Interspeech 2022
18-22 September 2022, Incheon, Korea

Contribution of the Glottal Flow Residual in Affect-Related Voice Transformation

Zihan Wang, Christer Gobl

Phonetics and Speech Laboratory
School of Linguistic, Speech and Communication Sciences
Trinity College Dublin, Ireland
zwang7@tcd.ie, cegobl@tcd.ie

Abstract

This paper explores the contribution of the glottal flow residual in affect-related voice transformation. This signal, which is defined as the difference between the output of the inverse filter estimating the glottal flow signal and the modelled source signal, was analysed using multiple regression analysis. Results show that the strength of the residual varies as a function of the source parameters and this variation is frequency dependent: low frequency energy in the residual is mainly determined by the glottal excitation strength, whereas mid to high frequencies are more influenced by the glottal pulse shape. A method for modelling the residual is presented, which enables modifications based on the changes in source parameters used for voice transformation. This method makes it possible to use the residual as part of the voice source signal when transforming the voice quality in expressive speech synthesis. The result of a listening test, involving the transformation of a neutral voice to an angry or a sad voice, shows that including the glottal flow residual can improve the perceived naturalness of the synthesis. However, the fact that the transformed utterances are still relatively degraded indicates that other factors also need to be considered.

Index Terms: glottal flow, residual, voice source, voice quality, affect, voice modification, voice transformation.

1. Introduction

This paper investigates the contribution of the *glottal flow residual* to the quality of the synthetic speech output in affect-related voice modification. The glottal flow residual is here defined as the difference between the output of the inverse filter (i.e. the estimated glottal flow derivative) and the modelled glottal flow derivative. Thus, this residual signal is conceptually related to the linear prediction (LP) residual but defined within an acoustic-phonetic source-filter framework.

The contribution of this residual signal has not been extensively studied, but Nukaga et al. [1] found that it carries important information relating to voice quality and naturalness. Specifically, they suggested that the glottal flow residual during the open phase of the glottal cycle is important for cueing breathiness, while the residual during the return phase after the main glottal excitation may be particularly important for the naturalness of the output.

In a somewhat different vein, but related to the current study, Cabral and Oliveira [2] proposed a method involving time-scale modification of the LP residual to modify three source parameters. Informal listening tests suggested that this method produced synthetic speech with higher naturalness compared to standard time domain methods such as LP-PSOLA [3], particularly for large modifications to the pitch.

Although the broad aim of this paper is to explore the general properties of the glottal flow residual, it is done in the context of a recently developed analysis-resynthesis system, GlórCáil [4-6]. This system has been used in the ABAIR Irish synthesis project [7, 8] to facilitate voice transformations and more expressive voices.

The analysis part of the GlórCáil system initially performs a source-filter decomposition of the speech signal, whereby estimates of the vocal tract (VT) and voice source parameters of an utterance are captured. For voiced speech, inverse filtering provides an estimate of the glottal flow signal as well as a parametric representation of the VT function. The inverse filtering is done automatically by using the Glottal Flow Model-based Iterative Adaptive Inverse Filtering (GFM-IAIF) algorithm [9], which is a modified version of the well-established Iterative Adaptive Inverse Filtering (IAIF) method [10]. The estimated glottal flow signal is modelled using the Liljencrants-Fant (LF) model [11] and the source parameters which define the glottal pulse shape of the model are automatically derived by using the DyProg-LF algorithm [12, 13]. Furthermore, aspiration noise is added to the modelled glottal flow pulses according to the principles described in [14, 15].

The synthesis part of the GlórCáil system takes the modelled source signal and filters it using the VT filter function from the inverse filtering to generate the resynthesised speech signal. However, the synthesised source signal is not identical to the original: the difference is the glottal flow residual, i.e. any aspect of the source signal not captured by the model (as well as any error in the vocal tract/source estimation). It is generally assumed that if the modelling of the VT filter and the voice source is successful, the residual should be small, and thus relatively unimportant.

However, the quality of the resynthesised speech is in fact often noticeably degraded. This can even be the case when there is no transformation involved, but the quality is generally worse as the degree of modification is increased [5]. For that reason, this study focusses on the contribution of the glottal flow residual to the quality of the speech output in the context of such voice modifications. A method for modelling the residual is proposed that would enable modifications to the residual corresponding to changes in voice source parameters in affect transformation. The method was tested on sample utterances and listeners' evaluations were elicited.

  10.21437/Interspeech.2022-11038

2. Analysis methods

In order to analyse the glottal flow residual, it needs to be extracted from the speech signal. Although the GlórCáil system can be used to do this, we decided here to opt instead for carefully analysed speech using the manual interactive techniques described in [16-18]. By doing so, any artefacts in the source-filter decomposition were minimised, and therefore the obtained glottal flow residual would represent a realistic estimate of this signal.

Apart from the accuracy of the source-filter decomposition, the estimated residual would to a certain extent also depend on the voice source model used. In this study, the LF model was chosen, mainly because it is used in GlórCáil, but also since it is a well-established model that has been used extensively. Nevertheless, it is expected that many other source models, e.g., [19-24] would produce similar results.

2.1. Speech data

The speech materials used are a set of recordings of the sentence "We were away a year ago", originally used in [25]. The recordings were carried out in a semi-anechoic studio using high-quality equipment to ensure phase linear speech data, which is required for this type of voice source analysis. The utterances were produced by a male speaker of Irish English, portraying different affects, namely *angry*, *bored*, *sad*, *surprised* as well as a neutral version.

For the voice source analysis, the recordings were downsampled to a sampling rate of 10 kHz. The five utterances contained a total of 609 glottal cycles, which were analysed by the second author using the interactive inverse filtering and model matching techniques described in [15-17].

2.2. Voice source parameters

In this study we focus on the three voice source parameters which are used as control parameters in the GlórCáil system, namely: the strength of the glottal excitation E_e, the global waveshape parameter R_d and the fundamental frequency f_0.

However, since the LF model is used to synthesise the glottal waveform, the full set of LF parameters is required. The LF model is defined according to the piecewise function in (1), where the first and second parts define the glottal flow derivative in the open phase and in the return phase respectively (see also the lower panel of Fig. 1).

$$U_g'(t) = \begin{cases} E_0\, e^{\alpha t} \sin \omega_g t & 0 \le t < t_e \\ -\dfrac{E_e}{\varepsilon T_a}\left(e^{-\varepsilon (t - t_e)} - e^{-\varepsilon T_b} \right) & t_e \le t < t_c \end{cases} \tag{1}$$

Note that E_0 and T_a are auxiliary parameters: E_0 can be derived from E_e, α, T_e, and ω_g, and T_a from ε and T_b as shown in (2) and (3) respectively.

$$E_0 = -E_e \frac{e^{-\alpha T_e}}{\sin \omega_g T_e} \tag{2}$$

$$T_a = \frac{1 - e^{-\varepsilon T_b}}{\varepsilon} \tag{3}$$

Hence, the LF model parameters are E_e, T_e, ω_g, α, ε and T_b, and the pulse duration is $T_e + T_b$. Note that analysis data for α are not needed since α is set to meet the LF model requirement of a given net flow gain. As is normally done, this net flow gain was here set to 0, which means that the α value is set so that the 'area' of the positive and negative parts of the flow derivative are the same.

Normally, analysis data are obtained for T_a (or R_a, see Fig. 1) and therefore ε will need to be derived from T_a and T_b.

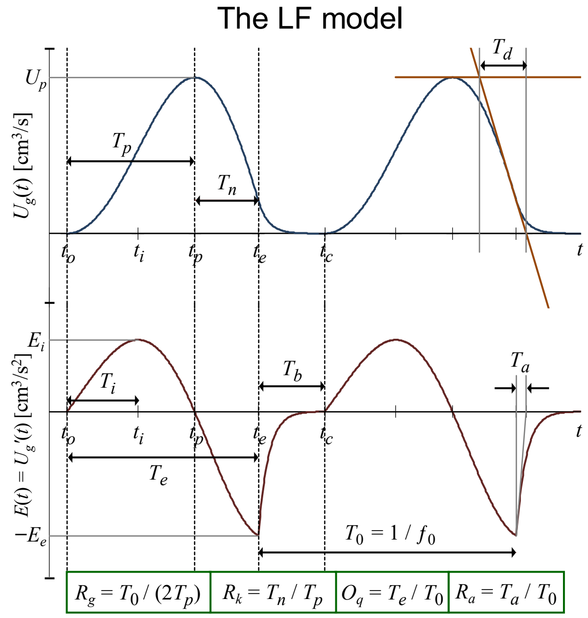

Figure 1: *Two LF model pulses and parameter definitions. Flow derivative (bottom) and corresponding glottal flow (top).*

In [26] Fant recast the LF model using a different set of parameters for a more intuitive control of voice quality, and this transformed LF modelled includes the global waveshape parameter R_d as the central parameter.

Note that $R_d = 0.11^{-1} f_0\, U_p / E_e$, which means that the R_d value also depends on the peak glottal flow U_p (see Fig. 1). Consequently, we also need data for U_p. Although not a parameter of the LF model, U_p can be obtained from the LF parameters according to (4).

$$U_p = -E_e \frac{\omega_g \left(e^{-\alpha T_e} + e^{-\alpha \left(T_e - \pi / \omega_g \right)} \right)}{\left(\omega_g^2 + \alpha^2 \right) \sin \omega_g T_e} \tag{4}$$

2.3. Analysis and modelling of the glottal flow residual

Since E_e is a major determinant of the sound pressure level (SPL) of the source, it is hypothesised that the SPL of the residual signal will be positively correlated with E_e – that is, the assumption is that the higher the E_e value, the greater the overall intensity of the residual. Furthermore, since E_e is likely to be higher where an utterance has higher f_0 values [27], it might also be the case that the SPL of the residual and f_0 are positively correlated. As for R_d, the distribution of energy in different frequency regions of the residual might be correlated with R_d, since variation in R_d affects the source spectral slope, where a higher R_d tends to correlate with a laxer phonatory setting, and a lower R_d with tenser settings.

To investigate the relationships between the voice source parameters and the intensity of the residual, multiple linear regression analysis was carried out. The SPL data for the residual signals were derived using Praat [28], and the SPL trajectories were carefully time-aligned with the voice source data.

Multiple linear regression analysis was initially carried out between the overall SPL of the residual signals and E_e, R_d and f_0 of the different utterances, and the results indicate a high level of correlation ($R^2 = 0.72$). The modelling function of the change in SPL of the residual as function of the dB-change in E_e (ΔE_e), the f_0 change (Δf_0) and the change in R_d (ΔR_d) is shown in (5).

$$\Delta L_{res} = 0.280\Delta E_e + 0.077\Delta f_0 - 2.69\Delta R_d \quad (5)$$

The analysis thus suggests that there is a positive correlation between the SPL of the residual and the parameters E_e and f_0, whereas the correlation with R_d is negative.

Multiple linear regression analysis was also carried out between the change in SPL of the residual in three different frequency bands and the three source parameters. The frequency bands were 0-1 kHz, 1-3 kHz and 3-5 kHz. The modelling functions of the change in SPL of the residual in the different frequency bands are shown in (6a), (6b) and (6c), and the R^2 values were 0.70, 0.75 and 0.69 respectively.

$$\Delta L_{res0-1} = 0.289\Delta E_e + 0.072\Delta f_0 - 2.42\Delta R_d \quad (6a)$$

$$\Delta L_{res1-3} = 0.141\Delta E_e + 0.131\Delta f_0 - 7.23\Delta R_d \quad (6b)$$

$$\Delta L_{res3-5} = 0.024\Delta E_e + 0.127\Delta f_0 - 9.33\Delta R_d \quad (6c)$$

The results from this analysis show that variation in the strength of the residual is frequency dependent. Notably, (5) and (6a) are very similar suggesting that the scaling of the low frequencies of the residual is very similar to the overall amplitude scaling. However, for the scaling in the higher frequency ranges, E_e becomes much less important, whereas f_0 and R_d (in particular) contribute more to the scaling.

This would suggest that low frequency energy in the residual is mainly determined by the glottal excitation, with a stronger excitation resulting in stronger low frequency energy in the residual. Mid to high frequency energy of the residual is more influenced by the glottal pulse shape: a glottal pulse generating more high frequency energy also results in more high frequency energy in the residual. Since the analysis shows that results differ among the frequency bands, the approach adopted for the modifications used here, are based on the expressions in (6a), (6b) and (6c).

The first step of the modification of the residual is to calculate the scale factors for the three different frequency bands of the residual signal with the changes in E_e, R_d and f_0. The residual signal is then rescaled by the scale factor of the 0-1 kHz frequency band and filtered by two pairs of first order low-pass and pre-emphasis filters.

The purpose of applying the low-pass and pre-emphasis filters in pairs is to achieve smooth changes in the amplitude levels of the spectrum of the residual across the frequency bands. The level difference between two adjacent bands can be realised by setting two symmetric cut-off frequencies (symmetric in terms of log frequency) for the filters around the edge of the adjacent frequency bands, i.e. around 1 kHz and 3 kHz for each pair of filters respectively. If there is no difference in the level between two adjacent frequency bands, e.g., between 1-3 kHz and 3-5 kHz, then the cut-off frequencies of both the low-pass filter and the pre-emphasis filter is set to 3 kHz and the two filters will cancel each other out.

3. Voice transformations

The relationship between the SPL of the residual and the source parameters obtained from the regression analysis enables voice transformation, which includes a modified glottal flow residual. In this section, we describe the process used to transform the original *neutral* utterance into utterances with a perceived angry or sad affect. For the purpose of testing and evaluating this process, we used the analysis of the original *angry* and *sad* utterances as targets for these transformations. Note, however, that having a target utterance is not necessarily required to carry out modifications involving the glottal flow residual.

Initially, the source parameters of the target were obtained (i.e. for the *angry* and the *sad* utterances in this case). To do this, alignment of the target parameter trajectories is necessary since each utterance has its unique glottal excitation timepoints (t_e, see Fig. 1). The alignment was here realised by scaling the overall duration of the target utterance to be the same as the duration of the original utterance. Note that since the overall duration of the utterances were relatively similar, as were the individual segment durations, non-linear time-scaling was not deemed necessary for these modifications.

The parameter trajectories for the source parameters E_e, U_p, O_q, T_a and f_0 were derived by linearly interpolating the source data for the target utterance. These trajectories were then resampled, and the target values for the parameters were extracted at the timepoints matching the excitation timepoints of the original utterance. Using the new values for E_e, U_p and f_0, the target R_d values could also be obtained, which were needed for the modification of the original glottal flow residual.

Using the modified source parameter data, a new LF voice source signal was generated, but at this point the f_0 contour of the original utterance was not changed. To this new source signal, the modified glottal flow residual was added, as derived from the expression in (6a), (6b), (6c) and the filtering process described in Section 2.4. Since f_0 is still identical to the original utterance, perfect alignment between the modified residual and the synthesised source signal was maintained.

At this stage the modified source signal, which now included the modified residual, was passed through the original VT filter derived from the inverse filtering, by which a transformed synthetic speech signal was generated, but still without f_0 modifications. As a final step, f_0 was modified using the TD-PSOLA method [3] to achieve the f_0 contour of the target utterance.

4. Listening test

To evaluate the effect of the filtered glottal flow residual on the quality of the transformed speech, a listening test involving the resynthesised speech utterances was performed. The test was administered online, and participants were instructed to take the test in a quiet room using high-quality headphones. They were also informed of the approximate duration of the test and that they could listen to each stimulus as many times as they wished. Before the main test, a practice file was presented, which contained all the test stimuli. This would allow the participants to familiarise themselves with the stimuli and to adjust the volume to a comfortable level. Participants would then proceed to the main test and judge the naturalness of the stimuli on a scale from 1 to 7, where 1 denotes extremely unnatural (i.e. the stimulus sounds very artificial) and 7 perfectly natural (i.e. the stimulus is likely to be produced by a human). The stimuli were presented in a randomised order, and included three versions of *neutral*, *angry* and *sad* voice: the original recording and the resynthesised signal with and without the modified glottal flow residual.

In total, 25 participants completed the test, and the results in terms of the mean opinion score are presented in Fig. 2. The mean values of the naturalness increased in all cases thus showing that the glottal flow residual can potentially contribute to the naturalness of the synthetic output in voice transformations. Nevertheless, the amount of improvement was rather less than expected for the transformed voices. This suggests that there are also other factors contributing to the degrading of the quality as the signal is transformed.

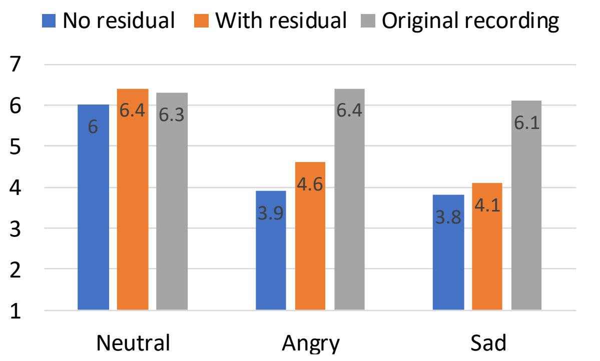

Figure 2: *Mean opinion score of the naturalness of utterances synthesised with or without the glottal flow residual.*

Note that, as expected, the resynthesis of the neutral utterance with the residual received almost identical ratings to the original recording. This simply shows that by adding back the residual to the modelled source signal, one should in principle obtain the original waveform.

5. Conclusions

This paper explores the potential contribution of the glottal flow residual for improving the quality of the synthetic speech output in the context of affect-related voice transformation.

The properties of this residual signal were analysed and modelled using multiple regression analysis, and the results suggest that there is a strong correlation between the SPL of the residual and the source parameters E_e, R_d and f_0. The results also show that the variation is frequency dependent, where low frequency energy in the residual is mainly determined by the glottal excitation and mid to high frequency energy of the residual is more influenced by the glottal pulse shape.

A method for modelling the residual is presented, which enables modifications of the residual based on the changes in the voice source parameters used in voice transformations. Consequently, this modelling makes it possible to use the modified residual as part of the voice source signal when transforming the voice quality in expressive speech synthesis.

The three voice source parameters used as control parameters in the GlórCáil system (i.e. E_e, R_d, and f_0) and the SPL of the residual in different frequency bands were used in the modelling and in the transformation of the residual. The results of the listening test support the proposition that when modified appropriately, the residual signal can improve the quality of the synthesis. However, they also indicate that even with the inclusion of the residual signal thus modified, the quality of the transformed voice is judged to be degraded relative to the original recordings of the targeted affective utterances.

As a next step we plan to incorporate the glottal flow residual into the GlórCáil system. However, the residual modelling will have to be tuned to the system, since the automatic analysis will most likely yield a larger residual with, potentially, a different correlation pattern. We further envisage on-the-fly adaptation of the modelling to optimise the residual also to the speaker's voice characteristics.

A limitation in the modelling presented here is that the voice transformation did not include any VT filter modifications or changes to the durations. Although the current modifications can transform the utterances and convey different affects, such additional modifications should contribute to the expressiveness of the utterances, and therefore also improve the naturalness. For instance, the lax voice quality associated with the sad utterance also results in substantially increased formant bandwidths – a source related effect on the VT filter settings that was not included here.

For the transformed angry voice, the voice quality was on the other hand quite tense. This combined with very narrow formant bandwidths for this speaker produced strong higher harmonics which may have caused some distortion due to aliasing. The LF model, and other source models based on sampled piecewise elementary functions, can introduce noticeable aliasing distortion when the sampling frequency is relatively low [29-31]. We are therefore planning a rerun of the experiment using instead the aliasing free version of the LF model presented in [32].

Additionally, while the GlórCáil system includes aspiration noise as part of the voiced excitation, the synthetic voice signal used here did not take this into account. It might therefore be interesting to compare the relative contribution of this noise source with that of the glottal flow residual in future studies.

An important aspect of the modelling of the residual is the capacity to capture frequency dependent changes. There was evidence from the present analysis that changes in the residual are indeed frequency-dependent when source parameters are modified to synthesise different affects. However, the frequency-depended model that was developed here needs to be explored in more detail in order to clarify precisely how the spectrum of the residual is affected by changes to the source parameters. The perceptual consequences of these spectral changes to the residual also need further investigation. It would be interesting to establish to what extent the more complex frequency modelling contributes to the naturalness and the affect perception of the synthetic speech compared to the simpler modelling involving only modification of the amplitude level of the residual.

6. Acknowledgements

This research was carried out in the Róbóglór and ABAIR projects, supported by the Irish Govt. Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media, with funding from the National Lottery, supporting An Stráitéis 20 bliain don Ghaeilge 2010-2030.

7. References

[1] N. Nukaga, C. Gobl and A. Ní Chasaide, "Experimental study on voice quality control based on source-filter decomposition," *Proceedings of the Workshop on Innovation and Applications in Speech Technology (IAST)*, Dublin, Ireland, pp. 49-52, 2012.

[2] J. P. Cabral and L. C. Oliveira, "Pitch-synchronous time-scaling for prosodic and voice quality transformations," *INTERSPEECH 2005*, Lisbon, Portugal, 2005, pp. 1137-1140.

[3] E. Moulines and F. Charpentier, "Pitch synchronous waveform processing techniques for text- to-speech synthesis using diphones," *Speech Communication*, vol. 9, pp. 453-467, 1990.

[4] A. Murphy, *Controlling the Voice Quality Dimension of Prosody in Synthetic Speech using an Acoustic Glottal Model*, unpublished PhD thesis, Trinity College Dublin, 2020.

[5] A. Murphy, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "Testing the GlórCáil System in a Speaker and Affect Voice Transformation Task," *Proc. 10th International Conference on Speech Prosody,* Tokyo, Japan, pp. 950-954, 2020.

[6] A. Murphy, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl (2021), "Integrating a voice analysis-synthesis system with a TTS Framework for controlling affect and speaker identity. *2021 32nd Irish Signals and Systems Conference (ISSC)*, 2021, pp. 1-6.

[7] A. Ní Chasaide, N. Ní Chiaráin, C. Wendler, H. Berthelsen, A. Murphy, and C. Gobl, "The ABAIR initiative: bringing spoken Irish into the digital space," *INTERSPEECH 2017*, Stockholm, Sweden, 2017, pp. 2113-2117.

[8] ABAIR.ie – The Synthesiser for Irish, https://abair.ie

[9] O. Perrotin and I. McLoughlin, "A spectral glottal flow model for source-filter separation of speech," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, pp. 7160-7165, 2019.

[10] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Communication*, vol. 11, nos. 2-3, pp. 109-118, 1992.

[11] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 26 (4), pp. 1-13, 1985.

[12] J. Kane, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "Exploiting time and frequency domain measures for precise voice source parameterisation," in *Speech Prosody 2012*, Shanghai, China, 2012, pp. 143-146.

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

[14] C. Gobl, "Aspiration noise generation based on glottal pulse characteristics," *Proceedings of the 9th Western Pacific Acoustics Conference*, Seoul, Korea, paper 274, pp. 8, 2006.

[15] Gobl, C., "Modelling aspiration noise during phonation using the LF voice source model," *Proceedings of the 8th International Conference on Spoken Language Processing, INTERSPEECH 2006*, Pittsburgh, Pennsylvania, 2006, pp. 965-968.

[16] P. Monahan, *Systems for voice source analysis*, unpublished PhD thesis, Trinity College Dublin, 1996.

[17] C. Gobl and and A. Ní Chasaide, "Techniques for analysing the voice source," in *Coarticulation: Theory, Data and Techniques*, eds. William J. Hardcastle and Nigel. Hewlett, Cambridge University Press, Cambridge, 300-321, 1999.

[18] C. Gobl and A. Ní Chasaide, "Voice source variation and its communicative functions," in *The Handbook of Phonetic Sciences (Second Edition)*, eds. William J. Hardcastle, John Laver and Fiona E. Gibbon, Oxford, Blackwell, pp. 378-423, 2010.

[19] H. Fujisaki and M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Tokyo, Japan, pp. 1605-1608, 1986.

[20] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 583-590, 1971.

[21] T. V. Ananthapadmanabha, "Acoustic analysis of voice source dynamics," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 25 (2-3), pp. 1-24, 1984.

[22] G. Fant, "Glottal source and excitation analysis," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 20 (1), pp. 85-107, 1979.

[23] Y.-L. Shue and A. Alwan, "A new voice source model based on high-speed imaging and its application to voice source estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, pp. 5134-5137, 2010.

[24] R. Veldhuis, "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation," *Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 566-571, 1998.

[25] I. Yanushevskaya, C. Gobl, and a. Ní Chasaide, "Voice parameter dynamics in portrayed emotions," *Proceedings of the 6th International Workshop: Models and analysis of vocal emissions for biomedical applications*, Firenze University Press, Florence, Italy, pp. 21-24, 2009.

[26] G. Fant, "The LF-model revisited. Transformations and frequency domain analysis," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 36 (2-3), pp. 119-156, 1995.

[27] G. Fant, "The voice source in connected speech," *Speech Communication*, vol. 22, pp. 125-139, 1997.

[28] Boersma, P. and Weenink, D. (2021), Praat: Doing phonetics by computer (Version 6.1.40). Software available from https://www.fon.hum.uva.nl/praat/

[29] P. H. Milenkovic, "Voice source model for continuous control of pitch period," *Journal of the Acoustical Society of America*, vol. 93, no. 2, pp. 1087-1096, 1993.

[30] H. Kawahara, K.-I. Sakakibara, H. Banno, M. Morise, T. Toda, and T. Irino, "Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation," *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Hong Kong, pp. 520-529, 2015.

[31] K. Hideki Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, T. Toda, and T. Irino, A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis", *INTERSPEECH 2017*, Stockholm, Sweden, 2017, pp. 1358-1362.

[32] C. Gobl, "The LF model in the frequency domain for glottal airflow modelling without aliasing distortion," *INTERSPEECH 2021*, Brno, Chechia, 2021, pp. 2651-2655.