INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

A System for Generating Voice Source Signals that Implements the Transformed LF-model Parameter Control

Zihan Wang[1], *Christer Gobl*[2]

[1,2]Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland

zwang7@tcd.ie, cegobl@tcd.ie

Abstract

This paper describes a system which fully implements the transformed LF glottal flow model, incorporating also the often-overlooked *k*-factors. A problem with the original proposal is that the global waveshape parameter R_d, central to the transformed model and used to predict some of the other parameters, cannot be directly controlled. Instead, a stylisation of the model was used to indirectly control R_d. However, this approach can yield substantial errors in the R_d value of the pulse, thus undermining the usefulness of the model. To overcome this problem, an iterative algorithm is presented, which ensures that for a given R_d input value, a pulse with that R_d value will be produced. Using this new R_d control, it transpired that some of the original parameter predictions give rise to combinations of values that are incompatible with the LF model. Modifications to the original predictions of the R_k parameter were incorporated to ensure model conformity.

Index Terms: LF model, R_d parameter, glottal flow, voice source, Newton-Raphson, iteration.

1. Introduction

A software system is presented for voice source waveform generation, which is based on the principles of what is commonly referred to as the transformed Liljencrants-Fant (LF) model [1, 2]. The aim was to develop a system which can produce source signals that adhere to the original concept of the transformed model as closely as possible. Since the often-overlooked *k*-factors are an integral part of the transformed model, they have also been incorporated in the system.

The original LF model as described in [3] is a well-established parametric model of glottal flow, which has been extensively used in research on the glottal source and voice quality in a range of studies, involving analysis, synthesis and perception (e.g., [4-25]). In the transformed model, the model itself is in fact identical to the original LF model: it differs only with respect to the parameter control. In the original version, the pulse shape is typically determined by the *R*-parameters: R_g (normalised frequency of the glottal formant), R_k (degree of glottal pulse symmetry) and R_a (normalised effective duration of the return phase) while the amplitude is controlled by the E_e parameter, i.e. the negative amplitude of the main glottal excitation (see Fig. 1). Albeit useful, the *R*-parameters show a high level of correlation and thus convey redundant information.

In the transformed control of the model, the waveshape parameter R_d [1, 2] is used to capture some of the naturally occurring covariation in the *R*-parameters by combining the values of E_e, U_p (peak glottal flow) and f_0 (voice fundamental frequency). The definition of R_d is shown in (1), where the scale factor 0.11^{-1} makes the R_d value the same as the declination time T_d (in ms) for f_0 = 110 Hz. Except for this constant, R_d is essentially the same as the popular normalised amplitude quotient (NAQ) parameter proposed in [26]. It has been shown that R_d captures the degree of tension in the voice, where a low value corresponds to a tense voice and a high value to a lax or breathy voice.

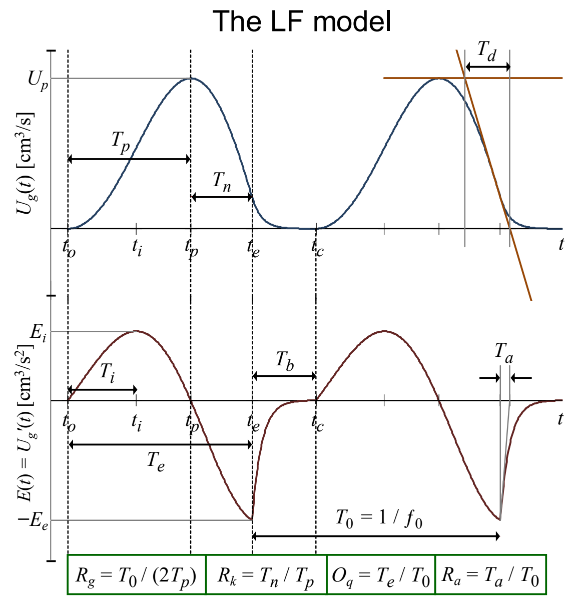

Figure 1: *The LF model: glottal flow (top), flow derivative (bottom).*

To synthesise the glottal waveform, R_d is used to predict the *R*-parameters, which in turn are converted into actual LF model parameters (see Section 2). These predictions are based on regression analysis of source data obtained from inverse filtering of natural speech. R_{kp} and R_{ap}, the predicted values of R_k and R_a respectively, are calculated from R_d according to (2) and (3).

$$R_d = 1000 \frac{U_p}{E_e} \frac{f_0}{110} = \frac{1}{0.11} \frac{T_d}{T_0} \quad (1)$$

$$R_{kp} = 0.118\, R_d + 0.224 \quad (2)$$

$$R_{ap} = 0.048\, R_d - 0.01 \quad (3)$$

Although there is also a similar expression for R_{gp} presented in [1], this is not used, as it would lead to an overspecification of the model: R_g is therefore determined by the values of R_d, R_{kp} and R_{ap}.

10.21437/Interspeech.2023-2204

Since there is no analytic expression to convert R_d into an LF parameter, a stylisation of the LF model was originally used to get R_g given R_d, R_{kp} and R_{ap}. This method, however, leads to errors in the R_d value of the synthesised glottal pulse, which can be quite substantial unless the potential range of parameter values is restricted.

In [27] an algorithm is presented which enables direct control of R_d in the LF model by using an iterative technique based on the Newton-Raphson method for two variables. Although it provides for accurate control of R_d, this algorithm does not conform with the transformed LF model. Since it requires that the duration of the open phase, T_e, be specified in order to generate the LF pulse, R_k is not specified by R_{kp} in (2). Furthermore, the algorithm in [27] does not allow for the *k*-factors to be applied to the predicted *R*-parameters for a more exact modelling. Therefore, a new algorithm was developed which provides accurate control of R_d without the need for prior specification of T_e. Furthermore, separate algorithms were developed to cater for the *k*-factors.

2. Deriving the LF parameters from the transformed model parameter set

The differentiated glottal flow pulse of the LF model (Fig. 1) is defined by the two functions in (4) for the open phase (t_o to t_e) and the return phase (t_e to t_c), where E_0 and T_a are according to (5) and (6) respectively.

$$U_g'(t) = \begin{cases} E_0\, e^{\alpha t} \sin\left(\omega_g (t - t_o)\right) & t_o \le t < t_e \\ \dfrac{-E_e}{\varepsilon T_a}\left(e^{-\varepsilon (t - t_e)} - e^{-\varepsilon T_b} \right) & t_e \le t < t_c \end{cases} \tag{4}$$

$$E_0 = -E_e e^{-\alpha T_e} \csc \omega_g T_e \tag{5}$$

$$T_a = \varepsilon^{-1}\left(1 - e^{-\varepsilon T_b}\right) \tag{6}$$

From (4) - (6) we see that the following six parameters are required to generate the LF pulse: E_e, T_e, ω_g, α, ε and T_b, which are the actual LF parameters. When synthesising consecutive pulses, T_0 is used to control pitch, where T_0 is T_b of the 'current' pulse plus T_e of the following pulse (see Fig. 1).

T_e and ω_g can be obtained directly from the *R*-parameters and T_0: $T_e = T_0(1+R_k)/(2R_g)$ and $\omega_g = 2\pi R_g / T_0$. However, ε and α require iterative estimation: ε is calculated from T_a and T_b, and α is determined from the LF model requirement of a given net flow gain, which is typically set to 0 [3]: therefore, there should be 'area balance', i.e. the area of the positive part of the model should equal the area of its negative part (Fig. 1). To find ε and α, the Newton-Raphson method as shown in (7) is often used:

$$x_{n+1} = x_n - \mathrm{f}(x_n)/\mathrm{f}'(x_n) \tag{7}$$

For calculating ε, $\mathrm{f}(\varepsilon)$ and $\mathrm{f}'(\varepsilon)$ are as follows:

$$\mathrm{f}(\varepsilon) = \varepsilon T_a - 1 + e^{-\varepsilon T_b} \tag{8}$$

$$\mathrm{f}'(\varepsilon) = T_a - T_b e^{-\varepsilon T_b} \tag{9}$$

For α, we get $\mathrm{f}(\alpha)$ and $\mathrm{f}'(\alpha)$, which are shown in (10) and (11).

$$\begin{aligned} \mathrm{f}(\alpha) = {} & \varepsilon\left(1 - e^{-\varepsilon T_b}\right)\left[\alpha \sin \omega_g T_e - \omega_g \cos \omega_g T_e + \omega_g e^{-\alpha T_e}\right] \\ & + \left(\omega_g^2 + \alpha^2\right) \sin \omega_g T_e \left[1 - e^{-\varepsilon T_b}\left(1 + \varepsilon T_b\right)\right] \end{aligned} \tag{10}$$

$$\begin{aligned} \mathrm{f}'(\alpha) = {} & \varepsilon\left(1 - e^{-\varepsilon T_b}\right)\left[\sin \omega_g T_e - \omega_g T_e e^{-\alpha T_e}\right] + \\ & 2\alpha \sin \omega_g T_e \left[1 - e^{-\varepsilon T_b}\left(1 + \varepsilon T_b\right)\right] \end{aligned} \tag{11}$$

As mentioned, the R_d value is calculated from E_e, U_p and f_0. E_e is an LF parameter and f_0 is used to determine T_b as $f_0^{-1} - T_e$. However, for U_p there is no analytic expression to convert it into LF parameters.

In the original LF model, U_p is allowed to 'float', i.e. it may attain whatever value is required to fulfil the condition of area balance. However, in the transformed LF model, two conditions must be fulfilled: (i) conformity with the specific U_p value as determined by R_d and f_0, and (ii) area balance.

To solve this, we take a similar approach to that presented in [27] where the Newton-Raphson method for two variables is used. However, in [27] the values of ω_g and α are allowed to float to obtain the specified R_d value and area balance, which requires a given T_e (or O_q, which is the glottal open quotient, here defined as T_e normalised to T_0, see Fig. 1). Hence, R_k is derived from the obtained ω_g and T_e rather than by R_{kp} in (2).

For the new algorithm, we instead let T_e and α float, and R_k is given by R_{kp} in (2). By integrating (4) from t_o to t_p and by expressing ω_g in terms of T_e and R_k as $(\pi/T_e)(1+R_k)$, U_p is according to (12) from which function f_1 in (13) is obtained:

$$U_p = \frac{-E_e \pi T_e^{-1}\left(1 + R_k\right)\left(e^{-\alpha T_e} + e^{-\alpha T_e R_k (1+R_k)^{-1}}\right)}{\left(\pi^2 T_e^{-2}\left(1 + R_k\right)^2 + \alpha^2\right) \sin\left(\pi\left(1 + R_k\right)\right)} \tag{12}$$

$$\begin{aligned} \mathrm{f}_1(T_e, \alpha) = {} & U_p\left(\pi^2 T_e^{-2}\left(1 + R_k\right)^2 + \alpha^2\right) \sin\left(\pi\left(1 + R_k\right)\right) \\ & + E_e \pi T_e^{-1}\left(1 + R_k\right)\left(e^{-\alpha T_e} + e^{-\alpha T_e R_k (1+R_k)^{-1}}\right) \end{aligned} \tag{13}$$

The second condition, i.e. area balance, provides us with the second function, f_2, which is the same as $\mathrm{f}(\alpha)$ in (10) except that ω_g is replaced by $(\pi/T_e)(1+R_k)$. The four partial derivatives required for the iterative calculations are as follows:

$$\begin{aligned} \frac{\partial \mathrm{f}_1}{\partial T_e} = {} & -2\pi^2 T_e^{-3}\left(1 + R_k\right)^2 U_p \sin\left(\pi\left(1 + R_k\right)\right) - \\ & E_e \pi T_e^{-2}\left[\begin{aligned} & \left(1 + \alpha T_e\right)\left(1 + R_k\right) e^{-\alpha T_e} + \\ & \left(1 + R_k\left(1 + \alpha T_e\right)\right) e^{-\alpha T_e R_k (1+R_k)^{-1}} \end{aligned}\right] \end{aligned} \tag{14}$$

$$\begin{aligned} \frac{\partial \mathrm{f}_1}{\partial \alpha} = {} & 2\alpha U_p \sin\left(\pi\left(1 + R_k\right)\right) - \\ & \pi E_e\left(\left(1 + R_k\right) e^{-\alpha T_e} + R_k e^{-\alpha T_e R_k (1+R_k)^{-1}}\right) \end{aligned} \tag{15}$$

$$\begin{aligned} \frac{\partial \mathrm{f}_2}{\partial T_e} = {} & \varepsilon\left(1 - e^{-\varepsilon T_b}\right) \times \\ & \pi T_e^{-2}\left(1 + R_k\right)\left[\cos\left(\pi\left(1 + R_k\right)\right) - \left(1 + \alpha T_e\right) e^{-\alpha T_e}\right] - \\ & 2\pi^2 T_e^{-3}\left(1 + R_k\right)^2 \sin\left(\pi\left(1 + R_k\right)\right)\left[1 - e^{-\varepsilon T_b}\left(1 + \varepsilon T_b\right)\right] \end{aligned} \tag{16}$$

$$\begin{aligned} \frac{\partial \mathrm{f}_2}{\partial \alpha} = {} & \varepsilon\left(1 - e^{-\varepsilon T_b}\right)\left[\sin\left(\pi\left(1 + R_k\right)\right) - \pi\left(1 + R_k\right) e^{-\alpha T_e}\right] + \\ & 2\alpha \sin\left(\pi\left(1 + R_k\right)\right)\left[1 - e^{-\varepsilon T_b}\left(1 + \varepsilon T_b\right)\right] \end{aligned} \tag{17}$$

The two formulas in (18) are then used for iteratively calculating T_e and α, where the Jacobian determinant is shown in (19). Note that all functions in (18) are functions of T_{ek} and α_k. For further details on this iterative procedure, see [27].

$$\begin{aligned} T_{e_{k+1}} &= T_{e_k} - \left(\det\left(\mathbf{J}\left(T_{e_k},\alpha_k\right)\right)\right)^{-1}\left[\frac{\partial \mathrm{f}_2}{\partial \alpha}\mathrm{f}_1 - \frac{\partial \mathrm{f}_1}{\partial \alpha}\mathrm{f}_2\right] \\ \alpha_{k+1} &= \alpha_k - \left(\det\left(\mathbf{J}\left(T_{e_k},\alpha_k\right)\right)\right)^{-1}\left[-\frac{\partial \mathrm{f}_2}{\partial T_e}\mathrm{f}_1 + \frac{\partial \mathrm{f}_1}{\partial T_e}\mathrm{f}_2\right] \end{aligned} \quad (18)$$

$$\det\left(\mathbf{J}\left(T_e,\alpha\right)\right) = \begin{vmatrix} \mathrm{f}'_{1_{T_e}} & \mathrm{f}'_{1_\alpha} \\ \mathrm{f}'_{2_{T_e}} & \mathrm{f}'_{2_\alpha} \end{vmatrix} = \frac{\partial \mathrm{f}_1}{\partial T_e}\frac{\partial \mathrm{f}_2}{\partial \alpha} - \frac{\partial \mathrm{f}_1}{\partial \alpha}\frac{\partial \mathrm{f}_2}{\partial T_e} \quad (19)$$

3. The k-factors k_p, k_a and k_k

For a more detailed modelling, the k-factors are used to modify the predicted R-parameters according to (20) - (22).

$$R_g = k_g \times R_{gp} \quad (20)$$

$$R_a = k_a \times R_{ap} \quad (21)$$

$$R_k = k_k \times R_{kp} \quad (22)$$

By default these factors are set to 1, but they may be changed to create nuanced modifications to the voice characteristics. For instance, an increase in k_a will produce a breathier voice, beyond what would be implied by the R_d specification.

Only two of the three factors can be changed, the third being uniquely determined by the other parameters. In [2], the focus is on k_g and k_a, with k_k being redundant. However, in the current system, any two of the three k-factors may be modified, making the remaining factor redundant.

k_g+k_a: When k_g and k_a are changed, R_k needs to be recalculated, while maintaining the correct R_d value. Since R_k is not given, the algorithm described in Section 2 cannot be directly applied and needs to be modified. Again, we let T_e and α float, but as ω_g is now specified, and thus independent of T_e, we get $\mathrm{f}_{1k_{ga}}$ as shown in (23) while $\mathrm{f}_{2k_{ga}}$ is identical to f(α) in (10):

$$\mathrm{f}_{1k_{ga}}\left(T_e,\alpha\right) = U_p\left(\omega_g^2+\alpha^2\right)\sin\omega_g T_e + E_e\,\omega_g\left(e^{-\alpha T_e} + e^{-\alpha\left(T_e-\pi\omega_g^{-1}\right)}\right) \quad (23)$$

Three of the four partial derivatives are shown in (24) - (26), while the fourth, $\partial\mathrm{f}_{2k_{ga}}/\partial\alpha$, is the same as f'($\alpha$) in (11). T_e and α can then be obtained from (18) and (19) by substituting f_1 and f_2 with $\mathrm{f}_{1k_{ga}}$ and $\mathrm{f}_{2k_{ga}}$ respectively.

$$\frac{\partial \mathrm{f}_{1k_{ga}}}{\partial T_e} = U_p\left(\omega_g^2+\alpha^2\right)\omega_g\cos\omega_g T_e - E_e\,\omega_g \alpha e^{-\alpha T_e}\left(1+e^{\alpha\pi\omega_g^{-1}}\right) \quad (24)$$

$$\frac{\partial \mathrm{f}_{1k_{ga}}}{\partial \alpha} = 2U_p\alpha\sin\omega_g T_e - E_e\,\omega_g e^{-\alpha T_e}\left(T_e + \left(T_e - \pi\omega_g^{-1}\right)e^{\alpha\pi\omega_g^{-1}}\right) \quad (25)$$

$$\begin{aligned} \frac{\partial \mathrm{f}_{2k_{ga}}}{\partial T_e} = {} & \varepsilon\left(1-e^{-\varepsilon T_b}\right)\omega_g\left[\alpha\cos\omega_g T_e + \omega_g\sin\omega_g T_e - \alpha e^{-\alpha T_e}\right] \\ & + \omega_g\left(\omega_g^2+\alpha^2\right)\cos\omega_g T_e\left[1 - e^{-\varepsilon T_b}\left(1+\varepsilon T_b\right)\right] \end{aligned} \quad (26)$$

k_g+k_k: When k_g and k_k are changed, R_a needs to be recalculated, while maintaining the correct R_d value. Again, the algorithm requires modification: since R_a depends on ε, in this case we let ε and α float. Although the two functions $\mathrm{f}_{1k_{gk}}$ and $\mathrm{f}_{2k_{gk}}$ are the same as $\mathrm{f}_{1k_{ga}}$ and $\mathrm{f}_{2k_{ga}}$ respectively, two of the four partial derivatives are different, the ones with respect to ε: $\partial\mathrm{f}_{1k_{gk}}/\partial\varepsilon = 0$ and $\partial\mathrm{f}_{2k_{gk}}/\partial\varepsilon$ is shown in (27). $\partial\mathrm{f}_{1k_{gk}}/\partial\alpha$ and $\partial\mathrm{f}_{2k_{gk}}/\partial\alpha$ are shown in (25) and (11) respectively.

$$\begin{aligned} \frac{\partial \mathrm{f}_{2k_{gk}}}{\partial \varepsilon} = {} & \left(1-\left(1-\varepsilon T_b\right)e^{-\varepsilon T_b}\right)\left[\alpha\sin\omega_g T_e - \omega_g\cos\omega_g T_e + \omega_g e^{-\alpha T_e}\right] \\ & + \varepsilon T_b^2 e^{-\varepsilon T_b}\left(\omega_g^2+\alpha^2\right)\sin\omega_g T_e \end{aligned} \quad (27)$$

By substituting T_e with ε and f_1 and f_2 with the corresponding $\mathrm{f}_{1k_{gk}}$ and $\mathrm{f}_{2k_{gk}}$ in (18) and (19), the resulting equations are used to derive ε and α.

k_k+k_a: In the third option, when k_k and k_a are changed, R_g needs to be recalculated, while maintaining the correct R_d value. In this case, R_k and R_a are given while R_g is unknown, and therefore the algorithm described in Section 2 is directly applicable.

4. Parameter conflicts

When applying the correct calculation of R_d as described above, it became clear that for R_d values above 1.71, the predicted R_k values according to R_{kp} in (2) are incompatible with the LF model, being slightly too low to produce a possible LF pulse.

The purpose of this system is to stay true to the concept of the transformed LF control, but this outcome was clearly never intended. Therefore, it was decided to retain the original prediction as defined in (2), but for $R_d > 1.71$, R_k values should be set to the lowest possible R_k that would produce an LF pulse.

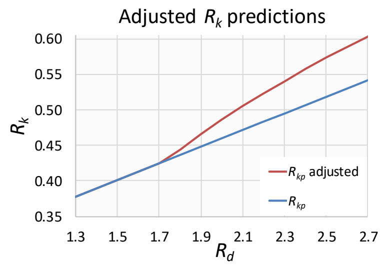

Figure 2: *For $R_d > 1.71$, R_k predictions are adjusted to the minimum R_k that will produce a valid LF pulse (red line).*

To find this R_k value, we need to find the O_q (or T_e) value which produces the minimum R_k for a specific $R_d > 1.71$. Regression analysis was carried out to find the relationship between R_k and O_q for 11 R_d values between 1.7 and 2.7 (R_d step = 0.1). The results of this analysis show that the relationship between R_k and O_q can be accurately modelled by a second order polynomial (R^2=0.9994 or higher) according to (28).

$$R_k = a_2 O_q^2 + a_1 O_q + a_0 \quad (28)$$

The coefficients of this polynomial, which depend on R_d, can also be modelled by second order polynomials (R^2=0.998 or higher). By setting the derivative of R_k with respect to O_q in (28) to 0, we get a function for finding the O_q value that will produce the minimum R_k for a given R_d, as shown in (29).

$$O_{q\to Rk\,\min} = \frac{-a_1}{2a_2} = \frac{0.11669R_d^2 + 0.0801R_d - 2.178}{2\times\left(0.09594R_d^2 + 0.61389R_d - 4.1333\right)} \quad (29)$$

Since O_q (and thus T_e) is now given and R_k is unknown, the iterative algorithm presented in [27] is used to find α and ω_g, from which the minimum R_k can be derived as $\omega_g\,\pi^{-1} O_q\, T_0 - 1$. The adjusted R_k predictions are shown by the red line in Fig. 2.

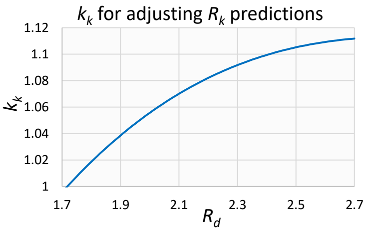

Figure 3: *Minimum k_k factor for $R_d > 1.71$ that needs to be applied to R_{kp} to produce a valid LF pulse.*

Alternatively, the data for the adjusted R_k can be used to derive an expression for calculating the minimum k_k required for $R_d > 1.71$ (see Fig. 3). It is important here for the predictions to be highly accurate, and a third order polynomial produces a nearly perfect fit ($R^2 = 1 - 1.04{\times}10^{-7}$). By using the formula in (30) together with (22) to get the new R_k value, the algorithm for k_k+k_a can be applied, i.e. the one described in Section 2.

$$k_k = 0.031978 R_d^3 - 0.32428 R_d^2 + 1.0706 R_d - 0.044299 \qquad (30)$$

5. Voice source signal generation

The discussion so far has been concerned mainly with the generation of one single LF pulse. The purpose of the system, however, is to allow for the generation of realistic voice source signals, which include the dynamics as reflected by the temporal variation in the source parameters. To this end, a system for synthesising source signals was developed, referred to as the voice source generator (VSG), using the MATLAB App Designer [28] (Fig. 4).

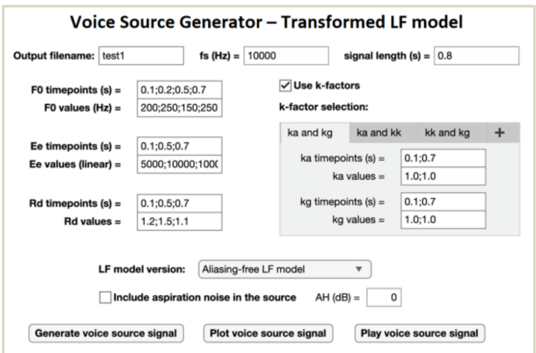

Figure 4: *The voice source generator user interface for the transformed LF parameter control.*

An important aspect of the VSG is how the fundamental period, T_0, is defined. To obtain the best correspondence between the entered f_0 data and the perceived pitch, we here define T_0 as the duration from one main excitation to the next (see Fig. 1), rather than as the duration of the LF pulse (for further details, see [29, pp. 11-13]). This means that two consecutive T_0 values are required to specify the LF pulse, one T_0 value for the open phase and the following T_0 value for the return phase. For the initial open phase, T_0 is by default set to be the same as the first T_0 value, whereas the T_0 value used for the final return phase, is the same as T_0 of the previous pulse.

Due to this definition of T_0, special consideration is needed for the derivation of the parameter values: the calculation of the values of one pulse is dependent on the values of the following pulse, which initially, of course, are unknown. Therefore, the parameter values need to be calculated in reverse order, starting with the last pulse, finishing with the first. Furthermore, for the last pulse, T_b is initially unknown and needs to be included in the iterations and updated as $T_0 - T_e$ for every iteration of T_e to converge to its correct value. Once this T_e value has been obtained, T_b of the preceding pulse is known, calculated as $T_0 - T_e$ of the final pulse.

The VSG currently incorporates the two different ways of discretising the LF model waveform, i.e. the standard method of sampling the functions in (4) and the aliasing-free version described in [30]. Since the latter implementation produces noticeably better sound quality [31], it is used here as the default. Furthermore, amplitude modulated aspiration noise can be added to the voice source pulses, where the modulation is determined by the glottal pulse shape as described in [32].

6. Conclusions

A system is presented which fully implements the transformed LF model, incorporating the often-overlooked *k*-factors. It is hoped that it will provide a useful tool for generating realistic voice source signals in a simple and straightforward way to facilitate further voice research.

To avoid errors in R_d, caused by the original stylisation of the LF pulse used in [1], an algorithm was implemented which involves four different versions of the Newton-Raphson iterative method for two variables. By using the new algorithm to obtain the correct R_d values, an anomaly was highlighted in the original R_k predictions, which leads to parameter combinations incompatible with the LF model. A remedy is proposed, which rectifies this while at the same time maintaining consistency with the original concept of the transformed LF model.

It should be noted that the prediction formulas in (2) and (3) are only meant to be valid up to $R_d = 2.7$ [1, 2]. Although the system works for higher R_d values as well, there are inconsistencies regarding the O_q values, which tend to decrease rather than increase as R_d increases beyond 2.7. Even when including the extended predictions proposed in [33] for $R_d > 2.7$ (see also [34, 35]), this issue remains. It seems that when R_d is very high, the O_q value is highly sensitive to small changes in R_k. Therefore, even a small inconsistency in the R_k prediction can lead to a large change in O_q. This is at odds with the concept of the transformed parameter control and would require further investigation. A different strategy may need to be adopted when it comes to predicting LF parameters from R_d when R_d is very large.

To ensure that a possible LF pulse is always produced from the input parameters, basic constraints are imposed on the input values to the VSG, according to the specifications in [29]. However, a more complex set of constraints will still be needed for the *k*-factors to prevent potential parameter conflicts. When these issues have been resolved, it is envisaged that the VSG app be made available from the ABAIR website [36].

7. Acknowledgements

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

8. References

[1] 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.

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

[3] 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.

[4] Y. Agiomyrgiannakis and O. Rosec, "Towards Flexible Speech Coding for Speech Synthesis: an LF + Modulated Noise Vocoder" in *Proc. INTERSPEECH 2008*, Brisbane, Australia, 2008, pp. 1849-1852.

[5] Y. Agiomyrgiannakis and O. Rosec, "ARX-LF-based source-filter methods for voice modification and transformation", in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 3589-3592.

[6] J. P. Cabral, S. Renals, K. Richmond, and J. Yamagishi, "Glottal spectral separation for parametric speech synthesis," in *Proc. INTERSPEECH 2008*, Brisbane, Australia, 2008, pp. 1829-1832.

[7] J. P. Cabral, S. Renals, J., Yamagishi, and K. Richmond, "HMM-based speech synthesiser using the LF-model of the glottal source," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011, pp. 4704-4707.

[8] G. Degottex, P. Lanchantin, A. Roebel, and X. Rodet, "Mixed source model and its adapted vocal tract filter estimate for voice transformation and synthesis," *Speech Communication*, vol. 55, no. 2, pp. 278-294, 2013.

[9] R. van Dinther, R. Veldhuis, and A. Kohlrausch, "Perceptual aspects of glottal-pulse parameter variations," *Speech Communication*, vol. 46, no. 1, pp. 95-112, 2005.

[10] A. del Pozo and S. Young, "The linear transformation of LF glottal waveforms for voice conversion," in *Proc. INTERSPEECH 2008*, Brisbane, Australia, 2008, pp. 1457-1460.

[11] C. Gobl and A. Ní Chasaide, "Acoustic characteristics of voice quality," *Speech Communication*, vol. 11, nos. 4-5, pp. 481-490, 1992.

[12] C. Gobl and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude", *Speech Communication*, vol. 40, pp. 189-212, 2003.

[13] S. Huber and A. Roebel, "On glottal source shape parameter transformation using a novel deterministic and stochastic speech analysis and synthesis system," in *Proc. INTERSPEECH 2015,* Dresden, Germany, 2015, pp. 289-293.

[14] J. Kane, M. Aylett, I. Yanushevskaya, and C. Gobl, "Phonetic feature extraction for context-sensitive glottal source processing. *Speech Communication*, vol. 59, pp. 10-21, 2014.

[15] I. Karlsson, "Glottal waveform parameters for different speaker types," in *Proc. 7th FASE Symposium*, Edinburgh, 1988, pp. 225-231.

[16] I. Karlsson, "Voice source dynamics for female speakers," in *Proc. Int. Conf. on Spoken Language Processing*, Kobe, Japan, 1990, pp. 225-231.

[17] A. Murphy, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "Voice source contribution to prominence perception: Rd implementation," in *Proc. INTERSPEECH 2018*, Hyderabad, India, 2018, pp. 217-221.

[18] A. Murphy, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "Testing the GlórCáil system in a speaker and affect voice transformation task," in *Proc. 10th Int. Conf. on Speech Prosody,* Tokyo, Japan, 2020, pp. 950-954.

[19] A. Ní Chasaide and C. Gobl, "Contextual variation of the vowel voice source as a function of adjacent consonants," *Language and Speech*, vol. 36, pp. 303-330, 1993.

[20] A. Ní Chasaide and C. Gobl, "Decomposing linguistic and affective components of phonatory quality," in *Proc. INTERSPEECH 2004*, Jeju Island, Korea, 2004, pp. 901-904.

[21] A. Ó Cinnéide, D. Dorran, M. Gainza, and E. Coyle, "A frequency domain approach to ARX-LF voiced speech parameterization and synthesis," in *Proc. INTERSPEECH 2011*, Florence, Italy, 2011, pp. 57-60.

[22] J. B. Pierrehumbert, "A preliminary study of the consequences of intonation for the voice source," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 30 (4), pp. 23-36, 1989.

[23] D. Vincent, O. Rosec, and T. Chonavel, "A new method for speech synthesis and transformation based on an ARX-LF source-filter decomposition and HNM modeling," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, 2007, pp. 525–528.

[24] I. Yanushevskaya, A. Murphy, C. Gobl, and A. Ní Chasaide. Global waveshape parameter R_d in signaling focal prominence: Perceptual salience in the absence of f0 variation. *Front. Commun.*, vol. 7, pp. 1-23, 2022.

[25] Z. Wang and C. Gobl. Contribution of the glottal flow residual in affect related voice transformation. in *Proc. INTERSPEECH 2022*, Incheon, Korea, 2022, pp. 5288–5292.

[26] P. Alku, T. Bäckström, and E. Vilkman, E., "Normalized amplitude quotient for parametrization of the glottal flow," *Journal of the Acoustical Society of America*, vol. 112, no. 2, pp. 701-710, 2002.

[27] C. Gobl, "Reshaping the transformed LF model: generating the glottal source from the waveshape parameter R_d," in *Proc. INTERSPEECH 2017*, Stockholm, Sweden, 2017, pp. 3008-3012.

[28] MATLAB 2022. Version 9.12.0.2170939 (R2022a). Natick, Massachusetts: The MathWorks Inc.

[29] C. Gobl, "The Voice Source in Speech Communication: Production and Perception Experiments Involving Inverse Filtering and Synthesis," PhD thesis, KTH, Stockholm, Sweden, 2003.

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

[31] Z. Wang and C. Gobl. "The perceptual effects of aliasing distortion in glottal flow modelling," in *Proc. 20th Int. Congress of Phonetic Sciences*, Prague, Czechia, forthcoming 2023.

[32] C. Gobl, "Modelling aspiration noise during phonation using the LF voice source model," in *Proc. INTERSPEECH 2006*, Pittsburgh, Pennsylvania, 2006, pp. 965-968.

[33] G. Fant, A. Kruckenberg, J. Liljencrants, and M. Båvegård, "Voice source parameters in continuous speech. Transformation of LF-parameters," in *Proc. of the Int. Conf. on Spoken Language Processing*, Yokohama, Japan, 1994, pp. 1451-1454.

[34] S. Huber and A. Roebel, "On the use of voice descriptors for glottal source shape parameter estimation," *Computer Speech & Language*, vol. 28, no. 5, pp. 1170-1194, 2014.

[35] S. Huber, A. Roebel, and G. Degottex, "Glottal source shape parameter estimation using phase minimization variants," in *Proc. INTERSPEECH 2012,* Portland, Oregon, 2012, pp. 1644-1647.

[36] ABAIR.ie – Speech and Language Technologies for Irish, https://abair.ie