Interspeech 2024
1-5 September 2024, Kos, Greece

Glottal inverse filtering and vocal tract tuning for the numerical simulation of vowel /a/ with different levels of vocal effort

Marc Freixes, Marc Arnela, Joan Claudi Socoró, Luis Joglar-Ongay, Oriol Guasch, Francesc Alías-Pujol

HER - Human-Environment Research, c/Sant Joan de la Salle, 42, 08022 Barcelona - La Salle - URL

marc.freixes@salle.url.edu

Abstract

Voice production models provide valuable information about the human voice generation. However, providing them with expressiveness remains a challenge. This work proposes a methodology to modify vocal effort (VE) in the numerical simulation of vowels using a glottal source Liljencrants-Fant (LF) model and a one-dimensional acoustic model based on the finite element method. Vowels recorded with high, mid, and low VE are inverse-filtered to obtain a glottal source signal, used to estimate the LF model Rd parameter. A tuning algorithm adjusts the vocal tract geometry to match the formants of the analysed vowel. Preliminary results for the vowel /a/ are presented. Objective analyses indicate the relevance of both glottal source and vocal tract changes in reproducing VE. They are also perceptually relevant for low VE, while the glottal source predominates in high VE. Perceptual assessment validates the methodology can convey different levels of VE, particularly low and medium.
Index Terms: numerical voice production, expressive speech, glottal source modelling, LF model, finite element method

1. Introduction

Voice production models offer valuable insights into the mechanisms of human voice generation. However, reproducing the nuances of expressive speech remains a significant challenge. Achieving expressiveness in a source-filter-based voice production model requires careful adjustment of both glottal source (GS) and vocal tract (VT) characteristics. These adjustments are essential for capturing primary prosodic features such as pitch and energy, as well as secondary prosodic features related to voice quality [1].

The characteristics of the glottal source and vocal tract in expressive speech have been studied using inverse filtering and copy-synthesis techniques. For instance, in [2] the impact of phonation types on emotion perception in vowels was analysed using a 1D articulatory synthesizer. Parametric glottal flow models, such as the Liljencrants-Fant (LF) model [3], have been introduced in copy-synthesis schemes. For instance, in [4] an LF model controlled by the Rd glottal shape parameter was used to investigate the tense-lax continuum and its emotional implications. Similarly, an Auto-Regressive eXogenous variant of the LF model was proposed in [5] to explore the contributions of glottal source and vocal tract to emotion perception. Finally, in [6] authors explored the correlation between F0 contours, voice quality, and affect across languages by modifying modal stimuli parameters using the KLSYN88 synthesizer.

Preliminary attempts have also been made to incorporate expressiveness in vowel simulation using the finite element method (FEM) with the LF model and realistic geometries obtained from magnetic resonance imaging. In [7], happy and aggressive vowels were generated by modifying the glottal flow signal according to averaged spectral tilt increments relative to their neutral counterparts. On the other hand, a vocal tract tuning method was used in [8] to mimic the singing formant based on predefined formant values.

This work presents a methodology to simulate vowels with different levels of vocal effort (VE) using an LF model and a one-dimensional acoustic model based on the finite element method. This involves modifying both the vocal tract geometry and the glottal source based on the analysis of real expressive speech. To this aim, /a/ vowels from the Zurich Voice Quality database [9] are analysed to obtain: i) their formant frequencies, and ii) the Rd parameter of the LF-model from the glottal flow estimated through inverse filtering. These parameters are then used to respectively adjust the vocal tract geometry and the glottal source model employed in the synthesis. The source code can be found at the git repository.[1]

The paper is structured as follows. Section 2 outlines the methodology proposed for the analysis, tuning and synthesis of expressive vowels. Section 3 details the conducted experiments, and the obtained results. Conclusions and future work are presented in Section 4.

2. Methodology

The objective of this methodology is to enhance a voice production model by analysing expressive utterances to understand the mechanisms that convey emotion and nuance in vocal production. As depicted in Figure 1, our approach involves setting up a voice production model through a systematic source-filter analysis of real speech, incorporating voice quality (VoQ) aspects related to the VT (e.g. formants positions) and the GS (e.g. spectral tilt). These elements are extracted from real expressive utterances specifically selected to transfer both GS and VT spectral cues to the synthetic signals. This work particularly focuses on the numerical generation of sustained vowels with different vocal efforts, excluding aspects like prosody (e.g. F0 and energy curves) for future research.

The procedure starts with the source-filter analysis of expressive utterances (see Figure 1), aiming to extract GS and VT features that can transfer relevant VoQ-based information to the synthesis. On the one hand, the expressive GS signal is obtained using a glottal inverse-filtering (GIF) technique, and the result is used to estimate parameters of a GS model. On the other hand, VT features that convey expressiveness are estimated to tune an original VT geometry, resulting in an expressive VT geometry.

In this work, two GS features are estimated to control the

[1] https://github.com/SpeechSalleBcn/inverse-filtering-evaluation/releases/tag/InterSpeech24

10.21437/Interspeech.2024-1835

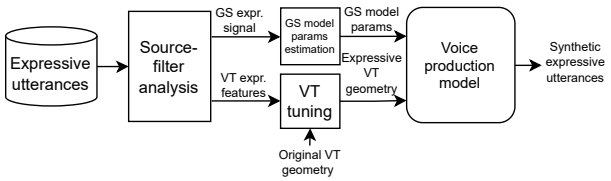


Figure 1: *Block diagram of the methodology* .

LF glottal flow model [3]: the Rd coefficient, which characterises the glottal pulse in terms of width and asymmetry and is highly correlated with GS spectral tilt (lower values indicate a tense/loud voice while higher values suggest a lax/soft voice) [10]; and the fundamental frequency (F0) curve, which determines the timing of LF glottal pulses for synthesis.

For the VT, the first five formants are estimated from the expressive utterance using Praat [11]. The formant frequencies are next input into a vocal tract tuning algorithm based on sensitivity functions [12], which modifies the original VT geometry to obtain the desired target formants. The VT impulse response is then computed using a one-dimensional (1D) FEM-based acoustic model [13]. The synthetic GS signal is generated with an LF model, using the estimated expressive F0 curve and the corresponding Rd value. This signal is finally convolved with the VT impulse response derived from the expressive VT transfer function, resulting in the synthetic expressive utterance.

2.1. Source-filter analysis

2.1.1. Glottal source

The analysis of glottal source of the speech signals was conducted using Quasi Closed Phase (QCP) [14], a GIF technique based on the principle of Closed Phase analysis [15]. QCP obtains accurate glottal source estimates by introducing an Attenuated Main Excitation (AME) weighting function that emphasises the closed phase time regions, where the voiced speech signal is solely influenced by VT. After obtaining glottal closure instants, using the SEDREAMS speech event detection technique based on the residual excitation and mean-based signal [16], a weighted linear prediction technique [17] was applied to a preemphasised short-time windowed speech frame of T_w seconds,using a Hanning window and a lip radiation factor $d \simeq 0.99$. The AME weighting function is defined through the three control parameters: i) position quotient, which refers to the relative initial position of the non-attenuated segment; ii) duration quotient, which denotes the relative length of the non-attenuated section; iii) and ramp quotient, which characterises the relative duration of the transition ramp connecting the attenuated and non-attenuated segments within each voiced speech period. The LF model parameter Rd was estimated using a dynamic programming algorithm [18].

2.1.2. Vocal tract

Speech signals with sustained /a/ vowels were analysed with Praat [11] to extract the first five formants (F1, F2, F3, F4 and F5), using the following setup: frame length of 25 ms, time step of 5 ms, preemphasis cutoff frequency of $F_c = 50$ Hz (using a preemphasis factor $\alpha = \exp(2\pi F_c/F_s)$, being F_s the sampling frequency in Hz). The formant value of each speech signal is obtained as the median value of the formants computed along a 1 s time window centred in the speech signal.

2.2. Synthesis

2.2.1. Glottal source model

The LF model is used as glottal source model to generate the waveform that represents the flow produced by the opening between the vocal folds in the larynx. The dynamic control of its parameters enables variations in voice quality and tension in synthetic speech. This work uses the LF implementation presented in [19], using the Rd parameter as the main parameter for controlling voice tension.

2.2.2. Vocal tract tuning

Starting from the discrete VT geometry for vowel /a/ in [20], we tune its shape to move formants F_i, $i = 1 \ldots 5$, to the target frequencies determined from the source-filter analysis. For this purpose, use is made of the iterative algorithm in [12]. Each one of the $n = 1 \ldots 40$ slices of the discrete VT is characterised by its cross-sectional area $A(n)$ and length $l(n)$. These become progressively modified until the error of the frequency value of each target formant is less than a prescribed tolerance. At each iteration, $k+1$, the values of $A_{k+1}(n)$ and $l_{k+1}(n)$ are related to the previous ones, $A_k(n)$ and $l_k(n)$, through

$$A_{k+1}(n) = A_k(n) + \alpha^A \sum_{i=1}^{5} z_{i_k} S^A_{i_k}(n), \tag{1a}$$

$$l_{k+1}(n) = l_k(n) + \alpha^l \sum_{i=1}^{5} z_{i_k} S^l_{i_k}(n), \tag{1b}$$

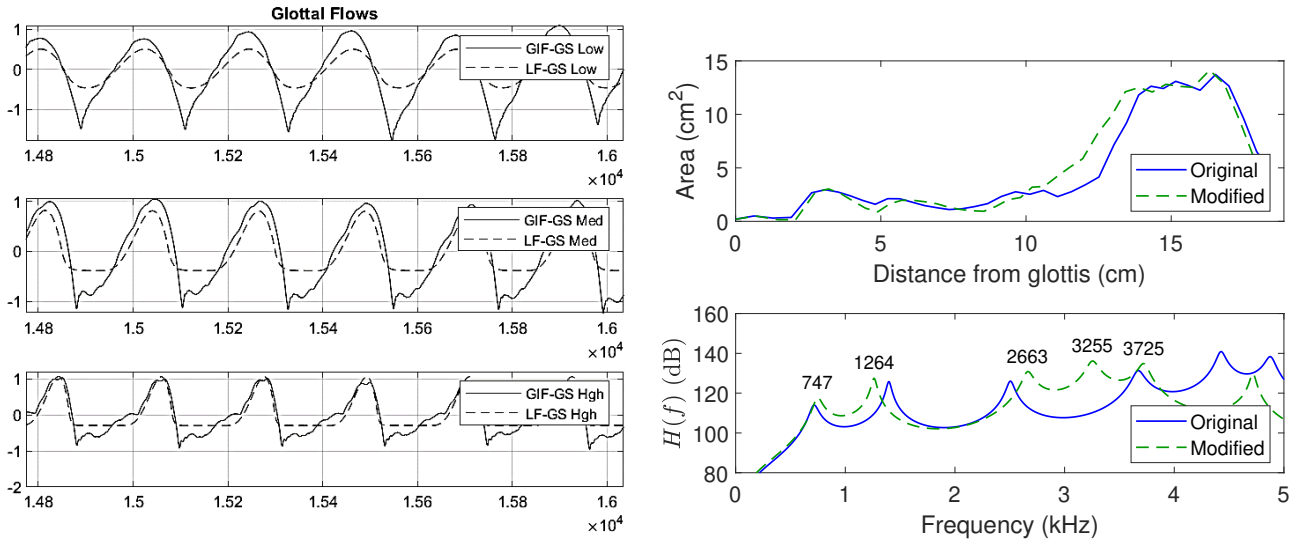
with $z_{i_k} = (\mathrm{F}_i^{\mathrm{tg}} - \mathrm{F}_{i_k})/\mathrm{F}_{i_k}$ and $\alpha^A = 15$, $\alpha^l = 10$ being the speed-up factors. The key to the algorithm are the sensitivity functions $S^A_{i_k}(n)$ and $S^l_{i_k}(n)$ that account for the responsiveness of the i-th formant to area and length perturbations, respectively. These sensitivity functions have been derived in literature using very different approaches, from circuit-based analogies [21, 22] to non-linear radiation pressure [12, 23], and, more recently, by means of linear perturbation analysis [24]. For the n-th slice at the k-th iteration it is shown that $S^A_{i_k}(n)$ is proportional to the difference between the kinetic and potential energy of the slice over the total mechanical energy of the VT, while $S^l_{i_k}(n)$ is proportional to minus the total energy of the slice over the total energy of the VT.

2.2.3. Numerical simulations

For the numerical simulation of vowel /a/ with the VT tuned to meet the formant values of a given vocal effort, we have proceeded as follows. We solved the 1D Webster horn equation

Table 1: *Formant frequency values (in Hz) and Rd estimated values obtained for the stimuli selected for the perceptual test.*

	F1	**F2**	**F3**	**F4**	**F5**	**Rd**
Original VT	714	1398	2506	3670	4428	
Low VE	633	1143	2877	3734	4133	1.92
Med VE	711	1262	2823	3492	4219	1.00
High VE	747	1264	2663	3255	3725	0.57
Low VE	594	1146	2787	3449	4049	2.16
Med VE	670	1171	2764	3399	3974	1.50
High VE	719	1145	2773	3279	4228	0.85



(a) *Glottal flows* (b) *Vocal tract tuning*

Figure 2: *Glottal flow signals obtained from GIF analysis and their corresponding generated LF-based signals for the three VE levels; and vocal tract tuning to modify an area function so that it produces a vocal tract transfer function $H(f)$ of a vowel /a/ with high VE.*

for the acoustic pressure using a Finite Element approach [24]. This equation in the frequency domain reads (see e.g., [25])

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(A\frac{\mathrm{d}p}{\mathrm{d}x}\right) + A\kappa^2 p = 0, \tag{2}$$

with $p(\boldsymbol{x}, f)$ being the acoustic pressure. Wall losses are introduced using the complex wave number κ, which can be computed as [26, 27]

$$\kappa^2 = k_0^2\left(1 - \mathrm{j}\frac{\mu\mathcal{P}}{k_0 A}\right), \tag{3}$$

with $\mu = 0.005$ being the boundary admittance coefficient, $k_0 = \omega/c_0$ the real wave number, $c_0 = 350$ m/s the speed of sound, $\mathcal{P}(x)$ the perimeter of the vocal tract, and $\mathrm{j} = \sqrt{-1}$ the imaginary unit. Radiation losses are neglected by imposing $p = 0$ at the mouth exit. A numerical simulation is run from 1 Hz to 12 kHz with a frequency resolution of 1 Hz, imposing an input volume velocity of $Q_i(f) = 1\ \mathrm{m}^3/\mathrm{s}$. The acoustic pressure $P_o(f)$ is computed at 0.03 m from the mouth exit. The vocal tract transfer function is next calculated as $H(f) = P_o(f)/Q_i(f)$. Finally, the vowel sound is computed as the convolution between the vocal tract impulse response $h(t)$ and the train of the glottal pulses $u_g(t)$ coming from the LF model, namely $p_o(t) = h(t) * u_g(t)$. In this way, only one FEM simulation is required for each vocal tract geometry and multiple vowel sounds with different excitation signals can be generated at a reduced computational cost.

3. Experiments and results

3.1. Selection of speech materials

Speech materials were obtained from the Zurich corpus of Vowel and Voice Quality, which contains isolated vowel sounds recorded by non-professional and professional actors and singers, with varying basic production parameters for phonation (voiced, breathy, creaky, whisper), vocal effort (low, medium and high) and fundamental frequencies [9].

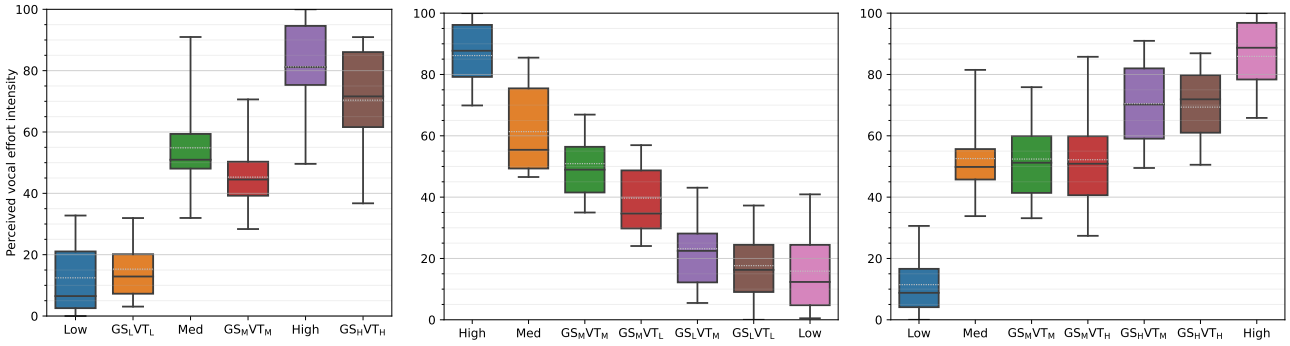
The objective evaluation was performed on a subset of 180 recordings selected according to the following criteria: adult male speakers; sustained vowel /a/ with F0 between 73 and 247Hz; and voiced phonation type with three levels of vocal effort. From this subset, 6 signals were chosen for the perceptual assessment. These examples correspond to a vowel /a/ uttered by a professional theatre actor and a contemporary singer at 110Hz with three levels of vocal effort. The formant frequencies and estimated Rd values for this signals are depicted in Table 1. Figure 2a shows an example of glottal flow signals (both the obtained from GIF analysis and the generated versions with LF model) for the three VE levels.

3.2. Vocal tract tuning

Figure 2b shows as an example the area function and the vocal tract transfer function $H(f)$ obtained to generate a vowel /a/ with high VE in the vocal tract tuning module. The algorithm is asked to obtain for the first five formant frequencies the values of 747, 1264, 2663, 3255, and 3725 Hz. Following the iterative algorithm based on sensitivity functions that is described in Section 2.2.2, it finds the area that produces this output.

3.3. Synthetic configurations

In order to evaluate the contribution of GS and VT in the generation of expressive vowels, nine synthetic configurations were considered following the methodology proposed in Figure 1. These configurations are denoted as $\mathrm{GS}_X\mathrm{VT}_Y$, where X and Y indicate the origin of the GS and VT used in the synthesis: L for low VE; M for medium VE; and H for high VE (hereafter, this subindex notation is applied for all the variables). For the GS model, the Rd parameter was estimated to match the target expressive utterance, while the VT geometry was tuned so their first five formants match the expressive utterance formants.



(a) *Real vs synthetic.* (b) *Synthetic configurations from Med to Low.* (c) *Synthetic configurations from Med to High.*

Figure 3: *Perceived vocal effort intensity for the synthetic vowels (GS_XVT_Y where X, Y correspond to **L**ow, **M**ed and **H**igh VE) and the real vowels (Low, Med, High).*

3.4. Objective evaluation

Each synthetic vowel was compared with the corresponding real expressive vowel to assess its proximity to the target. To this end, the similarity between two waveforms was computed as the symmetrical Kullback-Leibler spectral distance [28] between their long term average spectrums (LTAS).

The obtained Kullback-Leibler distances are depicted in Figure 4. The differences between all the synthetic configurations are statistically significant according to the Wilcoxon signed-rank test (with $p < 0.05$). When compared with the $\mathrm{GS}_M\mathrm{VT}_M$, it can be observed that both GS and VT have a relevant contribution to the generation of low and high vocal effort, significantly reducing the distance to the expressive target.

3.5. Perceptual assessment

The perceived vocal effort intensity for the different synthesis configurations was assessed through a MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) perceptual test [29]. The real reference signals (noted as Low, Med and High) were also included as anchor points. Thirty-four participants completed the test. Figure 3a shows the results for the first comparison, that includes the three reference signals and the three synthetic versions $\mathrm{GS_L VT_L}$, $\mathrm{GS_M VT_M}$ and $\mathrm{GS_H VT_H}$.

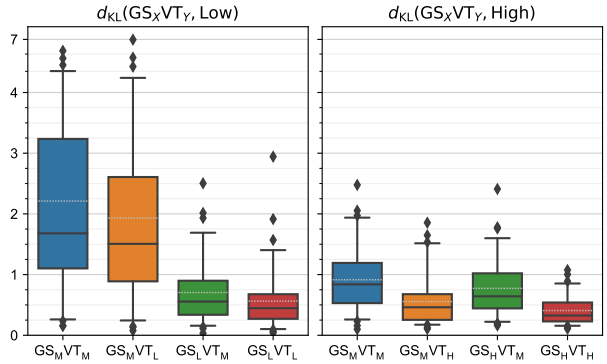


Figure 4: *Spectral Kullback-Leibler distances between the synthetic vowels (GS_XVT_Y where X, Y correspond to **L**ow, **M**ed and **H**igh VE) and the target real vowels (with Low VE in left subfigure and with High VE at right subfigure). Dotted lines represent the mean of the distributions, and whiskers are set to 5th and 95th percentiles.*

The perceived vocal effort of the synthetic signals is close to the reference for low and medium VE, with mean scores of 15 and 45 of the synthetic vs 12 and 55 for the real signals. Conversely, the synthetic high VE stimuli is perceived with lower VE than the high VE reference (70 vs 81).

The intermediate configurations from medium to low VE, and from medium to high VE were also evaluated. The results are depicted in Figure 3b and Figure 3c, respectively. In the med to low scenario, the differences between the evaluated configurations are statistically significant according to the Wilcoxon signed-rank test (with $p < 0.05$), except between $\mathrm{GS_L VT_M}$ and Low, and between $\mathrm{GS_L VT_L}$ and Low signals.

Conversely, in the configurations from med to high, there are non statistically significant differences between Med, $\mathrm{GS_M VT_M}$ and $\mathrm{GS_M VT_H}$ signals, as well as between $\mathrm{GS_H VT_M}$ and $\mathrm{GS_H VT_H}$ signals.

4. Conclusions

A methodology to introduce different levels of vocal effort in the simulation of vowels using the finite element method has been proposed and validated on vowel /a/. Both the vocal tract geometry and the glottal source model are adjusted based on the analysis of expressive speech.

Objective results demonstrate that modifications to both GS and VT are significantly relevant in approximating the expressive target. The results from the perceptual test show that this approach effectively conveys different degrees of vocal effort, particularly low and medium. Both GS and VT modifications yield statistically significant perceptual changes when transitioning from med to low VE, while GS predominates in the transformation from medium to high VE.

Future work will focus on validating the approach with other vowels and/or voice quality effects. Moreover, the proposed method will be adapted to be applied into the numerical simulation of vowels with 3D FEM-based models.

5. Acknowledgements

This research was partially funded by the Agencia Estatal de Investigación (AEI) through the FEMVoQ project (PID2020-120441GB-I00/AEI/10.13039/501100011033) and supported by the Departament de Recerca i Universitats (Generalitat de Catalunya) under Grant Ref. 2021 SGR 01396. The authors also thank Zihan Wang from Trinity College Dublin for providing access to their LF model implementation, and the participants of the perceptual test.

6. References

[1] P. Birkholz, L. Martin, Y. Xu, S. Scherbaum, and C. Neuschaefer-Rube, "Manipulation of the prosodic features of vocal tract length, nasality and articulatory precision using articulatory synthesis," *Computer Speech and Language*, vol. 41, pp. 116–127, 2017.

[2] P. Birkholz, L. Martin, K. Willmes, B. J. Kröger, and C. Neuschaefer-Rube, "The contribution of phonation type to the perception of vocal emotions in German: An articulatory synthesis study," *The Journal of the Acoustical Society of America*, vol. 137, no. 3, pp. 1503–1512, 2015.

[3] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)*, vol. 26, no. 4, pp. 1–13, 1985.

[4] A. Murphy, I. Yanushevskaya, A. N. Chasaide, and C. Gobl, "Rd as a Control Parameter to Explore Affective Correlates of the Tense-Lax Continuum," in *Proc. InterSpeech*, Stockholm, Sweden, Aug. 2017, pp. 3916–3920.

[5] Y. Li, J. Li, and M. Akagi, "Contributions of the glottal source and vocal tract cues to emotional vowel perception in the valence-arousal space." *The Journal of the Acoustical Society of America*, vol. 144, no. 2, p. 908, 2018.

[6] I. Yanushevskaya, C. Gobl, and C. A. Ní, "Cross-language differences in how voice quality and f_0 contours map to affect," *The Journal of the Acoustical Society of America*, vol. 144, no. 5, p. 2730, 2018.

[7] M. Freixes, M. Arnela, F. Alías, and J. C. Socoró, "GlottDNN-based spectral tilt analysis of tense voice emotional styles for the expressive 3D numerical synthesis of vowel [a]," in *Proc. 10th ISCA Speech Synthesis Workshop (SSW)*, Vienna, Austria, Sep. 2019, pp. 132–136.

[8] M. Arnela, O. Guasch, and M. Freixes, "Finite element generation of sung vowels tuning 3D MRI-based vocal tracts," in *Proc. of the 27th International Congress on Sound and Vibration (ICSV27)*. Graz, Austria - online: Silesian University Press, 11–16 July 2021, pp. 1–8.

[9] D. Maurer, C. d'Heureuse, H. Suter, V. Dellwo, D. Friedrichs, and T. Kathiresan, "The zurich corpus of vowel and voice quality, version 1.0," in *Interspeech*. International Speech Communication Association (ISCA), September 2018, pp. 1417–1421. [Online]. Available: https://doi.org/10.5167/uzh-156786

[10] B. Doval, C. d'Alessandro, and B. Diard, "Spectral methods for voice source parameters estimation," in *Proc. 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, 1997, pp. 533–536.

[11] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot international*, vol. 5, pp. 341–345, 01 2001.

[12] B. H. Story, "Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 715–718, 2006.

[13] O. Guasch, M. Arnela, and A. Pont, "Resonance tuning in vocal tract acoustics from modal perturbation analysis instead of nonlinear radiation pressure," *Journal of Sound and Vibration*, vol. 493, p. 115826, 2021.

[14] M. Airaksinen, B. Story, and P. Alku, "Quasi closed phase analysis for glottal inverse filtering," in *Proceedings of INTERSPEECH'2013*, 08 2013.

[15] D. Wong, J. Markel, and A. Gray, "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 4, pp. 350–355, 1979.

[16] T. Drugman and T. Dutoit, "Glottal closure and opening instant detection from speech signals," in *INTERSPEECH 2009, 10th Annual Conference of the International Speech*, Brighton, United Kingdom, 01 2009, pp. 2891–2894.

[17] C. Ma, Y. Kamp, and L. Willems, "Robust signal selection for linear prediction analysis of voiced speech," *Speech Communication*, vol. 12, no. 1, pp. 69–81, 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016763939390019H

[18] J. Kane and C. Gobl, "Automating manual user strategies for precise voice source analysis," *Speech Communication*, vol. 55, no. 3, pp. 397–414, 2013.

[19] Z. Wang and C. Gobl, "A System for Generating Voice Source Signals that Implements the Transformed LF-model Parameter Control," in *Proc. INTERSPEECH 2023*, 2023, pp. 4738–4742.

[20] M. Arnela, S. Dabbaghchian, O. Guasch, and O. Engwall, "Mri-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, pp. 2173–2182, 2019.

[21] G. Fant and S. Pauli, "Spatial characteristics of vocal tract resonance modes," *Speech Comm. Sem. Stockholm, Aug. 1974*, pp. 121–132, 1974.

[22] G. Fant, "Vocal-tract area and length perturbations," *STL-QPSR*, vol. 4, no. 1975, pp. 1–14, 1975.

[23] S. Adachi, H. Takemoto, T. Kitamura, P. Mokhtari, and K. Honda, "Vocal tract length perturbation and its application to male-female vocal tract shape conversion," *J. Acoust. Soc. Am.*, vol. 121, no. 6, pp. 3874–3885, 2007.

[24] O. Guasch, M. Arnela, and A. Pont, "Resonance tuning in vocal tract acoustics from modal perturbation analysis instead of nonlinear radiation pressure," *J. Sound Vib.*, vol. 493, p. 115826, 2021.

[25] M. M. Sondhi, "Model for wave propagation in a lossy vocal tract," *J. Acoust. Soc. Am.*, vol. 55, no. 5, pp. 1070–1075, 1974.

[26] L. J. Sivian, "Sound propagation in ducts lined with absorbing materials," *J. Acoust. Soc. Am.*, vol. 9(1), pp. 77–77, 1937.

[27] M. Arnela and O. Guasch, "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 4197–4209, 2013.

[28] E. Klabbers and R. Veldhuis, "Reducing audible spectral discontinuities," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 1, pp. 39–51, 2001.

[29] R. ITU, "ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems," *International Telecommunication Union*, 2003.