +/- 5% range of representing 20 cm H_2O across the four participants. Therefore, all raw P_{IO} values were converted to be expressed in cm H_2O for all seven participants using the following formula:

$$P_s \text{ cm } H_2O \approx (20 * n) / 14000 \quad (1)$$

The data was pre-processed using the FonaDyn software (Ternström et al., 2018) for voice mapping analysis to obtain EGG and acoustic metrics. Each metric, as seen in Table 1, is mapped across a f_o and SPL plane where each cell contains its own distribution and average (Ternström, 2025). Each cell is 1 dB high and 1 semitone wide (measured on the MIDI scale). For each participant, this resulted in a voice map with at least 300 visited cells, and metric averages in every cell based on an average of 200 cycles in each cell. The cell information of each metric was used for the prediction of P_s in this study.

Table 1. P_s and EGG/acoustic metrics obtained from FonaDyn pre-processing with description.

Metric	Description
P_s	Subglottal pressure (cm H_2O)
SPL	Sound pressure level (dB)
Q_{ci}	Quotient of contact by integration
$dEGG_{max}$	Normalized peak EGG derivative
HRF_{egg}	Harmonic richness factor
MIDI	f_o measured in semitones
Clarity	The f_o periodicity
Entropy	Cycle-rate sample entropy
SB	Spectrum balance (dB)
CPP	Cepstral peak prominence (dB)
Crest factor	Indication of high frequency content

Statistical analysis

The analysis followed the method of Lin et al. (2019), using backwards step-wise regression modeling, and an additional random forest regression analysis. The analysis was completed within the RStudio environment (RStudio team, 2025) using R programming (R Core team, 2024).

Backwards step-wise regression modeling (BSR)

Backwards step-wise regression modeling (BSR) was completed by utilizing the R library *leaps* to conduct BSR based on model optimization by exhaustive search. All predictors were standardized. Predictions were made using a subject-specific leave-one-out (LOO) cross validation strategy. The training data consisted of all participants, excluding the participant meant for the test data. The model selection on the training data was made based on *Bayesian information criterion* (BIC) and then evaluated using *root-mean-square-error* (RMSE*)* on the test data through a constructed test matrix. This process was repeated across all seven participants and resulted in seven test folds for cross validation.

Random forest regression analysis (RF)

Random forest regression analysis was completed using the *RandomForest* R library. Training and test data was divided based on the LOO cross validation approach, where one participant was left out for the testing stage, resulting in seven cross validation folds. The model was fitted based on the training data, where 500 trees were grown and 10 variables were tried at each split. Prediction was made based on the test data and the hierarchical importance of the predictors was obtained based on a % increase in mean-square error (MSE), where a higher value means a greater importance of the predictor for the prediction modeling. Similarly to the BSR approach, all predictors were standardized and RMSE was used to evaluate the prediction performance based on the test data.

Results

Figure 1 shows the optimal model selection using BSR for the training data, as indicated by the diamond shapes in the plot. The models with 6, 7 or 8 predictors were chosen based on BIC to be the optimal training models.

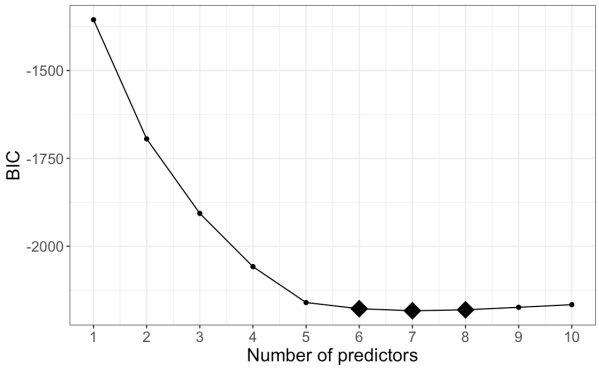

Figure 1. BIC - Model selection for training data.

The core predictors that were consistently chosen in the same hierarchical structure across the LOO cross validations were 1. sound pressure level (SPL), 2. fundamental frequency measured in semitones, 3. normalized peak EGG derivative ($dEGG_{max}$) and 4. quotient of contact by integration (Q_{ci}). The other predictors were in different hierarchical structures across the cross validations.

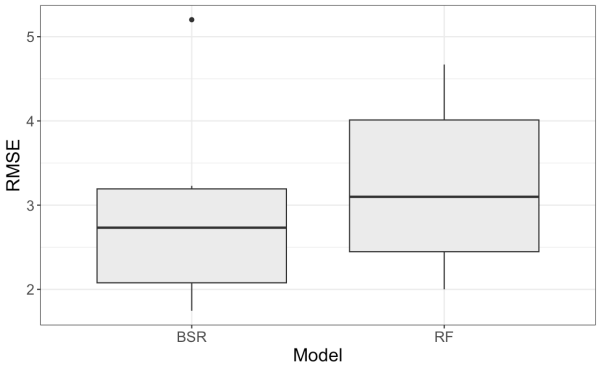

Figure 2. RMSE of P_s in cm H_2O based on test data and compared between the BSR and RF approach.

The P_s data points across participants ranged between 1.81 cm H_2O to 36.41 cm H_2O. Figure 2 shows the RMSE

results for P_s prediction across the LOO cross validation for the BSR modeling and RF analysis. The median RMSE for the BSR modeling across the LOO cross validation folds was 2.73 cm H_2O from target and a mean RMSE at 2.89 cm H_2O from target. The RMSE for P_s prediction from the RF regression modeling show a median RMSE at 3.10 cm H_2O from target, with a similar mean at 3.24 cm H_2O from target, which is higher than the BSR models.

Table 2. R^2 across training data LOO folds for BSR and RF.

LOO	BSR R^2	RF R^2
1	0.59	0.75
2	0.57	0.82
3	0.60	0.81
4	0.59	0.83
5	0.59	0.82
6	0.59	0.83
7	0.62	0.86

Table 2 shows the R^2 results of each LOO cross validation fold for the training data across both modeling methods. The RF approach has a R^2 within the range of 0.75-0.86 in comparison to the BSR approach that has a R^2 range of 0.57-0.62, indicating that more variance is explained across the regression models using the RF approach.

The hierarchical importance obtained from the RF analysis can be seen in Figure 3. The most important predictor for the prediction of P_s was sound pressure level (SPL) (dB_z in the diagram) with a score of over a 100% increase in MSE. The second order of the most important predictors were fundamental frequency in semitones (MIDI_z in the diagram), quotient of contact by integration (Qcontact_z in the diagram), normalized peak EGG derivative ($dEGG_{max}$) and harmonic richness factor (HRF_{egg}) at a lower % increase in MSE. The weakest predictors were crest factor, cepstral peak prominence (CPP), spectrum balance (SB), entropy and clarity.

Discussion

The results suggest that electroglottographic and acoustic metrics may be able to predict subglottal pressure considering that the average RMSE of P_s prediction was around 2.89 cm H_2O from target using BSR modeling and 3.24 cm H_2O from target in the RF analysis, which is similar to Lin et al. (2019) where the average RMSE across subjects were 2.6 cm H_2O in non-modal phonation. Importantly, the data used in this study consisted of singing phonation, which is not typical speech or specific non-modal phonation such as breathy or creaky voice. Therefore, the results are perhaps not as easily compared to previous studies. Lin et al. (2019) also used a five-fold cross validation approach across subjects and for each subject, whereas this investigation used a subject-specific leave-one-out cross validation for prediction. This ensures that subjects are separated between training and test data, which makes the prediction more reliable. However, since the conversion from raw intraoral pressure P_{IO} values to units of cm H_2O had to be approximated based on calibrated data from four out of the seven participants, it is possible that one could obtain even more precise RMSE values than what was achieved in the present study.

Interestingly, the RF regression modeling had a R^2 range of 0.75-0.86 compared to the BSR approach that had a R^2 range of 0.57-0.62 across the training data, which could indicate that the RF approach captured more of the complexity of the data. However, since the median and average RMSE were higher using the RF model compared to the BSR model, it is also possible that the RF model was overfitted.

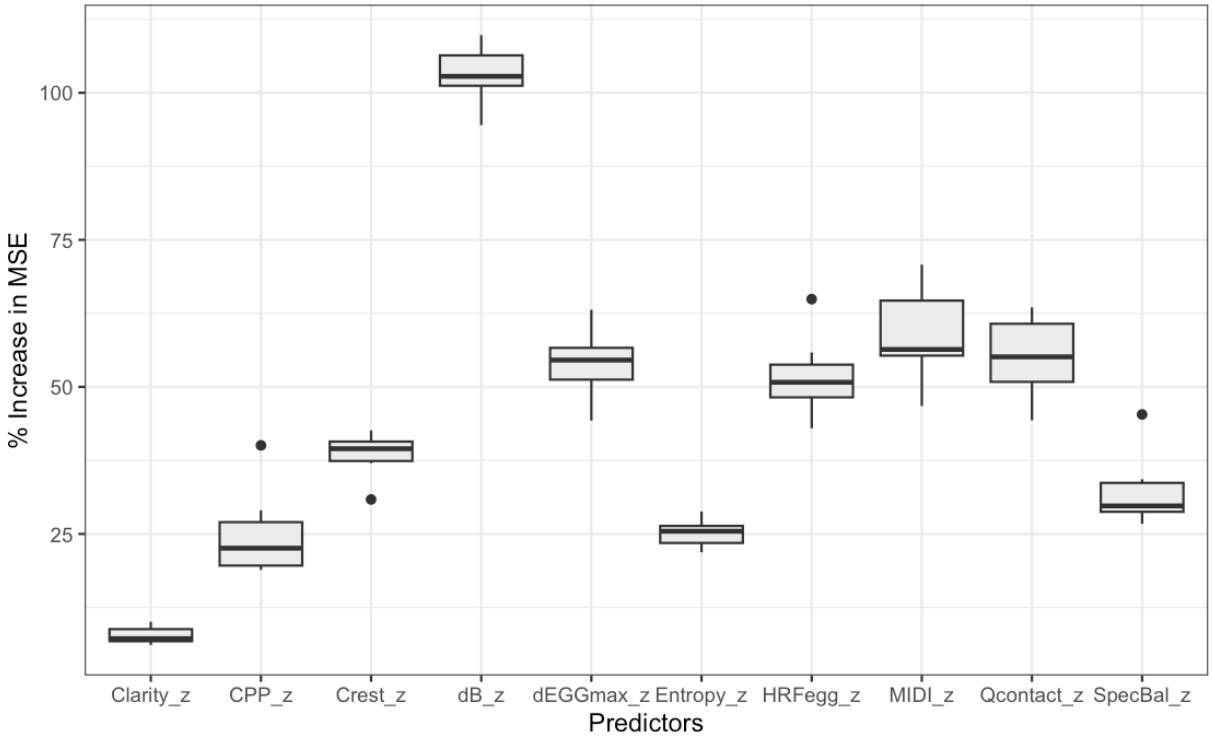

Figure 3. Random forest predictor importance.

Based on the BSR approach, models with 6, 7 or 8 predictors were deemed to be the best training models according to BIC, and the core predictors chosen were similar to the hierarchical importance analysis from the RF approach, with the exception of harmonic richness factor (HRF_{egg}). The importance of the predictors for P_s prediction as shown in Figure 3 indicate that sound pressure level (SPL) is the most important predictor followed by fundamental frequency measured in semitones (MIDI), normalized peak EGG derivative ($dEGG_{max}$), harmonic richness factor (HRF_{egg}) and quotient of contact by integration (Q_{ci}). This is in line with the findings of Lin et al. (2019), where the best performing models incorporated similar measurements based on the accelerometer signal, such as open quotient (OQ), which is similar to the Q_{ci} metric used in this study. Fundamental frequency was also chosen to be of importance for the models in the study by Lin et al. (2019), as well as other metrics based on glottal information such as harmonic richness factor, maximum flow declination rate (MFDR) and normalized amplitude quotient. However, the most important metric found in this study was sound pressure level which makes sense considering how much subglottal pressure correlates with sound pressure level (Björlund & Sundberg, 2016).

Although the data size in this study was quite small with only seven participants, the results suggest that P_s prediction is feasible using a combination of acoustic metrics, SPL, f_o and electroglottographic metrics obtained through the FonaDyn voice mapping software (Ternström et al., 2018). Therefore, electroglottography may be of similar use as accelerometers for the indirect estimation of subglottal pressure. It would be of future interest to both evaluate the methods presented in this work on a larger dataset as well as explore other methods for prediction such as the mixed effects modeling approach by McKenna et al. (2017).

To conclude, although this investigation is just a first attempt at understanding the feasibility of using electroglottographic and acoustic metrics for the prediction of P_s, this work may offer guidance for future research.

Acknowledgements

The first author thanks the *Speech* seminar participants of the 2025 autumn semester course *Project Course in AI and Language* at Stockholm University for the helpful discussions and feedback surrounding the project.

References

Björklund, S., & Sundberg, J. (2016). Relationship between subglottal pressure and sound pressure level in untrained voices. *Journal of Voice, 30*(1), 15-20. doi: 10.1016/j.jvoice.2015.03.006

Fryd, A. S., Van Stan, J. H., Hillman, R. E., & Mehta, D. D. (2016). Estimating subglottal pressure from neck-surface acceleration during normal voice production. *Journal of Speech, Language, and Hearing Research, 59*(6), 1335-1345. doi: 10.1044/2016_JSLHR-S-15-0430

Lin, J. Z., Espinoza, V. M., Marks, K. L., Zañartu, M., & Mehta, D. D. (2019). Improved subglottal pressure estimation from neck-surface vibration in healthy speakers producing non-modal phonation. *IEEE Journal of Selected Topics in Signal Processing, 14*(2), 449-460. doi: 10.1109/JSTSP.2019.2959267

McKenna, V. S., Llico, A. F., Mehta, D. D., Perkell, J. S., & Stepp, C. E. (2017). Magnitude of neck-surface vibration as an estimate of subglottal pressure during modulations of vocal effort and intensity in healthy speakers. *Journal of Speech, Language, and Hearing Research, 60*(12), 3404-3416. doi: 10.1044/2017_JSLHR-S-17-0180

R Core team. (2024). *R: A Language and Environment for Statistical Computing*.

Rothenberg, M. (1992). A multichannel electroglottograph. *Journal of Voice, 6*(1), 36-43. doi: 10.1016/S0892-1997(05)80007-4

RStudio Team. (2025). *RStudio: Integrated Development Environment for R*.

Ternström, S., Johansson, D., & Selamtzis, A. (2018). FonaDyn—A system for real-time analysis of the electroglottogram, over the voice range. *SoftwareX, 7*, 74-80. doi: 10.1016/j.softx.2018.03.002

Ternström, S., D'Amario, S., & Selamtzis, A. (2020). Effects of the lung volume on the electroglottographic waveform in trained female singers. *Journal of Voice, 34*(3), 485-e1. doi: 10.1016/j.jvoice.2018.09.006

Ternström, S. (2025). *The FonaDyn handbook* [3.4.2].