*INTERSPEECH 2017*
August 20–24, 2017, Stockholm, Sweden

# Reshaping the Transformed LF Model: Generating the Glottal Source from the Waveshape Parameter $R_d$

*Christer Gobl*

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

## Abstract

Precise specification of the voice source would facilitate better modelling of expressive nuances in human spoken interaction. This paper focuses on the transformed version of the widely used LF voice source model, and proposes an algorithm which makes it possible to use the waveshape parameter $R_d$ to directly control the LF pulse, for more effective analysis and synthesis of voice modulations. The $R_d$ parameter, capturing much of the natural covariation between glottal parameters, is central to the transformed LF model. It is used to predict the standard *R*-parameters, which in turn are used to synthesise the LF waveform. However, the LF pulse that results from these predictions may have an $R_d$ value noticeably different from the specified $R_d$, yielding undesirable artefacts, particularly when the model is used for detailed analysis and synthesis of non-modal voice. A further limitation is that only a subset of possible $R_d$ values can be used, to avoid conflicting LF parameter settings. To eliminate these problems, a new iterative algorithm was developed based on the Newton-Raphson method for two variables, but modified to include constraints. This ensures that the correct $R_d$ is always obtained and that the algorithm converges for effectively all permissible $R_d$ values.

**Index Terms**: transformed LF model, $R_d$ parameter, glottal, voice source, Newton-Raphson method

## 1. Introduction

The Liljencrants-Fant (LF) model [1] has in many ways become a reference model in voice source related research. Its extensive use in wide-ranging studies involving speech analysis, synthesis and perception (e.g., [2-31]) has made important contributions to our understanding of voice quality and its role in spoken communication.

Typically, the shape of the LF model pulse is represented by the *R*-parameters $R_g$, $R_k$ and $R_a$. The amplitude of the flow pulse is not an explicit parameter; instead a measure of the strength of the glottal excitation, $E_e$, is used to capture the source amplitude (see Fig. 1). Despite being useful descriptors of voice characteristics, these parameters often display strong covariation [8]. For instance, $R_k$ and $R_a$ are typically negatively correlated with $E_e$, and there is typically a positive correlation between $R_k$ and $R_a$, at least as long as $R_k$ values are relatively small [8].

In the transformed LF model [5, 6], a global waveshape parameter $R_d$ is introduced, with the aim of reducing parameter redundancy in the glottal pulse description. As can be seen from equation (1), $R_d$ is defined by $U_p$, $E_e$, $f_0$ and a scale factor. The ratio of the amplitudes $U_p$ and $E_e$ is equivalent to the declination time of the glottal pulse, $T_d$ (see Fig. 1). Thus, $R_d$ is the declination time normalised to the fundamental period and scaled by the factor $0.11^{-1}$. The scale factor has been chosen so as to make the numerical value of $R_d$ the same as the declination time in milliseconds for $f_0 = 110$ Hz.

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

From $R_d$, default $R_k$ and $R_a$ values can be predicted ($R_{kp}$ and $R_{ap}$) using empirical formulas derived from linear regression analysis. Equations (2) and (3) from [5] are based on voice source analysis of vowel and consonant data, mainly from [6, 19, 20]. (These equations are only meant to be valid for $R_d$ values between 0.3 and 2.7, but for predictions based on an extended range of $R_d$ values, see [32, 33].)

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

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

A similar empirical formula for predicting $R_g$ has also been proposed by [5]. However, as the LF model pulse shape is fully determined by four parameters (three pulse shape parameters and $E_e$), $R_g$ should be set by the values of $R_d$, $R_{kp}$ and $R_{ap}$. Based on a geometrical simplification of the LF model waveform, Fant proposed the following approximate formula for calculating $R_d$ from $R_g$, $R_k$ and $R_a$ [5, 6]:

$$R_d = \frac{1}{0.11}(0.5 + 1.2\, R_k)\left(\frac{R_k}{4\, R_g} + R_a\right) \quad (4)$$

By using (2) and (3) above to predict $R_k$ and $R_a$, and by rearranging equation (4), $R_g$ is predicted according to (5):

$$R_{gp} = \frac{R_{kp}}{4}\left(\frac{0.5 + 1.2\, R_{kp}}{0.11\, R_d - R_{ap}\,(0.5 + 1.2\, R_{kp})}\right) \quad (5)$$

For a more detailed analysis, where there are deviations from the predicted values of $R_k$ and $R_a$, the deviations can be specified by a corresponding correction coefficient [5].

The transformed LF model provides an efficient way of modelling the glottal waveform, and we have recently incorporated it into our voice analysis system GlóRí [34]. A problem, however, with the current definition of the transformed LF model is that it relies on an approximation of $R_d$ as shown in (4). According to [5] the maximum error of this approximation is 1.7 dB, for $R_d$ values less than 2.7 and 0.5 dB when $R_d$ is less than 1.4. Although relatively small, at least for voices of modal voice quality, for precise modelling of for instance lax and breathy voice, these errors are clearly undesirable.



http://dx.doi.org/10.21437/Interspeech.2017-1140

Furthermore, the errors reported are for $R_d$ values obtained using only a limited range of possible values for the $R$-parameters. For instance, $R_k$ and $R_a$ were restricted to be below 0.6 and 0.12 respectively [7]. Using the full range of parameter values, the errors in the estimated $R_d$ are often considerably greater. For more extreme voice qualities, the modelling may therefore become less effective. Note that when deriving the LF pulse using $R_d$, $E_e$ and the predicted $R$-parameters, the specified $R_d$ value is not used directly in the synthesis. The actual $R_d$ of the LF pulse will be different from what was specified, and we can only approximately infer its value from expression (4).

The reason for the approximation of $R_d$ is that the peak glottal flow, $U_p$, which is used in the calculation of $R_d$, cannot be expressed in analytic form for the LF model. Hence, a new iterative algorithm was developed, which eliminates this problem with the transformed LF model. The algorithm allows the $R_d$ parameter to effectively control the generation of the LF pulse for the full range of permissible parameter values.

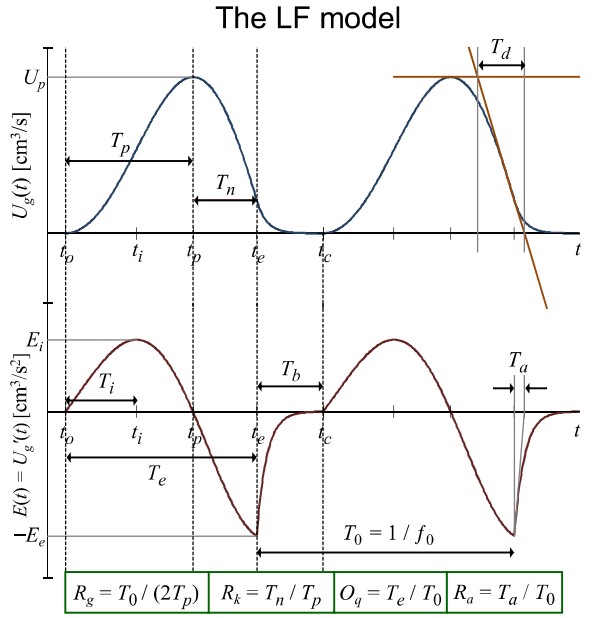


Figure 1. *Two LF model pulses and parameter definitions (for details, see [35]). Glottal flow (top), flow derivative (bottom).*

## 2. The LF model parameters

Often the parameters $E_e$, $R_g$, $R_k$ and $R_a$ are referred to as the "LF parameters". In fact, out of those four, only $E_e$ is an actual LF parameter. The differentiated glottal flow pulse of the LF model [1] is defined by (6), see also Fig. 1:

$$U_g'(t) = \begin{cases} E_0\, e^{\alpha t} \sin\omega_g t & t_o \le t \le t_e \\ \dfrac{-E_e}{\varepsilon\, T_a}\left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon T_b}\right) & t_e < t < t_c \end{cases} \quad (6)$$

Note that $E_0$ and $T_a$ in (6) are only ancillary parameters: $E_0$ is determined by $E_e$, $\alpha$, $T_e$, and $\omega_g$ according to:

$$E_0 = -E_e\, e^{-\alpha T_e} \csc\omega_g T_e \quad (7)$$

$T_a$ is determined by $\varepsilon$ and $T_b$ according to:

$$T_a = \varepsilon^{-1}\left(1 - e^{-\varepsilon T_b}\right) \quad (8)$$

Thus, the LF model parameters are $E_e$, $T_e$, $\omega_g$, $\alpha$, $\varepsilon$ and $T_b$. Naturally, when synthesising consecutive LF pulses, $f_0$ (or $T_0$) enters as a parameter, which is used to determine $T_b$. Note that $T_b = T_0 - T_e$, but for best possible synthesis quality, $T_b$ should ideally be determined by the $T_0$ and $T_e$ values of the following glottal pulse (for further discussion on this point, see [10, pp. 11-14]).

$T_e$ and $\omega_g$ are easily derived from the $R$-parameters and $f_0$: $T_e = (1+R_k)/(2R_g f_0)$ and $\omega_g = 2\pi R_g f_0$. However, $\alpha$ and $\varepsilon$ cannot be calculated analytically, but have to be estimated iteratively: $\varepsilon$ is derived from $T_a$ and $T_b$, and $\alpha$ is implicitly determined by the other parameters and from the LF model requirement of area balance [1], i.e. that the area of the positive part of the differentiated glottal flow pulse should be equal to area of the negative part (see the lower panel of Fig. 1).

In a typical implementation of the LF model, $\varepsilon$ and $\alpha$ are solved using the Newton-Raphson method as shown in (9):

$$x_{n+1} = x_n - \frac{\mathrm{f}(x_n)}{\mathrm{f}'(x_n)} \quad (9)$$

In the case of $\varepsilon$, f($\varepsilon$) and f '($\varepsilon$) are as follows:

$$\mathrm{f}(\varepsilon) = \varepsilon T_a - 1 + e^{-\varepsilon T_b} \quad (10)$$

$$\mathrm{f}'(\varepsilon) = T_a - T_b\, e^{-\varepsilon T_b} \quad (11)$$

In the case of $\alpha$, we get:

$$\mathrm{f}(\alpha) = A_o + A_r \quad (12)$$

$$\mathrm{f}'(\alpha) = \left(1 - \frac{2\alpha A_r}{E_e}\right)\sin\omega_g T_e - \omega_g T_e\, e^{-\alpha T_e} \quad (13)$$

$A_o$ is the area of the open phase (i.e. from $t_o$ to $t_e$) of the LF pulse and $A_r$ is the area of the return phase (i.e. from $t_e$ to $t_c$), as in (14) and (15) respectively.

$$A_o = \frac{E_0\, e^{\alpha T_e}}{\sqrt{\omega_g^2 + \alpha^2}} \sin\left(\omega_g T_e - \arctan\frac{\omega_g}{\alpha}\right) + \frac{E_0\,\omega_g}{\omega_g^2 + \alpha^2} \quad (14)$$

$$A_r = -\frac{E_e}{\varepsilon^2 T_a}\left(1 - e^{-\varepsilon T_b}\left(1 + \varepsilon T_b\right)\right) \quad (15)$$

Initial values of $\varepsilon$ and $\alpha$ are not particularly critical for these functions, and values typically converge very quickly.

## 3. Determining $R_d$ of the LF model

As we can see from the above, $U_p$ is not a parameter of the LF model. Nevertheless, this value has to be set, along with the values of $E_e$ and $f_0$, in order to achieve the correct $R_d$.

In the original LF model, the peak flow is allowed to 'float', i.e. it may attain any value as long as the condition of area balance is met. However, in the transformed LF model, two conditions have to be fulfilled: a specific $U_p$ value is required (as determined by $R_d$) as well as area balance.

Instead of using the approximate formula in (5) for setting $R_g$ (and thus $\omega_g$), we will use the Newton-Raphson method for two variables to find the correct values of $\omega_g$ and $\alpha$, given the specified $R_d$ value. The Newton-Raphson iterative method for two variables is shown in (16), where $\mathbf{J}(\omega_g, \alpha)^{-1}$ is the inverse Jacobian matrix defined in (17).

$$\begin{bmatrix}\omega_{gk+1}\\ \alpha_{k+1}\end{bmatrix} = \begin{bmatrix}\omega_{gk}\\ \alpha_{k}\end{bmatrix} - \mathbf{J}\left(\omega_{gk},\alpha_k\right)^{-1}\begin{bmatrix}\mathrm{f}_1\left(\omega_{gk},\alpha_k\right)\\ \mathrm{f}_2\left(\omega_{gk},\alpha_k\right)\end{bmatrix} \quad (16)$$

$$\mathbf{J}\left(\omega_g,\alpha\right)^{-1} = \frac{1}{\det\left(\mathbf{J}\left(\omega_g,\alpha\right)\right)}\times\begin{bmatrix}\dfrac{\partial \mathrm{f}_2}{\partial\alpha} & -\dfrac{\partial \mathrm{f}_1}{\partial\alpha}\\ -\dfrac{\partial \mathrm{f}_2}{\partial\omega_g} & \dfrac{\partial \mathrm{f}_1}{\partial\omega_g}\end{bmatrix} \quad (17)$$

By integrating equation (6) from $t_o$ to $t_p$, and replacing $E_0$ with the expression in (7), we get the following formula for $U_p$:

$$U_p = -\frac{E_e\,\omega_g\left(e^{-\alpha T_e} + e^{-\alpha\left(T_e-\pi\omega_g^{-1}\right)}\right)}{\left(\omega_g^2+\alpha^2\right)\sin\omega_g T_e} \quad (18)$$

From (18), we derive function $\mathrm{f}_1$:

$$\mathrm{f}_1\left(\omega_g,\alpha\right) = U_p\left(\omega_g^2+\alpha^2\right)\sin\omega_g T_e + E_e\omega_g\left(e^{-\alpha T_e} + e^{-\alpha\left(T_e-\pi\omega_g^{-1}\right)}\right) \quad (19)$$

The second condition, i.e. area balance, provides us with the second function, $\mathrm{f}_2$, which is identical to function (12). The four partial derivatives required for calculating the inverse Jacobian matrix in (16) are as follows:

$$\frac{\partial \mathrm{f}_1}{\partial\omega_g} = U_p\left(2\omega_g\sin\omega_g T_e + T_e\left(\omega_g^2+\alpha^2\right)\cos\omega_g T_e\right) + E_e\,e^{-\alpha T_e}\left[1 + e^{\alpha\pi\omega_g^{-1}}\left(1-\alpha\pi\omega_g^{-1}\right)\right] \quad (20)$$

$$\frac{\partial \mathrm{f}_1}{\partial\alpha} = 2U_p\,\alpha\sin\omega_g T_e - E_e\,\omega_g\,e^{-\alpha T_e}\left[T_e + e^{\alpha\pi\omega_g^{-1}}\left(T_e - \pi\,\omega_g^{-1}\right)\right] \quad (21)$$

$$\frac{\partial \mathrm{f}_2}{\partial\omega_g} = \omega_g T_e\sin\omega_g T_e + \left(\alpha T_e - 1\right)\cos\omega_g T_e + e^{-\alpha T_e} - A_r\,E_e^{-1}\left(2\omega_g\sin\omega_g T_e + T_e\left(\omega_g^2+\alpha^2\right)\cos\omega_g T_e\right) \quad (22)$$

$$\frac{\partial \mathrm{f}_2}{\partial\alpha} = \left(1-\frac{2\alpha\,A_r}{E_e}\right)\sin\omega_g T_e - \omega_g T_e\,e^{-\alpha T_e} \quad (23)$$

By expanding (16), we get the following two formulas for iteratively calculating $\omega_g$ and $\alpha$:

$$\begin{aligned}\omega_{gk+1} &= \omega_{gk} - \frac{1}{\det\left(\mathbf{J}\left(\omega_{gk},\alpha_k\right)\right)}\left[\frac{\partial \mathrm{f}_2}{\partial\alpha}\mathrm{f}_1 - \frac{\partial \mathrm{f}_1}{\partial\alpha}\mathrm{f}_2\right]\\ \alpha_{k+1} &= \alpha_k - \frac{1}{\det\left(\mathbf{J}\left(\omega_{gk},\alpha_k\right)\right)}\left[-\frac{\partial \mathrm{f}_2}{\partial\omega_g}\mathrm{f}_1 + \frac{\partial \mathrm{f}_1}{\partial\omega_g}\mathrm{f}_2\right]\end{aligned} \quad (24)$$

The Jacobian determinant in (17) and (24) is:

$$\det\left(\mathbf{J}\left(\omega_g,\alpha\right)\right) = \begin{vmatrix}\mathrm{f}'_{1\omega g} & \mathrm{f}'_{1\alpha}\\ \mathrm{f}'_{2\omega g} & \mathrm{f}'_{2\alpha}\end{vmatrix} = \frac{\partial \mathrm{f}_1}{\partial\omega_g}\frac{\partial \mathrm{f}_2}{\partial\alpha} - \frac{\partial \mathrm{f}_1}{\partial\alpha}\frac{\partial \mathrm{f}_2}{\partial\omega_g} \quad (25)$$

Note that the functions and the partial derivatives in (24) are all functions of $\omega_{gk}$ and $\alpha_k$.

This provides the basic solution to the problem of achieving the correct $U_p$ and $R_d$ as well as area balance. Unfortunately, this direct implementation of the Newton-Raphson method does not guarantee that values will converge. If initial values are not sufficiently close to the correct targets, they will fail to converge. Furthermore, occasionally when values converge they are incorrect, $\omega_g$ being too high. Therefore, despite working properly in many cases, in effect these limitations render the basic implementation of the algorithm unusable.

### 3.1. Predicting initial $\omega_g$ and $\alpha$ values

By studying the properties of the two functions (19) and (12), we find that both are reasonably well behaved when $\alpha$ is varying while keeping $\omega_g$ constant at its correct value (see the two top panels of Fig. 2). They are decreasing monotonically and there is one unique zero. However, looking at the lower two panels of Fig. 2, where $\omega_g$ is changing while $\alpha$ is set to its correct value, we find a complex pattern with multiple zeros: hence the potential for $\omega_g$ converging to an incorrect value.

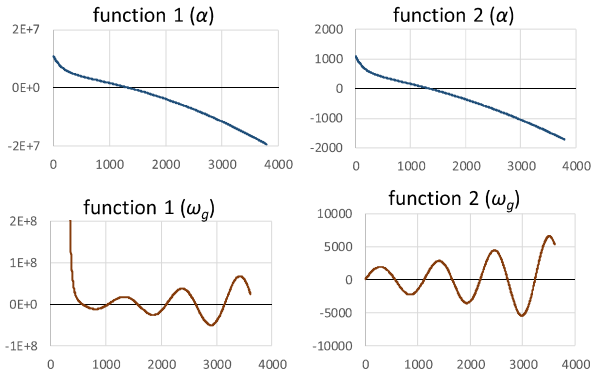


Figure 2. *The two functions used in the iterative algorithm: there is no unique solution for $\omega_g$ and $\alpha$.*

It is clear that the correct $\omega_g$ is found at the first zero (the lowest value); the $\omega_g$ from a higher zero would produce two or more pulses within the glottal cycle, where the first pulse would have the correct $U_p$ value, as illustrated in Fig. 3.

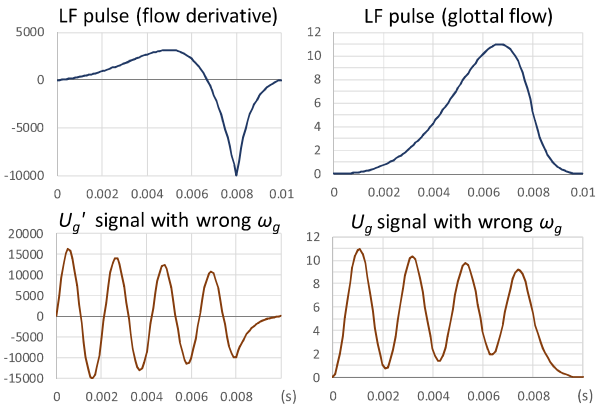


Figure 3. *LF pulse (flow derivative and flow) with correct $\omega_g$ and $\alpha$ (top) and incorrect values (bottom) due to the algorithm converging to the wrong $\omega_g$* and *$\alpha$.*

In an attempt to remedy these problems, two complementary approaches were explored. First, regression analysis was carried out to derive expressions which allow us to better guess the initial values of $\omega_g$ and $\alpha$, thereby minimising the risk of

values not converging. Secondly, constraints are imposed on potential values of $\omega_g$. It can easily be shown that $R_g$ values of the LF model have to be greater than $0.5/O_q$ and smaller than $1/O_q$, where $O_q = T_e/T_0$. In terms of the LF parameters, the permissible $\omega_g$ values are as follows:

$$\frac{\pi}{T_e} < \omega_g < \frac{2\pi}{T_e} \tag{26}$$

Regression analysis was carried out correlating $R_g$ with $R_d$. LF pulses with $R_g$ values ranging from 0.55 to 5.05 in steps 0.1 were generated (46 different settings). This range was repeated with eight different $O_q$ values ranging from 0.25 to 0.95 in steps of 0.1. In turn, for each of the eight $O_q$ values, eight different $R_a$ values were used, ranging from 0 to 0.14 in steps of 0.02. Note that not all parameter combinations would be consistent with the LF model. In case of a conflict between the $R_g$ and $O_q$ values, $O_q$ was adjusted to the closest possible value. In case of a conflict between $O_q$ and $R_a$, $R_a$ was adjusted to the nearest possible value. A total of $46\times8\times8 = 2944$ different pulses were generated, but for the regression analysis only the unadjusted combinations were used.

The results from the analysis presented here are based on the correlation between $R_g$ with $R_d$ for different $O_q$ values, while keeping $R_a = 0$. This produced the following formula for predicting $R_g$:

$$\hat{R}_g = 0.684\, O_q^{-1.145}\, R_d^{0.155} \tag{27}$$

By multiplying the expression in (27) by $2\pi f_0$ we get an initial estimate of $\omega_g$ as a function of the $R_d$ and $O_q$ values. If a predicted value is outside the permissible range, as determined by (26), the initial $\omega_g$ value is adjusted to the nearest possible value. The average normalised error in the predicted $R_g$ values for the 2944 LF pulses was 0.044, the maximum error was 0.71 and $R^2$=0.985. By also including the $R_a$ variation in the regression analysis, a much more complex prediction formula is obtained, but the overall predictive power was only marginally improved. In comparison, using formula (5) on the same data resulted in an $R^2$=0.29 when no constraints were applied and an $R^2$=0.66 when values were adjusted to the nearest possible, if outside the permissible range.

A similar regression analysis could be carried out for the prediction of the initial $\alpha$ value. However, it is perhaps more effective to use the initial $\omega_g$ prediction and then calculate the matching $\alpha$ value using (9) and the functions (12) and (13). Applying this method to the 2944 LF pulses described above, by replacing the original $R_g$ values with $R_g$ values predicted from (27), produced an $R^2$=0.885 for the predicted $\alpha$ values.

### 3.2. Constraining the parameter values

Although the initial predictions of $\omega_g$ and $\alpha$ are reasonably good and help speed up the rate of convergence, they still do not necessarily guarantee convergence in every instance. To achieve that, we need to constrain the iterative process of (24) by evaluating every new estimate of $\omega_g$. If a value violates the constraints of (26), i.e. is outside the range of possible $\omega_g$ values, the $\omega_g$ estimate is changed to a value that is permitted.

In this case, the nearest possible value should not be used, since this would result in a perpetual, non-converging loop. Instead, $\omega_g$ values are constrained according to (28), where $\hat{\omega}_g$ represents a particular output value of the $\omega_g$-iteration in (24). The absolute difference between $\hat{\omega}_g$ and the nearest boundary is calculated, and $\pi/T_e$ (or multiples of $\pi/T_e$) is subtracted if the difference is greater than $\pi/T_e$. If $\hat{\omega}_g$ is too low, the difference is subtracted from the upper boundary, if $\hat{\omega}_g$ is too high it is added to the lower boundary.

$$\begin{cases} \omega_g = \dfrac{2\pi}{T_e} - \mathrm{mod}\left(\dfrac{\pi}{T_e} - \hat{\omega}_g, \dfrac{\pi}{T_e}\right), & \hat{\omega}_g \le \dfrac{\pi}{T_e} \\ \omega_g = \dfrac{\pi}{T_e} + \mathrm{mod}\left(\hat{\omega}_g - \dfrac{\pi}{T_e}, \dfrac{\pi}{T_e}\right), & \hat{\omega}_g \ge \dfrac{2\pi}{T_e} \end{cases} \tag{28}$$

This constraint imposed on the normal iterative process of the Newton-Raphson method ensures that values will always converge, with the exception of $R_d$ values that are very close to the boundary of permissible $R_d$ values.

The range of theoretically possible $R_d$ values produced by the LF model is only constrained by the area of the return phase. There is in principle no upper $R_d$ boundary, something which can be deduced from (18) by letting $\omega_g$ approach its upper boundary, $2\pi/T_e$.

The lower $R_d$ boundary is determined by the area of the return phase. Since area balance is a requirement of the LF model, the negative area of the return phase needs to be matched by a positive area. The area from $t_p$ to $t_e$ cannot be positive, and consequently the $R_d$ minimum is when $U_p$ (i.e. the area from $t_o$ to $t_p$) is equal to the absolute area of the return phase. Therefore, the minimum $R_d$ is according to (29), where $A_r$ is the area of the return phase (15):

$$R_{d_{\min}} = -\frac{A_r\, f_0}{0.11 E_e} \tag{29}$$

It should be noted that this is a theoretical limit, which can only be achieved when α tends to infinity. Obviously, such pulses cannot be produced in reality, and for $R_d$ values very close to this limit, the algorithm will not converge. This problem can easily be avoided by adding a small value to the theoretical limit in (29), e.g., by constraining $R_d$ so that it is always greater than $R_{d_{\min}} + 0.1$.

A constraint is also imposed on $\alpha$: it is important to cap $\alpha$ values to avoid numerical overflow errors. A maximum of 20,000 and a minimum of −1,000 seem to work well. If outside these boundaries, values are modified according to (30) where $\hat{\alpha}$ is a particular output value of the $\alpha$-iteration in (24).

$$\begin{cases} \alpha = \alpha_{\max} - \mathrm{mod}\left(\hat{\alpha} - \alpha_{\max},\ \alpha_{\max}/2\right), & \hat{\alpha} > \alpha_{\max} \\ \alpha = \alpha_{\min} + \mathrm{mod}\left(\alpha_{\min} - \hat{\alpha},\ \left|\alpha_{\min}/2\right|\right), & \hat{\alpha} < \alpha_{\min} \end{cases} \tag{30}$$

## 4. Conclusions

Due to approximations in the original definition of the transformed LF model, errors in the $R_d$ parameter are introduced, which can lead to undesirable artefacts. A new iterative algorithm is presented, which eliminates these errors. The algorithm ensures that the correct $R_d$ is always obtained and it converges for effectively all permissible $R_d$ values. Ongoing work on improving the efficiency of the algorithm involves refinement of the initial $\omega_g$ predictions and the constraints. When the constraints are imposed, the number of iterations required for the algorithm to converge can increase considerably. Thus, the better the initial predictions, the less likely this is to happen. We are currently also working on extending the transformed LF model to incorporate aspiration noise according to the principles in [36, 37].

# 5. References


[1] 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, 4, pp. 1-13, 1985.

[2] J. P. Cabral, S. Renals, J. Yamagishi, and K. Richmond, "HMM-based speech synthesiser using the LF-model of the glottal source," pp. 4704-4707, 2011.

[3] R. Carlson, G. Fant, C. Gobl, B. Granström, I. Karlsson, and Q. Lin, "Voice source rules for text-to-speech synthesis," *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Glasgow, vol. 1, pp. 223-226, 1989.

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

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

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

[7] G. Fant, A. Kruckenberg, J. Liljencrants, and M. Båvegård, "Voice source parameters in continuous speech. Transformation of LF-parameters," *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, pp. 1451-1454, 1994.

[8] C. Gobl, "Voice source dynamics in connected speech," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 1, pp. 123-159, 1988.

[9] C. Gobl, "A preliminary study of acoustic voice quality correlates," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 4, pp. 9-21, 1989.

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

[11] C. Gobl, E. Bennett, and A. Ní Chasaide, "Expressive synthesis: how crucial is voice quality?," *Proceedings of the IEEE Workshop on Speech Synthesis*, Santa Monica, California, paper 52, 4 pp., 2002.

[12] C. Gobl and A. Ní Chasaide, "The effects of adjacent voiced/voiceless consonants on the vowel voice source: a cross language study," *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 2-3, pp. 23-59, 1988.

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

[14] C. Gobl and A. Ní Chasaide, "Voice source variation in the vowel as a function of consonantal context," in W. J. Hardcastle and N. Hewlett (Eds.) *Coarticulation: Theory, Data and Techniques*, Cambridge University Press, Cambridge, pp. 122-143, 1999.

[15] 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.

[16] C. Gobl and A. Ní Chasaide, "Amplitude-based source parameters for measuring voice quality," *Proceedings of the ISCA VOQUAL'03 Workshop on Voice Quality: Functions, Analysis and Synthesis*, Geneva, Switzerland, pp. 151-156, 2003.

[17] C. Gobl and J. Mahshie, "Inverse filtering of nasalized vowels using synthesized speech," *Journal of Voice*, vol. 27, no. 2, pp. 155-169, 2013.

[18] J. Kane, M. Kane, and C. Gobl, "A spectral LF model based approach to voice source parameterisation," *INTERSPEECH 2010*, Makuhari, Japan, pp. 2606-2609, 2010.

[19] I. Karlsson, "Glottal waveform parameters for different speaker types," *Proceedings of 7 FASE Symposium*, Edinburgh, pp. 225-231, 1988.

[20] I. Karlsson, "Voice source dynamics for female speakers," *Proceedings of the International Conference on Spoken Language Processing*, Kobe, Japan, pp. 225-231, 1990.

[21] 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.

[22] A. Ní Chasaide, A. and C. Gobl, "Decomposing linguistic and affective components of phonatory quality," *Proceedings of the 8th International Conference on Spoken Language Processing, INTERSPEECH 2004*, Jeju Island, Korea, vol. 2, pp. 901-904, 2004.

[23] 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.

[24] 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, 4, pp. 23-36, 1989.

[25] C. Ryan, A. Ní Chasaide, and C. Gobl, "Voice quality variation and the perception of affect: continuous or categorical?" *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS-15)*, Barcelona, pp. 2409-2412, 2003.

[26] 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.

[27] I. Yanushevskaya, C. Gobl, and A. Ní Chasaide, "Voice quality and $f_0$ cues for affect expression: implications for synthesis. *Proceedings of the 9th European Conference on Speech Communication and Technology, INTERSPEECH 2005*, Lisbon, Portugal, pp. 1849-1852, 2005.

[28] I. Yanushevskaya, C. Gobl, and A. Ní Chasaide, "Mapping Voice to Affect: Japanese listeners." *Proceedings of the 3rd International Conference on Speech Prosody,* Dresden, Germany, paper OS4-4-265, 4 pp., 2006.

[29] I. Yanushevskaya, C. Gobl, and A. Ní Chasaide, Voice quality in affect cueing: does loudness matter? *Frontiers in Psychology*, vol. 4, article 335, pp. 1-14, 2013.

[30] I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "The interaction of long-term voice quality with the realisation of focus," *Proceedings of the 8th International Conference on Speech Prosody,*, Boston, Massachusetts, pp. 931-935, 2016.

[31] I. Yanushevskaya, A. Murphy, C. Gobl and A. Ní Chasaide, "Perceptual salience of voice source parameters in signaling focal prominence," *INTERSPEECH 2016*, San Francisco, California, pp. 3161-3165, 2016.

[32] 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.

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

[34] J. Dalton, J. Kane, I. Yanushevskaya, A. Ní Chasaide, and C. Gobl, "GlóRí - the Glottal Research Instrument", *Proceedings of the 7th International Conference on Speech Prosody,* Dublin, Ireland, pp. 944-948, 2014.

[35] 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.

[36] 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.

[37] 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, pp. 965-968, 2006.